

Llicenciatura en Ciències i Tècniques Estadístiques

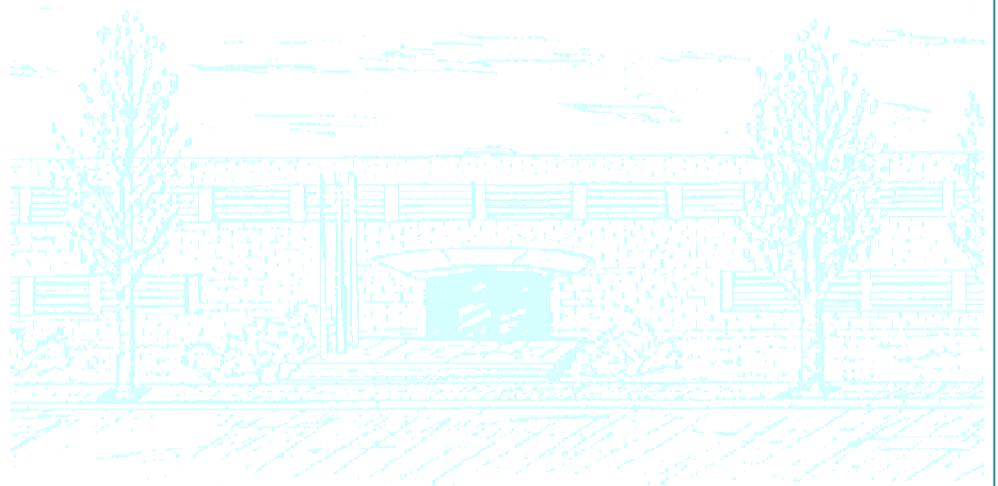
Títol: Anàlisi Estadístic de dats de Atletismo

**Autors: Dolores Cano Jiménez
Antonio Jiménez Vecina**

Director: Pau Fonseca Casas

Departament: Statistics and Operations Research

Convocatori: Junio 2012



**Facultat de Matemàtiques
i Estadística**

UNIVERSITAT POLITÈCNICA DE CATALUNYA

Índice del documento

ÍNDICE DEL DOCUMENTO	2
1 INTRODUCCIÓN	6
1.1 PRESENTACIÓN.....	6
1.2 OBJETIVOS	6
1.3 MOTIVACIONES	6
2 MARCO TEÓRICO.....	7
2.1 CONSIDERACIONES.....	7
2.1.1 <i>Cualidades Físicas.....</i>	<i>7</i>
2.1.2 <i>Factores Determinantes.....</i>	<i>7</i>
2.2 EFECTOS DEL ENTRENAMIENTO	7
2.3 FACTORES LIMITADORES.....	8
2.4 FASES DEL ENTRENAMIENTO.....	8
2.5 EVALUACIÓN DEL ESTADO FÍSICO.....	8
2.6 METODOLOGÍA ESTADÍSTICA.....	12
2.6.1 <i>Metodología general utilizada.....</i>	<i>12</i>
2.6.2 <i>Modelos lineales generalizados.....</i>	<i>13</i>
2.6.3 <i>Criterio de Akaike.....</i>	<i>13</i>
2.6.4 <i>Análisis de la varianza.....</i>	<i>13</i>
2.6.5 <i>Pruebas de normalidad.....</i>	<i>14</i>
2.6.6 <i>Pruebas de hipótesis.....</i>	<i>14</i>
2.6.7 <i>Análisis de los residuos.....</i>	<i>15</i>
2.6.8 <i>Detección de valores influyentes a priori y a posteriori.....</i>	<i>15</i>
2.6.9 <i>Buscar mejor modelo.....</i>	<i>16</i>
2.6.10 <i>Contraste de varianza incremental.....</i>	<i>16</i>
2.6.11 <i>Test de Durbin-Watson o ACF.....</i>	<i>17</i>
2.7 ESTADO DEL ARTE	18
2.7.1 <i>Índice de dureza.....</i>	<i>18</i>
2.7.2 <i>Ejercicios respiración.....</i>	<i>19</i>
2.7.3 <i>Ejercicios equilibrio muscular.....</i>	<i>20</i>
3 MATERIAL Y MÉTODOS.....	21
3.1 SOFTWARE.....	21
3.2 INDIVIDUOS	21
3.2.1 <i>Individuos para el experimento de afectación de entrenamiento.....</i>	<i>21</i>

3.2.1.1	Completar individuos para experimento	22
3.2.1.1.1	Tratamiento datos ficheros HRM.	22
3.2.2	<i>Individuos para la determinación del grado de dureza de una prueba atlética.</i>	27
3.2.2.1	Datos técnicos carreras.....	27
3.2.2.2	Datos facilitados	27
3.2.2.3	Cálculos realizados.....	28
3.2.2.3.1	Cálculo Edad	28
3.2.2.3.2	Cálculo Cooper	28
3.2.2.3.3	Cálculo VO ₂ max.....	29
3.2.2.3.4	Recodificación edad.....	29
3.3	EXPERIMENTOS PROPUESTOS PARA LA DETERMINACIÓN DE LA AFECTACIÓN DEL ENTRENAMIENTO.	30
4	ANÁLISIS DE DATOS	32
4.1	ANÁLISIS DATOS EXPERIMENTO AFECTACIÓN DEL ENTRENAMIENTO.	32
4.1.1	<i>Datos obtenidos de voluntarios.....</i>	32
4.1.2	<i>Ampliación datos corredores eolymp.</i>	34
4.1.3	<i>Repetición del estudio haciendo tratamiento de missings.</i>	35
4.2	ANÁLISIS DE DATOS PARA EL ANÁLISIS DEL ÍNDICE DE DUREZA.	38
4.2.1	<i>Juegos de datos</i>	39
4.2.1.1	Agrupado por carrera	39
4.2.1.2	Agrupado por corredor.....	41
4.2.2	<i>Resultados obtenidos</i>	44
4.2.2.1	Resultados juego de datos cursas	44
4.2.2.2	Resultados juego de datos cursas sin la carrera de la maquinista 5 km	47
4.2.2.3	Resultados juego de datos cursas sin las carrera de Behovia y selva marítima ...	51
4.2.2.4	Resultados juego de datos cursas sin la carrera de Guanta.....	55
4.2.2.5	Modelo tomado como "mejor modelo".	57
4.2.2.6	Resultados Carreras individuales.....	58
4.2.2.6.1	Cursa de Bombers.....	58
4.2.2.6.2	Cursa de la Mercè.	59
4.3	APLICACIÓN PRÁCTICA DE RESULTADOS.	61
5	CONCLUSIONES	63
5.1	DATOS EXPERIMENTOS.....	63
5.2	DATOS DE ATLETAS EN CARRERAS.	63
5.3	LÍNEAS FUTURAS.....	64
5.3.1	<i>Vincular los dos estudios.</i>	64
5.3.2	<i>Mejora VO₂Max con respirador.</i>	64
5.3.3	<i>Efecto condiciones climatológicas.....</i>	64
5.3.4	<i>Efecto Altura.</i>	64

5.3.5	<i>Efecto latitud.</i>	65
5.3.6	<i>Efecto Calendario (Inicio – Fin temporada).</i>	66
5.3.7	<i>Analizar clusters.</i>	66
5.3.7.1	Grupos de edad.	66
5.3.7.2	Análisis Cuartil.	66
5.3.8	<i>Análisis de componentes principales.</i>	66
5.4	ASPECTOS MEJORABLES.	66
5.4.1	<i>Captación voluntarios para experimentos.</i>	66
5.4.2	<i>Mayor duración del experimento.</i>	67
5.4.3	<i>Afinamiento índice de dureza.</i>	67
6	APENDICES	68
6.1	APENDICE A. GLOSARIO	68
6.2	APENDICE B. ENTRENAMIENTO EQUILIBRIO MUSCULAR.	68
6.3	APENDICE C. ENTRENAMIENTO RESPIRATORIO.	72
6.4	APENDICE D. FORMATO FICHEROS HRM.	72
6.4.1	<i>Parámetros generales.</i>	72
6.5	APENDICE E. RESULTADOS R (LISTADOS SALIDA).	74
6.5.1	<i>Resultados Experimentos.</i>	74
6.5.1.1	Análisis descriptivo.	74
6.5.1.2	Búsqueda modelo Cooper_2	75
6.5.1.3	Búsqueda modelo Mejora	78
6.5.2	<i>Resultados Experimentos (Ampliación de datos).</i>	81
6.5.2.1	Análisis descriptivo.	81
6.5.2.2	Normalidad variable respuesta	82
6.5.2.3	Búsqueda modelo	82
6.5.2.4	Estudio residuos	86
6.5.2.5	Detección valores influyentes	88
6.5.3	<i>Resultados Experimentos (Tratamiento de missings).</i>	89
6.5.3.1	Análisis descriptivo.	89
6.5.3.2	Búsqueda modelo	91
6.5.3.3	Estudio de residuos	96
6.5.3.4	Detección valores influyentes	100
6.5.4	<i>Resultados juego de datos cursas</i>	102
6.5.4.1	Análisis descriptivo.	102
6.5.4.2	Estudio de normalidad de los datos	105
6.5.4.3	Búsqueda del modelo	105
6.5.5	<i>Resultados juego de datos cursas sin la carrera de la maquinista 5 km....</i>	107
6.5.5.1	Análisis descriptivo.	107
6.5.5.2	Estudio de normalidad de los datos	109

6.5.5.3	Búsqueda del modelo	110
6.5.5.4	Estudio de residuos	112
6.5.5.5	Detección de valores influyentes a priori y a posteriori.....	115
6.5.6	<i>Resultados juego de datos cursas sin las carrera de Behovia y selva marítima</i>	117
6.5.6.1	Análisis descriptivo.....	117
6.5.6.2	Búsqueda del modelo	118
6.5.6.3	Estudio de residuos	122
6.5.6.3.1	Diagrama bivalente de residuos estandarizados vs previstos	123
6.5.6.3.2	Independencia de los residuos studentizados: Test de Durbin-Watson o ACF (modelo completo).....	124
6.5.6.3.3	Diagrama bivalente de los residuos estandarizados vs X_i (modelo completo)	125
6.5.6.3.4	Detección de valores influyentes a priori y a posteriori (modelo completo) ..	125
6.5.6.4	Análisis de los residuos y valores influyentes del modelo resultante del step....	126
6.5.6.4.1	Diagrama bivalente de los residuos estandarizados vs predichos	128
6.5.6.4.2	Diagrama bivalente de los residuos estandarizados vs regresoras	129
6.5.7	<i>Resultados juego de datos cursas sin la carrera de Guanta</i>	134
6.5.7.1	Análisis descriptivo.....	134
6.5.7.2	Estudio de normalidad de los datos	135
6.5.7.3	Búsqueda del modelo	135
6.5.7.4	Estudio de residuos	136
6.5.7.4.1	Diagrama bivalente de residuos studentizados vs previstos.....	137
6.5.7.4.2	Diagrama bivalente de los residuos estandarizados vs X_i	138
6.5.7.4.3	Independencia de los residuos studentizados: Test de Durbin-Watson o ACF	138
6.5.7.5	Detección de valores influyentes a priori y a posteriori.....	139
6.5.8	<i>Resultados juego de datos completo.</i>	140
6.5.8.1	Cursa de Bombers.....	140
6.5.8.2	Cursa de la Mercè.	142
7	BIBLIOGRAFIA.....	146

1 Introducción

1.1 Presentación

Actualmente los atletas, aprovechando la tecnología de los pulsímetros o las pruebas específicas, generan una gran cantidad de información que puede ser analizada con el fin de mejorar su rendimiento deportivo.

1.2 Objetivos

Determinar si una determinada prueba atlética es más dura que otra es una tarea compleja dado que son diferentes los factores que intervienen. Así mismo conocer cuál es el estado de forma de un atleta y ver cuáles son las acciones que pueden llevarse a cabo para mejorar en un tiempo acotado, es también complejo.

El objetivo del proyecto es dual, por un lado, se desea analizar la información existente de una determinada prueba deportiva con el objetivo de determinar su grado de dureza. Dicho grado de dureza se ha asociado al VO_2Max , variable que representa la cantidad máxima de oxígeno que el organismo puede absorber, transportar y consumir por unidad de tiempo determinado. Esta variable es indicativo de la capacidad aeróbica del atleta (cuanto mayor sea el VO_2Max , mayor capacidad aeróbica tiene el atleta).

Por otro lado, se desea también analizar la afectación de un tipo de entrenamiento adicional en la mejora física del atleta. Para ello se ha definido un experimento con un grupo de control y dos tratamientos específicos destinados a la mejora de la forma física. Estos tratamientos son:

- Mejora de la capacidad aeróbica mediante el entrenamiento de los músculos respiratorios. Se le pide al atleta que además de su entrenamiento habitual realice unos sencillos ejercicios de respiración (ver detalle de ejercicios en el punto 6.3) cada día durante unos 15 minutos. Con la realización de estos ejercicios deberíamos comprobar el crecimiento del VO_2Max .
- Mejora del equilibrio muscular. Se le pide al atleta que además de su entrenamiento habitual realice unos ejercicios de propiocepción (ver detalle de ejercicios en 6.2) basados en ejercitar el equilibrio muscular. Con la realización de estos ejercicios se pretende observar una mejora en la velocidad del atleta.

De esta forma habremos analizado:

- El estado de forma de un atleta.
- Dureza de una carrera

La combinación de estos dos objetivos permitirá a la larga, predecir el comportamiento de un atleta en una carrera concreta.

1.3 Motivaciones

Por medio del análisis de los datos de los entrenamientos puede obtenerse muchos factores que ayuden a la progresión del mismo:

- Es posible aconsejar un entrenamiento adicional al atleta para mejorar su forma física, velocidad, etc....
- Es posible detectar algún problema físico del atleta.
- Es posible calcular el riesgo de lesión o sobre-entrenamiento del atleta.
- Facilidad de seguimiento de los atletas por parte del entrenador.

2 Marco teórico

2.1 Consideraciones

2.1.1 Cualidades Físicas

Existen ciertos factores determinantes de la condición física de cada persona que pueden ayudar a decidir un plan de entrenamiento adecuado que nos puede ayudar a conseguir el máximo desarrollo del potencial físico de la persona.

- ✓ **Resistencia.** Capacidad de realizar un esfuerzo durante el mayor tiempo posible.
- ✓ **Fuerza.** Capacidad de vencer una resistencia por medio de un esfuerzo muscular
- ✓ **Velocidad.** Capacidad de realizar movimientos en el menor tiempo posible
- ✓ **Flexibilidad.** Capacidad de estirar los músculos sin sentir dolor.

2.1.2 Factores Determinantes

- ✓ **Resistencia.**
 - **Resistencia aeróbica.** Nos permite prolongar un esfuerzo de intensidad leve existiendo equilibrio entre el aporte y el gasto de oxígeno por el organismo
 - **Resistencia anaeróbica.** Nos permite resistir una deuda de oxígeno manteniendo un esfuerzo intenso.
- ✓ **Fuerza.**
 - **Contracción isotónica concéntrica.** La tensión muscular es suficiente para superar la resistencia. El músculo se acorta acercando sus intersecciones.
 - **Contracción isotónica excéntrica.** La tensión muscular no es suficiente para vencer la resistencia. Supone un alargamiento del músculo separando sus dos puntos de intersección.
 - **Contracción isométrica.** El músculo no se acorta ni se alarga sólo genera tensión.
- ✓ **Velocidad.** 4 tipos:
 - **Velocidad de Traslación.** Directamente relacionada con el tiempo.
 - **Frecuencia.**
 - **Velocidad de reacción.** Tiempo que el deportista tarda en reaccionar ante un estímulo
 - **Velocidad mental.** Rapidez de decisión mental.
- ✓ **Flexibilidad.**
 - **Movilidad articular.** Depende de la configuración de cada articulación.
 - **Elasticidad.** Capacidad motriz que permite al músculo estirarse hasta cierto límite.

2.2 Efectos del entrenamiento

Efectos sobre el sistema cardíaco:

- ✓ Mayor longitud y grosor de las fibras cardíacas
- ✓ Mayor fuerza de contracción
- ✓ Menor fuerza cardíaca en reposo
- ✓ Menor frecuencia cardíaca en ejercicio
- ✓ Menor tensión arterial
- ✓ Mejor recuperación

Efectos sobre el sistema respiratorio:

- ✓ Mayor capacidad vital pulmonar
- ✓ Mayor eficacia en músculos respiratorios
- ✓ Menor consumo de oxígeno para un mismo trabajo
- ✓ Menor frecuencia respiratoria y más profunda
- ✓ Mayor poder de consumo de oxígeno

Efectos sobre el sistema nervioso:

- ✓ Mejora de la rapidez de conducción de los estímulos
- ✓ Mejora de los mecanismos de respuesta
- ✓ Economiza movimientos en la misma tarea

Efectos sobre el sistema óseo y muscular:

- ✓ Refuerzo de ligamentos y tendones en especial en las inserciones óseas y musculares. Aumento en el espesor de los cartílagos en todas las articulaciones
- ✓ Aumento del tejido conectivo como consecuencia del entrenamiento de la fuerza, del tejido muscular y del porcentaje mineral óseo.

2.3 Factores limitadores

- ✓ **Deuda de oxígeno.** La actividad física provoca un gasto de oxígeno directamente proporcional a la intensidad del esfuerzo y duración. Cantidad de oxígeno que se gasta por encima de los valores basales después del esfuerzo.
- ✓ **Consumo de oxígeno y capacidad de absorción.** Mayor cantidad de O₂ que una persona puede consumir en un trabajo físico.

2.4 Fases del entrenamiento

- ✓ **Fase de alarma.** Frente al estímulo del organismo se produce una alerta de forma general y se genera la homeostasis celular. Se produce un aumento de las funciones cardiovasculares y metabólicas.
- ✓ **Fase de resistencia.** Si el estímulo continúa, el organismo lucha por volver al equilibrio, para ello se adapta y adquiere un estado de resistencia.
- ✓ **Fase de agotamiento.** Se trastornan los ajustes del individuo disminuyendo su resistencia ante cualquier estímulo.
- ✓ **Fase de readaptación.** El organismo reacciona y sustituye las pérdidas sufridas en las anteriores fases.
- ✓ **Fase de asimilación compensatoria.** La eficacia del entrenamiento está ligada a la reposición o compensación de las energías perdidas durante el esfuerzo.
- ✓ **Fase de restauración ampliada.** Período en que se aumenta la capacidad del esfuerzo.
- ✓ **Umbral.** Existe un umbral por debajo del cual el estímulo no produce respuesta adaptativa. La adaptación funcional se logra como consecuencia de la asimilación de estímulos sucesivamente crecientes.

2.5 Evaluación del estado físico

Para ver el estado físico de un individuo pueden realizarse diferentes tipos de test. Entre los más famosos tenemos:

- ✓ **Test de Cooper (Test de los 12 minutos)**

Objetivo: Valorar la resistencia aeróbica. Determinar el VO₂ máximo

Desarrollo: Recorrer la máxima distancia en el tiempo de 12 minutos en un terreno llano (pista de atletismo). Una carga constante que provoca el agotamiento a los 12 minutos de

iniciarse, correlaciona significativamente con el valor de VO₂ máximo, por lo que obtenemos la siguiente relación:

PARAMETROS PARA HOMBRES			PARAMETROS PARA MUJERES				
CATEGORIA	DISTANCIA (metros)	VO ₂ máx (ml.kg ⁻¹ .min ⁻¹)	CATEGORIA	DISTANCIA (metros)			
				< 30 años	30 a 39 años	40 a 49 años	> 49 años
Muy Baja	< 1.600	< 28,0	Muy Baja	< 1525	< 1375	< 1200	< 1050
Baja	1600 a 2000	28,1 a 34,0	Baja	1525 a 1825	1375 a 1675	1200 a 1525	1050 a 1375
regular	2000 a 2400	34,1 a 42,0	regular	1826 a 2150	1676 a 2000	1526 a 1825	1376 a 1675
Buena	2400 a 2800	42,1 a 52,0	Buena	2151 a 2625	2001 a 2475	1826 a 2300	1676 a 2150
Excelente	> 2800	> 52,0	Excelente	> 2625	> 2475	> 2300	> 2150

CLASIFICACION PARA FONDISTAS (Jorge de Hegeüs)		POTENCIA AEROBICA MAXIMA	
CATEGORIA	DISTANCIA (metros)	ESTIMACION DEL VO ₂ máx	
De Nivel Internacional	> 4000	VO ₂ máx = $\frac{\text{metros} - 504,0941662}{44,78265098}$	
De Nivel Nacional	3500 a 4000		
De Nivel Regional	< 3500		

Cooper, K: *The new aerobics*, New York, M. Evans Co.

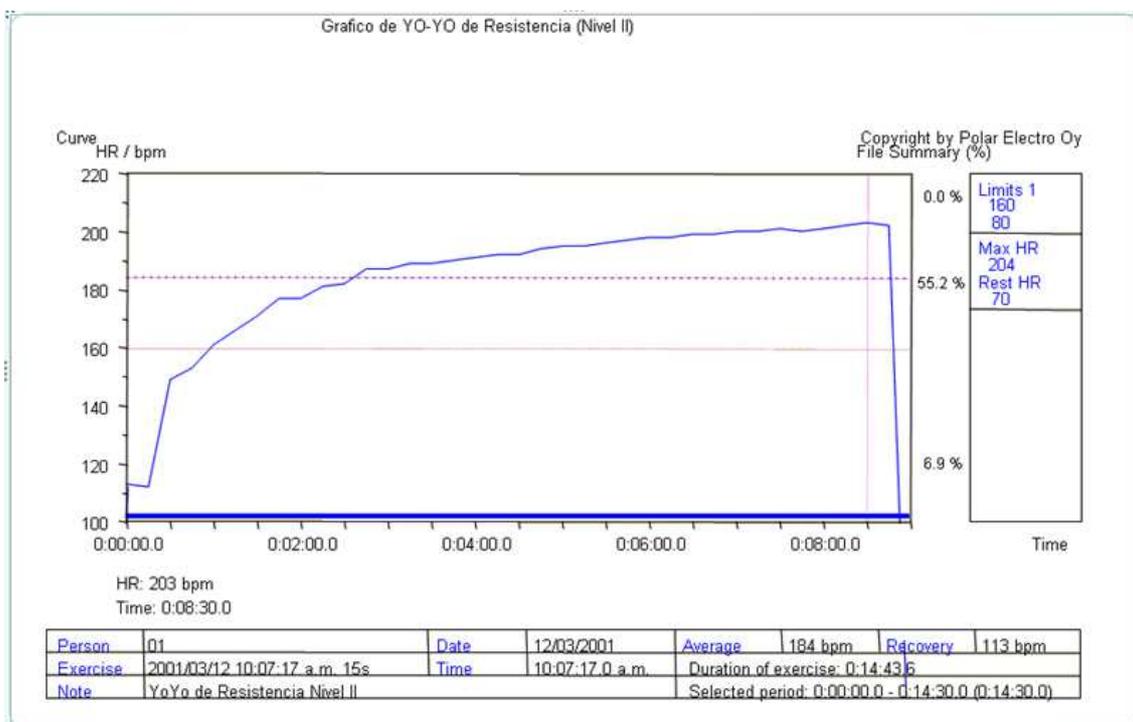
✓ **Yo-yo Test**

Objetivo: Medir el consumo máximo de oxígeno (VO₂ máximo)

Desarrollo: La prueba consiste en realizar carreras de ida y vuelta de 20 metros demarcados por dos líneas que los deportistas deben tocar en el momento de escuchar la señal sonora. Por medio de una grabación se le va marcando al deportista el ritmo de la carrera. Los deportistas deben abandonar la prueba cuando ya no pueden llegar al final de los 20 metros al ritmo que marca la grabación.

Existen 3 variantes de esta prueba:

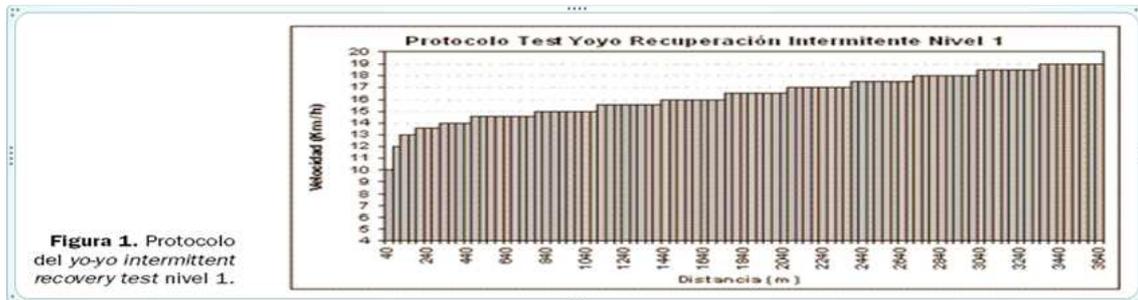
- Resistencia.
- Resistencia intermitente.



Bangsbo, J.: *Yo - Yo Test*, Buenos Aires, ARGENTINA, 1999.

Bangsbo, J.: *La Fisiología de Fútbol*. Copenhagen, DINAMARCA 1993.

➤ Recuperación intermitente.



✓ Test incremental en cinta

Objetivo: Medición indirecta del consumo máximo de oxígeno ante un esfuerzo maximal
Desarrollo: La prueba consiste en correr en una cinta controlando la velocidad. Se comienza a 8 Km/h y se aumenta en 1 Km/h cada 2 minutos. Cuando se llega al agotamiento, se realiza una recuperación poniendo la velocidad a 5 Km/h durante 3 minutos.

PROTOCOLO

En CINTA se comienza a 8 km/h, y se aumenta cada 2 minutos la velocidad en 1km/h.

VALORES REFERENCIALES DE RECUPERACION CARDIACA

calificación	R 1' (%)	R 2' (%)	R 3' (%)
MALA	< 9.4	< 20.6	< 25.7
REGULAR	9.4 a 11.1	20.6 a 23.1	25.7 a 27.8
BUENA	11.2 a 13.4	23.1 a 25.9	27.8 a 29.6
MUY BUENA	13.5 a 15.8	26.0 a 27.9	29.6 a 33.4
EXCELENTE	> 15.8	> 27.9	> 33.4

VALORES DE RECUPERACION CARDIACA

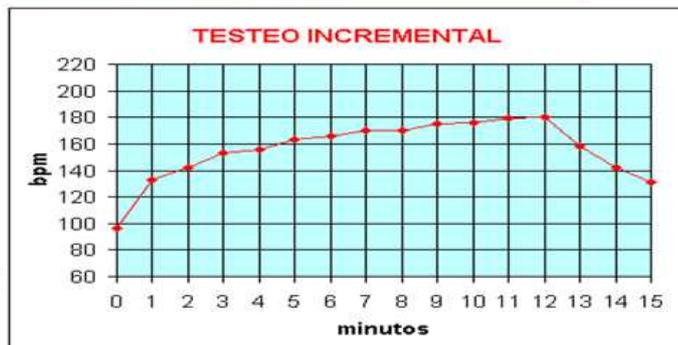
R 1' (%): R 2' (%): R 3' (%):

CONSUMO MAXIMO DE OXIGENO SEGUN PUGH

$VO_2 \text{ máx (ml/kg/min)} = 3.656 \times 13 - 3.99 = 43.5$

VALORES FUNCIONALES

minutos	km/h	bpm
0	0	96
1	8	133
2	8	142
3	9	153
4	9	156
5	10	163
6	10	166
7	11	170
8	11	170
9	12	175
10	12	176
11	13	179
12	13	180
13	5	158
14	5	142
15	5	131



✓ Test de Tokmakidis

Objetivo: Medición indirecta del consumo máximo de oxígeno.
Desarrollo: La prueba consiste en una cierta distancia y medir el tiempo que tarda en recorrerla.

Test general de campo (distancias en atletismo) de Tokmakidis**Evaluación indirecta del VO₂ máx y PO₂ para deportistas**

DISTANCIA metros	CONSUMO MAXIMO DE OXIGENO RELATIVO ml/kg/min
200	[(-3.6300 + (0.77038 X km/h))] X 3.5
400	[(-1.3010 + (0.78600 X km/h))] X 3.5
600	[(0.3862 + (0.79320 X km/h))] X 3.5
800	[(0.8964 + (0.81460 X km/h))] X 3.5
1000	[(1.2730 + (0.83250 X km/h))] X 3.5
1500	[(2.4388 + (0.83430 X km/h))] X 3.5
1609	[(2.5043 + (0.84000 X km/h))] X 3.5
2000	[(2.7297 + (0.85270 X km/h))] X 3.5
3000	[(2.9226 + (0.89000 X km/h))] X 3.5
5000	[(3.1747 + (0.91390 X km/h))] X 3.5
10000	[(4.7226 + (0.86980 X km/h))] X 3.5
15000	[(4.8619 + (0.88720 X km/h))] X 3.5
20000	[(4.9574 + (0.89950 X km/h))] X 3.5
42195	[(6.9021 + (0.82460 X km/h))] X 3.5

EJEMPLO:

Un deportista corrió 3000 metros en 11' 30", pesa 60.5 kg y terminó la prueba con 184 latidos por minuto.

- a) Pasaje del tiempo a segundos = $11 \times 60 + 30 = 690$
- b) Velocidad en m/s = $3000 / 690 = 4.35$
- c) Velocidad en km/h = $4.35 \times 3.6 = 15.7$
- d) Estimación del VO₂ máx en **ml/kg/min** = $[(2.9226 + (0.89000 \times 15.7))] \times 3.5 = \mathbf{59.1}$
- e) Estimación del VO₂ máx en l/min = $59.1 \times 60.5 = 3.6$
- f) Estimación del VO₂ máx en ml/min = $59.1 \times 80.5 = 3577$
- g) Valoración del PO₂ (Pulso de Oxígeno) en ml/sistole = $3577 / 184 = 19.4$

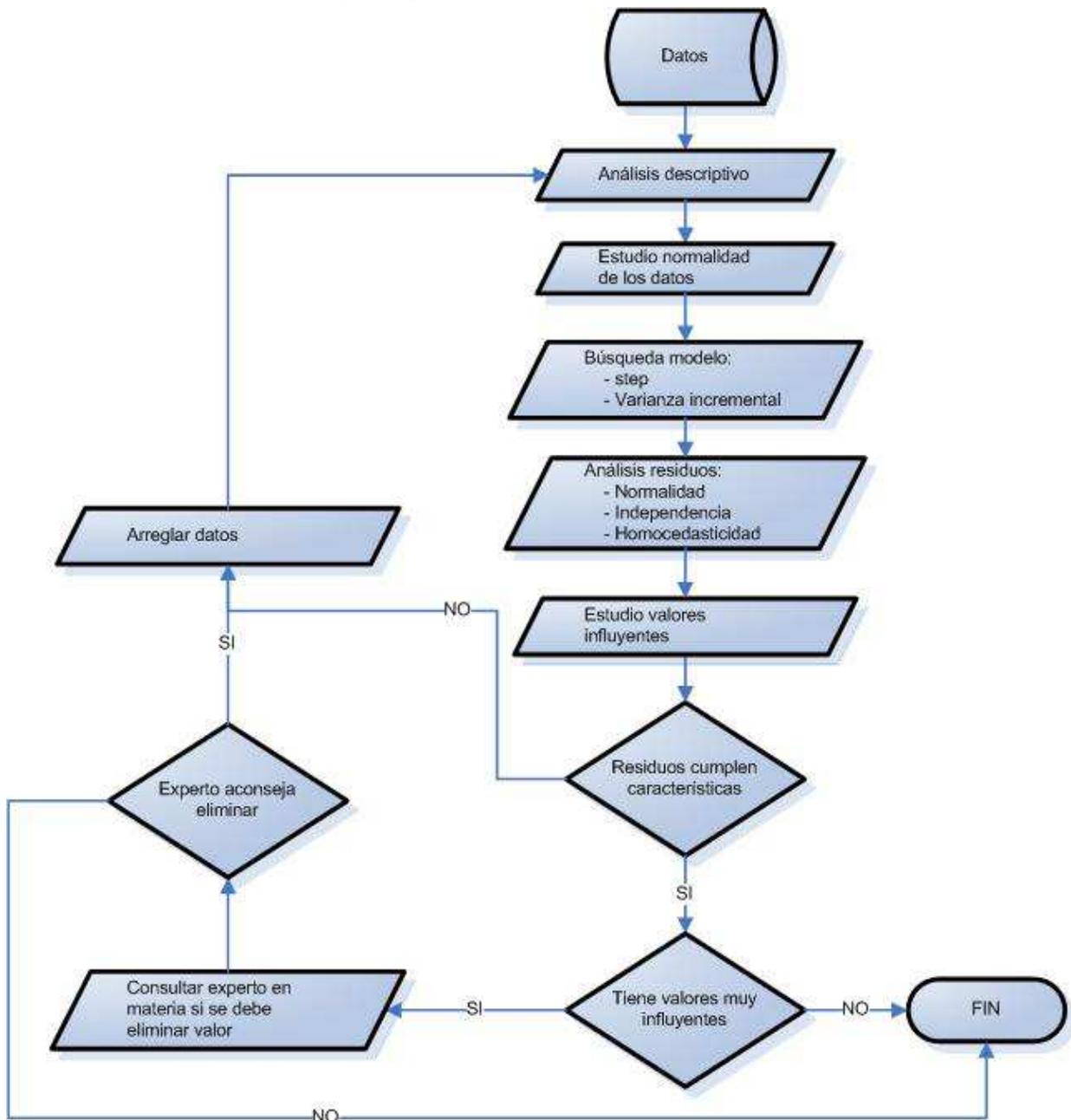
Bibliografía: Pancorbo Sandoval, A.E.: Medicina del deporte y ciencias aplicadas al alto rendimiento y la salud, pp. 215-219, 2002

En este experimento trabajaremos con el test de Cooper por su facilidad para utilizarlo tanto para los corredores como para nosotros, ya que no disponemos del material necesario para usar otros tipos de test, tampoco todo el mundo tiene acceso a una cinta de correr. El test de Cooper solamente necesita de una superficie llana y un cronómetro para medir los 12 minutos.

También es sencillo el cálculo de Cooper a través de los datos resultantes de las carreras, tomando la distancia recorrida y el tiempo empleado para ello.

2.6 Metodología estadística

2.6.1 Metodología general utilizada.



Con los datos obtenidos se realizan un análisis descriptivo con el fin de visualizar posibles correlaciones o cualquier aspecto que pudiera ser relevante en el estudio. Se estudia la normalidad de los datos para ver si es necesaria alguna transformación en los datos o se puede continuar con el estudio tal cual.

Posteriormente se realiza la búsqueda del mejor modelo ya sea usando la metodología step o con varianza incremental.

Sobre el modelo encontrado, se realiza análisis de residuos estudiando si estos cumplen los estándares de normalidad, independencia y homocedasticidad. En caso de no cumplirlos, se deberá estudiar la posibilidad de transformar los datos o ajustarlos de alguna forma para volver a analizar.

Se debe además realizar un estudio de datos influyentes a priori y a posteriori, los datos resultantes, se deben contrastar con un experto en la materia para que nos indique si es coherente que ese valor continúe estando en el estudio o debemos eliminarlo de los datos y volver a analizar.

2.6.2 Modelos lineales generalizados

El objetivo de los MLGz es el estudio de los métodos estadísticos implicados en el análisis de las relaciones entre medidas realizadas en grupos de estudio. Éstos se clasifican en función de cómo sean las variables Respuesta o Explicativas. La idea es conseguir un modelo que permita sustituir los datos por los valores estimados de un modelo $\hat{\mu}$, de tal forma que éstos minimicen algún criterio, ya sea Mínimos Cuadrados, Máxima Verosimilitud, etc.

A la hora de escoger el modelo, éste tiene que estar relacionado con el tipo de datos para que podamos predecir el comportamiento de las variables. Por otro lado, a la hora de seleccionar las variables que deben formar parte de nuestro modelo hay que ser bastante cautelosos, ya que se desea una mejora en el ajuste pero sin sobreestimar el modelo.

2.6.3 Criterio de Akaike

A la hora de buscar el mejor modelo que se adecúe a los datos, se ha aplicado el procedimiento step (paso a paso) del paquete estadístico R. En el paso 1) entra la variable que satisfaga que su AIC sea inferior al AIC del modelo nulo.

$$AIC = 2(-l(\hat{\mu}, y) + p) \quad l \text{ es el máximo de log-verosimilitud}$$

p es el número de parámetros

El Criterio de Información de Akaike equivale a un contraste F para probar si la reducción de varianza aportada por la introducción de nuevos parámetros es significativa. En los sucesivos pasos se plantea si las variables introducidas en el nuevo modelo deben permanecer o no y si puede entrar una nueva variable. El procedimiento finaliza cuando ninguna variable entra o sale, es decir, el AIC es superior al del modelo anterior.

El algoritmo paso a paso tiene las ventajas del algoritmo de introducción progresiva pero lo mejora al no mantener fijas en el modelo las variables que ya entraron en una etapa, evitando de esta forma problemas de multicolinealidad. En la práctica, es un algoritmo bastante utilizado que proporciona resultados razonables cuando se tiene un número grande de variables regresoras.

2.6.4 Análisis de la varianza

El análisis de la varianza es un método para el análisis de datos procedentes del diseño experimental frecuentemente empleado. La extensión de los modelos de regresión lineal para el tratamiento del análisis de la varianza y de la covarianza se denomina Modelo Lineal General.

- Los modelos de análisis de la varianza pueden tratarse mediante regresión lineal (múltiple) aunque siendo cautelosos ya que los modelos ANOVA resultan sobrep parametrizados.
- El tratamiento del análisis de la varianza mediante regresión lineal enfatiza el modelo de Análisis de la varianza.
- La diagnosis y la validación de los modelos ANOVA (y ANCOVA) pueden realizarse como el análisis de los residuos en regresión múltiple, mediante una serie de gráficos que nos permiten detectar posibles anomalías.
- Otro motivo más para utilizar la formulación por regresión es que facilita el trabajo con experimentos con un número de réplicas distintas (diseños no balanceados, o balanceados inicialmente pero con problemas posteriores) o bien tratarse de diseños complejos.

2.6.5 Pruebas de normalidad

Una de las razones por las cuales es importante verificar las condiciones de Normalidad es porque, en ausencia de ésta, el estimador no sesgado de mínima varianza no coincide con el estimador de β por máxima verosimilitud. Los estimadores por mínimos cuadrados no son eficientes, tienen una varianza superior a la varianza de los estimadores MV.

A la hora de verificar la normalidad de la variable respuesta, ésta se puede observar bien sea mediante un histograma de frecuencias o mediante test de Kolmogorov-Smirnov (KS). En caso de ausencia de normalidad, se recomendaría una transformación Box-Cox para corregirla. El test KS compara nuestros valores observados de VO_2Max, y_i , donde $i=1, \dots, n$ con n valores generados de una Normal con μ, σ . Si el p-value es superior a 0.05, entonces no rechazaremos la hipótesis de Normalidad.

La Normalidad de los residuos se puede ver de dos formas: mediante histograma de residuos y mediante un Normal P-P plot.

En la parte Experimental se aplica el test de Shapiro-Wilk en vez de KS por ser más adecuado con tan pocas observaciones como las que se disponen (menos de 50 observaciones). Dicho test permite verificar la ausencia de normalidad dentro de cada grupo de estudio.

La prueba de Shapiro-Wilk pone a prueba la hipótesis nula de que una muestra de x_1, \dots, x_n viene de una distribución normal de la población. Fue publicado en 1965 por Samuel Shapiro y Wilk Martín. [1]

$$W = \frac{\left(\sum_{i=1}^n a_i x_{(i)}\right)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \text{ donde}$$

- $x_{(i)}$ el i -ésimo de menor número en la muestra;
- $\bar{x} = (x_1 + \dots + x_n) / n$ es la media de la muestra;
- las constantes de a_i están dadas por [2]

$$(a_1, \dots, a_n) = \frac{m^T V^{-1}}{(m^T V^{-1} V^{-1} m)^{1/2}} \text{ donde}$$

$$m = (m_1, \dots, m_n)^T$$

y m_1, \dots, m_n son los valores esperados de las estadísticas de orden de variables aleatorias independientes e idénticamente distribuida la muestra de la distribución normal estándar, y V es la matriz de covarianza de los estadísticos de orden.

El usuario puede rechazar la hipótesis nula si W es demasiado pequeño o bien si su p-valor es inferior a 0.05.

2.6.6 Pruebas de hipótesis

Suponiendo que tenemos dos poblaciones distribuidas normalmente con medias desconocidas μ_1 y μ_2 , respectivamente. Se puede aplicar una prueba t de Student para comparar las medias de dichas poblaciones basándonos en dos muestras independientes tomadas de ellas. La primera muestra es de tamaño m , con media \bar{x}_1 y varianza s_1^2 y la segunda muestra es de tamaño n , con media \bar{x}_2 y varianza s_2^2 .

Si las varianzas de las poblaciones son iguales $\sigma_1^2 = \sigma_2^2 = \sigma^2$ entonces se puede mostrar que $\frac{\bar{x}_1 - \bar{x}_2 - (\mu_1 - \mu_2)}{\sqrt{s_p^2 \left(\frac{1}{m} + \frac{1}{n}\right)}}$ se distribuye como una t con $m+n-2$ grados de libertad. En

este caso la varianza poblacional σ^2 es estimada por una varianza combinada de las varianzas de las dos muestras tomadas, dada por la siguiente fórmula: $s_p^2 = \frac{(m-1)s_1^2 + (n-1)s_2^2}{m+n-2}$

Un intervalo de confianza del 100 $(1-\alpha)$ % para la diferencia $\mu_1 - \mu_2$ de las medias poblacionales será de la forma: $\bar{x} - \bar{y} \pm t_{(\alpha/2, n+m-2)} s_p \sqrt{\frac{1}{m} + \frac{1}{n}}$

Las fórmulas para las pruebas de hipótesis son las siguientes:

Caso I

$$H_0: \mu_1 = \mu_2$$

$$H_a: \mu_1 < \mu_2$$

Caso II

$$H_0: \mu_1 = \mu_2$$

$$H_a: \mu_1 \neq \mu_2$$

Caso III

$$H_0: \mu_1 = \mu_2$$

$$H_a: \mu_1 > \mu_2$$

Prueba estadística: $t = \frac{(\bar{x} - \bar{y})}{s_p \sqrt{\frac{1}{m} + \frac{1}{n}}}$ con $m+n-2$ grados de libertad

Decisión:

- si $t_{cal} < -t_{\alpha}$ entonces se rechaza H_0
- si $t_{cal} < t_{\alpha/2}$ o $t_{cal} > t_{1-\alpha/2}$ entonces se rechaza H_0
- si $t_{cal} > t_{1-\alpha}$ entonces se rechaza H_0

2.6.7 Análisis de los residuos

Se define el residuo asociado a una observación muestral como:

$$e_i = y_i - \hat{y}_i \sim N(0; \sigma^2(1 - h_{ii})), \quad i = 1, \dots, n,$$

Como $\sigma^2(e_i)$ no es constante, es usual tipificar los residuos y se definen los residuos estandarizados.

$$r_i = \frac{e_i}{s_{R_i} \sqrt{1 - h_{ii}}} \sim N(0; 1), \quad i = 1, \dots, n.$$

Para evitar la dependencia entre numerador y denominador de la expresión anterior, también se utilizan los residuos estudentizados

$$t_i = \frac{e_i}{s_{R_i(i)} \sqrt{1 - h_{ii}}}, \quad i = 1, \dots, n.$$

Donde $s_{R_i(i)}^2$ es la varianza residual calculada sin el dato (x_i, Y_i) .

El análisis descriptivo y gráfico de los residuos estandarizados da una buena idea acerca de si se verifican o no las hipótesis del modelo de regresión. Gráficos de interés son: el histograma y el gráfico de residuos frente a predicciones.

2.6.8 Detección de valores influyentes a priori y a posteriori

Resulta fácil desarrollar ejemplos que ponen de manifiesto que existen observaciones que tienen mucha mayor influencia en las propiedades del modelo que otras, hasta el extremo que en presencia de, por ejemplo, 100 valores observados, las propiedades de los estimadores dependan únicamente de unos pocos de esos valores.

- Este aspecto está relacionado con la fiabilidad del modelo en la realización de predicciones, y parece más conveniente un modelo que venga avalado por la totalidad de la muestra empleada para su estimación que no aquel otro que sólo dependa de unas pocas observaciones.
- El estudio de los valores influyentes a priori determinará la robustez del diseño de recogida de los datos y el estudio de los valores influyentes a posteriori determinará la robustez de los parámetros estimados.
- No hay que confundir observaciones influyentes con residuos atípicos: una observación influyente puede tener o no un residuo estadísticamente grande, y viceversa, un residuo atípico no implica que la observación sea influyente.

Se define el valor de influencia a priori (leverage) de la observación (x_i, Y_i) por

$h_{tt} = \frac{1}{n} \left(1 + \frac{(\bar{x}_t - \bar{x})^2}{s_x^2} \right)$, $t=1, \dots, n$, h_{tt} es una medida de la distancia del punto x_i al centro y es un valor comprendido entre $1/n$ y 1. Se verifica que $E(h_{tt}) = \frac{2}{n}$ y se puede considerar que

una observación tiene un valor de influencia grande si se verifica que $h_{tt} > \bar{n}$. Las observaciones con residuos estandarizados grandes se denominan datos atípicos.

Una observación influyente a posteriori es aquella cuya inclusión modifica sustancialmente la estimación del modelo. Es importante identificar y estudiar las observaciones influyentes a posteriori. Para ello se utiliza el estadístico Distancia de Cook

$D(i) = \frac{r_i^2}{2} \left(\frac{h_{ii}}{1-h_{ii}} \right)$, $i=1, \dots, n$ donde r es el i -ésimo residuo estandarizado y h_{ii} es el valor de influencia a priori de la i -ésima observación. Bajo la hipótesis nula, no es una observación influyente a priori no es una observación influyente a posteriori se verifica que

$$D(i) \sim F_{2;n-2}, \quad i = 1, \dots, n,$$

$$DFFITS(i) = \left(\frac{h_{ii}}{1-h_{ii}} \right)^{1/2} t_i, \quad i = 1, \dots, n,$$

Otro estadístico análogo es el siguiente:

El estadístico DFFITS (i) se considera significativo si es mayor que $2/\sqrt{n}$.

2.6.9 Buscar mejor modelo

La importancia de X_i y X_j cuando se usan como predictoras separadamente está ligada a su coeficiente de correlación con la variable dependiente, además juegan las correlaciones entre los predictores.

$$Y_i = \beta_1 + \beta_2 X_{2,i} + \dots + \beta_p X_{p,i} + \varepsilon_i \text{ on } \varepsilon_i \approx N(0, \sigma^2 I) \quad \text{Independientes}$$

Para escoger la mejor ecuación se requiere del análisis de residuos:

- Histograma de los residuos y/o normal QQ Plot de los residuos.
- Diagrama bivalente de los residuos estandarizados vs \hat{Y} .
- Diagrama bivalente de los residuos estandarizados vs X_i .
- Detección de observaciones influyentes a priori y a posteriori

Las transformaciones para mejorar las propiedades de la regresión se deben introducir si el análisis de los residuos indica la necesidad y valorar la dificultad introducida en la interpretación del modelo.

Las observaciones influyentes posiblemente se tienen que suprimir y repetir el proceso. Un residuo estandarizado es un posible outlier si tiene magnitud absoluta superior a 2-3 en función de su tamaño muestral.

Una distancia de Cook es elevada si se encuentra por encima de 0,5 y 1 según el número de estimadores (p) y tamaño de la muestra (n). Hay autores que prefieren obviar la interpretación a partir de la distribución teórica y considerar solamente como observaciones influyentes a posteriori las que tienen una distancia de Cook inusualmente elevada respecto de las magnitudes del resto de observaciones. Chatterjee i Hadi proponen emplear el límite $4/(n-p)$.

Un leverage (h_{ii}) se considera elevado si excede el valor $2p/n$. El máximo teórico es 1.

2.6.10 Contraste de varianza incremental

Este contraste, nos sirve, al igual que el procedimiento Step, para buscar un modelo que sea factible para nuestros datos.

Sea $Y = \mu + \varepsilon = X\beta + \varepsilon$ modelo con variable respuesta continua y Normal, donde Y , $\eta = \mu$ son $n \times 1$, X es $n \times p$ de rango y β es $p \times 1$ y los errores son no sesgados y no correlacionados de varianza constante. Los estimadores ordinarios por mínimos cuadrados se denotan como $\hat{\beta}$.

Sea $Y = \mu + \varepsilon = X\beta + \varepsilon$ modelo con variable respuesta continua y Normal, donde Y , $\eta = \mu$ son $n \times 1$, X es $n \times p$ de rango y β es $p \times 1$ y los errores son no sesgados y no correlacionados de varianza constante sujeto a un conjunto de restricciones lineales $A\beta = c$ que definen una hipótesis a contrastar denominada H_0 , donde A es una matriz $q \times p$ de rango $q < p$. Los estimadores ordinarios por mínimos cuadrados se denotan como $\hat{\beta}_H$.

El contraste de la hipótesis H_0 se puede realizar a partir de un estadístico distribuido según la ley de Fisher y relacionado con la variación de la suma de cuadrados residual y los grados de libertad del modelo. A dicho contraste se le llama contraste de la varianza incremental. Dicho contraste nos permite ver, en otras palabras, si añadir una nueva variable a nuestro modelo su aporte es tan significativo que reduce bastante la varianza residual aunque perdamos algún grado de libertad.

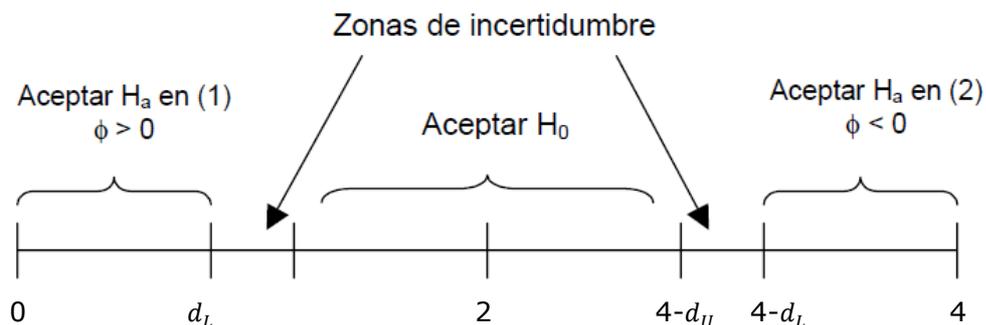
$$\text{Test V.I } F = \frac{(SCR_H - SCR)/q}{SCR/(n - P)} \approx F_{p-1, n-p}$$

2.6.11 Test de Durbin-Watson o ACF

Dado el modelo de regresión lineal múltiple $Y = \beta_1 + \beta_2 \cdot X_2 + \dots + \beta_k X_k + u_t$, el test de Durbin-Watson permite contrastar si el término de perturbación está autocorrelacionado según un esquema AR (1), i.e., la hipótesis nula indica que si el término de perturbación es de la forma $u_t = \phi \cdot u_{t-1} + \varepsilon_t$ entonces necesariamente $\phi = 0$ (con lo que no habría autocorrelación según esquema AR (1)):

$$(1) \begin{cases} H_0: u_t \approx AR(1) \text{ con } \phi = 0 \\ H_1: u_t \approx AR(1) \text{ con } \phi > 0 \end{cases} \quad (2) \begin{cases} H_0: u_t \approx AR(1) \text{ con } \phi = 0 \\ H_1: u_t \approx AR(1) \text{ con } \phi < 0 \end{cases}$$

El estadístico que se utiliza para realizar dicho test es el estadístico Durbin-Watson (el cual obtendremos con ayuda de R mediante paquete lmtest). El estadístico DW es un valor comprendido entre 0 y 4. Valores "próximos" a 2 nos dicen que no debemos rechazar la hipótesis nula. Por el contrario, para valores de DW "alejados" de 2, sí rechazaremos H_0 (i.e., aceptaremos la existencia de autocorrelación de tipo AR (1)):



En el gráfico anterior, d_L y d_U son valores tabulados que dependen del número de observaciones (n), del número de regresores (k), y del nivel de significación (α).

Añadir, finalmente, que si $0 < DW < 2$, el coeficiente ϕ será positivo (estaremos en el contraste unilateral (1)), mientras que si $2 < DW < 4$, dicho coeficiente ϕ será negativo (estaremos en el contraste (2)).

El test de Durbin-Watson presenta una serie de problemas:

- Sólo es útil para contrastar si el término de perturbación está autocorrelacionado según un esquema AR (1). El test no serviría para detectar otros tipos de autocorrelación.
- Las cotas d_L y d_U sólo son válidas si el modelo de regresión inicial contiene término independiente (que habitualmente hemos denotado por β_1)
- Presenta "zonas de incertidumbre", de forma que si el estadístico DW se encuentra entre d_L y d_U , o entre $4 - d_U$ y $4 - d_L$, no podremos concluir nada.
- Sólo es válido si todos los regresores del modelo son deterministas, es decir, si el modelo incluye un regresor que dependa del término de perturbación u_{t-1} .

Dado un proceso puramente aleatorio $\{Z_t\}$ con media 0 y varianza σ_Z^2 , diremos que un proceso $\{X_t\}$ es un proceso autoregresivo de orden 1, AR (1), si $X_t = \phi X_{t-1} + Z_t$ donde ϕ es

una constante. Siempre que $|\phi| < 1$, la media del proceso es nula y su varianza es finita, positiva y constante. También cumple que su función de autocovarianza entre X_{t_1} y X_{t_2} depende únicamente del intervalo t_1 y t_2 y no de t . Es decir, es un proceso estacionario de segundo orden. La función de autocorrelación ACF, decrece de forma exponencial según el signo del parámetro del modelo. En el caso que se tuviese una serie no estacionaria, que por ejemplo, presente una tendencia, la autocorrelación muestral presenta una disminución muy lenta ya que cada una de sus observaciones toma un valor similar a la anterior.

2.7 Estado del arte

2.7.1 Índice de dureza

Existe un cálculo del índice de dureza de las rutas de bicicleta (IBP). Este índice sirve para darle al ciclista una idea clara de la dificultad de la ruta antes de realizarla.

HISTORIA: El índice nace en el año 2003, cuando un grupo de aficionados al ciclismo de montaña y de carretera empieza a guardar los tracks de sus salidas grabados con GPS en una hoja de cálculo. La posición en 4 dimensiones en numerosos puntos de una ruta da una cantidad de información que se empieza a comparar. Observando estos datos surge la idea de encontrar un índice para valorar la dureza de una salida en bicicleta.

Se barajan 19 diferentes parámetros y se empiezan a combinar, valorar y modificar para obtener una primera puntuación de referencia. Durante unos meses se comparan las sensaciones de las salidas que se van realizando con la puntuación que se obtiene con la puntuación obtenida y se van corrigiendo los factores para adaptarlos.

De esa forma se obtiene una primera fórmula y en los siguientes 3 años se va modificando en función de los valores medios que se van obteniendo de las salidas realizadas.

A finales de 2006 un equipo técnico desarrolla la programación necesaria para poner a disposición de la comunidad ciclista la posibilidad de utilizar el índice IBP a través de una página web www.ibpindex.com y una aplicación de escritorio. En pocos meses pasa a ser una aplicación muy valorada entre los ciclistas convirtiéndose en un índice de referencia para una gran cantidad de clubs.

ASPECTOS TÉCNICOS: el índice IBP es una puntuación que valora el grado de dificultad de una ruta recorrida en bicicleta de montaña o de carretera. Se obtiene analizando los datos de latitud, longitud y altura en multitud de puntos del recorrido. Estos puntos se obtienen del track grabado con un GPS, que los recoge de manera automática con una cadencia entre 20 y 40 puntos por kilómetro en función de las variaciones del terreno, los cambios de dirección y el tipo de GPS. A partir de estos puntos se calculan las distancias recorridas en los diferentes tramos de subidas y bajadas (1%, 5%, 10%, etc...) se computan los % que representan sobre el total, los metros totales ascendidos, descendidos, los ratios medios de subida y de bajada, los kilómetros totales y la distribución de los tramos de subida, también se tiene en cuenta los cambios de dirección vertical y horizontal y la altitud por donde discurre el recorrido. Se aplica una función matemática no lineal y se obtiene una puntuación entre 0 e infinito. En la fórmula intervienen 19 parámetros que fueron ajustados durante un período de 3 años, analizando y comparando multitud de rutas por un equipo de ciclistas.

El índice es útil para poder conocer el grado de dificultad de diferentes rutas comparándolas entre sí, al no tener en cuenta factores subjetivos, el valor será significativo para cada cual.

Ejemplo: un IBP de 60 será una ruta muy dura para un no iniciado, normal para alguien con una preparación media o muy fácil para un profesional.

Sufrimiento de una ruta = estado de forma físico + condiciones climatológicas
+ ritmo de la carrera + IBP

ADAPTACIÓN A CARRERAS DE ATLETISMO: del mismo modo que es posible contar un índice de dureza para las salidas de bicicleta, debe ser posible encontrar un índice de dureza para carreras de atletismo tanto de montaña como de asfalto. En este proyecto intentaremos realizar una primera aproximación, aunque adaptada a la cantidad de datos de que disponemos.

Mediante este índice de dureza y teniendo en cuenta las características del atleta, se puede cumplir uno de los objetivos propuestos en el proyecto que se trata de prever el comportamiento de cada una de las carreras. De momento y por falta de datos y tiempo no se tendrán en cuenta los datos del track de la carrera.

2.7.2 Ejercicios respiración.

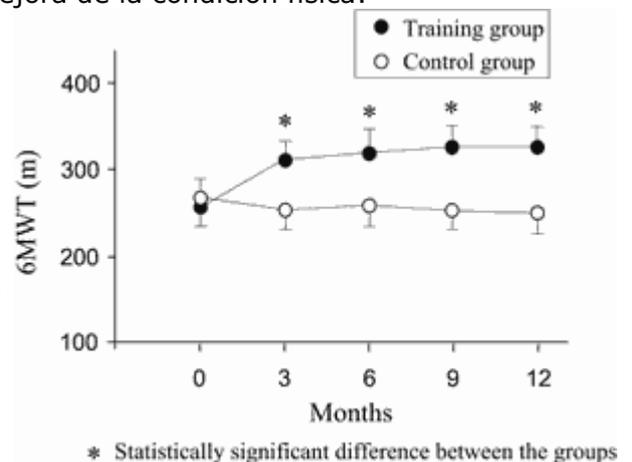
Está demostrado que la utilización de un aparato respirador (espirómetro) genera un entrenamiento de la musculatura respiratoria, lo cual puede ayudar al usuario a mejorar su respiración tanto en personas sedentarias como personas afectadas por una afección respiratoria.

Se publicó un artículo en la revista Chest titulado "Los efectos de 1 año de entrenamiento específico de la musculatura inspiratoria en pacientes con BNCO" [consultar](#) realizado por M. Beckerman y colaboradores realizado con personas afectas de Bronconeumopatía Crónica Obstructiva (con evidencia espirométrica de limitación crónica al flujo de aire). En este se comparan los resultados de un grupo que realiza entrenamiento inspiratorio en relación a un grupo control.

El entrenamiento respiratorio consistía en 2 sesiones de 15 minutos al día, seis días por semana durante 12 meses. El grupo de entrenamiento realizaba los ejercicios con un aparato en el que iban aumentando la resistencia a medida que avanzaba el estudio, mientras que el grupo de control realizaba los mismos ejercicios con un aparato que no ofrecía apenas resistencia.

En base a los diferentes test realizados cada 3 meses a todos los participantes en el estudio se obtiene entre otros resultado de mejora de la condición física:

Para valorar la condición física, los pacientes realizaban el llamado 6MWT que consiste en recorrer la máxima distancia posible en 6 minutos. En la imagen adjunta se observa la mejora significativa de la distancia recorrida en 6 minutos que se da en el grupo de pacientes que realizan entrenamiento de la musculatura inspiratoria, mientras que no hay diferencias en el grupo control. Hay que significar, que si bien se produce una mejora importante en los 3 primeros meses de entrenamiento, la mejora (aunque pequeña como es lógico) se mantiene durante los 12 meses del estudio.



Dada la dificultad para la utilización de espirómetros para el experimento planteado en este proyecto, se plantea la mejora mediante unos ejercicios realizados sin utilización de aparato, a base de la realización de unos ejercicios de respiración.

2.7.3 Ejercicios equilibrio muscular.

El desequilibrio muscular es la situación en la que algunos músculos se inhiben y debilitan mientras otros quedan acortados y tensos perdiendo elasticidad. Para ver el grado de no simetría de los músculos se deben tener en cuenta la postura, simetría, modo de andar, longitud del músculo y patrones de movimiento. Esto trae consigo consecuencias como alteraciones de los mecanismos articulares, amplitud limitada de los movimientos, etc....

En atletismo, sobre todo para controlar la velocidad es imprescindible controlar el estado de la musculatura. Teniendo en cuenta el entrenamiento y características del atleta, podríamos prever la mejora que tendrá el atleta que era uno de los objetivos propuestos en el proyecto.

3 Material y métodos

3.1 Software

El software utilizado durante el proyecto es:

- Actualmente existen muchos tipos de pulsímetros que ayudan a los atletas a tomar medidas sobre su estado físico:
 - ✓ Polar
 - ✓ Suunto
 - ✓ Garmin
 - ✓ etc....

Estos aparatos suelen obtener un fichero donde se indican diversas variables obtenidas durante el entrenamiento a intervalos concretos. Los ficheros obtenidos en este proyecto son provenientes de pulsímetros polar y son ficheros en formato HRM.

- Paquete R para el tratamiento estadístico de los datos
- Microsoft Excel para la toma de datos y cálculos auxiliares como edad, Cooper, VO₂Max,... También utilizamos el Excel para la interpretación de los ficheros HRM provenientes de los pulsímetros.

3.2 Individuos

3.2.1 Individuos para el experimento de afectación de entrenamiento.

Se plantea la búsqueda de individuos que formen parte de nuestro experimento de forma voluntaria, para ello confeccionamos un formulario que el voluntario deberá rellenar con sus datos, así como un folleto explicativo de los experimentos planteados. Se realiza la visita a distintos clubs deportivos para hablar con el responsable, explicarle en qué consiste el experimento e intentar que los atletas que él entrena participen. Con este fin se visitan clubs como:

- Club Natació Terrassa
- Poliesportiu Bonaire (Terrassa)
- Gavà
- Viladecans
- Canaletes
- Cornellà
- ...

Aparte de intentar captar voluntarios entre conocidos no pertenecientes a ningún club e intentar que cuelguen la información en páginas de aficionados como 10de1000.

Se prepara un cartel que se cuelga en diferentes clubs por si alguna persona, aparte de los que entrenan con los responsables con los que se habla estuviese interesada.

La idea es captar atletas de diferentes edades (mayores de edad), sexo, grado de entrenamiento y en muestra representativa para realizar 3 grupos:

- Grupo que realice de forma adicional entrenamiento respiratorio.
- Grupo que realice de forma adicional entrenamiento de equilibrio muscular
- Grupo de control que no realice ningún entrenamiento adicional

El intento de buscar voluntarios suficientes fracasa pese al gran esfuerzo invertido, al conseguir solamente la participación del Club Natació Terrassa que aporta 6 corredores, aunque nos comentan que alguno más tenía intención de realizarlo pero no ha podido ser debido a lesiones o alguna circunstancia personal que le impide entrenar durante una temporada. A éstos sumamos 6 atletas entre los conocidos donde también hubiese participado alguna persona más que por lesiones finalmente no pudieron hacerlo. Total 12 corredores.

Muchos atletas no participaron en el estudio debido a que les resultaba molesto añadir a su rutina la realización de un esfuerzo adicional aunque fuese la de anotar los datos resultantes del entrenamiento para que se añada en el estudio, además de medir su forma física mediante un test de Cooper y estar pendiente de cuando llega el final del experimento para volver a realizar ese test de Cooper

De los corredores participantes, uno de ellos fue descartado por detectar que la frecuencia cardíaca que indica es en todos los entrenamientos la misma. De algún otro corredor tenemos datos missing por ejemplo porque saliera a correr sin llevar pulsímetro o no indicaron la distancia recorrida, por lo que no nos puede indicar la frecuencia cardíaca media de esa sesión.

El hecho de haber conseguido tan pocos corredores hace que para el estudio planteado no sea una muestra suficientemente representativa, por lo que los resultados obtenidos no son concluyentes.

3.2.1.1 Completar individuos para experimento

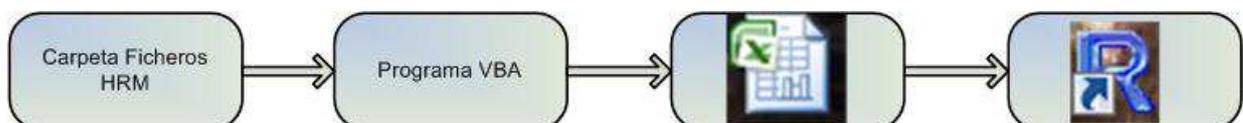
Como se ha mencionado en el punto anterior, el número de individuos a estudiar en experimentos es insuficiente, así que con afán de mejorar este punto en la medida de lo posible obtenemos datos de algunos corredores más que usaremos como el grupo de control. Estos datos los obtendremos a través del estudio de ficheros con formato HRM obtenidos a través de su pulsímetro durante los entrenamientos. El formato de los datos queda explicado en el documento de Polar (20/07/2004) (ver 6.4).

Como variables de los ficheros tendremos:

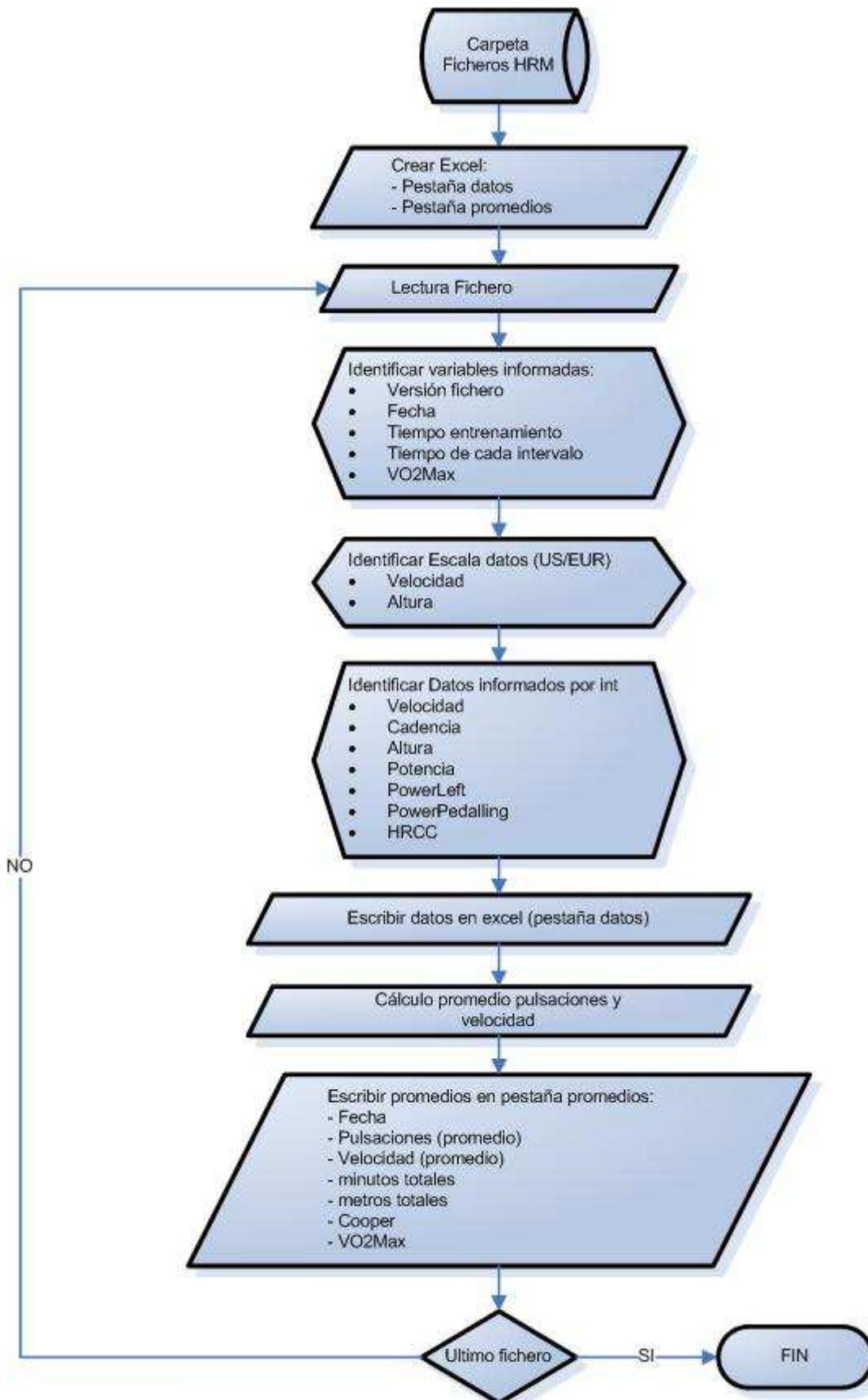
Speed (0=off, 1=on)
Cadence (0=off, 1=on)
Altitude (0=off, 1=on)
Power (0=off, 1=on)
Power Left Right Balance (0=off, 1=on)
Power Pedalling Index (0=off, 1=on)
HR/CC data 0 = HR data only, 1 = HR + cycling data
US / Euro unit / 0 = Euro (km, km/h, m, °C) / 1 = US (miles, mph, ft, °F)

La lectura de estos datos se ha controlado mediante un programa VBA.

3.2.1.1.1 Tratamiento datos ficheros HRM.



A partir de una carpeta contenedora de archivos con formato HRM obtendremos los datos en un Excel de modo que habremos conseguido un fichero para su estudio en R.



El programa leerá un archivo y buscará la información referente a:

- Versión fichero. Donde nos indica los datos que se están informando de cada intervalo (velocidad, altura, cadencia, ...)
- Fecha de entrenamiento
- Tiempo entrenamiento
- Tiempo de cada intervalo informado en el fichero
- VO₂Max

Se debe identificar la escala de los datos (US / EUR) para interpretación de velocidad y altura.

Se debe averiguar (mediante la versión) los datos que vienen informados en cada intervalo.

- Velocidad
- Cadencia
- Altura
- Potencia
- PowerLeft
- PowerPedalling
- HRCC

En la pestaña de datos del Excel se escribirán todos los datos provenientes del fichero. Se calculan los promedios de pulsaciones y de velocidad.

En la pestaña promedios se escribirá una línea por cada fichero donde se informará:

- Fecha
- Pulsaciones (promedio)
- Velocidad (promedio)
- Minutos totales
- Metros totales (Calculado a partir de la velocidad)
- Cooper (Calculado a partir de los metros y los minutos)

```

Sub Importacio()
Dim Mode           As String
Dim Speed          As Boolean
Dim Cadence        As Boolean
Dim Altitude       As Boolean
Dim Power          As Boolean
Dim PowerLeft      As Boolean
Dim PowerPedalling As Boolean
Dim HRCC           As Boolean
Dim US_EUR         As String '0 = Euro (km, km/h, m, °C)
                        '1 = US (miles, mph, ft, °F)

Dim File           As String
Dim xFile          As String
Dim Valores        As Byte
Dim xColumna      As String
Dim Columna       As String
Dim Fila          As Long
Dim FilaDatos     As Long
Dim fso            As Object
Dim carpeta       As Object
Dim fichero       As Object
Dim strPath       As String
Dim filaFicheroDatos As Long
Dim filaMedias    As Integer
Dim avgPulsaciones As Double
Dim avgVelocidad  As Double
Dim data          As String
Dim tempsTotal    As String
Dim interval      As String

```

```

'Indicar la carpeta donde tenemos los ficheros de datos
strPath = "D:\UPC\Projecte\Dades\HRM(5)\07123103"
Workbooks.Open Filename:=strPath + "\Datos.xlsx"

```

```
Set fso = CreateObject("Scripting.FileSystemObject")
Set carpeta = fso.GetFolder(strPath)
filaFicheroDatos = 1
filaMedias = 1

For Each fichero In carpeta.Files
    File = strPath + "\" + fichero.Name

    xFile = fichero.ShortName
    xFile = Left(xFile, Len(xFile) - 4)
    xColumna = "EFGHIJKLMNOP"
    'Carregar arxiu
    Workbooks.OpenText Filename:=File, DataType:=xlDelimited, Tab:=True

    Mode = Right(Range("A4").Value, 9)
    data = Right(Range("A5").Value, 8)
    tempsTotal = Right(Range("A7").Value, 10)
    interval = Right(Range("A8").Value, 1)
    Valores = 1
    Sheets.Add
    Sheets("Hoja1").Name = "datos"
    Range("A1").Value = "Fecha"
    Range("B1").Value = "Intervalo"
    Range("C1").Value = "Tiempo Total"
    Range("D1").Value = "Tiempo Intervalo"
    If Mid(Mode, 8, 1) = 1 Then
        US_EUR = "US"
    Else
        US_EUR = "EUR"
    End If
    If Mid(Mode, 1, 1) = 1 Then
        Speed = True
        Valores = Valores + 1
        Columna = Mid(xColumna, Valores, 1)
        If US_EUR = "EUR" Then
            Range(Columna & "1").Value = "Speed_Kmh"
        Else
            Range(Columna & "1").Value = "Speed_mph"
        End If
    Else
        Speed = False
    End If
    If Mid(Mode, 2, 1) = 1 Then
        Cadence = True
        Valores = Valores + 1
        Columna = Mid(xColumna, Valores, 1)
        Range(Columna & "1").Value = "Cadence"
    Else
        Cadence = False
    End If
    If Mid(Mode, 3, 1) = 1 Then
        Altitude = True
        Valores = Valores + 1
        Columna = Mid(xColumna, Valores, 1)
        If US_EUR = "EUR" Then
            Range(Columna & "1").Value = "Altitude_m"
        Else
            Range(Columna & "1").Value = "Altitude_ft"
        End If
    Else
        Altitude = False
    End If
    If Mid(Mode, 4, 1) = 1 Then
        Power = True
        Valores = Valores + 1
        Columna = Mid(xColumna, Valores, 1)
        Range(Columna & "1").Value = "Power"
    Else
        Power = False
    End If
    If Mid(Mode, 5, 1) = 1 Then
        PowerLeft = True
        Valores = Valores + 1
    End If
End For
```

```

        Columna = Mid(xColumna, Valores, 1)
        Range(Columna & "1").Value = "PowerLeft"
    Else
        PowerLeft = False
    End If
    If Mid(Mode, 6, 1) = 1 Then
        PowerPedalling = True
        Valores = Valores + 1
        Columna = Mid(xColumna, Valores, 1)
        Range(Columna & "1").Value = "PowerPedalling"
    Else
        PowerPedalling = False
    End If
    If Mid(Mode, 7, 1) = 1 Then
        HRCC = True
        Valores = Valores + 1
    Else
        HRCC = False
    End If
    Range("E1").Value = "Pulsaciones"
    Fila = 1
    While Worksheets(xFile).Range("A" & Fila).Value <> "[HRData]"
        Fila = Fila + 1
    Wend
    FilaDatos = 1
    avgPulsaciones = 0
    avgVelocidad = 0
    While Worksheets(xFile).Range("A" & Fila).Value <> ""
        Fila = Fila + 1
        FilaDatos = FilaDatos + 1
        Worksheets("datos").Range("A" & FilaDatos).Value = data
        Worksheets("datos").Range("B" & FilaDatos).Value = FilaDatos - 1
        Worksheets("datos").Range("C" & FilaDatos).Value = "" + tempsTotal
        Worksheets("datos").Range("D" & FilaDatos).Value = interval
        Worksheets("datos").Range("E" & FilaDatos).Value = Worksheets(xFile).Range("A" & Fila).Value
        If (Worksheets("datos").Range("E1").Value = "Pulsaciones") Then avgPulsaciones = avgPulsaciones +
Worksheets("datos").Range("E" & FilaDatos).Value
        Worksheets("datos").Range("F" & FilaDatos).Value = Worksheets(xFile).Range("B" & Fila).Value
        If (Worksheets("datos").Range("F1").Value = "Speed_Kmh") Then avgVelocidad = avgVelocidad +
Worksheets("datos").Range("F" & FilaDatos).Value
        Worksheets("datos").Range("G" & FilaDatos).Value = Worksheets(xFile).Range("C" & Fila).Value
        Worksheets("datos").Range("H" & FilaDatos).Value = Worksheets(xFile).Range("D" & Fila).Value
    Wend

    'Calculo de promedios del fichero
    avgPulsaciones = avgPulsaciones / (FilaDatos - 2)
    avgVelocidad = avgVelocidad / (FilaDatos - 2)

    'Copiar los datos en un unico fichero de datos
    Range("A1:H" & (FilaDatos - 1)).Select
    Selection.Copy
    Windows("Datos.xlsx").Activate
    Worksheets("Datos").Range("A" & filaFicheroDatos).Select
    ActiveSheet.Paste
    'Los promedios los copiaremos en otra pestaña
    Worksheets("Promedios").Range("A" & filaMedias).Value = data
    Worksheets("Promedios").Range("B" & filaMedias).Value = avgPulsaciones
    Worksheets("Promedios").Range("C" & filaMedias).Value = avgVelocidad
    Worksheets("Promedios").Range("D" & filaMedias).Value = tempsTotal

    filaFicheroDatos = filaFicheroDatos + FilaDatos - 1
    filaMedias = filaMedias + 1

    'ActiveWorkbook.SaveAs Filename:=Left(File, Len(File) - 4)
    Windows(xFile & ".hrm").Activate
    ActiveWindow.Close

    Next fichero

End Sub

```

3.2.2 Individuos para la determinación del grado de dureza de una prueba atlética.

Obtenemos una muestra de corredores suficientemente amplia de una base de datos existente en la web <http://www.eolymp.cat>. Estos datos se bajan de la web en formato plano y se transforma para su estudio.

3.2.2.1 Datos técnicos carreras

Para poder relacionar los datos obtenidos con las carreras en las que han participado, realizamos la búsqueda de los siguientes datos:

- Distancia
- Tipo de suelo
- Desnivel

Para ello se realizan búsquedas por internet. En muchos casos las páginas consultadas provienen de datos informados por atletas, obtenidos de medidas con GPS. Durante la búsqueda era muy frecuente encontrar varios usuarios que informando de la misma carrera, discrepen en sus datos, por lo que hubo que contrastar la información para quedarnos con la más fiable.

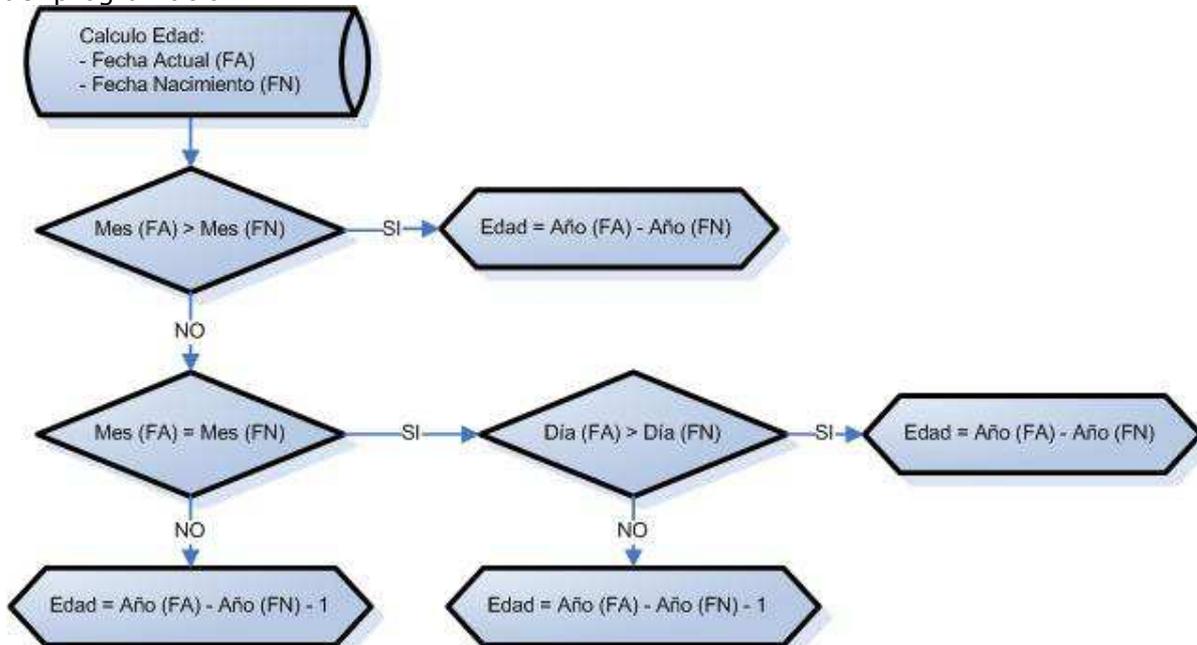
3.2.2.2 Datos facilitados

- Data. Fecha de la carrera
- Esdeveniment esportiu. Nombre de la carrera
- Official time miliseconds. Milisegundos que emplea el corredor en esta cursa concreta
- Lap. Resultado parcial
- Lap position. Posición parcial
- Posició. Posición general en la que finaliza la carrera
- Posició per categoria. Posición en la que finaliza la carrera dentro de la categoría donde se ha inscrito
- Lap time. Tiempo parcial
- Temps oficial. Tiempo en el que finaliza la carrera (Formato hora)
- Personal: Birthday. Día de nacimiento
- Personal: Gènere. Género del corredor (Hombre, Mujer)
- Uid. Id de corredor

3.2.2.3 Cálculos realizados

3.2.2.3.1 Cálculo Edad

Los cálculos expuestos en este apartado se han realizado en Excel mediante código de programación VBA.



```

Sub edad()
Dim filaDatos As Integer
Dim fechaNac As String
Dim nacimiento As Date
Dim edad As Byte
filaDatos = 2
While (Sheets("eolymp_cat").Range("L" & filaDatos).Value <> "")
fechaNac = Sheets("eolymp_cat").Range("J" & filaDatos).Value
nacimiento = Right(fechaNac, 2) & "/" & Mid(fechaNac, 6, 2) & "/" & Left(fechaNac, 4)
If Month(Date) > Month(nacimiento) Then
edad = Year(Date) - Year(nacimiento)
ElseIf Month(Date) = Month(nacimiento) Then
If Day(Date) > Day(nacimiento) Then
edad = Year(Date) - Year(nacimiento)
Else
edad = Year(Date) - Year(nacimiento) - 1
End If
Else
edad = Year(Date) - Year(nacimiento) - 1
End If
Sheets("eolymp_cat").Range("R" & filaDatos).Value = edad
filaDatos = filaDatos + 1
Wend
End Sub
  
```

3.2.2.3.2 Cálculo Cooper

Se realiza el cálculo de Cooper para cada registro del fichero Excel mediante la siguiente fórmula:

$$\text{Cooper} = 720 * \text{metros/segundos}$$

3.2.2.3.3 Cálculo VO₂max

Se utilizan fórmulas de Excel para el cálculo del VO₂Max. Para dicho cálculo se utiliza el test de Tokmakidis que nos indicará los coeficientes a aplicar en la fórmula según la distancia recorrida. Las distancias que no se encontraban en el cuadro adjunto se buscaron utilizando interpolación lineal.

$$\text{VO}_2\text{Max} = (\text{Coeficiente1} + \text{Coeficiente2} * (\text{km/hora})) * 3,5$$

Test general de campo (distancias en atletismo) de Tokmakidis

Evaluación indirecta del VO₂ máx y PO₂ para deportistas

<u>DISTANCIA</u>	<u>CONSUMO MAXIMO DE OXIGENO RELATIVO</u>
metros	ml/kg/min
200	[(-3.6300 + (0.77038 X km/h)] X 3.5
400	[(-1.3010 + (0.78600 X km/h)] X 3.5
600	[(0.3862 + (0.79320 X km/h)] X 3.5
800	[(0.8964 + (0.81460 X km/h)] X 3.5
1000	[(1.2730 + (0.83250 X km/h)] X 3.5
1500	[(2.4388 + (0.83430 X km/h)] X 3.5
1609	[(2.5043 + (0.84000 X km/h)] X 3.5
2000	[(2.7297 + (0.85270 X km/h)] X 3.5
3000	[(2.9226 + (0.89000 X km/h)] X 3.5
5000	[(3.1747 + (0.91390 X km/h)] X 3.5
10000	[(4.7226 + (0.86980 X km/h)] X 3.5
15000	[(4.8619 + (0.88720 X km/h)] X 3.5
20000	[(4.9574 + (0.89950 X km/h)] X 3.5
42195	[(6.9021 + (0.82460 X km/h)] X 3.5

Bibliografía: Pancorbo Sandoval, A.E. : Medicina del deporte y ciencias aplicadas al alto rendimiento y la salud, pp. 215-219, 2002

3.2.2.3.4 Recodificación edad.

Se realizarán recodificaciones de la variable edad según los segmentos de edad marcados por la federación catalana de atletismo:

```
Edad_r<-factor(Edad_r, levels=c("m35", "m40", "m45","m50", "m55","m60", "m65"))
Edad_r[Edad<=17]<-NA
Edad_r[Edad>17 & Edad<=34]<-"m35"
Edad_r[Edad>34 & Edad<=39]<-"m40"
Edad_r[Edad>39 & Edad<=44]<-"m45"
Edad_r[Edad>44 & Edad<=49]<-"m50"
Edad_r[Edad>49 & Edad<=54]<-"m55"
Edad_r[Edad>54 & Edad<=59]<-"m60"
Edad_r[Edad>59]<-"m65"
```

El sentido de la codificación de edad, se debe a la diferencia del rendimiento y mejora que pueden reflejarse según la edad. Se desarrollan diferentes características según la edad:

- Hasta los 25 años se desarrolla velocidad.
- Desde 25 a 40 años se desarrolla fuerza y se pierde velocidad
- A partir de los 45 años se desarrolla la resistencia.

Todos los datos son tratados con carácter confidencial por tratarse de datos reales tal como marca la LOPD 15/99.

3.3 Experimentos propuestos para la determinación de la afectación del entrenamiento.

La duración del experimento planteado es de 6 semanas y consiste en pasarnos los siguientes datos:

- Al inicio del experimento
 - Nombre o alias de corredor
 - Club
 - Edad
 - Sexo
 - Altura
 - Peso
 - Horas de entrenamiento semanal habitual
 - Grado de entrenamiento (cuánto tiempo lleva entrenando)
 - Experimento inscrito
 - Resultado de test de Cooper
- De cada entrenamiento
 - Día entrenamiento
 - Hora inicio
 - Tiempo entrenamiento
 - Tipo suelo (Asfalto / Montaña)
 - Frecuencia cardíaca media
 - Velocidad media
 - Distancia recorrida
- Al final del experimento
 - Resultado de test de Cooper

Los experimentos planteados son:

- **Experimento 1.** Grupo de control. Los atletas que se apunten en este experimento realizarán sus entrenamientos con normalidad y nos pasarán los datos de los mismos.
- **Experimento 2.** Los atletas que se apunten en este experimento realizarán sus entrenamientos con normalidad además de realizar unos ejercicios de equilibrio muscular y nos pasarán los datos de los mismos. La descripción de los ejercicios se puede consultar en el Apéndice B.

Comparando los resultados de los inscritos en este experimento con el grupo de control debería observarse una mejora en la velocidad del atleta.

Tal como se indica en el punto 2.7.3, esta mejora ya ha sido demostrada y existen publicaciones con los resultados obtenidos.

- **Experimento 3.** Los atletas que se apunten en este experimento realizarán sus entrenamientos con normalidad además de realizar unos ejercicios de entrenamiento de los músculos respiratorios y nos pasarán los datos de los mismos. La descripción de los ejercicios se puede consultar en el Apéndice C.

Comparando los resultados de los inscritos en este experimento con el grupo de control debería observarse una mejora en la frecuencia cardíaca media.

Tal como se ha indicado en el punto 2.7.2, la mejora de la capacidad aeróbica que supone estos ejercicios está demostrada con el uso de respiradores, pero no conocemos ningún estudio que haya demostrado la mejoría sin la utilización de los mismos.

En la obtención de los datos hubo gente que por no disponer de un pulsímetro no indicó la frecuencia cardíaca media, otros no indicaron la distancia recorrida, encontramos otra persona que para todos los entrenamientos nos indica la misma frecuencia media. Todos estos inconvenientes dificultan el análisis.

4 Análisis de datos

4.1 Análisis datos experimento afectación del entrenamiento.

En este experimento se busca estudiar la afectación que tiene la realización de unos ejercicios extra para la mejora del rendimiento aeróbico de un atleta, para ello se tendrán en cuenta los datos de los voluntarios tomados en cada uno de los entrenamientos. Se buscan los factores más influyentes en la mejora de dicho rendimiento.

4.1.1 Datos obtenidos de voluntarios.

Se obtiene una muestra inicial con 12 personas donde después de observar el contenido se decide quitar una persona del estudio por sospechar que los datos contenidos en el campo de frecuencia cardíaca no son fiables ya que repetía el mismo valor de frecuencia cardíaca media para muchos de los entrenamientos, teniendo distancia y duración del entrenamiento diferentes.

Como se puede observar, se disponen de muy pocos individuos para realizar el estudio, además los datos están muy correlacionados entre sí, ya que los valores de Cooper y VO₂Max se calculan a partir de la velocidad. Esta situación hace que sea más conveniente el estudio de estos datos agrupándolos por corredor.

De todos los datos que llegan, se toman las medidas agregadas para cada uno de los corredores, para ellos se calcula el promedio de la frecuencia cardíaca, distancia recorrida, tiempo empleado y velocidad. Para los datos hora de inicio y tipo de suelo se toma el valor predominante.

Con estos datos agregados se realiza un análisis descriptivo de los mismos.

Variables:

- Cooper_1. Cooper al inicio del experimento
- Cooper_2. Cooper al final del experimento
- Millora: Cooper_2 - Cooper_1

Se toma como variable respuesta el Cooper_2 y se realiza una búsqueda de las variables que mejor explican la respuesta, para ello se hace comparación de los modelos resultantes. El mejor modelo es el que nos da un AIC más pequeño. R nos hace esta comparación a través de la instrucción step.

En la búsqueda del modelo se indica que no se puede trabajar con toda la base de datos ya que no se disponen de todos los campos de alguno de los corredores (en uno falta la distancia y en otro falta la frecuencia cardíaca media).

Se Plantea el modelo sin Cooper_1 para ver el resultado.

```
step2_cp2<-  
step(corredors1.coop2,~experimento_r+Frec_Media+Sexo+Edad+Altura+Peso+H_entr_semana+  
m_minuto+t_experiencia+Tipo_suelo,data=corredors)  
Start: AIC=150.14
```

El mejor modelo sería incluyendo m_minuto pero al ser una variable que no todos los corredores han informado, se tiene que trabajar con un subconjunto de los datos descartando 2 corredores.

Step: AIC=22.99

Cooper_2 ~ m_minuto + Frec_Media + H_entr_semana + experimento_r +
Tipo_suelo + Sexo + t_experiencia

	Df	Sum of Sq	RSS	AIC
+ Edad	1	16	0	-Inf
+ Altura	1	16	0	-Inf
+ Peso	1	16	0	-Inf
<none>			16	23
- t_experiencia	1	2275	2292	70
- Sexo	1	12170	12187	87
- Tipo_suelo	1	88198	88214	107
- experimento_r	2	180409	180426	112
- H_entr_semana	1	184757	184774	114
- Frec_Media	1	229818	229834	116
- m_minuto	1	393730	393747	122

Step: AIC=-Inf

Cooper_2 ~ m_minuto + Frec_Media + H_entr_semana + experimento_r +
Tipo_suelo + Sexo + t_experiencia + Edad

	Df	Sum of Sq	RSS	AIC
<none>			0	-Inf
- Edad	1	16	16	23
- t_experiencia	1	2260	2260	72
- Sexo	1	7855	7855	85
- Tipo_suelo	1	88135	88135	109
- experimento_r	2	174356	174356	114
- H_entr_semana	1	151486	151486	114
- Frec_Media	1	170473	170473	115
- m_minuto	1	254929	254929	119

Mensajes de aviso perdidos

- 1: attempting model selection on an essentially perfect fit is nonsense
- 2: attempting model selection on an essentially perfect fit is nonsense

Cooper_2 ~ m_minuto + Frec_Media + H_entr_semana + experimento_r +
Tipo_suelo + Sexo + t_experiencia

```
> cp2_scpl.0<-lm(Cooper_2~1,data=corredors2)
> cp2_scpl.1<-lm(Cooper_2~m_minuto,data=corredors2)
> anova(cp2_scpl.0,cp2_scpl.1)
Analysis of Variance Table
```

Model 1: Cooper_2 ~ 1

Model 2: Coop

er_2 ~ m_minuto

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	9	1836389				
2	8	632614	1	1203775	15.223	0.004534 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
> cp2_scpl.2<-lm(Cooper_2~m_minuto+Frec_Media,data=corredors2)
> anova(cp2_scpl.1,cp2_scpl.2)
Analysis of Variance Table
```

Model 1: Cooper_2 ~ m_minuto

Model 2: Cooper_2 ~ m_minuto + Frec_Media

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	8	632614				
2	7	424843	1	207771	3.4234	0.1067

Se busca un modelo con la variable respuesta mejora.

Step: AIC=98.28

millora ~ t_experiencia + Peso + Altura + experimento_r

	Df	Sum of Sq	RSS	AIC
+ Tipo_suelo	1	407.7	351	49.593
+ Sexo	1	163.2	596	54.876
+ m_minuto	1	111.1	648	55.713
+ Frec_Media	1	58.9	700	56.488

+ H_entr_semana	1	58.9	700	56.488
+ Edad	1	16.6	743	57.075
- t_experiencia	1	2237.2	18147	97.856
<none>			15910	98.277
- experimento_r	2	15680.6	31590	102.508
- Altura	1	19596.2	35506	105.911
- Peso	1	28820.8	44730	108.682

Por el criterio de Akaike, el mejor modelo es la variable mejora en función de tiempo de experiencia, peso, altura y experimento.

4.1.2 Ampliación datos corredores eolymp.

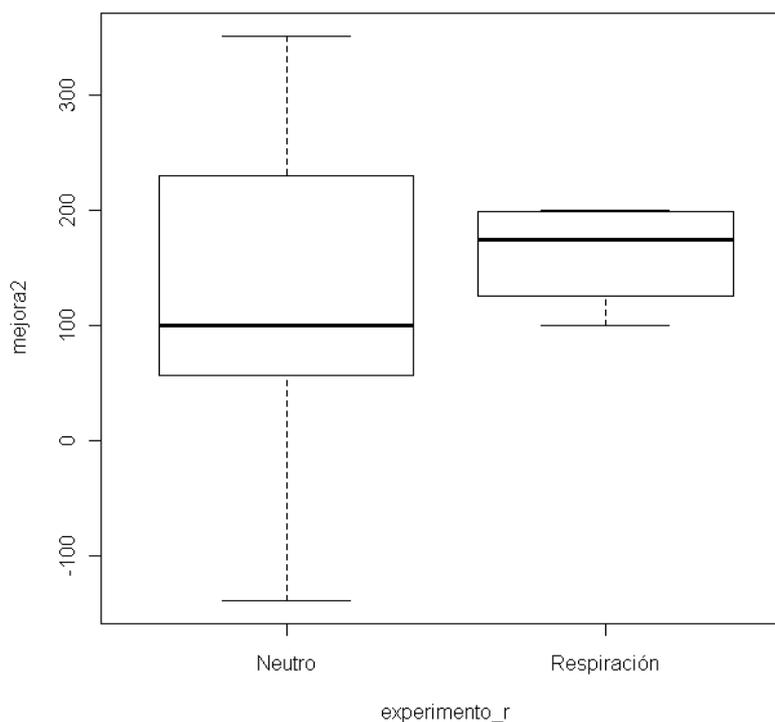
Con el fin de completar un poco el estudio, se buscan los datos de entrenamientos de algunos corredores. Estos atletas se incluirán dentro del grupo de control ya que no han realizado ninguno de los experimentos planteados.

Tenemos 3 grupos de corredores:

- Grupo de control
- Realización ejercicios de respiración
- Realización ejercicios de equilibrio

Tras evaluar los datos de los que se dispone, y viendo que queda un solo corredor con datos válidos del grupo de ejercicios de equilibrio, se decide realizar el estudio con la comparación de los grupos control y respiración. Como variable respuesta tomamos la mejora.

Realizando análisis descriptivo, se denota una gran diferencia en la mejora entre los dos grupos. Se observa que la media de mejora es mayor en el grupo de respiración que en el grupo de control, además tenemos menor variabilidad. La variabilidad tiene el sentido de que en el grupo de control han entrado atletas con un nivel de entrenamiento previo al experimento bastante bajo, por lo que estas personas han de presentar una mejora alta por el solo hecho de entrenar durante el tiempo que dura el experimento. Estas observaciones se realizan a nivel gráfico, aunque al intentar reforzar la hipótesis con el test de comparación de medias aparece ésta como no representativa



```
> oneway.test(mejora2 ~ experimento_r, data=corredores)

One-way analysis of means (not assuming equal variances)

data:  mejora2 and experimento_r
F = 0.4419, num df = 1.000, denom df = 9.178, p-value = 0.5225
```

Se estudia la normalidad de los datos tanto en el conjunto global como para cada uno de los grupos de estudio con resultados positivos.

Por la metodología step, el mejor modelo resultante sería el siguiente, pese a que no termina de realizar todas las comparaciones debido a que tenemos corredores a los que les falta algún dato como distancia o frecuencia cardíaca. El mismo resultado lo obtenemos por el método de varianza incremental.

```
Step: AIC=102.72
mejora ~ t_experiencia + Peso

      Df Sum of Sq  RSS   AIC
+ H_entr_semana  1    6062  5208  59.828
+ experimento_r  1    3635  7635  62.889
+ Edad          1     994 10276  65.265
+ Frec_Media    1     919 10351  65.323
+ Tipo_suelo    1     653 10617  65.526
+ Sexo          1     410 10860  65.708
+ m_minuto      1     370 10900  65.737
+ Altura        1     294 10976  65.792
<none>                    37973 102.717
- Peso          1    21898 59871 106.180
- t_experiencia 1   123900 161872 118.116
```

En el estudio de residuos, aparece el segundo corredor con un valor residual atípico, pero en el estudio de normalidad de los residuos no aparece ningún problema.

Se realiza el estudio de los valores influyentes donde aparecen los corredores 9 y 12 como influyentes a priori y el corredor 10 como influyente a posteriori.

4.1.3 Repetición del estudio haciendo tratamiento de missings.

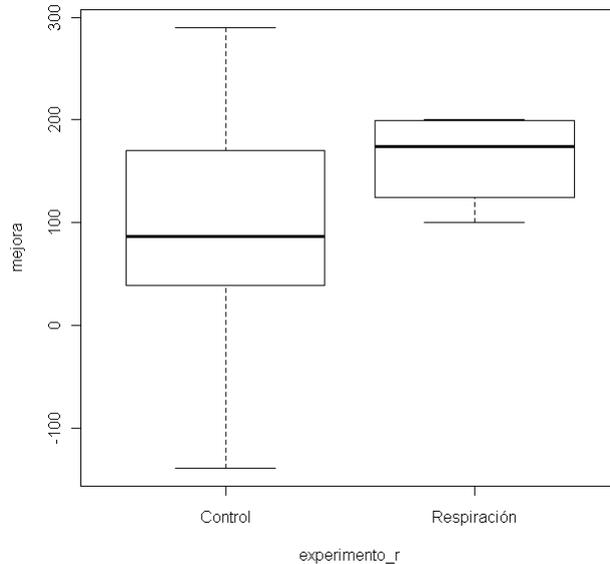
En el estudio realizado, se tienen muchos problemas debido a que hay corredores a los que les faltan datos importantes, como la distancia recorrida o la frecuencia cardíaca. Estos datos son de 2 individuos que se eliminarán del estudio. Estos dos individuos son de sexo femenino. Observando los datos restantes, el resultado es que queda un solo individuo de sexo femenino. A priori, se piensa si este valor podría desvirtuar el estudio el hecho de quedar con una sola mujer. Debido a las diferencias de fisonomía entre hombres y mujeres (<http://www.saludalia.com>), las mujeres precisan el doble de tiempo para desarrollar igual fuerza y trabajar con mayor regularidad para mantener su volumen muscular. Esto hace que las mujeres sean más lentas en pruebas rápidas (100, 200 m,...), aunque la diferencia deja de ser importante a medida que aumentamos la distancia. La variable que estamos midiendo en nuestro estudio es mejora, por lo que con menor motivo influirá esta situación ya que la mejora es una variable relativa a la persona.

Para realizar el análisis se disponen de 10 corredores entre los cuales 6 actuarán dentro del grupo de control y 4 al grupo experimental (realizan ejercicios de respiración). Se observa gráficamente una diferencia en las medias de ambos grupos, aunque si se realiza el test de comparación de medias, nos indica que no existe tal diferencia.

```
> oneway.test(mejora ~ experimento_r, data=corredores)

One-way analysis of means (not assuming equal variances)

data:  mejora and experimento_r
F = 1.3536, num df = 1.000, denom df = 6.493, p-value = 0.2856
```



Para poder continuar con el estudio, se comprueba que la variable mejora cumple con los test de normalidad, el test que usaremos es el de Shapiro que responde mejor al tener tan pocas observaciones. Se realiza el test tanto a nivel global como intergruparal y el resultado obtenido es que cumple las condiciones de normalidad.

```
> shapiro.test(mejora)

      Shapiro-Wilk normality test

data:  mejora
W = 0.9326, p-value = 0.4742

> shapiro.test(mejora[experimento_r=="Control"])

      Shapiro-Wilk normality test

data:  mejora[experimento_r == "Control"]
W = 0.9744, p-value = 0.9208

> shapiro.test(mejora[experimento_r=="Respiración"])

      Shapiro-Wilk normality test

data:  mejora[experimento_r == "Respiración"]
W = 0.8746, p-value = 0.3162
```

Al buscar el modelo por la metodología step, el proceso no llega a finalizarse mostrando el error *"attempting model selection on an essentially perfect fit is nonsense"* que parece indicar que la suma de cuadrados es 0, lo que hace que el AIC se nos vaya a infinito. Se intenta realizar la búsqueda de modelo por varianza incremental por el procedimiento Forward.

Entra en el modelo la variable tiempo de experiencia y posteriormente experimento, es decir, el grupo al que pertenece (control, respiración). Teniendo 10 observaciones, estimar un modelo con más variables supondría sobreestimarlos, los datos que se obtienen no son concluyentes por este mismo motivo. El modelo que se estima es el siguiente:

```
> summary (mejora.lm14)

Call:
lm(formula = mejora ~ t_experiencia + experimento_r, data = corredores)

Residuals:
    Min       1Q   Median       3Q      Max
-72.867 -29.177   7.444  22.617  76.666

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    215.1789    28.5990   7.524 0.000135 ***
t_experiencia  -1.8445     0.2984  -6.182 0.000453 ***
experimento_rRespiración  74.0889    31.6292   2.342 0.051660 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 49 on 7 degrees of freedom
Multiple R-squared:  0.8616,    Adjusted R-squared:  0.822
F-statistic: 21.78 on 2 and 7 DF,  p-value: 0.0009872
```

Grupo de control

$$\text{mejora} = 215.1789 - 1.8445 * t_experiencia$$

Grupo de respiración

$$\text{mejora} = 289.2678 - 1.8445 * t_experiencia$$

El modelo encontrado muestra una clara diferencia de la mejora entre los dos grupos, lo que significaría que los ejercicios de respiración ayudan positivamente a aumentar la capacidad aeróbica y por tanto el rendimiento deportivo. Explica un 82,2 % los datos, pero aun siendo una explicación muy alta, no se toma como concluyente debido a los pocos individuos que componen el estudio. Según este modelo, el tiempo de experiencia tiene una influencia negativa en la mejora, lo que es lógico ya que cuanto más entrenado esté un atleta, más difícil es obtener una mejora.

Para la validación del modelo, se realiza un estudio de los residuos estudentizados y los valores previstos y se encuentra que el corredor 7 presenta un valor residual algo elevado (2.58237992). Analizando este corredor, se observa que se trata de un individuo con muy poco tiempo de experiencia, por lo que es normal que tenga un valor de mejora alto aun estando en el grupo de control.

```
> lm14_student
      1          2          3          4          5          6
0.43061220 -2.19428591  0.78724294 -0.09359263  0.47660500 -1.04116443
      7          8          9         10
2.58237992 -0.67782152  0.50434663 -0.71537308
```

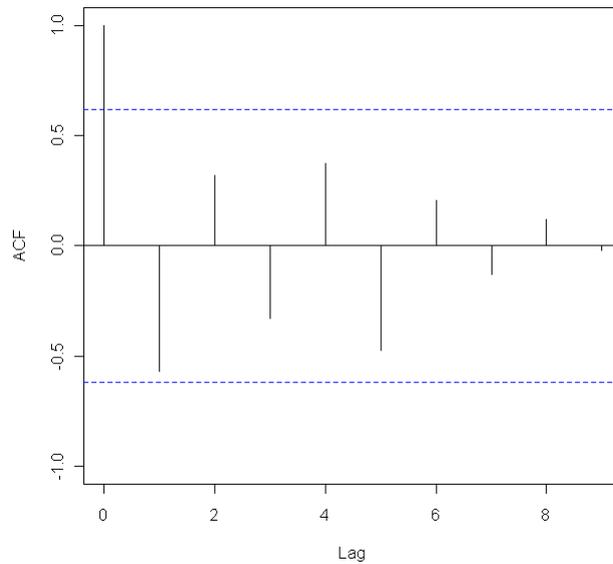
Se observa que los residuos no son autocorrelacionados mediante un test de Durbin-Watson y con un ACF.

```
> dwtest(mejora.lm14)

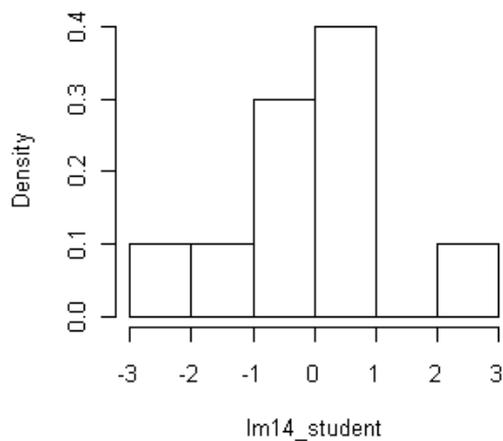
Durbin-Watson test

data:  mejora.lm14
DW = 3.3421, p-value = 0.9942
alternative hypothesis: true autocorrelation is greater than 0
```

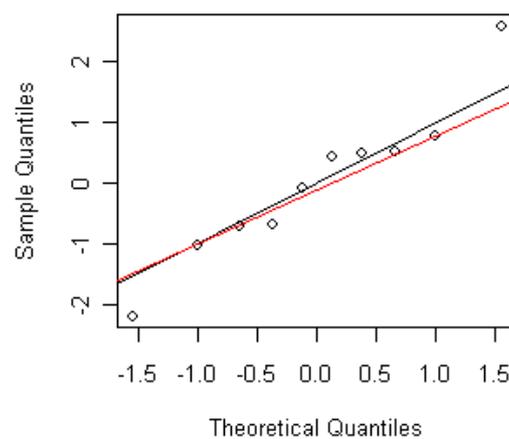
Series lm14_student



Histogram of lm14_student



Normal Q-Q Plot



Los residuos parecen seguir una normalidad.

En el estudio de valores influyentes se encuentra un valor influyente a priori. Se trata de un corredor del grupo de control que la mejora le sale negativa. Se encuentra además que el corredor 7 es influyente a posteriori, este corredor es el mismo que aparecía con un valor residual elevado.

4.2 Análisis de datos para el análisis del índice de dureza.

En este capítulo tratará de medir la dureza de una carrera con el fin de conocer a priori lo que costará realizarla en función de su distancia y su desnivel. De esta manera, un atleta podrá hacerse una idea de la marca que conseguirá antes de disputarse la carrera.

4.2.1 Juegos de datos

En el estudio de estos datos, se miran los siguientes juegos de datos:

- Agrupado por carrera
- Agrupado por corredor

En el juego de datos agrupado por carreras se realizan los siguientes subconjuntos:

- Subconjunto de datos agrupado por carreras sin contar con la carrera de la maquinista 5 Km
- Subconjunto de datos agrupado por carreras sin contar con la carrera de la Behovia y Selva marítima
- Subconjunto de datos agrupado por carreras sin contar con la carrera de Guanta
- Subconjunto de datos de la cursa de Bombers
- Subconjunto de datos de la cursa de la Mercè

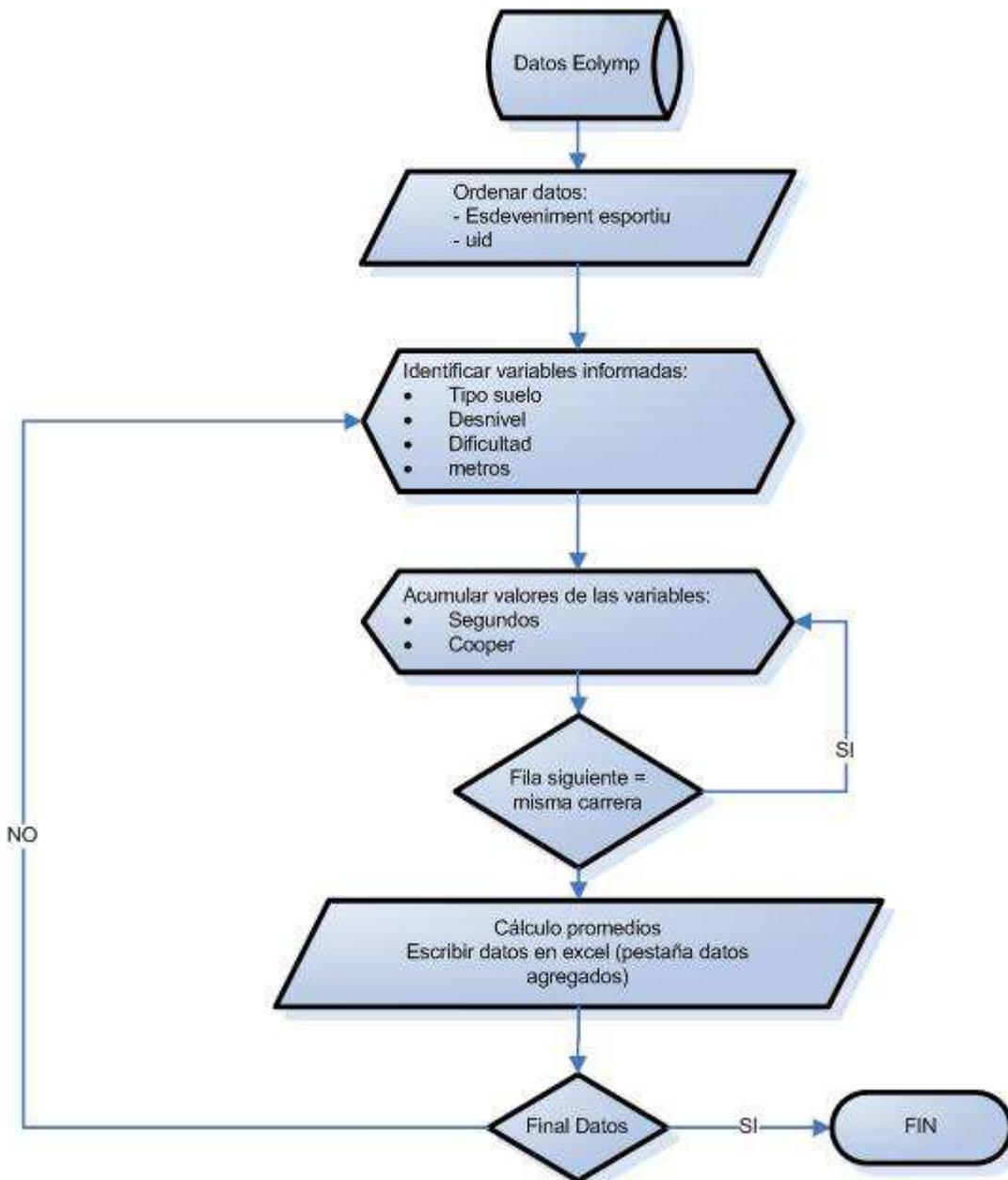
4.2.1.1 Agrupado por carrera

Hemos realizado la búsqueda de los datos genéricos de cada carrera:

- Distancia de la carrera (en metros)
- Tipo de suelo (Asfalto, montaña)
- Desnivel
- Dificultad. Este dato es subjetivo, depende del usuario del cual se han tomado los datos.

Realizamos los siguientes cálculos:

- Promedio de tiempo en segundos que emplean los corredores para terminar esta carrera.
- Velocidad media. Dato calculado según la distancia de la carrera y el promedio de tiempo anterior
- VO₂Max. Calculando el promedio de todos los registros de esta cursa.
- Promedio de Cooper. Calculando el promedio de todos los registros de esta cursa



```

Sub CooperCursa()
Dim cursa As String
Dim filaDatos As Integer
Dim filaAgregados As Integer
Dim sexo As String
Dim metros As Double
Dim segundos As Double
Dim cooper As Double
Dim cursas As Integer
Dim edad As Byte
Dim tipoSuelo As String
Dim desnivel As Integer
Dim dificultad As String

'ORDEN POR ESDEVENIMENT ESPORTIU Y POR UID
ActiveWorkbook.Worksheets("eolymp_cat").Sort.SortFields.Clear
ActiveWorkbook.Worksheets("eolymp_cat").Sort.SortFields.Add Key:=Range( _
"B2:B3414"), SortOn:=xlSortOnValues, Order:=xlAscending, DataOption:= _
xlSortNormal
ActiveWorkbook.Worksheets("eolymp_cat").Sort.SortFields.Add Key:=Range( _
"L2:L3414"), SortOn:=xlSortOnValues, Order:=xlAscending, DataOption:= _
xlSortTextAsNumbers
  
```

```

With ActiveWorkbook.Worksheets("eolymp_cat").Sort
    .SetRange Range("A1:U3414")
    .Header = xlYes
    .MatchCase = False
    .Orientation = xlTopToBottom
    .SortMethod = xlPinYin
    .Apply
End With
' FIN ORDENACION

filaDatos = 2
filaAgregados = 2
cursa = Sheets("eolymp_cat").Range("B" & filaDatos).Value
While (cursa <> "")
    metros = 0
    cursas = 0
    segundos = 0
    cooper = 0
    While (cursa = Sheets("eolymp_cat").Range("B" & filaDatos).Value)
        sexo = Sheets("eolymp_cat").Range("K" & filaDatos).Value
        tipoSuelo = Sheets("eolymp_cat").Range("S" & filaDatos).Value
        desnivel = Sheets("eolymp_cat").Range("T" & filaDatos).Value
        dificultad = Sheets("eolymp_cat").Range("U" & filaDatos).Value
        metros = metros + Sheets("eolymp_cat").Range("M" & filaDatos).Value
        edad = Sheets("eolymp_cat").Range("R" & filaDatos).Value
        segundos = segundos + Sheets("eolymp_cat").Range("N" & filaDatos).Value
        cooper = cooper + Sheets("eolymp_cat").Range("Q" & filaDatos).Value
        filaDatos = filaDatos + 1
        cursas = cursas + 1
    Wend
    metros = metros / cursas
    segundos = segundos / cursas
    cooper = cooper / cursas

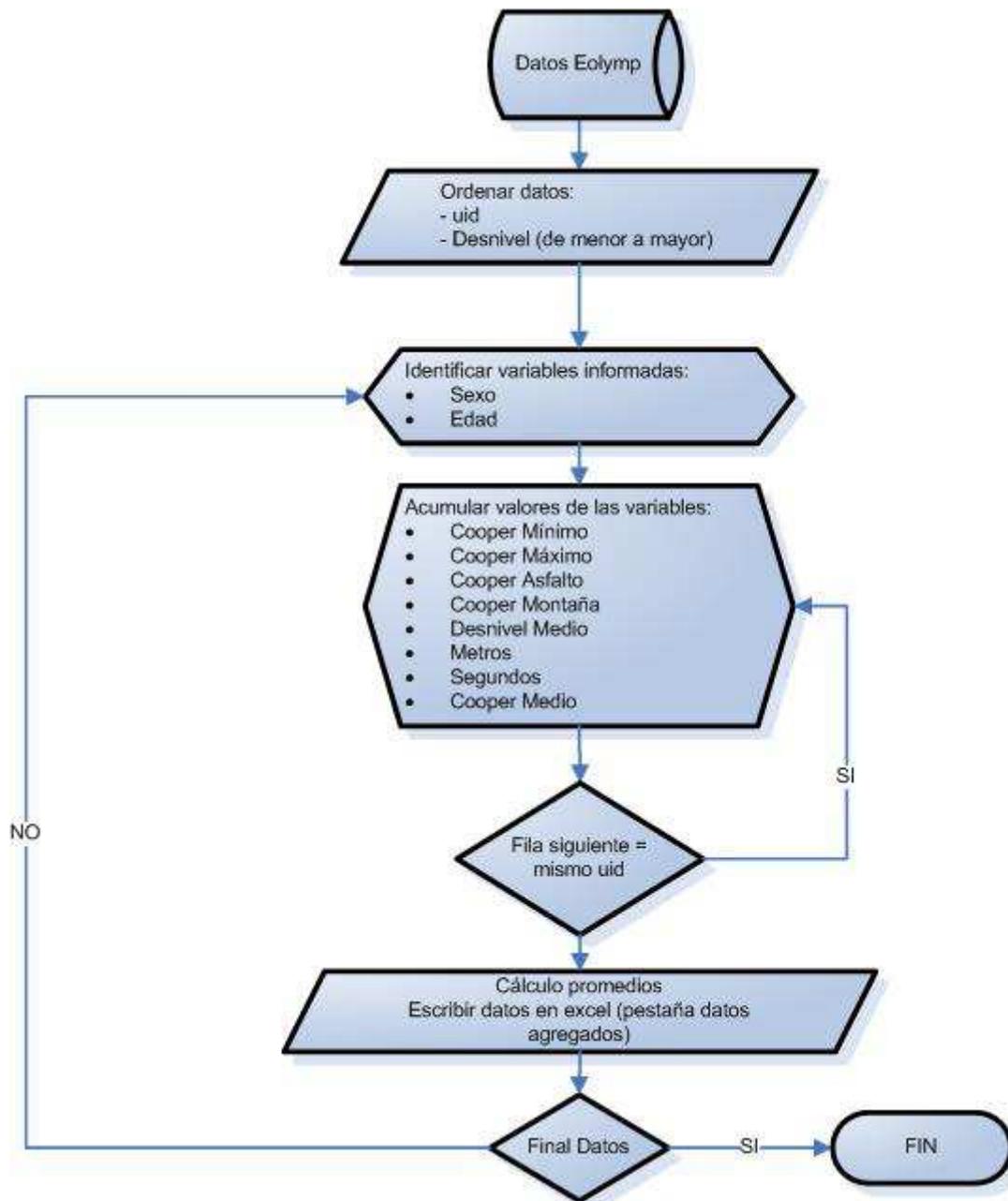
    Sheets("CooperCursa").Range("A" & filaAgregados).Value = cursa
    Sheets("CooperCursa").Range("B" & filaAgregados).Value = metros
    Sheets("CooperCursa").Range("C" & filaAgregados).Value = segundos
    Sheets("CooperCursa").Range("F" & filaAgregados).Value = cooper
    Sheets("CooperCursa").Range("G" & filaAgregados).Value = tipoSuelo
    Sheets("CooperCursa").Range("H" & filaAgregados).Value = desnivel
    Sheets("CooperCursa").Range("I" & filaAgregados).Value = dificultad
    filaAgregados = filaAgregados + 1
    cursa = Sheets("eolymp_cat").Range("B" & filaDatos).Value
Wend
End Sub

```

4.2.1.2 Agrupado por corredor

Se realizan los siguientes cálculos:

- Edad del corredor según su fecha de nacimiento.
- Promedio distancia. Calculo sobre todas las carreras en las que participa dicho corredor
- Promedio tiempo (en segundos)
- Promedio de VO₂Max.
- Promedio Cooper.
- Cooper mínimo. Cooper calculado para la carrera con menor desnivel en la que participa el corredor
- Cooper máximo. Cooper calculado para la carrera con mayor desnivel en la que participa el corredor
- Cooper Asfalto. Promedio de Cooper calculado para todas las carreras con tipo de suelo asfalto en la que participa el corredor
- Cooper Montaña. Promedio de Cooper calculado para todas las carreras con tipo de suelo montaña en la que participa el corredor
- Desnivel medio. Promedio de desnivel de las carreras en las que participa el corredor



```

Sub agregarDatosUID()
Dim uid As String
Dim filaDatos As Integer
Dim filaAgregados As Integer
Dim sexo As String
Dim edad As String
Dim metros As Double
Dim segundos As Double
Dim cooperMedio As Double
Dim cooperMinimo As Double
Dim cooperMaximo As Double
Dim cooperAsfalt As Double
Dim cooperMuntanya As Double
Dim cursas As Byte
Dim cursasAsfalt As Byte
Dim cursasMuntanya As Byte
Dim desnivelMedio As Double
Dim desnivel As Integer
Dim vo2 As Double
Dim cursaMinimo As String
  
```

```
Dim cursaMaximo As String
```

```
' ORDENAR PRIMERO POR UID Y LUEGO POR DESNIVEL DE MENOR A MAYOR
```

```
ActiveWorkbook.Worksheets("eolymp_cat").Sort.SortFields.Clear
ActiveWorkbook.Worksheets("eolymp_cat").Sort.SortFields.Add Key:=Range( _
    "L2:L3414"), SortOn:=xlSortOnValues, Order:=xlAscending, DataOption:= _
    xlSortTextAsNumbers
ActiveWorkbook.Worksheets("eolymp_cat").Sort.SortFields.Add Key:=Range( _
    "T2:T3414"), SortOn:=xlSortOnValues, Order:=xlAscending, DataOption:= _
    xlSortNormal
With ActiveWorkbook.Worksheets("eolymp_cat").Sort
    .SetRange Range("A1:U3414")
    .Header = xlYes
    .MatchCase = False
    .Orientation = xlTopToBottom
    .SortMethod = xlPinYin
    .Apply
End With
```

```
' FIN ORDENACION
```

```
filaDatos = 2
filaAgregados = 2
uid = Sheets("eolymp_cat").Range("L" & filaDatos).Value
While (uid <> "")
    metros = 0
    cursas = 0
    cursesAsfalt = 0
    cursesMuntanya = 0
    segundos = 0
    cooperMedio = 0
    cooperMinimo = 0
    cooperMaximo = 0
    cooperAsfalt = 0
    cooperMuntanya = 0
    desnivel = 0
    vo2 = 0
    While (uid = Sheets("eolymp_cat").Range("L" & filaDatos).Value)
        If (cooperMinimo = 0) Then
            cooperMinimo = Sheets("eolymp_cat").Range("Q" & filaDatos).Value
            cooperMaximo = Sheets("eolymp_cat").Range("Q" & filaDatos).Value
            cursaMinimo = Sheets("eolymp_cat").Range("B" & filaDatos).Value
            cursaMaximo = Sheets("eolymp_cat").Range("B" & filaDatos).Value
        End If
        If (desnivel > Sheets("eolymp_cat").Range("T" & filaDatos).Value) Then
            cooperMinimo = Sheets("eolymp_cat").Range("Q" & filaDatos).Value
            cursaMinimo = Sheets("eolymp_cat").Range("B" & filaDatos).Value
        ElseIf (desnivel < Sheets("eolymp_cat").Range("T" & filaDatos).Value) Then
            cooperMaximo = Sheets("eolymp_cat").Range("Q" & filaDatos).Value
            cursaMaximo = Sheets("eolymp_cat").Range("B" & filaDatos).Value
        End If
        If (Sheets("eolymp_cat").Range("S" & filaDatos).Value = "Asfalt") Then
            cooperAsfalt = cooperAsfalt + Sheets("eolymp_cat").Range("Q" & filaDatos).Value
            cursesAsfalt = cursesAsfalt + 1
        ElseIf (Sheets("eolymp_cat").Range("S" & filaDatos).Value = "Muntanya") Then
            cooperMuntanya = cooperMuntanya + Sheets("eolymp_cat").Range("Q" & filaDatos).Value
            cursesMuntanya = cursesMuntanya + 1
        End If

        desnivel = Sheets("eolymp_cat").Range("T" & filaDatos).Value
        desnivelMedio = desnivelMedio + Sheets("eolymp_cat").Range("T" & filaDatos).Value
        vo2 = vo2 + Sheets("eolymp_cat").Range("P" & filaDatos).Value
        sexo = Sheets("eolymp_cat").Range("K" & filaDatos).Value
        edad = Sheets("eolymp_cat").Range("R" & filaDatos).Value
        metros = metros + Sheets("eolymp_cat").Range("M" & filaDatos).Value
        segundos = segundos + Sheets("eolymp_cat").Range("N" & filaDatos).Value
        cooperMedio = cooperMedio + Sheets("eolymp_cat").Range("Q" & filaDatos).Value
        cursas = cursas + 1
        filaDatos = filaDatos + 1
    Wend
    If cursas = 0 Then
        metros = 0
        segundos = 0
```

```

Else
  metros = metros / cursas
  segundos = segundos / cursas
  desnivelMedio = desnivelMedio / cursas
  cooperMedio = cooperMedio / cursas
  vo2 = vo2 / cursas
End If
If (cursesAsfalt > 0) Then
  cooperAsfalt = cooperAsfalt / cursesAsfalt
End If
If (cursesMuntanya > 0) Then
  cooperMuntanya = cooperMuntanya / cursesMuntanya
End If

Sheets("CooperUId").Range("A" & filaAgregados).Value = uid
Sheets("CooperUId").Range("B" & filaAgregados).Value = sexo
Sheets("CooperUId").Range("C" & filaAgregados).Value = edad
Sheets("CooperUId").Range("D" & filaAgregados).Value = metros
Sheets("CooperUId").Range("E" & filaAgregados).Value = segundos
Sheets("CooperUId").Range("G" & filaAgregados).Value = vo2
Sheets("CooperUId").Range("H" & filaAgregados).Value = cooperMedio
Sheets("CooperUId").Range("I" & filaAgregados).Value = cooperMinimo
Sheets("CooperUId").Range("J" & filaAgregados).Value = cooperMaximo
Sheets("CooperUId").Range("K" & filaAgregados).Value = cooperAsfalt
Sheets("CooperUId").Range("L" & filaAgregados).Value = cooperMuntanya
Sheets("CooperUId").Range("M" & filaAgregados).Value = desnivelMedio
Sheets("CooperUId").Range("N" & filaAgregados).Value = cursaMinimo
Sheets("CooperUId").Range("P" & filaAgregados).Value = cursaMaximo
filaAgregados = filaAgregados + 1
uid = Sheets("eolymp_cat").Range("L" & filaDatos).Value
Wend
End Sub

```

4.2.2 Resultados obtenidos

4.2.2.1 Resultados juego de datos cursas

Se intenta buscar un índice de dureza de la carrera teniendo en cuenta alguno de los siguientes datos:

- Desnivel
- Metros
- Tipo Suelo (Asfalto / Montaña)

La variable analizada como respuesta es VO₂Max.

Se realiza un análisis descriptivo (ver APENDICE E. Resultados R (listados salida).) donde se observa como dato interesante una diferencia de medias según el tipo de suelo, que viene a indicar que el VO₂Max es menor debido a la dureza que supone una cursa de montaña donde normalmente hablamos de un desnivel mayor. Se observa además una mayor variabilidad en estas medias.

Se realiza un test de comparación de medias que apoya la teoría de que las medias tienen un comportamiento distinto según el tipo de suelo.

```

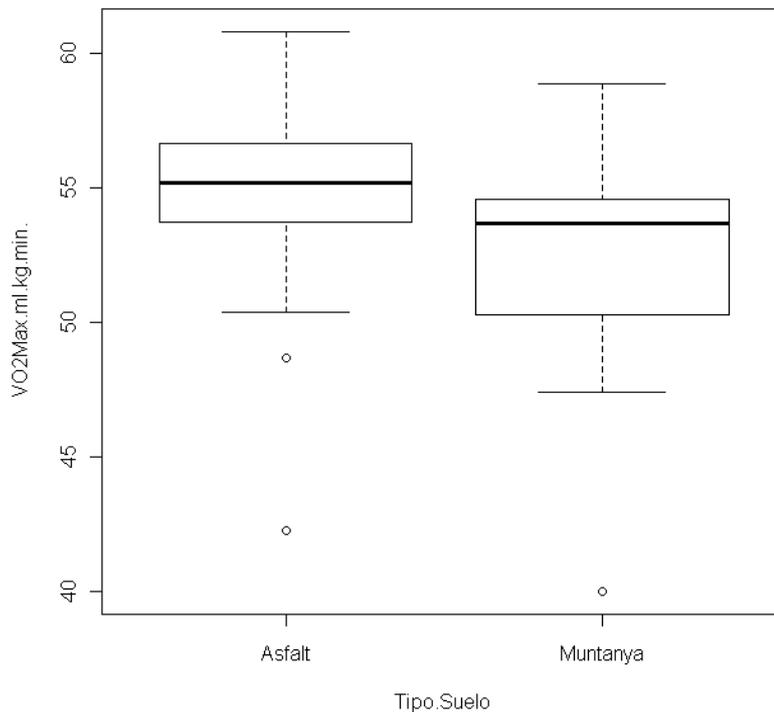
> t.test(VO2Max.ml.kg.min. ~ Tipo.Suelo, data=cooper_curses)

Welch Two Sample t-test

data:  VO2Max.ml.kg.min. by Tipo.Suelo
t = 2.1674, df = 16.367, p-value = 0.04528
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.06483038 5.40892640
sample estimates:
mean in group Asfalt mean in group Muntanya
 54.94757                52.21069

```

El p-valor obtenido indica que se rechaza la hipótesis de que las medias sean independientes del tipo de suelo, aunque esté muy justo.



Mirando la gráfica se puede observar que hay 3 puntos sospechosos, 2 de ellos en las carreras de asfalto y 1 en las carreras de montaña.

Una de estas carreras es la de la maquinista en su versión de 5km, cabe destacar que esta carrera tiene lugar en julio, es decir, hacia final de temporada, momento en que los corredores están en plena forma. El mismo día que se hace esta carrera se disputa también la versión de 10km Así que probablemente los corredores que disputan la versión de 5km son personas poco entrenadas y que deciden hacer algo de deporte de cara al verano y por eso se observa que el valor de $VO_2\max$ es considerablemente más bajo que el resto.

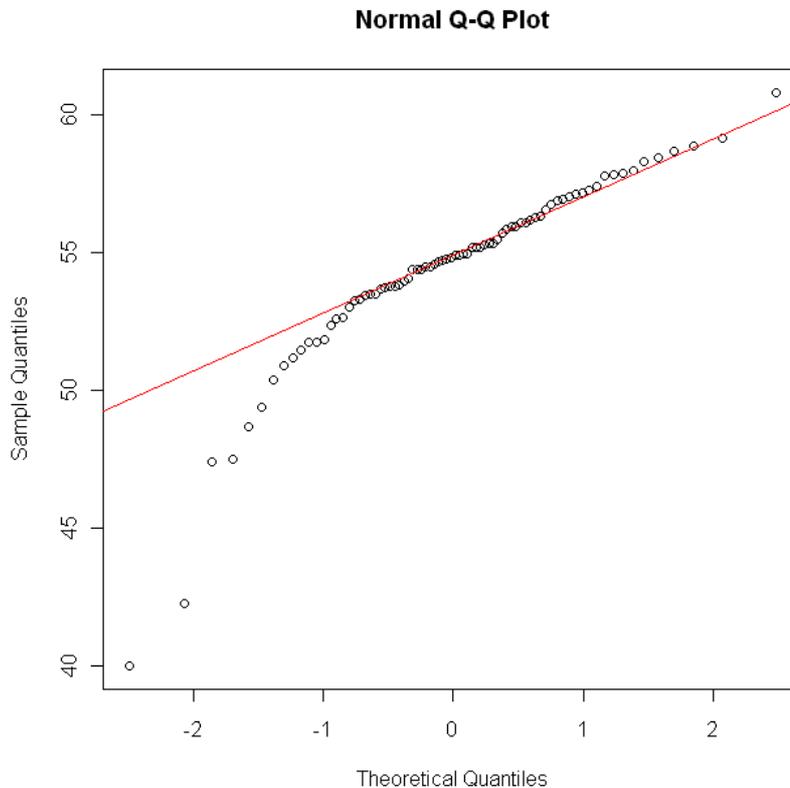
En lo que respecta a carreras de montaña los puntos que destacan corresponden a las carreras de "selva marítima", también denominada "Xtreme" debido a la gran dureza de la carrera. Se trata de una carrera muy exigente apta solamente para personas muy entrenadas. Se disputa en 3 días, aunque los datos de los que se dispone solamente son de 1 día.

A priori se desconoce si por la naturaleza de los datos, sería o no necesario realizar una transformación o por el contrario se puede continuar realizando análisis aplicando los métodos habituales. Para decidir esto se ha realizado un test de normalidad donde se puede observar que la cola izquierda está excesivamente alejada de la recta, lo que puede venir provocado por los outliers ya comentados. Pese a esto se realiza el test de normalidad donde por los resultados no muestran indicios para rechazar que los datos sean normales.

```
> VO2Max_teo<-rnorm(78,mean(VO2Max.ml.kg.min.),sd(VO2Max.ml.kg.min.))
> ks.test(VO2Max.ml.kg.min., VO2Max_teo)
```

Two-sample Kolmogorov-Smirnov test

```
data: VO2Max.ml.kg.min. and VO2Max_teo
D = 0.1026, p-value = 0.81
alternative hypothesis: two-sided
```



Se realiza por el método de Step la búsqueda del mejor modelo con variable respuesta VO₂Max. En este modelo entrarían todas las variables con sus interacciones. La variable desnivel por sí sola no aporta mucha información pero si su interacción con el tipo de suelo. Vemos que el modelo encontrado explicaría un 41,49 %

```
Step: AIC=155.11
VO2Max.ml.kg.min. ~ Tipo.Suelo + Desnivel + metros + Tipo.Suelo:Desnivel +
  Desnivel:metros + Tipo.Suelo:metros + Tipo.Suelo:Desnivel:metros

              Df Sum of Sq   RSS   AIC
<none>                464.11 155.11
- Tipo.Suelo:Desnivel:metros  1    32.507 496.62 158.39
> summary(step_cooper_curses)

Call:
lm(formula = VO2Max.ml.kg.min. ~ Tipo.Suelo + Desnivel + metros +
    Tipo.Suelo:Desnivel + Desnivel:metros + Tipo.Suelo:metros +
    Tipo.Suelo:Desnivel:metros, data = cooper_curses)

Residuals:
    Min       1Q   Median       3Q      Max
-11.3773  -1.2362   0.1597   1.4897   5.8677

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.187e+01  1.186e+00  43.750 < 2e-16 ***
Tipo.SueloMuntanya  1.485e+01  3.049e+00   4.869 6.70e-06 ***
Desnivel        6.232e-03  5.778e-03   1.079  0.2845
metros          2.061e-04  8.647e-05   2.384  0.0198 *
Tipo.SueloMuntanya:Desnivel -1.911e-02  8.143e-03  -2.347  0.0217 *
Desnivel:metros -2.851e-07  3.637e-07  -0.784  0.4357
Tipo.SueloMuntanya:metros -1.178e-03  2.838e-04  -4.152 9.16e-05 ***
Tipo.SueloMuntanya:Desnivel:metros  8.605e-07  3.886e-07   2.214  0.0301 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.575 on 70 degrees of freedom
```

Multiple R-squared: 0.4681, Adjusted R-squared: 0.4149
F-statistic: 8.8 on 7 and 70 DF, p-value: 1.014e-07

El modelo encontrado sería:

Carrera de asfalto:

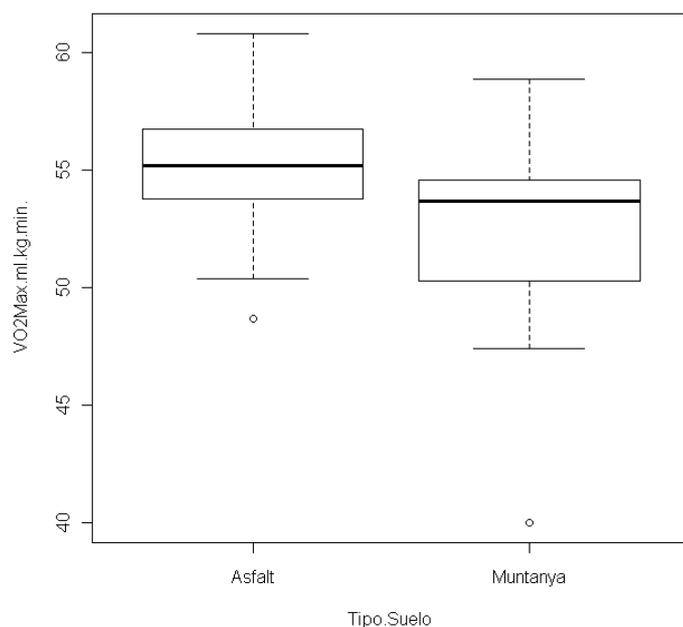
$$\text{VO}_2\text{Max} = 0.006232 * \text{desnivel} + 0.0002061 * \text{metros} - 0.0000002851 * \text{Desnivel} * \text{Metros} + 51.87$$

Carrera de montaña:

$$\text{VO}_2\text{Max} = -0.012878 * \text{desnivel} - 0.0009719 * \text{metros} + 0.0000005754 * \text{Desnivel} * \text{Metros} + 66.72$$

4.2.2.2 Resultados juego de datos cursas sin la carrera de la maquinista 5 km

La carrera de la maquinista se celebra hacia el final de la temporada y en el mismo día se celebra la misma carrera pero de 10 Km lo que hace que los atletas que llevan toda la temporada entrenando se apunten a la carrera larga, lo que deja para personas sin entrenamiento continuado la carrera de 5 Km.



A nivel gráfico se continúa observando que la media de VO₂Max en las carreras de asfalto es superior que en las carreras de montaña. Se realiza un test de comparación de medias.

```
> oneway.test(VO2Max.ml.kg.min. ~ Tipo.Suelo, data=cooper_curses2)
```

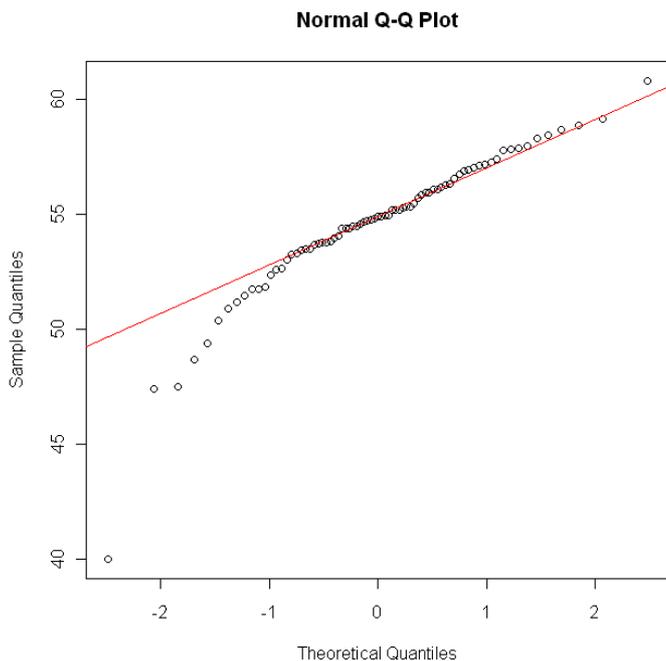
One-way analysis of means (not assuming equal variances)

data: VO2Max.ml.kg.min. and Tipo.Suelo

F = 5.5644, num df = 1.000, denom df = 15.582, p-value = 0.03175

El p-valor obtenido indica que se rechaza la hipótesis de que las medias sean independientes del tipo de suelo. Lo que viene a apoyar las deducciones a partir del gráfico.

Igual que con el juego de datos anterior se estudia la normalidad de los datos para comprobar que efectivamente cumplen la normalidad.



```
> ks.test(VO2Max.ml.kg.min., VO2Max_teo2)

Two-sample Kolmogorov-Smirnov test

data: VO2Max.ml.kg.min. and VO2Max_teo2
D = 0.1558, p-value = 0.3086
alternative hypothesis: two-sided
```

No se puede rechazar la hipótesis de normalidad.

Se busca ahora el modelo que mejor explique los datos. En este caso se observa que explica un 49.74 %, índice superior al anterior modelo (incluyendo la carrera de la maquinista 5km).

```
Step: AIC=127.98
VO2Max.ml.kg.min. ~ Tipo.Suelo + Desnivel + metros + Tipo.Suelo:Desnivel +
  Desnivel:metros + Tipo.Suelo:metros + Tipo.Suelo:Desnivel:metros

              Df Sum of Sq    RSS    AIC
<none>                329.66 127.98
- Tipo.Suelo:Desnivel:metros  1    29.527 359.18 132.58
> summary(step_cooper_curses2)

Call:
lm(formula = VO2Max.ml.kg.min. ~ Tipo.Suelo + Desnivel + metros +
    Tipo.Suelo:Desnivel + Desnivel:metros + Tipo.Suelo:metros +
    Tipo.Suelo:Desnivel:metros, data = cooper_curses2)

Residuals:
    Min       1Q   Median       3Q      Max
-7.3798 -1.3365  0.2148  1.3559  5.8677

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.253e+01  1.014e+00  51.801 < 2e-16 ***
Tipo.SueloMuntanya  1.419e+01  2.592e+00   5.475 6.63e-07 ***
Desnivel       5.857e-03  4.905e-03   1.194  0.23651
metros        1.676e-04  7.376e-05   2.273  0.02616 *
Tipo.SueloMuntanya:Desnivel -1.874e-02  6.913e-03  -2.711  0.00846 **
Desnivel:metros -2.449e-07  3.088e-07  -0.793  0.43037
Tipo.SueloMuntanya:metros -1.140e-03  2.411e-04  -4.729  1.16e-05 ***
Tipo.SueloMuntanya:Desnivel:metros  8.203e-07  3.300e-07   2.486  0.01534 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 2.186 on 69 degrees of freedom
 Multiple R-squared: 0.5437, Adjusted R-squared: 0.4974
 F-statistic: 11.74 on 7 and 69 DF, p-value: 9.67e-10

El modelo resultante sería:

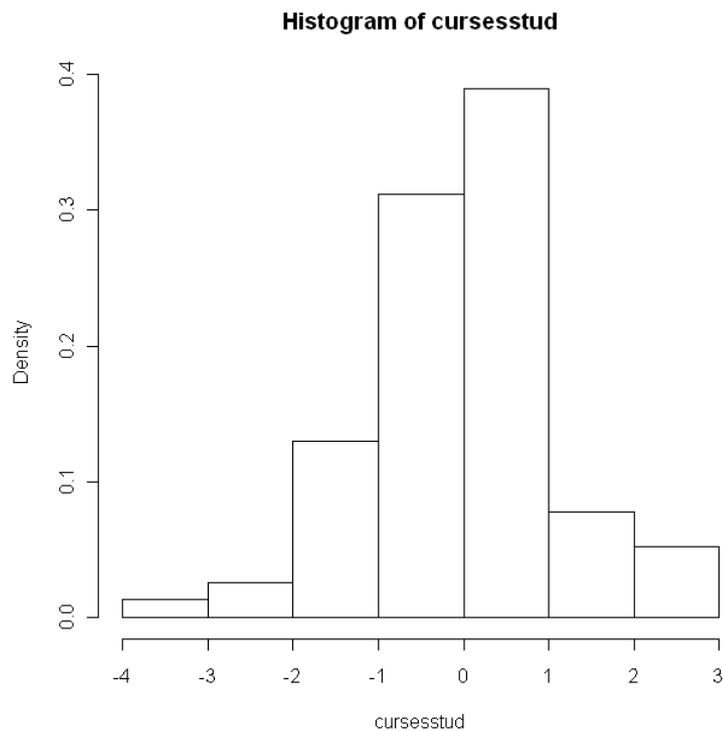
Carrera de asfalto:

$$\text{VO}_2\text{Max} = 0.005857 * \text{desnivel} + 0.0001676 * \text{metros} - 0.0000002449 * \text{Desnivel} * \text{metros} + 52.53$$

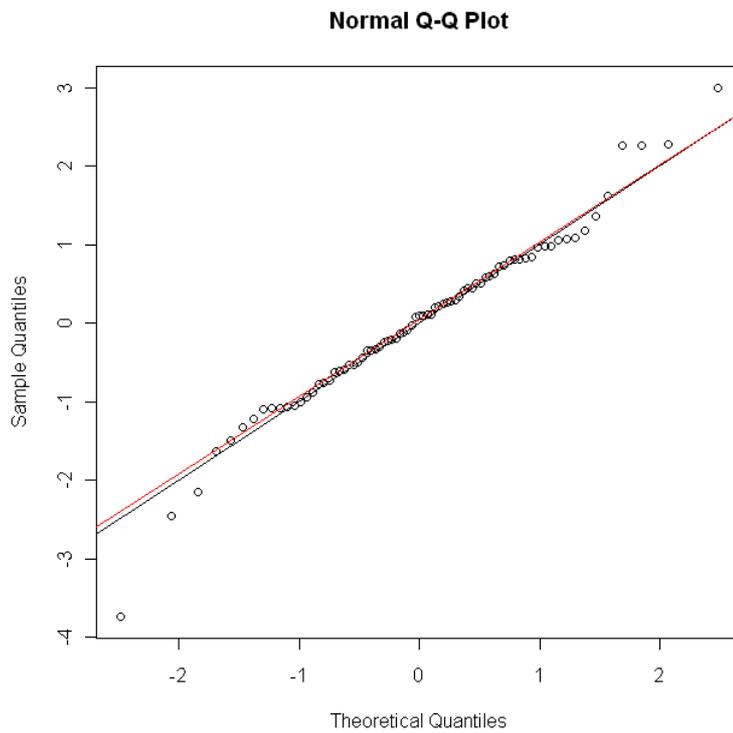
Carrera de montaña:

$$\text{VO}_2\text{Max} = -0.012883 * \text{desnivel} - 0.0009724 * \text{metros} + 0.0000005754 * \text{Desnivel} * \text{metros} + 66.72$$

Se hace un estudio de residuos para decidir si ya es bueno este modelo o bien es necesario hacer más ajustes.



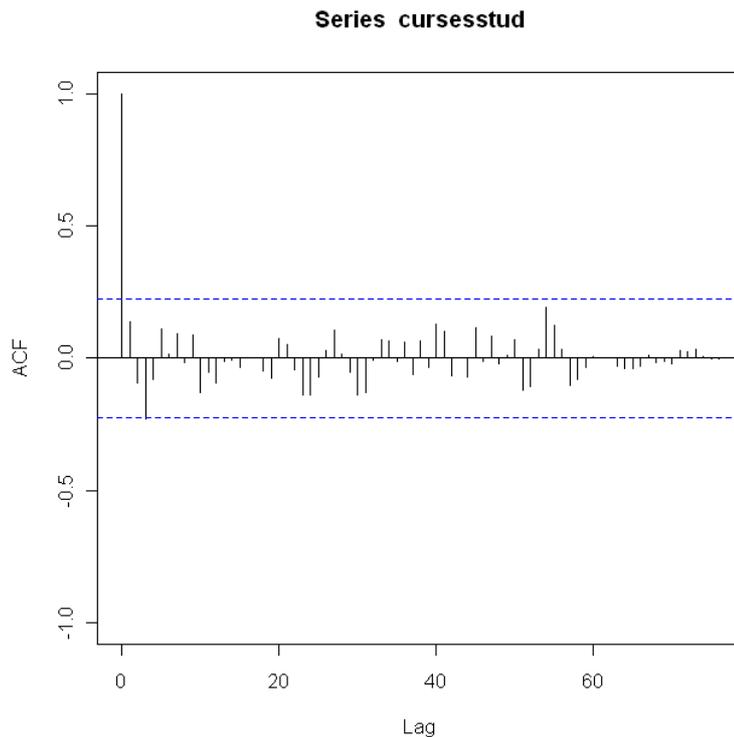
Se observan unos residuos estudentizados con valor absoluto mayor que 2-3. Se debe comprobar a qué carreras pertenecen estos valores influyentes.



Los residuos cumplen casi con el plot de normalidad excepto en que las colas son un poco pesadas. Hay dos valores atípicamente anormales:

Residuos estudentizados	
Pujada-Baixada a Guanta	2.9967
Behobia-San Sebastián	-3.7384

Se estudia la independencia de los residuos.



La carrera de Behovia- San Sebastián es un posible outlier.

En el estudio de los valores influyente se observa que la carrera más influyente resulta ser la maratón de Collserola.

4.2.2.3 Resultados juego de datos cursas sin las carrera de Behovia y selva marítima

Las carreras Behovia y selva marítima son también valores outliers debido a las características de las mismas.

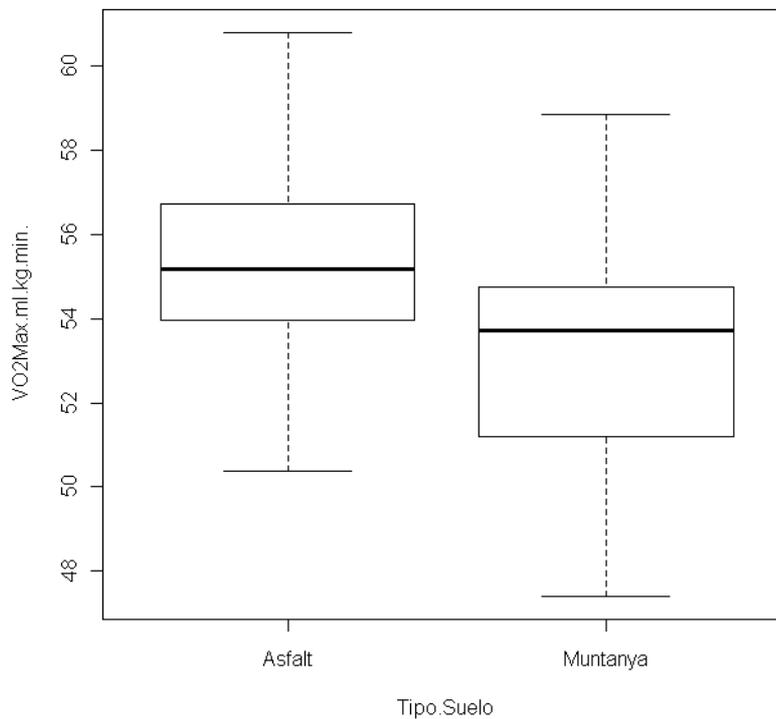
Behovia es una media maratón con recorrido lineal, cosa no habitual en este tipo de carreras y con un desnivel mayor a la mayoría de media maratón de asfalto, lo que la convierte en una carrera de una gran dureza.

Selva marítima es una carrera de 3 días de competición y con un desnivel muy pronunciado, lo que hace que también esta sea una carrera extremadamente dura.

Con este nuevo juego de datos continúa observándose una diferencia en las medias según el tipo de suelo. El VO_2Max es mayor en las carreras de asfalto que en las de montaña, además se observa mayor variabilidad en las carreras de montaña.

```
> oneway.test(VO2Max.ml.kg.min. ~ Tipo.Suelo, data=vo2max_sinbehobia)
One-way analysis of means (not assuming equal variances)

data: VO2Max.ml.kg.min. and Tipo.Suelo
F = 5.2786, num df = 1.00, denom df = 15.37, p-value = 0.03602
```



El mejor modelo obtenido con este juego de datos es el siguiente, que explica un 32,28% de los datos:

```

Step: AIC=113.33
VO2Max.ml.kg.min. ~ Tipo.Suelo + metros + Tipo.Suelo:metros

              Df Sum of Sq   RSS   AIC
<none>                305.48 113.33
+ Desnivel             1    0.937 304.54 115.10
- Tipo.Suelo:metros    1   94.169 399.64 131.48

> summary(step_vo2max_sinbehoia)

Call:
lm(formula = VO2Max.ml.kg.min. ~ Tipo.Suelo + metros + Tipo.Suelo:metros,
    data = vo2max_sinbehoia)

Residuals:
    Min       1Q   Median       3Q      Max
-5.7132 -1.3605  0.0166  1.3504  4.4827

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.332e+01  5.538e-01  96.281 < 2e-16 ***
Tipo.SueloMuntanya  2.720e+00  1.225e+00   2.221 0.029536 *
metros        1.460e-04  3.667e-05   3.980 0.000164 ***
Tipo.SueloMuntanya:metros -3.551e-04  7.591e-05  -4.678 1.34e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.074 on 71 degrees of freedom
Multiple R-squared:  0.3503,    Adjusted R-squared:  0.3228
F-statistic: 12.76 on 3 and 71 DF,  p-value: 9.26e-07

```

Si se realiza el estudio de forma manual, aplicando el test de varianza incremental, se pueden obtener otros modelos interesantes.

No se debe perder de vista que es necesario el principio de modelos encajados o jerárquicos.

$$Test V.I F = \frac{(SCR_H - SCR)/q}{SCR/(n - P)} \approx F_{p-1, n-p}$$

Se realiza contrastes teniendo en cuenta las variables individuales de lo que el que resulta como más significativo es el de tipo de suelo

```
> lm3<-lm(VO2Max.ml.kg.min. ~Tipo.Suelo)
> anovalm0,lm3)
Analysis of Variance Table

Model 1: VO2Max.ml.kg.min. ~ 1
Model 2: VO2Max.ml.kg.min. ~ Tipo.Suelo
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1      74 470.17
2      73 416.25  1    53.915 9.4553 0.002961 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Buscando más variables que sumen en el modelo, se encuentra que la variable metros tiene una significación superior al 5 %, aunque no habría que descartarla antes de mirar el comportamiento de las interacciones. Se introduce en el modelo y estudia el comportamiento de las interacciones

```
Model 1: VO2Max.ml.kg.min. ~ Tipo.Suelo
Model 2: VO2Max.ml.kg.min. ~ metros + Tipo.Suelo
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1      73 416.25
2      72 399.64  1    16.61 2.9924 0.08794 .

Model 1: VO2Max.ml.kg.min. ~ metros + Tipo.Suelo
Model 2: VO2Max.ml.kg.min. ~ metros + +Tipo.Suelo + Tipo.Suelo:metros
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1      72 399.64
2      71 305.48  1    94.169 21.887 1.344e-05 ***
```

Se observa que la interacción de metros con tipo de suelo es significativa. Se estudia la entrada en el modelo de la variable desnivel.

```
Model 1: VO2Max.ml.kg.min. ~ metros + +Tipo.Suelo + Tipo.Suelo:metros
Model 2: VO2Max.ml.kg.min. ~ metros + Desnivel + Tipo.Suelo + Tipo.Suelo:metros
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1      71 305.48
2      70 304.54  1    0.93674 0.2153 0.6441
```

Según este resultado, el desnivel no sería significativo, pero igualmente se estudia el comportamiento de las interacciones.

```
lm9<-lm(VO2Max.ml.kg.min. ~metros+Desnivel+Tipo.Suelo+Tipo.Suelo:metros+metros:Desnivel)
summary(lm9)
anova(lm7,lm9)
> summary(lm9)
```

```
Call:
lm(formula = VO2Max.ml.kg.min. ~ metros + Desnivel + Tipo.Suelo +
  Tipo.Suelo:metros + metros:Desnivel)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-4.9690 -1.0605  0.0119  1.2432  4.8273
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.393e+01  6.499e-01  82.983 < 2e-16 ***
metros       9.154e-05  4.347e-05   2.106 0.038876 *
Desnivel    -3.115e-03  2.459e-03  -1.267 0.209470
Tipo.SueloMuntanya  7.451e+00  2.431e+00   3.065 0.003103 **
metros:Tipo.SueloMuntanya -7.635e-04  2.016e-04  -3.788 0.000321 ***
metros:Desnivel  2.552e-07  1.161e-07   2.198 0.031304 *
---

```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.031 on 69 degrees of freedom
 Multiple R-squared: 0.3947, Adjusted R-squared: 0.3508
 F-statistic: 8.997 on 5 and 69 DF, p-value: 1.28e-06

```
> anova(lm7,lm9)
Analysis of Variance Table
```

```
Model 1: VO2Max.ml.kg.min. ~ metros + Desnivel + Tipo.Suelo + Tipo.Suelo:metros
Model 2: VO2Max.ml.kg.min. ~ metros + Desnivel + Tipo.Suelo + Tipo.Suelo:metros +
metros:Desnivel
Res.Df  RSS Df Sum of Sq    F Pr(>F)
1      70 304.54
2      69 284.61  1    19.928 4.8314 0.0313 *
```

Se comprueba que efectivamente la interacción de metros y desnivel es significativa, este modelo explica un 35,08 %.

Se observa que pasa con el modelo con todas las interacciones dobles:

```
Call:
lm(formula = VO2Max.ml.kg.min. ~ metros + Desnivel + Tipo.Suelo +
    Tipo.Suelo:metros + metros:Desnivel + Tipo.Suelo:Desnivel)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-4.4978 -1.3505  0.0348  1.3531  5.4050
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.383e+01  6.346e-01  84.817  <2e-16 ***
metros       7.601e-05  4.294e-05   1.770  0.0812 .
Desnivel    -1.803e-03  2.469e-03  -0.730  0.4678
Tipo.SueloMuntanya  6.102e+00  2.447e+00   2.493  0.0151 *

metros:Tipo.SueloMuntanya -4.002e-04  2.577e-04  -1.553  0.1251
metros:Desnivel          2.864e-07  1.140e-07   2.513  0.0144 *
Desnivel:Tipo.SueloMuntanya -9.083e-03  4.175e-03  -2.176  0.0331 *
```

Este modelo indicaría que el tipo de suelo no es significativo en el modelo, lo cual no parece lógico. Este desajuste puede estar provocado por tener pocas observaciones y por ello se podría estar sobreestimando el modelo. Los modelos plausibles serían el proveniente del step y el lm9.

Modelo proveniente de Step.

Carrera de asfalto:

$$\text{VO}_2\text{Max} = 0.000146 * \text{metros} + 53.32$$

Carrera de montaña:

$$\text{VO}_2\text{Max} = -0.0002091 * \text{metros} + 56.04$$

Modelo proveniente método varianza incremental (lm9).

Carrera de asfalto:

$$\text{VO}_2\text{Max} = 0.00009154 * \text{metros} - 0.003115 * \text{desnivel} + 0.0000002552 * \text{metros} * \text{desnivel} + 53.93$$

Carrera de montaña:

$$\text{VO}_2\text{Max} = -0.00067196 * \text{metros} - 0.003115 * \text{desnivel} + 0.0000002552 * \text{metros} * \text{desnivel} + 61.381$$

Realizando un estudio de residuos, se fijará el límite para considerar que un valor destaca en el valor residual igual a 3. Se encuentra un valor por encima de 3, aunque no es muy importante la diferencia.

```
> vo2max_sinbehobia[stud_lm_step_sosp,]
      curso metros segundos Km_hora
34 Cursa de muntanya Santa Coloma de Queralt 14000 5114.004 9.855291
   VO2Max.ml.kg.min. Cooper Tipo.Suelo Desnivel Dificultad
34          47.40175 2059.688  Muntanya      870  Moderat
```

Por lo demás, se puede considerar que el análisis de residuos no indica ninguna anomalía en los datos. La maratón de Collserola aparece como valor influyente tanto a priori como a posteriori.

4.2.2.4 Resultados juego de datos cursas sin la carrera de Guanta

Se realiza el test de normalidad de los datos con resultados positivos, por lo que se puede continuar con el estudio del modelo.

```
VO2Max_teo<-rnorm(length(VO2Max.ml.kg.min.),mean(VO2Max.ml.kg.min.),sd(VO2Max.ml.kg.min.))
ks.test(VO2Max.ml.kg.min., VO2Max_teo)

Two-sample Kolmogorov-Smirnov test

data: VO2Max.ml.kg.min. and VO2Max_teo
D = 0.0946, p-value = 0.8982
alternative hypothesis: two-sided
```

Aplicando la metodología step, el modelo que se encuentra explica un 1 % más que en el estudio realizado sin quitar la carrera de Guanta.

```
Step: AIC=107.46
VO2Max.ml.kg.min. ~ Tipo.Suelo + metros + Tipo.Suelo:metros

              Df Sum of Sq   RSS   AIC
<none>                283.75 107.46
+ Desnivel             1    0.458 283.29 109.34
- Tipo.Suelo:metros    1   89.071 372.82 125.66
> summary(lm_step)

Call:
lm(formula = VO2Max.ml.kg.min. ~ Tipo.Suelo + metros + Tipo.Suelo:metros,
    data = vo2max_curses)

Residuals:
    Min       1Q   Median       3Q      Max
-5.366 -1.278  0.174  1.272  4.836

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.332e+01  5.376e-01  99.193 < 2e-16 ***
Tipo.SueloMuntanya  2.243e+00  1.206e+00   1.859 0.067198 .
metros         1.460e-04  3.560e-05   4.101 0.000109 ***
Tipo.SueloMuntanya:metros -3.459e-04  7.379e-05  -4.688 1.32e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.013 on 70 degrees of freedom
Multiple R-squared:  0.3837,    Adjusted R-squared:  0.3573
F-statistic: 14.53 on 3 and 70 DF,  p-value: 1.875e-07
```

Carrera de asfalto:

$$VO_2Max = 0.000146 * metros + 53.32$$

Carrera de montaña:

$$VO_2Max = - 0.0001999 * metros + 55.563$$

Se observa que el modelo para las carreras de asfalto es el mismo que se ajustó en el apartado anterior, lo cual tiene lógica teniendo en cuenta que la carrera de Guanta es de montaña.

Se prueba de ajustar el modelo sin usar el método step, comparando por varianza incremental. Se mira el modelo completo:

```
> lm_completo<-lm(VO2Max.ml.kg.min. ~metros*Desnivel*Tipo.Suelo)
> summary(lm_completo)

Call:
```

```
lm(formula = VO2Max.ml.kg.min. ~ metros * Desnivel * Tipo.Suelo)
```

Residuals:

Min	1Q	Median	3Q	Max
-4.6271	-1.1364	0.0618	1.2789	4.3932

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.235e+01	8.314e-01	62.959	< 2e-16 ***
metros	1.908e-04	6.063e-05	3.146	0.002483 **
Desnivel	6.428e-03	4.018e-03	1.600	0.114401
Tipo.SueloMuntanya	9.537e+00	2.660e+00	3.585	0.000640 ***
metros:Desnivel	-2.830e-07	2.530e-07	-1.119	0.267360
metros:Tipo.SueloMuntanya	-6.449e-04	2.592e-04	-2.489	0.015357 *
Desnivel:Tipo.SueloMuntanya	-2.196e-02	5.710e-03	-3.846	0.000273 ***
metros:Desnivel:Tipo.SueloMuntanya	7.290e-07	2.774e-07	2.628	0.010662 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.789 on 66 degrees of freedom
Multiple R-squared: 0.541, Adjusted R-squared: 0.4923
F-statistic: 11.11 on 7 and 66 DF, p-value: 3.395e-09

Mediante procedimiento Backward se intenta ajustar el modelo con las interacciones dobles y observando el test de varianza incremental, indicando éste que no se puede prescindir de interacciones triples, quedando así como mejor modelo el modelo completo. Si se compara la R² ajustada en este modelo con el del apartado anterior (sin quitar la carrera de Guanta), se observa una ganancia de un 7 %.

```
Model 1: VO2Max.ml.kg.min. ~ metros + Desnivel + Tipo.Suelo + metros:Desnivel +
  metros:Tipo.Suelo + Desnivel:Tipo.Suelo
Model 2: VO2Max.ml.kg.min. ~ metros * Desnivel * Tipo.Suelo
  Res.Df    RSS Df Sum of Sq    F Pr(>F)
1      67 233.46
2      66 211.34 1      22.12 6.9081 0.01066 *
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

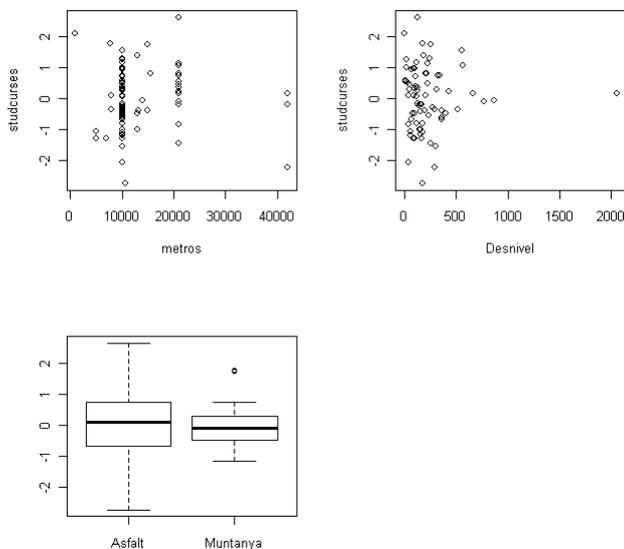
Carrera de asfalto:

$$VO_2Max = 0.0001908 * metros + 0.006428 * desnivel - 0.000000283 * metros * desnivel + 52.35$$

Carrera de montaña:

$$VO_2Max = - 0.0004541 * metros - 0.015532 * desnivel + 0.000000446 * metros * desnivel + 61.887$$

Se debe realizar un estudio de residuos par observar linealidad, independendencia, y homocedasticidad.



Se puede observar una carrera de montaña que se aleja bastante del resto, aunque su valor residual no es superior a 2.

```
> dwtest(lm_completo)
```

Durbin-Watson test

```
data: lm_completo
DW = 1.9185, p-value = 0.3455
alternative hypothesis: true autocorrelation is greater than 0
```

Se muestra independencia de los residuos estudentizados.

Si se miran los valores influyentes, se observa que la maratón de Collserola resulta influyente tanto a priori como a posteriori.

4.2.2.5 Modelo tomado como "mejor modelo".

Al eliminar la carrera de Guanta no se está ganando mucha explicación, además dado el bajo número de carreras sobre las que se tiene información, es importante asegurarse de no eliminar más carreras de las que son verdaderos outliers. Después de analizar los resultados obtenidos, se toma como mejor modelo el resultante del anterior resultado antes de eliminar Guanta.

Carrera de asfalto:

$$VO_2Max = 0.00009154 * metros - 0.003115 * desnivel + 0.0000002552 * metros * desnivel + 53.93$$

Carrera de montaña:

$$VO_2Max = - 0.00067196 * metros - 0.003115 * desnivel + 0.0000002552 * metros * desnivel + 61.381$$

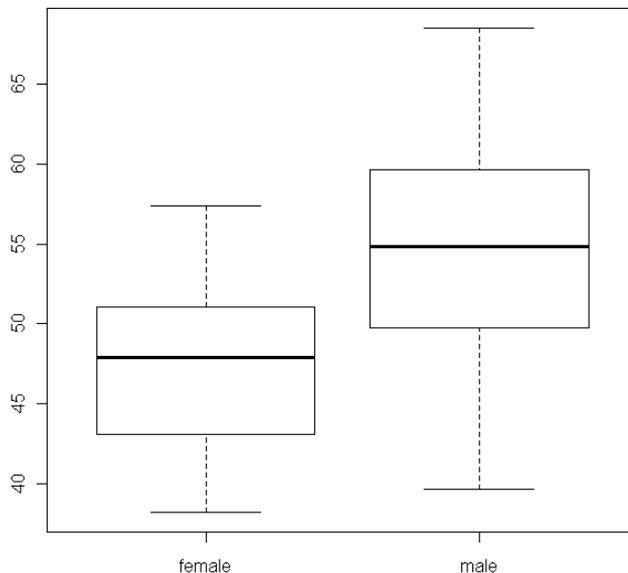
4.2.2.6 Resultados Carreras individuales.

4.2.2.6.1 Cursa de Bombers

Se realiza un estudio descriptivo observándose las siguientes correlaciones:

Variable	Correlación
Posición	-0,9494863
Posición por categoría	-0,6686935
Segundos	-0,9762118
Edad	-0,1543308

Se recodifica la edad y se busca diferencias entre los grupos codificados, no encontrando que exista una importante diferencia de medias entre estos grupos, aunque sí que se observan diferencias entre los diferentes sexos:



Se refuerza este plot con un test de diferencia de medias

```
> t.test(VO2Max.ml.kg.min.~Personal..Gènere)

Welch Two Sample t-test

data: VO2Max.ml.kg.min. by Personal..Gènere
t = -7.3795, df = 53.967, p-value = 1e-09
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -8.944642 -5.122741
sample estimates:
mean in group female   mean in group male
      47.56676           54.60045
```

Mediante el método step, se busca el mejor modelo para nuestras variables, resultando el siguiente:

```
Step: AIC=217.24
VO2Max.ml.kg.min. ~ Posició + Posició.per.categoria + Personal..Gènere +
  Edad + Posició:Posició.per.categoria + Posició:Personal..Gènere +
  Posició:Edad

Df Sum of Sq  RSS  AIC
<none>          558.71 217.24
```

```

+ Edad:Posició.per.categoria          1      2.58 556.14 218.09
+ Edad:Personal..Gènere                1      2.38 556.33 218.17
+ Posició.per.categoria:Personal..Gènere 1      0.08 558.63 219.20
- Posició:Edad                         1     21.97 580.68 224.84
- Posició:Personal..Gènere             1     66.87 625.58 243.39
- Posició:Posició.per.categoria        1    435.43 994.14 358.72
> summary(step3)

Call:
lm(formula = VO2Max.ml.kg.min. ~ Posició + Posició.per.categoria +
    Personal..Gènere + Edad + Posició:Posició.per.categoria +
    Posició:Personal..Gènere + Posició:Edad)

Residuals:
    Min       1Q   Median       3Q      Max
-6.5269 -0.9245 -0.1062  0.9727  5.0266

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.206e+01  1.002e+00  61.959 < 2e-16 ***
Posició      -1.528e-03  1.382e-04 -11.061 < 2e-16 ***
Posició.per.categoria -2.629e-03  2.291e-04 -11.477 < 2e-16 ***
Personal..Gènere  4.131e+00  7.089e-01  5.827 1.80e-08 ***
Edad         -6.599e-02  1.855e-02  -3.557 0.000451 ***
Posició:Posició.per.categoria  2.370e-07  1.729e-08  13.705 < 2e-16 ***
Posició:Personal..Gènere  -4.012e-04  7.470e-05 -5.371 1.85e-07 ***
Posició:Edad    8.078e-06  2.624e-06  3.078 0.002323 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.523 on 241 degrees of freedom
Multiple R-squared:  0.9517,    Adjusted R-squared:  0.9503
F-statistic: 678.1 on 7 and 241 DF,  p-value: < 2.2e-16

```

4.2.2.6.2 Cursa de la Mercè.

Igual que en la anterior carrera, se seleccionan los registros correspondientes a la cursa de la Mercè. Se eliminan aquellos registros donde la edad parece no ser correcta para evitar falsas conclusiones.

El estudio de correlaciones obtenido es el siguiente:

Variable	Correlación
Posición	-0, 9300197
Posición por categoría	-0, 7332571
Segundos	-0, 9723904
Edad	-0, 0595065

Al igual que en la anterior carrera, los grupos de edad no parecen presentar medias diferentes de VO₂Max, pero sí que se observa en los sexos:

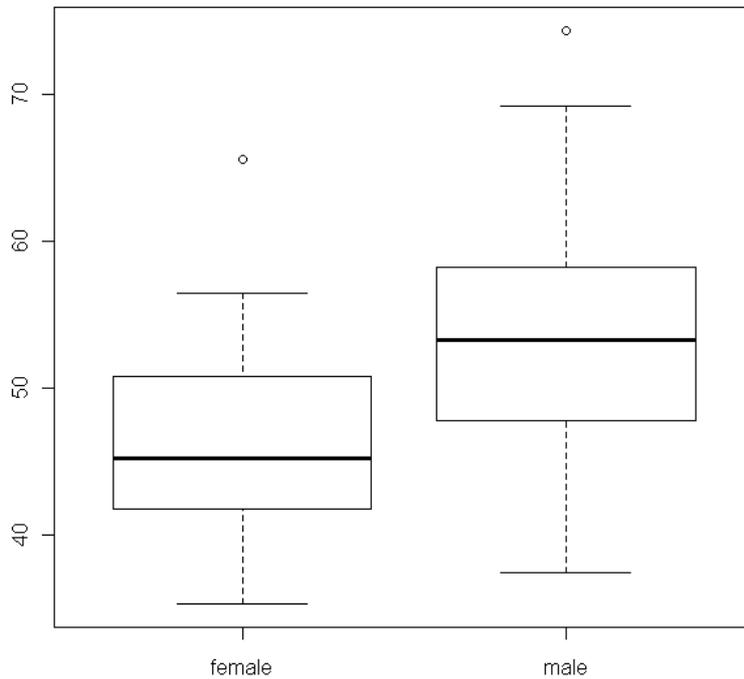
```

> t.test(VO2Max.ml.kg.min.~Personal..Gènere)

Welch Two Sample t-test

data:  VO2Max.ml.kg.min. by Personal..Gènere
t = -6.6678, df = 59.96, p-value = 9.221e-09
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -8.712655 -4.691463
sample estimates:
mean in group female    mean in group male
      46.55358              53.25564

```



Mediante el método step, se busca el mejor modelo para nuestras variables, resultando el siguiente:

```
>step_mer<step(VO2Max_lm,~Posició*Edad*Posició.per.categoria*Personal..Gènere)
Step:  AIC=376.44
VO2Max.ml.kg.min. ~ Posició + Posició.per.categoria + Personal..Gènere +
  Posició:Posició.per.categoria + Posició:Personal..Gènere +
  Posició.per.categoria:Personal..Gènere +
  Posició:Posició.per.categoria:Personal..Gènere
```

	Df	Sum of Sq	RSS	AIC
<none>			994.34	376.44
+ Edad	1	5.261	989.08	376.82
- Posició:Posició.per.categoria:Personal..Gènere	1	55.745	1050.08	391.08

```
> summary(step_mer)

Call:
lm(formula = VO2Max.ml.kg.min. ~ Posició + Posició.per.categoria +
  Personal..Gènere + Posició:Posició.per.categoria + Posició:Personal..Gènere +
  Posició.per.categoria:Personal..Gènere +
  Posició:Posició.per.categoria:Personal..Gènere)

Residuals:
    Min       1Q   Median       3Q      Max
-6.7173 -0.8462 -0.0799  1.2497 10.0667

Coefficients:
(Intercept)          6.379e+01  1.434e+00
Posició             -4.668e-03  8.476e-04
Posició.per.categoria  3.606e-02  1.150e-02
Personal..Gènere     5.388e-01  1.461e+00
Posició:Posició.per.categoria -2.414e-06  7.987e-07
Posició:Personal..Gènere     -1.766e-02  2.316e-03
```

```

Posició.per.categoria:Personal..Gèneremale      -1.977e-02  1.167e-02
Posició:Posició.per.categoria:Personal..Gèneremale  3.271e-06  8.017e-07
t value Pr(>|t|)
(Intercept)                                44.475 < 2e-16 ***
Posició                                     -5.507 7.90e-08 ***
Posició.per.categoria                       3.135 0.00189 **
Personal..Gèneremale                        0.369 0.71256
Posició:Posició.per.categoria              -3.023 0.00272 **
Posició:Personal..Gèneremale               -7.626 3.29e-13 ***
Posició.per.categoria:Personal..Gèneremale -1.693 0.09142 .
Posició:Posició.per.categoria:Personal..Gèneremale  4.080 5.78e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.83 on 297 degrees of freedom
Multiple R-squared:  0.9351,    Adjusted R-squared:  0.9335
F-statistic:    611 on 7 and 297 DF,  p-value: < 2.2e-16

```

4.3 Aplicación práctica de resultados.

Se obtienen los resultados de las carreras Selva marítima y Xtreme de la web oficial <http://www.costabravaxtremrunning.com/es/>. La carrera de Selva marítima tiene una distancia de 25 Km y presenta un desnivel de 807 metros. Con esta información y teniendo en cuenta el tiempo empleado en completar la carrera, se aplica Tokmakidis para el cálculo del VO₂Max para cada corredor. El resultado obtenido de realizar un promedio del VO₂Max para cada uno de los corredores es de 41,43.

La carrera Xtreme se compone de tres etapas, la primera corresponde a realizar el trayecto de Selva marítima, es decir, 25 Km con un desnivel de 807 metros, la segunda etapa tiene una distancia de 55 Km con un desnivel de 1.470 m, la tercera etapa tiene una distancia de 55 Km con un desnivel de 1.500 m. Se calcula el VO₂Max resultante para cada una de las etapas por separado y posteriormente se calcula un promedio de las 3. Realizando un promedio para todos los corredores, el resultado obtenido es:

- 1ª etapa: 25,85
- 2ª etapa: 33,39
- 3ª etapa: 33,71

Realizando el promedio de las 3 etapas, nos da un resultado de 30,98.

Observando estos resultados, se diría que la primera etapa es la más dura de todas. De hecho, si analizamos el desnivel de cada una respecto a su kilometraje, se obtiene una proporción mayor para la primera etapa:

Distancia	Desnivel	Proporción
25.000	807	0,032
55.000	1.470	0,027
55.000	1.500	0,027

Se observa además que la primera etapa es más rápida, la velocidad media calculada es de 2,25 Km/h, siendo 1,94 Km/h para la segunda y 2,06 Km/h para la tercera. Estos factores pueden ser indicativos de que la primera etapa será la más dura de las 3 disputadas.

Según el modelo encontrado para el índice de dureza, sustituyendo por los valores de las etapas obtenemos los siguientes resultados:

$$\text{VO}_2\text{Max} = - 0.00067196 * \text{metros} - 0.003115 * \text{desnivel} + 0.0000002552 * \text{metros} * \text{desnivel} + 61.381$$

INDICE TEÓRICO			SELVA MARITIMA		XTREME	
Distancia	Desnivel	Indice	Valor Real	Desviación	Valor Real	Desviación
25.000	807	47,217	41,433	-12,25%	25,850	-45,25%
55.000	1.470	40,477			33,390	-17,51%
55.000	1.500	40,805			33,710	-17,39%

En la estimación del modelo se había eliminado la carrera de Selva marítima por ser un claro outlier, se comprueba que efectivamente se aleja mucho el índice calculado de la realidad. Esta carrera es especialmente dura ya que presenta un desnivel considerable con bastantes subidas y bajadas, dificultando mucho la prueba en comparación con otras carreras con el mismo desnivel. Este es el motivo por el que el índice de dureza encontrado se adapte menos a los datos empíricos.

Se obtiene también los resultados de la carrera de Policías de Manlleu de 10 Km, disputada el pasado 13 de mayo. Se obtienen los resultados de la web oficial <http://www.cursapolicia.com/arxiu/classif12/Res2012.pdf>. Esta cursa tiene un desnivel de 118 m.

Según el modelo encontrado y teniendo en cuenta que se trata de una carrera de asfalto, el resultado teórico de VO₂Max sería de 54,78

$$\text{VO}_2\text{Max} = 0.00009154 * \text{metros} - 0.003115 * \text{desnivel} + 0.0000002552 * \text{metros} * \text{desnivel} + 53.93$$

Se realiza el cálculo de VO₂Max de los participantes de la carrera y se promedian dichos valores, obteniendo como resultado 52,84, resultado no muy alejado del teórico.

INDICE TEÓRICO			Policías	
Distancia	Desnivel	Índice	Valor Real	Desviación
10.000	118	54,780	52,840	-3,54%

5 Conclusiones

5.1 Datos experimentos

De toda la información recabada durante las seis semanas de tratamiento para cada corredor no se pueden sacar conclusiones por no disponer de información suficiente para extrapolar. Lamentablemente, no se ha conseguido suficiente apoyo debido a varios motivos: un poco de desconfianza, estar seis semanas cada atleta recogiendo sus datos modificando en algunos casos su entrenamiento (casos Equilibrio Muscular y Respiración), otros corredores decían que ellos hacían deporte para eliminar el estrés de la semana, no disponer de pulsímetro, etc. Pese a este serio contratiempo, no consideramos que hayamos perdido el tiempo ya que esta información recogida es importante para un futuro estudio que disponga de más tiempo, más corredores con un diseño a ser posible balanceado (mismo número de corredores para cada experimento). Con esta información y combinándola con el índice de dureza se podría realizar un seguimiento de un corredor y ver si su entrenamiento, bien sea para Asfalto o bien para Montaña, es el más acorde según su promedio de VO_2Max .

Habiendo realizado un estudio de la mejora con algunos corredores más y teniendo grupo de control y grupo experimental (ejercicios de respiración) se encuentra que el mejor modelo es el siguiente:

Grupo de control

$$\text{mejora} = 215.1789 - 1.8445 * t_{\text{experiencia}}$$

Grupo de respiración

$$\text{mejora} = 289.2678 - 1.8445 * t_{\text{experiencia}}$$

5.2 Datos de atletas en carreras.

El índice de dureza encontrado es:

$$VO_2Max = - 0.00067196 * \text{metros} - 0.003115 * \text{desnivel} + 0.0000002552 * \text{metros} * \text{desnivel} + 61.381$$

Los resultados presentados anteriormente son los obtenidos después de depurar la base de datos inicial de carreras de Eolymp. Contrastando los resultados obtenidos con una persona entendida en el mundo del atletismo, se han tenido que eliminar algunas carreras por los motivos que se describen a continuación:

- Selva Marítima es una carrera de montaña muy exigente que forma parte de un conjunto de carreras. La prueba se disputa en varios días llegando a disputarse un total de 135 km. Esta es una causa principal por la cual el VO_2Max obtenido en general sea tan pequeño (la gente no rinde al máximo reservando energías para los días siguientes). Como se quiere analizar las carreras disputadas en un día y dicha carrera no se ajusta a los parámetros del estudio, se procedió a su eliminación.
- 5 km La Maquinista. Esta carrera se disputa junto a otra carrera que es del doble de recorrido (2 vueltas al circuito) y coincide en unas fechas hacia el final de temporada. La gente preparada disputa la carrera de más distancia en lugar de esta carrera en la que participan personas con poco entrenamiento. Así pues, los atletas que se adaptan a nuestro estudio son los de 10 km *La Maquinista*.
- Behobia-San Sebastián. Se trata de una media maratón que se desarrolla en Asfalto pero que presenta un VO_2Max anormalmente bajo en comparación con otras medias maratones de Asfalto. Es una carrera que en vez de ser circular como suelen ser casi

todas las pruebas de estas características, es lineal pero con mucho desnivel. De hecho presenta dos importantes altos, Gaintxurizketa (Km. 8) y el Alto de Miracruz (Km. 17). Además de presentar muchas subidas y bajadas, cosa que es atípico en la mayoría de medias maratones de Asfalto que suelen ser bastante llanas.

Cabe mencionar que en el apartado de observaciones Influyentes a priori y posteriori, se detecta que la carrera Maratón de Collserola es muy influyente tanto a priori como a posteriori. Se decide volver a pedir asesoramiento para saber cómo actuar ante esta carrera. Se decide dejarla en el estudio ya que se trata de una prueba de montaña muy exigente, con un desnivel muy elevado con lo cual no es de extrañar que el VO_2Max sea bastante inferior al resto de carreras del mismo tipo de Suelo. Es una carrera muy interesante a la hora de determinar la dureza de una carrera de montaña.

5.3 Líneas futuras.

5.3.1 Vincular los dos estudios.

Conociendo el estado de forma de un corredor de forma individual, y conociendo el grado de dureza de una carrera, se podría conocer la marca que hará el atleta en esa carrera, es decir, en qué tiempo la completará.

El índice de dureza de una carrera se traduce como el VO_2Max medio de la carrera, así, si se multiplica el VO_2Max (atleta) por el índice de la dureza de la carrera obtendríamos el tiempo que el corredor tardará en completarla.

5.3.2 Mejora VO_2Max con respirador.

Pese a no tener datos suficientes para evaluar la eficacia del experimento, lo que sí que se aprecia observando los datos es que las personas que han realizado el experimento de los ejercicios de los músculos respiratorios han mejorado su VO_2max , incluso la apreciación de los propios corredores es favorable.

Conociendo este dato, sería interesante realizar un estudio para medir la mejora de los atletas que hacen ejercicios de respiración usando respirador y las que los hacen sin ayuda del mismo. A tal efecto, se harían dos grupos de estudio:

- Atletas que utilizan respirador
- Atletas que hacen los ejercicios de respiración sin ayuda de respirador

5.3.3 Efecto condiciones climatológicas.

Durante el ejercicio, el calor producido por los músculos activos excede el calor liberado por el cuerpo y la temperatura aumenta. La elevación de la temperatura corporal provoca un aumento en la sudoración y flujo sanguíneo a la piel. Como resultado, el calor es eliminado por la evaporación de sudor desde la piel, irradiando calor del cuerpo hacia zonas más frescas que rodean al deportista, el cual se pierde por el mecanismo de convección hacia el aire. Cuando la tasa de producción de calor durante el ejercicio iguala a la tasa de eliminación de calor por el cuerpo, la temperatura hará una meseta a un nivel elevado. Cuando se produce más calor que lo que el cuerpo puede perder, la temperatura se elevará a un nivel potencialmente perjudicial. Una alta temperatura y humedad ambientes contribuyen a aumentar el riesgo de lesiones térmicas, ya que producen la capacidad corporal de eliminar el calor.

5.3.4 Efecto Altura.

El aire está compuesto por dos gases, 21 % de oxígeno y 79 % de nitrógeno, siendo la composición del aire invariable, sin embargo varían las presiones de los gases, así la presión de oxígeno baja la misma proporción que la presión atmosférica en cuanto ganamos altura.

La presión atmosférica es una envoltura gaseosa que rodea la tierra con un espesor de aproximadamente 100 km y ejerce un determinado peso sobre la superficie de la tierra, la capa inferior de ésta se llama troposfera y condiciona la vida en la tierra, su espesor varía según la latitud, así es de 17.000 metros en el ecuador y solo de 6.000 metros en los polos, es decir que los síntomas de sufrimiento en la altura aparecen a menor altura en los polos que cerca del ecuador.

La presión de oxígeno disminuye con la altura y sabemos que la presión es la única fuerza que hace progresar al oxígeno desde el aire hacia las células del organismo, así la disminución de la presión con la altura es un problema o un obstáculo a la oxigenación de las células y es responsable de numerosas enfermedades de la altura e incluso la sensación de permanente asfixia y mal rendimiento físico e intelectual.

La baja concentración de oxígeno en la sangre produce una deficiencia de oxígeno dentro de las células del organismo llamada hipoxia celular. Esto activará una serie de mecanismos de adaptación que permite mantener un consumo de oxígeno adaptado a las necesidades en la altura.

En el transporte de oxígeno desde los pulmones hasta las células se producen adaptaciones que tienden a conservar la oxigenación. Las adaptaciones a realizar a 3 niveles:

- A nivel respiratorio, aumentando el volumen de aire respirado que constituye la hiperventilación.
- A nivel sanguíneo, multiplicando el número de transportadores que son los glóbulos rojos, produciendo la eritrocitosis.
- A nivel celular, permitiendo una mejor liberación de oxígeno en su transportador.

La altitud produce bastantes efectos secundarios a los que el hombre no está del todo adaptado, entre las más importantes son las siguientes:

- A mayor altitud menor presión del aire
- A mayor altitud menor cantidad de oxígeno
- A mayor altitud menor temperatura
- A mayor altitud menor humedad (baja higrometría)
- A mayor altitud menor protección contra la radiación solar
- A mayor altitud menor resguardo contra el viento

A mayor altitud menor presión en el aire quiere decir que cuanto más se asciende, la capa de aire que tenemos encima es menos espesa. Esto influye en nuestra respiración ya que aunque el porcentaje de oxígeno es el mismo en toda la troposfera (un 21 %) encontramos menor cantidad porque en la altura, al igual que los demás gases, el oxígeno se dilata al encontrar menor presión. Por cada inspiración que realizamos hay una disminución en la cantidad de oxígeno a la que estamos acostumbrados a respirar en una altitud más baja. Cuanto mayor es la altitud más grave es esta situación hasta llegar a un punto en que según la preparación y esfuerzo físico realizado no se puede continuar con la actividad.

5.3.5 Efecto latitud.

La fuerza de la gravedad en la superficie de la tierra no es exactamente igual en todos los sitios. Existen pequeñas variaciones de un lugar a otro. Hay varios factores que intervienen para que así ocurra como por ejemplo la latitud.

Debido a la rotación de la tierra, los cuerpos experimentan una fuerza centrífuga nula en los polos y máxima en el ecuador. Esta fuerza centrífuga hace disminuir el efecto de la atracción gravitatoria y la desvía de su dirección original hacia el centro de la tierra.

El campo gravitatorio aumenta también con la latitud debido a la forma de la tierra sobre todo en los polos donde la distancia hasta el centro de la tierra es menor que la del ecuador. La fuerza de gravedad es menor estando en el ecuador.

Cuanto mayor sea la fuerza de la gravedad el atleta pesará más, disminuyendo así el rendimiento en carrera.

5.3.6 Efecto Calendario (Inicio – Fin temporada).

El calendario de carreras está marcado por la llegada del calor ya que como está demostrado, esta época es la de menor rendimiento. Así pues agosto es el mes del año con menor número de carreras programadas y el mes en el que los atletas descansan de su actividad anual.

La llegada de septiembre significa la vuelta al calendario de carreras, y significa que la mayoría de atletas retoman su entrenamiento. Siendo ésta la época de menor estado de forma y por tanto donde se hacen peores marcas. A medida que vamos avanzando en la temporada, los atletas van mejorando también su forma física y por lo tanto sus marcas serán mejores.

5.3.7 Analizar clusters.

5.3.7.1 Grupos de edad.

Realizar análisis por grupos de edad marcando las diferencias de las características que desarrolla cada uno de los grupos:

- Hasta 25 años se desarrolla velocidad
- Hasta 40-45 años se va perdiendo velocidad y se desarrolla fuerza.
- A partir de 45 años se desarrolla resistencia.

5.3.7.2 Análisis Cuartil.

Analizando por cuartiles, tenemos los corredores con mejor forma física en el primer cuartil, por lo que supuestamente los resultados deberían ser diferentes.

5.3.8 Análisis de componentes principales.

Un camino alternativo, en caso de interesarnos reducir considerablemente el nº de variables del estudio, en función de 2-3 creadas a partir de éstas, es el análisis de componentes principales. Dicho análisis, difiere del modelo lineal, en que no tenemos definida una variable respuesta que queramos estudiar en función de unas variables explicativas. El criterio para saber qué número de factores o componentes principales cogemos se decide a través del gráfico de sedimentación o/y en función del porcentaje de variabilidad acumulada explicada. Cabe resaltar, que más de tres factores dificulta considerablemente la interpretación.

5.4 Aspectos mejorables.

5.4.1 Captación voluntarios para experimentos.

De haber podido captar un mayor número de voluntarios para nuestros experimentos, lo ideal hubiese sido disponer de una población homogénea y aleatoria, con el mismo número de participantes en cada uno de los experimentos con el fin de hacer un diseño balanceado. Con esta información se puede plantear una tipología de estudio diferente de la planteada, que sería en este caso, un análisis de modelos mixtos con medidas repetidas. En este tipo de estudios podríamos trabajar con VO₂Max o Frecuencia Media como variables respuesta. El objetivo sería cómo afecta si tal día se ha entrenado por Asfalto o bien por Montaña, la distancia recorrida dicho día, etc. Es decir, la particularidad del día de entreno controlando con las características generales de ese individuo (Sexo, Edad, Altura, Peso, etc.). En este tipo de diseños conviene distinguir dos tipos de varianzas:

- Varianza Intraindividual. Variación aleatoria de un mismo individuo.
- Varianza Interindividual. Variación aleatoria entre individuos.

Los modelos mixtos permiten describir un individuo, no sólo como miembro de una población, sino como un individuo con características propias.

5.4.2 Mayor duración del experimento.

Si se hubiese podido hacer que todos los individuos estuviesen un tiempo haciendo cada uno de los experimentos planteados podíamos haber realizado un estudio de datos repetidos, de manera que se puede observar el comportamiento de un solo corredor en todas las situaciones posibles.

5.4.3 Afinamiento índice de dureza.

Sería necesario más tiempo de análisis y más pruebas con el fin de ir afinando este índice. También sería necesario disponer de la orografía completa de las carreras, ya que el desnivel global a veces no es significativo de la gran dureza que presenta la prueba.

6 APENDICES

6.1 APENDICE A. Glosario

- **EPOC** Consumo de oxígeno post ejercicio. Indica la cantidad de oxígeno adicional que el cuerpo necesita para recuperarse después del ejercicio.
- **RR** Frecuencia cardíaca de reserva. Se obtiene restando la frecuencia cardíaca máxima de una persona de su frecuencia en reposo.
- **EXERTION** Evaluación del esfuerzo percibido. Proceso de adaptación individual a las cargas de entrenamiento.
- **TE (Training Effect)** Grado de mejora de condición aeróbica. Se determina comparando el nivel de EPOC medido durante entrenamiento con el nivel de rendimiento.

6.2 APENDICE B. Entrenamiento equilibrio muscular.

Ejercicios a realizar una vez por semana. Los atletas tienden a desarrollar más fuerza en una de las dos extremidades. Lo que se propone son unos ejercicios para intentar corregir este desajuste por medio de la capacidad del músculo para responder a posiciones y situaciones anormales, haciendo trabajo de equilibrio conjuntamente con movimientos explosivos de potencia. Este programa mejorará la velocidad del atleta.

Ejercicio 1. Backbridge con una Sola Pierna.

Echarse con el balón en la espalda y con una pierna en contacto firme con el suelo. Asegúrate de que la pierna que está en el suelo esté en un ángulo mayor a 90 grados y tu pie derecho puesto hacia adelante. Mantén la otra pierna en el aire por tres a cuatro series de 30 a 45 segundos dependiendo del nivel de desequilibrio de las piernas. Un compañero puede añadir un mayor reto creando inestabilidad con pequeños desplazamientos del balón.



Ejercicio 2. Equilibrio en Tabla de equilibrio.

Mantén el equilibrio en cada pierna sobre el tablero oscilante. Repite de tres a cuatro series de 30 a 45 segundos en cada pierna. Al final trata de llegar a 60 segundos en cada pierna.



Ejercicio 3. Equilibrio en el Tablero Oscilante con un Balón con Peso.

Si dominas el trabajo sobre la tabla de equilibrio, el siguiente nivel de progresión puedes ser agarrarse de un balón suave o balón medicinal. Repite de tres a cuatro series y de 30 a 45 segundos en cada pierna. Para un mayor reto, un asistente puede realizarte pases muy cortos desde diferentes ángulos (utilizar siempre poca carga para este ejercicio).



Ejercicio 4. Cajón Búlgaro.

Con una barra sobre la espalda (15Kg. aprox.), coloca el pie derecho en un cajón. Asegúrate de que el esfuerzo sea sobre la pierna colocada sobre el cajón. Súbete al cajón con la pierna posterior. Haz de tres a cuatro series de seis repeticiones en cada pierna.



Ejercicio 5. Levantamiento de Balón Suave.

Con un balón suave de 1,5 a 2 Kg., para principiantes en este ejercicio, para así permitir progresiones seguras hacia un balón suave de mayor peso (no utilizar balón medicinal, son demasiado duros). Coloca el balón entre los pies. Aprieta el balón con los pies y bájalo durante la fase de preparación. Luego, súbelo mientras lo levantas hacia arriba en el aire. Repite de tres a cuatro series de seis a diez repeticiones.



Ejercicio 6. Salto Sentadilla a una pierna.

Coloca una pierna en frente de la otra en posición de separación. Ve hacia abajo con las piernas y los brazos y expándete mientras mantienes la posición de sentadilla a una pierna. El aterrizaje debería realizar en la misma posición como en la fase de bajada. Haz de tres a cuatro series de seis saltos en cada pierna.



Ejercicio 7. Saltos L.

Coloca una pierna en una mesa haciendo el ángulo de la cadera aproximadamente a 90 grados del suelo. Asegúrate de tener una colchoneta o suelo suave de aterrizaje al realizar estas rutinas. Anda hacia abajo en una pierna y de forma rápida hacia arriba con la misma pierna. Haz de tres a cuatros series de seis a ocho repeticiones en cada pierna.



Ejercicio 8. Step Ups en Step.

Coloca un pie en un step y haz el movimiento hacia arriba con dicho pie mientras llevas la otra pierna hacia arriba. Asegúrate de mantener la misma acción del brazo como si estuvieras corriendo. Haz tres a cuatro series de seis a ocho repeticiones en cada pierna.



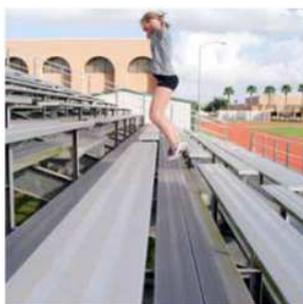
Ejercicio 9. Step Ups Rápidos de pie.

Coloca un step aeróbico en un suelo seguro para prevenir que se mueva. Tras la señal, el atleta saltará al step con la pierna/pie que usa generalmente para salir de los bloques. El atleta saltará sobre y fuera del step lo más rápido posible durante 30 segundos. Asegúrate de mantener la secuencia adecuada de los movimientos del brazo en el movimiento al correr.



Ejercicio 10. Saltos de Estadio.

Usa los steps de estadio de aluminio para hacer este ejercicio, de esta forma te proporcionará un aterrizaje mucho más suave. Coloca las manos detrás de la cabeza y ponte en cuclillas, luego de forma rápida salta hacia arriba al siguiente step. Realiza esto en aproximadamente 10 filas y baja andando y repite el mismo procedimiento de tres a cuatro veces. Step Ups Rápidos de Pie (Imagen 9) Coloca un step aeróbico en un suelo seguro para prevenir que se mueva. Tras la señal, el atleta saltará al step con la pierna/pie que usa generalmente para salir de los bloques. El atleta saltará sobre y fuera del step lo más rápido posible durante 30 segundos. Asegúrate de mantener la secuencia adecuada de los movimientos del brazo en el movimiento al correr.



Ejercicio 11. Saltos en una sola pierna.

Encuentra un área que sea suave y plana si estás fuera. Pon en fila seis conos y practica saltando sobre uno o dos para alinear la distancia apropiada entre conos. Inicia saltando en los conos y luego salta en una sola pierna sobre cada uno. Realiza este ejercicio de tres a cuatros veces por cada pierna.



Ejercicio 12. Lanzamientos a Cuchara.

Usa un balón suave o un balón medicinal de 8Kg. Ponte en cuclillas y levántate mientras lanzas el balón con la mayor fuerza posible. Realiza este ejercicio de 6 a 8 veces.



Ejercicio 13. Sprints inclinados.

Encuentra un área inclinada que no sea de mayor de 4 - 5%. La distancia necesitada debería ser de 25 a 35 metros de longitud. Tras la señal, comienza el sprint. Realiza este sprint seis a ocho veces.

6.3 APENDICE C. Entrenamiento respiratorio.

Primera fase. Tomar el aire durante 4 segundos, retener durante 16 segundos. Se expira el aire durante 8 segundos

Segunda fase. Se incorpora una fase de apnea, se toma el aire durante 4 segundos, se retiene durante 16 segundos, se suelta durante 8 segundos y se aguanta 4 segundos antes de volver a tomar aire.

Tercera fase. Se toma aire durante 4 segundos, retener durante 16 segundos soltar en 8 y retener 16 segundos más.

Practicar 15 minutos al día

6.4 APENDICE D. Formato ficheros HRM.

6.4.1 Parámetros generales.

DATA	COMMENTS
[Params]	Basic settings
Version=107	Exact hrm file version (1.02, 1.05*, 1.06#, 1.07&).
Monitor=1	Heart rate monitor type 1 = Polar Sport Tester / Vantage XL 2 = Polar Vantage NV (a.k.a. VNV) 3 = Polar Accurex Plus 4 = Polar Xtrainer Plus

	<p>7 = Polar Coach 8 = Polar S210 9 = Polar S410 10 = Polar S510 / S520 11 = Polar S610 / S610i 12 = Polar S710 / S710i / S720i 13 = Polar S810 / S810i 15 = Polar E600 20 = Polar AXN500 21 = Polar AXN700 22 = Polar S625X 23 = Polar S725</p>
<p>Mode=110 (abc) With versions <input type="checkbox"/> 1.05</p>	<p>Data types: a) Cad/Alt: 0 = Cad, 1 = Alt, 3 = None b) CC data 0 = HR data only, 1 = HR + cycling data c) US / Euro unit 0 = Euro (km, km/h, m) 1 = US (miles, mph, ft)</p> <p>All distance, speed and altitude values depend on US/Euro unit selection (km / miles, km/h / mph, m / ft).</p>
<p>SMode=11011010 (abcdefgh) With versions 1.06</p>	<p>Data type parameters a) Speed (0=off, 1=on) b) Cadence (0=off, 1=on) c) Altitude (0=off, 1=on) d) Power (0=off, 1=on) e) Power Left Right Balance (0=off, 1=on) f) Power Pedalling Index (0=off, 1=on) g) HR/CC data 0 = HR data only, 1 = HR + cycling data h) US / Euro unit 0 = Euro (km, km/h, m, °C) 1 = US (miles, mph, ft, °F)</p> <p>All distance, speed, altitude and temperature values depend on US/Euro unit selection (km / miles, km/h / mph, m / ft, °C / °F).</p>
<p>SMode=110110100 (abcdefghi) With versions 1.07 <input type="checkbox"/></p>	<p>i) Air pressure (0=off, 1=on)</p>
<p>Date=20010227</p>	<p>Date of exercise (yyyymmdd) For example 20010227 means 27th February 2001)</p>
<p>StartTime=06:23:36.0</p>	<p>Start time (hh:mm:ss.d)</p>
<p>Length=00:30:00.4</p>	<p>Length of exercise (hh:mm:ss.d)</p>
<p>Interval=5</p>	<p>Data type: 5 = 5 seconds recording interval 15 = 15 seconds recording interval 30 = 30 seconds recording interval</p>

	60 = 60 seconds recording interval
	120 = 120 seconds recording interval (dynamic)
	240 = 240 seconds recording interval (dynamic)
	300 = 5 minutes recording interval
	480 = 480 seconds recording interval (dynamic)
	238 = R - R data (VNV, S810, S810i)
	204 = intermediate times only (PST, VXL, VNV, XTr+, Acc+)
Upper1=160	Upper limit 1 (bpm)
Lower1=80	Lower limit 1 (bpm)
Upper2=160	Upper limit 2 (bpm)
Lower2=80	Lower limit 2 (bpm)
Upper3=160	Upper threshold / Upper limit 3 (bpm)
Lower3=80	Lower threshold / Lower limit 3 (bpm)
Timer1=00:00	Exercise timer 1 (mm:ss)
Timer2=00:00	Exercise timer 2 (mm:ss)
Timer3=00:00	Exercise timer 3 (mm:ss)
ActiveLimit=0	Limits in use in "File Summary": 0 = Limits 1 and 2, 1 = Treshold limits
MaxHR=195	Personal max heart rate (bpm)
RestHR=52	Personal resting heart rate (bpm)
StartDelay=300	RR Start delay (ms) (Vantage NV RR data only)
VO2max=50	VO2max at time of exercise (for calories calculation) ml/min/kg #
Weight=75	Weight at time of exercise (for calories calculation) kg#

6.5 APENDICE E. Resultados R (listados salida).

6.5.1 Resultados Experimentos

6.5.1.1 Análisis descriptivo

```
> summary(corredors)
  corredor      Edad      Sexo      Altura      Peso
Alba      :1  Min.    :11.00  H:8  Min.    :148.0  Min.    :41.00
Alejandro:1  1st Qu.:32.50  M:3  1st Qu.:161.0  1st Qu.:58.00
Alicia   :1  Median  :38.00           Median :172.0  Median  :65.00
Antonio  :1  Mean    :38.45           Mean   :170.5  Mean    :66.91
Bea      :1  3rd Qu.:48.00           3rd Qu.:180.5  3rd Qu.:75.00
Javi     :1  Max.    :55.00           Max.    :188.0  Max.    :99.00
(Other)  :5

H_entr_semana  t_experiencia  experimento  Cooper_1  Cooper_2
Min.    :1.000  Min.    : 1.00  Min.    :1.000  Min.    :1668  Min.    :2019
1st Qu.:2.000  1st Qu.: 4.00  1st Qu.:1.000  1st Qu.:2290  1st Qu.:2425
Median :7.000  Median : 60.00  Median :2.000  Median :3000  Median :3100
Mean   :5.091  Mean   : 64.27  Mean   :1.909  Mean   :2674  Mean   :2839
3rd Qu.:7.000  3rd Qu.: 90.00  3rd Qu.:3.000  3rd Qu.:3120  3rd Qu.:3244
Max.    :7.000  Max.    :240.00  Max.    :3.000  Max.    :3300  Max.    :3400

Hora_Inicio  Tipo_suelo  Frec_Media  Distancia  Minutos
M:2          Asphalt :5  Min.    :128.1  Min.    : 3228  Min.    :25.00
T:9          Mixto  :4  1st Qu.:137.4  1st Qu.: 4820  1st Qu.:32.00
           Montana:2  Median  :144.5  Median : 9697  Median :49.00
           Mean   :145.7  Mean   : 8519  Mean   :46.82
           3rd Qu.:154.6  3rd Qu.:11800  3rd Qu.:61.00
           Max.    :171.2  Max.    :12242  Max.    :65.00
           NA's   : 1.0  NA's    : 1
```

m_minuto	VO2Max	millora	experimento_r
Min. :107.6	Min. :17.08	Min. : 50.0	Neutre :5
1st Qu.:151.2	1st Qu.:28.00	1st Qu.:100.0	Equilibri :2
Median :181.5	Median :36.47	Median :150.0	Respiració:4
Mean :182.8	Mean :35.02	Mean :164.5	
3rd Qu.:211.4	3rd Qu.:41.24	3rd Qu.:199.0	
Max. :255.1	Max. :52.50	Max. :351.0	
NA's : 1.0	NA's : 1.00		

6.5.1.2 Búsqueda modelo Cooper_2

```
> corredors1.coop2<-lm(Cooper_2~1)
> step1_cp2<-
step(corredors1.coop2,~experimento_r+Frec_Media+Sexo+Edad+Altura+Peso+H_entr_semana+
m_minuto+t_experiencia+Cooper_1+Tipo_suelo)
Start: AIC=150.14
Cooper_2 ~ 1
```

	Df	Sum of Sq	RSS	AIC
+ Cooper_1	1	1803486	32903	84.987
+ m_minuto	1	1203775	632614	114.550
+ experimento_r	2	1148760	687629	117.384
+ H_entr_semana	1	627708	1208680	121.025
+ Edad	1	624726	1211663	121.049
+ t_experiencia	1	520969	1315420	121.871
<none>			1836389	123.207
+ Altura	1	257011	1579378	123.700
+ Tipo_suelo	1	218744	1617645	123.939
+ Sexo	1	207264	1629125	124.010
+ Frec_Media	1	114194	1722195	124.565
+ Peso	1	11	1836378	125.207

```
Step: AIC=103.31
Cooper_2 ~ Cooper_1
```

	Df	Sum of Sq	RSS	AIC
+ t_experiencia	1	10059	22844	83.339
+ experimento_r	2	12404	20499	84.255
+ Sexo	1	6862	26040	84.648
+ Frec_Media	1	3703	29200	85.793
+ Peso	1	3684	29218	85.800
+ m_minuto	1	2530	30373	86.187
+ Altura	1	1876	31027	86.400
+ Edad	1	1713	31190	86.453
+ H_entr_semana	1	805	32098	86.740
+ Tipo_suelo	1	290	32613	86.899
<none>			47142	103.312
- Cooper_1	1	2711564	2758706	150.144

```
Step: AIC=104.14
Cooper_2 ~ Cooper_1 + t_experiencia
```

Mensajes de aviso perdidos
 1: In add1.lm(fit, scope\$add, scale = scale, trace = trace, k = k, :
 using the 10/12 rows from a combined fit
 2: In add1.lm(fit, scope\$add, scale = scale, trace = trace, k = k, :
 using the 10/12 rows from a combined fit

Modelo sin Cooper_1

```
step2_cp2<-
step(corredors1.coop2,~experimento_r+Frec_Media+Sexo+Edad+Altura+Peso+H_entr_semana+
m_minuto+t_experiencia+Tipo_suelo,data=corredors)
Start: AIC=150.14
Cooper_2 ~ 1
```

	Df	Sum of Sq	RSS	AIC
+ m_minuto	1	1203775	632614	114.55
+ experimento_r	2	1148760	687629	117.38
+ H_entr_semana	1	627708	1208680	121.03
+ Edad	1	624726	1211663	121.05
+ t_experiencia	1	520969	1315420	121.87
<none>			1836389	123.21
+ Altura	1	257011	1579378	123.70

```

+ Tipo_suelo      1      218744 1617645 123.94
+ Sexo            1      207264 1629125 124.01
+ Frec_Media      1      114194 1722195 124.56
+ Peso            1           11 1836378 125.21
Error en step(corredors1.coop2, ~experimento_r + Frec_Media + Sexo + Edad + :
  number of rows in use has changed: remove missing values?
Además: Mensajes de aviso perdidos
In add1.lm(fit, scope$add, scale = scale, trace = trace, k = k, :
  using the 10/12 rows from a combined fit

```

Modelo Subconjunto corredores (con información m_minuto)

```

> corredors2.coop2<-lm(Cooper_2~1,data=corredors2)
> step3_cp2<-
step(corredors2.coop2,~experimento_r+Frec_Media+Sexo+Edad+Altura+Peso+H_entr_semana+
m_minuto+t_experiencia+Tipo_suelo,data=corredors2)
Start: AIC=123.21
Cooper_2 ~ 1

```

	Df	Sum of Sq	RSS	AIC
+ m_minuto	1	1203775	632614	114.55
+ experimento_r	2	1148760	687629	117.38
+ H_entr_semana	1	627708	1208680	121.03
+ Edad	1	624726	1211663	121.05
+ t_experiencia	1	520969	1315420	121.87
<none>			1836389	123.21
+ Altura	1	257011	1579378	123.70
+ Tipo_suelo	1	218744	1617645	123.94
+ Sexo	1	207264	1629125	124.01
+ Frec_Media	1	114194	1722195	124.56
+ Peso	1	11	1836378	125.21

Step: AIC=114.55
Cooper_2 ~ m_minuto

	Df	Sum of Sq	RSS	AIC
+ Frec_Media	1	207771	424843	112.57
+ t_experiencia	1	155533	477081	113.73
+ experimento_r	2	229374	403240	114.05
<none>			632614	114.55
+ Edad	1	82202	550411	115.16
+ Sexo	1	76063	556551	115.27
+ Tipo_suelo	1	71088	561526	115.36
+ H_entr_semana	1	70156	562458	115.38
+ Altura	1	36086	596527	115.96
+ Peso	1	244	632370	116.55
- m_minuto	1	1203775	1836389	123.21

Step: AIC=112.57
Cooper_2 ~ m_minuto + Frec_Media

	Df	Sum of Sq	RSS	AIC
+ H_entr_semana	1	177583	247260	109.16
+ Sexo	1	141962	282881	110.50
<none>			424843	112.57
+ experimento_r	2	120243	304600	113.24
+ Altura	1	20600	404243	114.07
+ t_experiencia	1	14488	410355	114.22
+ Tipo_suelo	1	3432	421411	114.49
+ Peso	1	920	423923	114.55
- Frec_Media	1	207771	632614	114.55
+ Edad	1	13	424830	114.57
- m_minuto	1	1297352	1722195	124.56

Step: AIC=109.16
Cooper_2 ~ m_minuto + Frec_Media + H_entr_semana

	Df	Sum of Sq	RSS	AIC
+ experimento_r	2	144240	103020	104.40
<none>			247260	109.16
+ Tipo_suelo	1	25294	221966	110.08
+ Sexo	1	22326	224934	110.21
+ Altura	1	13510	233750	110.59
+ t_experiencia	1	9165	238095	110.78
+ Peso	1	2645	244614	111.05

```
+ Edad          1          75  247185 111.15
- H_entr_semana 1  177583  424843 112.57
- Frec_Media    1  315198  562458 115.38
- m_minuto      1  830070 1077330 121.87
```

Step: AIC=104.4

Cooper_2 ~ m_minuto + Frec_Media + H_entr_semana + experimento_r

	Df	Sum of Sq	RSS	AIC
+ Tipo_suelo	1	88538	14482	86.781
<none>			103020	104.401
+ Sexo	1	12312	90708	105.128
+ Peso	1	4970	98050	105.906
+ Edad	1	3962	99058	106.009
+ Altura	1	3790	99230	106.026
+ t_experiencia	1	2480	100540	106.157
- experimento_r	2	144240	247260	109.156
- H_entr_semana	1	201580	304600	113.242
- Frec_Media	1	287523	390544	115.727
- m_minuto	1	509115	612136	120.221

Step: AIC=86.78

Cooper_2 ~ m_minuto + Frec_Media + H_entr_semana + experimento_r +
Tipo_suelo

	Df	Sum of Sq	RSS	AIC
+ Sexo	1	12191	2292	70.345
+ Edad	1	3238	11244	86.250
<none>			14482	86.781
+ t_experiencia	1	2296	12187	87.055
+ Peso	1	2185	12297	87.145
+ Altura	1	1550	12932	87.649
- Tipo_suelo	1	88538	103020	104.401
- experimento_r	2	207484	221966	110.077
- H_entr_semana	1	250541	265023	113.850
- Frec_Media	1	253656	268138	113.967
- m_minuto	1	448084	462566	119.420

Step: AIC=70.34

Cooper_2 ~ m_minuto + Frec_Media + H_entr_semana + experimento_r +
Tipo_suelo + Sexo

	Df	Sum of Sq	RSS	AIC
+ t_experiencia	1	2275	16	22.994
+ Altura	1	474	1818	70.027
<none>			2292	70.345
+ Peso	1	201	2090	71.425
+ Edad	1	32	2260	72.203
- Sexo	1	12191	14482	86.781
- Tipo_suelo	1	88416	90708	105.128
- experimento_r	2	182333	184625	110.235
- H_entr_semana	1	187110	189401	112.490
- Frec_Media	1	227553	229845	114.426
- m_minuto	1	398211	400502	119.979

Step: AIC=22.99

Cooper_2 ~ m_minuto + Frec_Media + H_entr_semana + experimento_r +
Tipo_suelo + Sexo + t_experiencia

	Df	Sum of Sq	RSS	AIC
+ Edad	1	16	0	-Inf
+ Altura	1	16	0	-Inf
+ Peso	1	16	0	-Inf
<none>			16	23
- t_experiencia	1	2275	2292	70
- Sexo	1	12170	12187	87
- Tipo_suelo	1	88198	88214	107
- experimento_r	2	180409	180426	112
- H_entr_semana	1	184757	184774	114
- Frec_Media	1	229818	229834	116
- m_minuto	1	393730	393747	122

Step: AIC=-Inf

Cooper_2 ~ m_minuto + Frec_Media + H_entr_semana + experimento_r +

```
Tipo_suelo + Sexo + t_experiencia + Edad
      Df Sum of Sq  RSS  AIC
<none>                0 -Inf
- Edad                1    16    16   23
- t_experiencia      1  2260  2260   72
- Sexo               1  7855  7855   85
- Tipo_suelo         1 88135 88135  109
- experimento_r      2 174356 174356  114
- H_entr_semana     1 151486 151486  114
- Frec_Media         1 170473 170473  115
- m_minuto           1 254929 254929  119
Mensajes de aviso perdidos
1: attempting model selection on an essentially perfect fit is nonsense
2: attempting model selection on an essentially perfect fit is nonsense
```

```
Cooper_2 ~ m_minuto + Frec_Media + H_entr_semana + experimento_r +
  Tipo_suelo + Sexo + t_experiencia
```

```
> cp2_scp1.0<-lm(Cooper_2~1,data=corredors2)
> cp2_scp1.1<-lm(Cooper_2~m_minuto,data=corredors2)
> anova(cp2_scp1.0,cp2_scp1.1)
Analysis of Variance Table
```

```
Model 1: Cooper_2 ~ 1
Model 2: Coop
er_2 ~ m_minuto
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1      9 1836389
2      8 632614  1  1203775 15.223 0.004534 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> cp2_scp1.2<-lm(Cooper_2~m_minuto+Frec_Media,data=corredors2)
> anova(cp2_scp1.1,cp2_scp1.2)
Analysis of Variance Table
```

```
Model 1: Cooper_2 ~ m_minuto
Model 2: Cooper_2 ~ m_minuto + Frec_Media
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1      8 632614
2      7 424843  1   207771 3.4234 0.1067
```

6.5.1.3 Búsqueda modelo Mejora

```
> corredors1.millora<-lm(millora~1)
> step1_millora<-
step(corredors1.millora,~experimento_r+Frec_Media+Sexo+Edad+Altura+Peso+H_entr_semana+
m_minuto+t_experiencia+Tipo_suelo)
Start: AIC=109.15
millora ~ 1
```

```
      Df Sum of Sq  RSS  AIC
+ t_experiencia  1  23481.7 24275 81.946
+ Edad           1 11534.6 36222 85.948
+ H_entr_semana  1  9312.5 38444 86.544
+ Frec_Media     1  8852.3 38905 86.663
<none>                47757 86.713
+ m_minuto       1  4387.7 43369 87.749
+ Peso           1  4335.9 43421 87.761
+ experimento_r  2 11622.1 36135 87.924
+ Tipo_suelo     1  3572.1 44185 87.936
+ Sexo           1  2064.0 45693 88.271
+ Altura         1     2.9 47754 88.712
```

```
Step: AIC=105.43
millora ~ t_experiencia
```

```
      Df Sum of Sq  RSS  AIC
+ Peso           1  10748 13527 78.098
+ Altura         1   6575 17701 80.788
+ Sexo           1  3340 20936 82.466
+ m_minuto       1   347 23928 83.802
```

```

+ Edad          1          248 24027  83.843
+ Tipo_suelo    1          203 24072  83.862
+ H_entr_semana 1           24 24251  83.936
+ Frec_Media    1           10 24265  83.942
+ experimento_r 2         3451 20824  84.413
<none>                    56249 105.432
- t_experiencia 1        34325 90574 109.148

```

Step: AIC=102.97
millora ~ t_experiencia + Peso

	Df	Sum of Sq	RSS	AIC
+ Altura	1	1853	11674	78.625
+ Edad	1	1845	11682	78.632
+ H_entr_semana	1	837	12689	79.459
+ Sexo	1	798	12728	79.490
+ experimento_r	2	2947	10580	79.641
+ Frec_Media	1	384	13143	79.810
+ Tipo_suelo	1	203	13324	79.947
+ m_minuto	1	86	13441	80.035
<none>			38789	102.972
- Peso	1	17460	56249	105.432
- t_experiencia	1	46887	85675	110.481

Step: AIC=102.51
millora ~ t_experiencia + Peso + Altura

	Df	Sum of Sq	RSS	AIC
+ experimento_r	2	10914.8	759	55.296
+ Sexo	1	1808.9	9865	78.942
+ H_entr_semana	1	1339.7	10334	79.406
+ m_minuto	1	979.4	10695	79.749
+ Edad	1	894.2	10780	79.828
+ Tipo_suelo	1	75.9	11598	80.560
+ Frec_Media	1	45.3	11629	80.586
<none>			31590	102.508
- Altura	1	7198.3	38789	102.972
- t_experiencia	1	8727.9	40318	103.436
- Peso	1	16345.4	47936	105.512

Step: AIC=98.28
millora ~ t_experiencia + Peso + Altura + experimento_r

	Df	Sum of Sq	RSS	AIC
+ Tipo_suelo	1	407.7	351	49.593
+ Sexo	1	163.2	596	54.876
+ m_minuto	1	111.1	648	55.713
+ Frec_Media	1	58.9	700	56.488
+ H_entr_semana	1	58.9	700	56.488
+ Edad	1	16.6	743	57.075
- t_experiencia	1	2237.2	18147	97.856
<none>			15910	98.277
- experimento_r	2	15680.6	31590	102.508
- Altura	1	19596.2	35506	105.911
- Peso	1	28820.8	44730	108.682

Step: AIC=99.18
millora ~ t_experiencia + Peso + Altura + experimento_r + Tipo_suelo

Mensajes de aviso perdidos

```

1: In add1.lm(fit, scope$add, scale = scale, trace = trace, k = k, :
  using the 10/12 rows from a combined fit
2: In add1.lm(fit, scope$add, scale = scale, trace = trace, k = k, :
  using the 10/12 rows from a combined fit
3: In add1.lm(fit, scope$add, scale = scale, trace = trace, k = k, :
  using the 10/12 rows from a combined fit
4: In add1.lm(fit, scope$add, scale = scale, trace = trace, k = k, :
  using the 10/12 rows from a combined fit
5: In add1.lm(fit, scope$add, scale = scale, trace = trace, k = k, :
  using the 10/12 rows from a combined fit

```

Analysis of Variance Table

Model 1: millora ~ 1

```

Model 2: millora ~ t_experiencia
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1      11 90574
2      10 56249  1    34325 6.1023 0.03309 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> anova(corredors1.millora,corredors1.millora3)
Analysis of Variance Table

Model 1: millora ~ 1
Model 2: millora ~ Peso
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1      11 90574
2      10 85675  1    4898.2 0.5717 0.467
> anova(corredors1.millora,corredors1.millora4)
Analysis of Variance Table

Model 1: millora ~ 1
Model 2: millora ~ Altura
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1      11 90574
2      10 89723  1    850.67 0.0948 0.7645
> anova(corredors1.millora,corredors1.millora5)
Analysis of Variance Table

Model 1: millora ~ 1
Model 2: millora ~ experimento_r
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1      11 90574
2       9 73665  2    16908 1.0329 0.3946
> corredors1.millora6<-lm(millora~t_experiencia+experimento_r)
> anova(corredors1.millora2,corredors1.millora6)
Analysis of Variance Table

Model 1: millora ~ t_experiencia
Model 2: millora ~ t_experiencia + experimento_r
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1      10 56249
2       8 53170  2    3079.4 0.2317 0.7984
> corredors1.millora7<-lm(millora~t_experiencia+Peso)
> anova(corredors1.millora2,corredors1.millora7)
Analysis of Variance Table

Model 1: millora ~ t_experiencia
Model 2: millora ~ t_experiencia + Peso
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1      10 56249
2       9 38789  1    17460 4.0513 0.07499 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> corredors1.millora8<-lm(millora~t_experiencia+Altura)
> anova(corredors1.millora2,corredors1.millora8)
Analysis of Variance Table

Model 1: millora ~ t_experiencia
Model 2: millora ~ t_experiencia + Altura
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1      10 56249
2       9 47936  1    8313.4 1.5609 0.2431
> corredors1.millora9<-lm(millora~t_experiencia+Peso+Altura)
> anova(corredors1.millora7,corredors1.millora9)
Analysis of Variance Table

Model 1: millora ~ t_experiencia + Peso
Model 2: millora ~ t_experiencia + Peso + Altura
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1       9 38789
2       8 31590  1    7198.3 1.8229 0.2139
> corredors1.millora10<-lm(millora~t_experiencia+Peso+experimento_r)
> anova(corredors1.millora7,corredors1.millora10)
Analysis of Variance Table

Model 1: millora ~ t_experiencia + Peso
Model 2: millora ~ t_experiencia + Peso + experimento_r

```

```

Res.Df  RSS Df Sum of Sq    F Pr(>F)
1      9 38789
2      7 35506  2    3282.7 0.3236 0.7338

```

6.5.2 Resultados Experimentos (Ampliación de datos)

6.5.2.1 Análisis descriptivo

```

> summary(corredores)
  corredor      Edad      Sexo      Altura      Peso
Alba      :1  Min.    :11.00  H:9  Min.    :148.0  Min.    :41.00
Alejandro:1 1st Qu.:32.50  M:3  1st Qu.:160.5  1st Qu.:59.25
Alicia    :1 Median  :36.50      Median  :169.0  Median  :64.50
Antonio   :1 Mean    :36.08      Mean    :168.8  Mean    :65.67
Bea       :1 3rd Qu.:44.25  3rd Qu.:177.2  3rd Qu.:70.00
eo73     :1 Max.    :50.00      Max.    :188.0  Max.    :99.00
(Other)   :6      NA's    : 2.0

H_entr_semana  t_experiencia  experimento  Cooper_1  Cooper_2
Min.    :1.000  Min.    : 1.00  Min.    :1.000  Min.    :1668  Min.    :2019
1st Qu.:2.000  1st Qu.: 5.00  1st Qu.:1.000  1st Qu.:2343  1st Qu.:2294
Median  :6.500  Median  :60.00  Median  :1.000  Median  :2479  Median  :2575
Mean    :5.083  Mean    :57.92  Mean    :1.667  Mean    :2582  Mean    :2718
3rd Qu.:7.000  3rd Qu.:72.00  3rd Qu.:3.000  3rd Qu.:3030  3rd Qu.:3173
Max.    :9.000  Max.    :180.00  Max.    :3.000  Max.    :3200  Max.    :3400

Hora_Inicio  Tipo_suelo  Frec_Media  Distancia  Minutos
M:4          Asfalto:6  Min.    :128.1  Min.    : 3228  Min.    :30.00
T:8          Otro   :6  1st Qu.:137.7  1st Qu.: 6358  1st Qu.:37.50
              Median  :148.0  Median  :11133  Median  :57.00
              Mean    :146.2  Mean    : 9254  Mean    :52.25
              3rd Qu.:153.0  3rd Qu.:11950  3rd Qu.:64.25
              Max.    :171.2  Max.    :12242  Max.    :75.00
              NA's    : 1.0  NA's    : 1

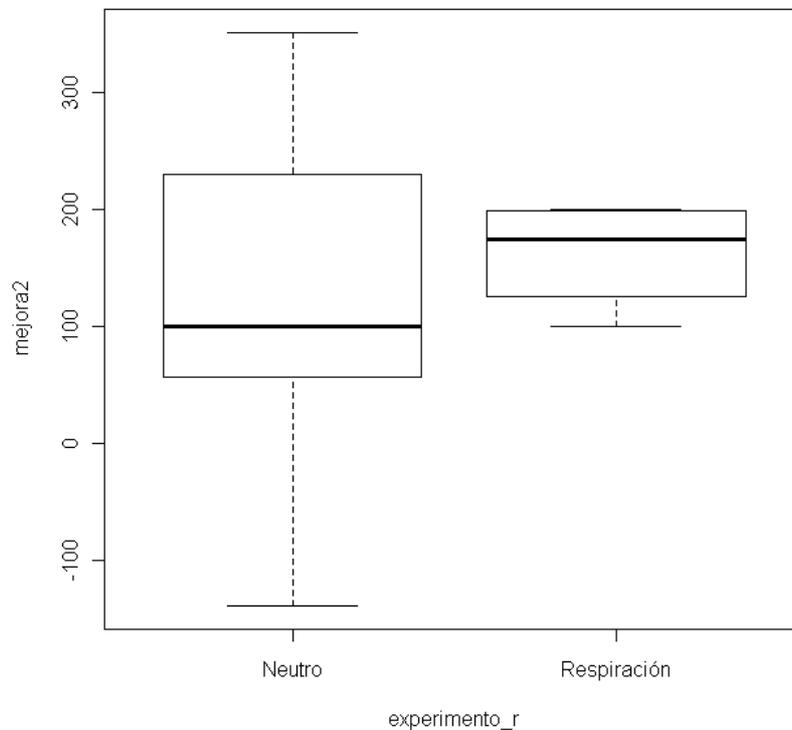
  m_minuto      VO2Max      experimento_r
Min.    :107.6  Min.    :17.08  Control      :8
1st Qu.:150.1  1st Qu.:28.13  Respiración:4
Median  :181.5  Median  :35.76
Mean    :177.7  Mean    :33.64
3rd Qu.:208.5  3rd Qu.:40.17
Max.    :255.1  Max.    :44.88
NA's    : 1.0  NA's    : 1.00

> oneway.test(mejora2 ~ experimento_r, data=corredores)

One-way analysis of means (not assuming equal variances)

data:  mejora2 and experimento_r
F = 0.4419, num df = 1.000, denom df = 9.178, p-value =
0.5225

```



6.5.2.2 Normalidad variable respuesta

```
> shapiro.test(mejora)

Shapiro-Wilk normality test

data:  mejora
W = 0.9532, p-value = 0.6847

> shapiro.test(mejora[experimento_r=="Control"])

Shapiro-Wilk normality test

data:  mejora[experimento_r == "Control"]
W = 0.9566, p-value = 0.7771

> shapiro.test(mejora[experimento_r=="Respiración"])

Shapiro-Wilk normality test

data:  mejora[experimento_r == "Respiración"]
W = 0.8746, p-value = 0.3162
```

6.5.2.3 Búsqueda modelo

```
> corredores.mejora<-lm(mejora~1, data=corredores)
> step1_mejora<-
step(corredores.mejora,~experimento_r+Frec_Media+Sexo+Edad+Altura+Peso+H_entr_semana+
+ +m_minuto+t_experiencia+Tipo_suelo,data=corredores)
Start:  AIC=116.85
mejora ~ 1

          Df Sum of Sq  RSS   AIC
+ t_experiencia  1  13190.7 21722  67.253
+ Peso           1   7789.3 27124  69.030
<none>                    34913  69.049
+ Edad          1   4431.5 30481  69.963
+ Sexo          1   4131.4 30781  70.042
```

```
+ m_minuto      1      3506.4 31407 70.203
+ H_entr_semana 1      3438.5 31474 70.220
+ Tipo_suelo    1      2415.1 32498 70.476
+ Frec_Media    1      2371.0 32542 70.487
+ Altura        1      1160.0 33753 70.779
+ experimento_r 1         28.1 34885 71.043
```

```
Step: AIC=106.18
mejora ~ t_experiencia
```

	Df	Sum of Sq	RSS	AIC
+ Peso	1	10452	11270	64.004
+ Altura	1	8753	12969	65.127
+ experimento_r	1	6260	15462	66.534
+ Sexo	1	2732	18990	68.178
+ H_entr_semana	1	2581	19141	68.241
+ m_minuto	1	313	21409	69.137
+ Tipo_suelo	1	299	21423	69.142
+ Edad	1	146	21576	69.199
+ Frec_Media	1	59	21664	69.232
<none>			59871	106.180
- t_experiencia	1	112253	172124	116.853

```
Step: AIC=102.72
mejora ~ t_experiencia + Peso
```

	Df	Sum of Sq	RSS	AIC
+ H_entr_semana	1	6062	5208	59.828
+ experimento_r	1	3635	7635	62.889
+ Edad	1	994	10276	65.265
+ Frec_Media	1	919	10351	65.323
+ Tipo_suelo	1	653	10617	65.526
+ Sexo	1	410	10860	65.708
+ m_minuto	1	370	10900	65.737
+ Altura	1	294	10976	65.792
<none>			37973	102.717
- Peso	1	21898	59871	106.180
- t_experiencia	1	123900	161872	118.116

```
Step: AIC=104.63
mejora ~ t_experiencia + Peso + H_entr_semana
```

Mensajes de aviso perdidos

```
1: In add1.lm(fit, scope$add, scale = scale, trace = trace, k = k, :
  using the 8/12 rows from a combined fit
2: In add1.lm(fit, scope$add, scale = scale, trace = trace, k = k, :
  using the 8/12 rows from a combined fit
3: In add1.lm(fit, scope$add, scale = scale, trace = trace, k = k, :
  using the 8/12 rows from a combined fit
```

```
corredores2<-na.omit(corredores)
mejora.lm00<-lm(mejora~1,data=corredores)
mejora.lm0<-lm(mejora~1,data=corredores2)
mejora.lm1<-lm(mejora~Frec_Media,data=corredores2)
mejora.lm2<-lm(mejora~Edad,data=corredores)
mejora.lm3<-lm(mejora~Altura,data=corredores2)
mejora.lm4<-lm(mejora~Peso,data=corredores)
mejora.lm5<-lm(mejora~H_entr_semana,data=corredores)
mejora.lm6<-lm(mejora~t_experiencia,data=corredores)
mejora.lm7<-lm(mejora~experimento_r,data=corredores)
mejora.lm8<-lm(mejora~Tipo_suelo,data=corredores)
mejora.lm9<-lm(mejora~m_minuto,data=corredores2)
anova(mejora.lm0,mejora.lm1)
anova(mejora.lm00,mejora.lm2)
anova(mejora.lm0,mejora.lm3)
anova(mejora.lm00,mejora.lm4)
anova(mejora.lm00,mejora.lm5)
anova(mejora.lm00,mejora.lm6)
anova(mejora.lm00,mejora.lm7)
anova(mejora.lm00,mejora.lm8)
anova(mejora.lm0,mejora.lm9)
```

```
> anova(mejora.lm0,mejora.lm1)
Analysis of Variance Table
```

```

Model 1: mejora ~ 1
Model 2: mejora ~ Frec_Media
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1      7 34913
2      6 32542  1    2371 0.4372 0.5331
> anova(mejora.lm00,mejora.lm2)
Analysis of Variance Table

Model 1: mejora ~ 1
Model 2: mejora ~ Edad
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1      11 172124
2      10 165925  1   6198.7 0.3736 0.5547
> anova(mejora.lm0,mejora.lm3)
Analysis of Variance Table

Model 1: mejora ~ 1
Model 2: mejora ~ Altura
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1      7 34913
2      6 33753  1   1160 0.2062 0.6657
> anova(mejora.lm00,mejora.lm4)
Analysis of Variance Table

Model 1: mejora ~ 1
Model 2: mejora ~ Peso
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1      11 172124
2      10 161872  1  10252 0.6333 0.4446
> anova(mejora.lm00,mejora.lm5)
Analysis of Variance Table

Model 1: mejora ~ 1
Model 2: mejora ~ H_entr_semana
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1      11 172124
2      10  90879  1   81245 8.94 0.01357 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> anova(mejora.lm00,mejora.lm6)
Analysis of Variance Table

Model 1: mejora ~ 1
Model 2: mejora ~ t_experiencia
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1      11 172124
2      10  59871  1  112253 18.749 0.001489 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> anova(mejora.lm00,mejora.lm7)
Analysis of Variance Table

Model 1: mejora ~ 1
Model 2: mejora ~ experimento_r
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1      11 172124
2      10 168068  1   4056 0.2413 0.6338
> anova(mejora.lm00,mejora.lm8)
Analysis of Variance Table

Model 1: mejora ~ 1
Model 2: mejora ~ Tipo_suelo
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1      11 172124
2      10 157979  1  14145 0.8954 0.3663
> anova(mejora.lm0,mejora.lm9)
Analysis of Variance Table

Model 1: mejora ~ 1
Model 2: mejora ~ m_minuto
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1      7 34913
2      6 31407  1   3506.4 0.6699 0.4444

```

```

mejoral.lm10<-lm(mejora~t_experiencia+Edad,data=corredores2)
mejoral.lm11<-lm(mejora~t_experiencia+Sexo,data=corredores2)
mejoral.lm12<-lm(mejora~t_experiencia+Altura,data=corredores2)
mejoral.lm13<-lm(mejora~t_experiencia+Peso,data=corredores2)
mejoral.lm14<-lm(mejora~t_experiencia+experimento_r,data=corredores2)
mejoral.lm15<-lm(mejora~t_experiencia+Tipo_suelo,data=corredores2)
mejoral.lm16<-lm(mejora~t_experiencia+m_minuto,data=corredores2)
anova(mejoral.lm6,mejoral.lm10)
anova(mejoral.lm6,mejoral.lm11)
anova(mejoral.lm6,mejoral.lm12)
anova(mejoral.lm6,mejoral.lm13)
anova(mejoral.lm6,mejoral.lm14)
anova(mejoral.lm6,mejoral.lm15)
anova(mejoral.lm6,mejoral.lm16)
> anova(mejoral.lm6,mejoral.lm10)
Analysis of Variance Table

Model 1: mejora ~ t_experiencia
Model 2: mejora ~ t_experiencia + Edad
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1       6 21722
2       5 21576  1    146.13 0.0339 0.8612
> anova(mejoral.lm6,mejoral.lm11)
Analysis of Variance Table

Model 1: mejora ~ t_experiencia
Model 2: mejora ~ t_experiencia + Sexo
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1       6 21722
2       5 18990  1     2732 0.7193 0.4351
> anova(mejoral.lm6,mejoral.lm12)
Analysis of Variance Table

Model 1: mejora ~ t_experiencia
Model 2: mejora ~ t_experiencia + Altura
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1       6 21722
2       5 12969  1    8753.1 3.3746 0.1256
> anova(mejoral.lm6,mejoral.lm13)
Analysis of Variance Table

Model 1: mejora ~ t_experiencia
Model 2: mejora ~ t_experiencia + Peso
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1       6 21722
2       5 11270  1    10452 4.6372 0.08388 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> anova(mejoral.lm6,mejoral.lm14)
Analysis of Variance Table

Model 1: mejora ~ t_experiencia
Model 2: mejora ~ t_experiencia + experimento_r
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1       6 21722
2       5 15462  1     6260 2.0243 0.2141
> anova(mejoral.lm6,mejoral.lm15)
Analysis of Variance Table

Model 1: mejora ~ t_experiencia
Model 2: mejora ~ t_experiencia + Tipo_suelo
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1       6 21722
2       5 21423  1    298.88 0.0698 0.8022
> anova(mejoral.lm6,mejoral.lm16)
Analysis of Variance Table

Model 1: mejora ~ t_experiencia
Model 2: mejora ~ t_experiencia + m_minuto
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1       6 21722
2       5 21409  1    313.03 0.0731 0.7977
---

```

```

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> anova(mejora.lm6,mejora.lm13)
Analysis of Variance Table

Model 1: mejora ~ t_experiencia
Model 2: mejora ~ t_experiencia + Peso
  Res.Df  RSS Df Sum of Sq   F Pr(>F)
1       10 59871
2        9 37973  1    21898 5.19 0.04871 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> anova(mejora.lm13,mejora.lm20)
Analysis of Variance Table

Model 1: mejora ~ t_experiencia + Peso
Model 2: mejora ~ t_experiencia + Peso + experimento_r
  Res.Df  RSS Df Sum of Sq   F Pr(>F)
1        9 37973
2         8 31794  1    6178.5 1.5546 0.2477
> anova(mejoral.lm13,mejoral.lm20)
Error en anova.lm(mejoral.lm13, mejoral.lm20) :
  objeto 'mejoral.lm20' no encontrado
> mejoral.lm20<-lm(mejora~t_experiencia+Peso+experimento_r,data=corredores2)
> anova(mejoral.lm13,mejoral.lm20)
Analysis of Variance Table

Model 1: mejora ~ t_experiencia + Peso
Model 2: mejora ~ t_experiencia + Peso + experimento_r
  Res.Df  RSS Df Sum of Sq   F Pr(>F)
1         5 11270.0
2         4  7635.1  1    3634.9 1.9043 0.2397
>

> studmejora2
      1          2          3          4          5
1.2864204691  2.2195295880 -0.6502604472  0.5583022076  0.0461022195
      6          7          8          9         10
0.8773001396 -1.7452738330  0.0007348387 -1.6936056723 -0.7069892193
      11         12
-0.1650728440 -0.7659126139

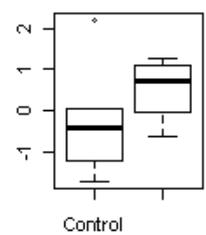
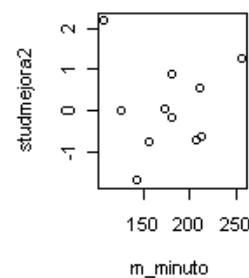
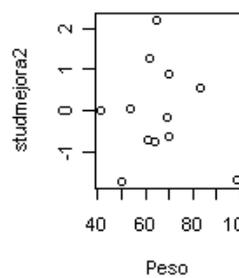
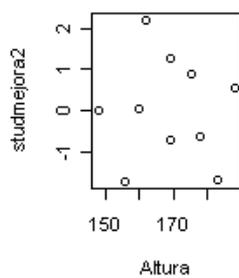
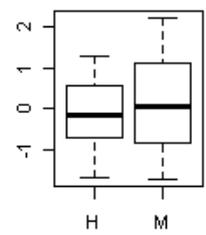
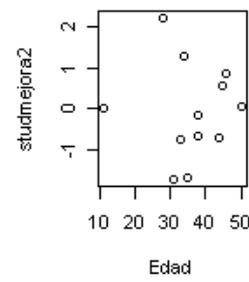
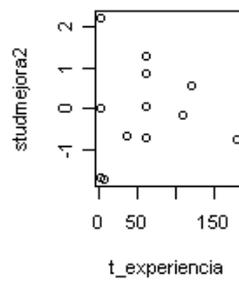
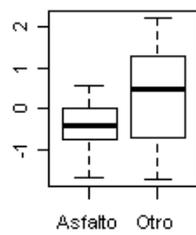
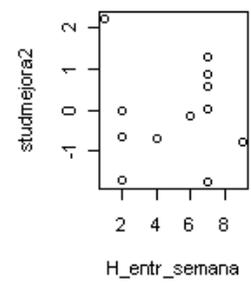
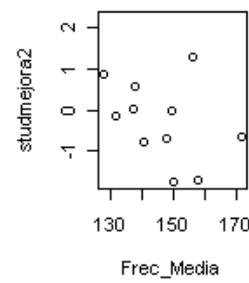
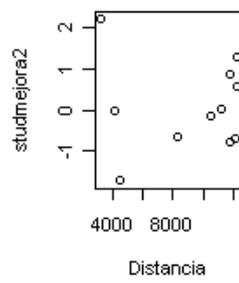
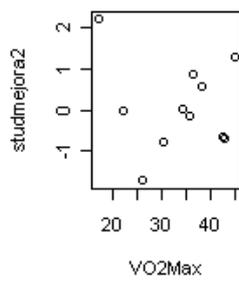
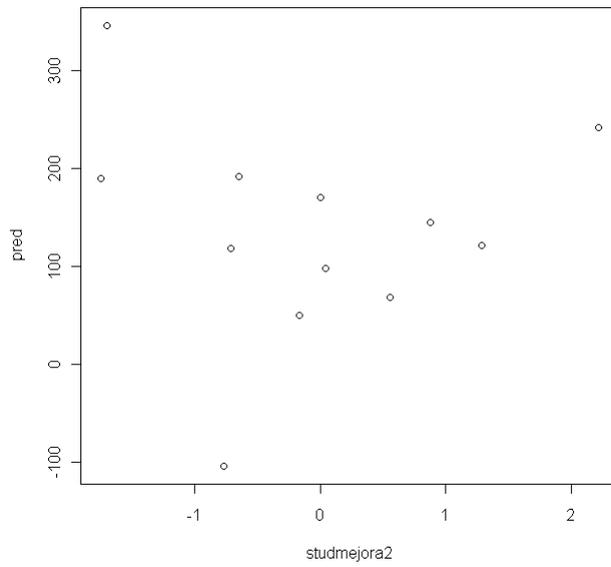
```

6.5.2.4 Estudio residuos

```

studmejora2<-rstudent(mejora.lm13)
noudf<-data.frame(corredores)
noudf
pred<-predict(mejora.lm13, newdata=noudf)
cbind(noudf, pred)
plot(studmejora2,pred)

```

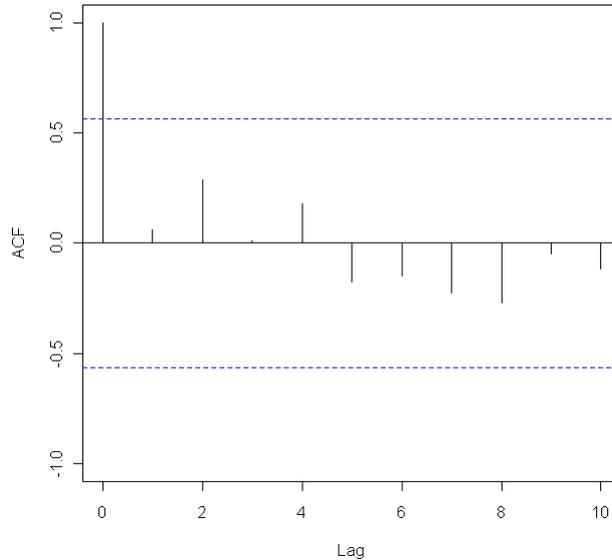


```
Cargar paquete en R
lmtest
dwtest(mejora.lm13)
> dwtest(mejora.lm13)
```

Durbin-Watson test

```
data: mejora.lm13
DW = 1.7458, p-value = 0.2837
alternative hypothesis: true autocorrelation is greater than 0
```

Series studmejora2



6.5.2.5 Detección valores influyentes

```
> anclatge_sosp<-which(lev_step>2*3/length(mejora))
> lev_mod13<-hatvalues(mejora.lm13)
> corredores<-data.frame(corredores,lev_mod13)
> corredores
```

corredor	Edad	Sexo	Altura	Peso	H_entr_semana	t_experiencia	experimento	
1	Javi	34	H	169	62	7	60	3
2	Alicia	28	M	162	65	1	2	1
3	Mario	38	H	178	70	2	36	3
4	Xavi	45	H	188	83	7	120	3
5	Bea	50	M	160	54	7	60	1
6	Paco	46	H	175	70	7	60	3
7	Alba	31	M	156	50	7	6	1
8	Alejandro	11	H	148	41	2	2	1
9	Antonio	35	H	183	99	2	1	1
10	Pepe	44	H	169	61	4	60	1
11	eo73	38	H	NA	69	6	108	1
12	eo78	33	H	NA	64	9	180	1

Cooper_1	Cooper_2	Hora_Inicio	Tipo_suelo	Frec_Media	Distancia	Minutos	
1	3090	3288	M	Otro	156.1500	12242.42	56
2	1668	2019	M	Otro	NA	3227.75	30
3	3010	3160	T	Asfalto	171.1800	8261.54	39
4	3000	3100	T	Asfalto	138.1740	12182.60	58
5	2450	2550	T	Otro	137.2000	11133.33	64
6	3200	3400	T	Otro	128.0700	11800.00	65
7	2500	2600	T	Otro	149.8500	NA	49
8	2130	2300	T	Asfalto	149.2200	4173.33	33
9	1920	2210	T	Asfalto	157.4500	4453.85	31
10	3138	3211	T	Otro	148.0000	12100.00	58
11	2458	2497	M	Asfalto	132.1284	10477.90	69
12	2414	2275	M	Asfalto	140.7598	11743.60	75

m_minuto	VO2Max	mejora	experimento_r	lev_mod13	
1	255.0500	44.88	198	Respiración	0.08919597
2	107.5900	17.08	351	Control	0.17620494
3	211.8300	42.31	150	Respiración	0.10828122
4	210.0400	38.04	100	Respiración	0.28965866

```

5 173.9600 34.26 100 Control 0.13990683
6 181.5400 36.47 200 Respiración 0.09087886
7 NA NA 100 Control 0.24025797
8 126.4600 22.22 170 Control 0.38480100
9 143.6700 25.92 290 Control 0.69227090
10 206.9756 42.81 73 Control 0.09266954
11 181.5600 35.76 39 Control 0.15812808
12 156.5300 30.34 -139 Control 0.53774602

> distcook_sosp<-which(cook_mejora>4/(length(mejora)-3))
> cook_mod13=cooks.distance(mejora.lml3)
> corredores<-data.frame(corredores,cook_mod13)
> corredores
  corredor Edad Sexo Altura Peso H_entr_semana t_experiencia experimento
1 Javi 34 H 169 62 7 60 3
2 Alicia 28 M 162 65 1 2 1
3 Mario 38 H 178 70 2 36 3
4 Xavi 45 H 188 83 7 120 3
5 Bea 50 M 160 54 7 60 1
6 Paco 46 H 175 70 7 60 3
7 Alba 31 M 156 50 7 6 1
8 Alejandro 11 H 148 41 2 2 1
9 Antonio 35 H 183 99 2 1 1
10 Pepe 44 H 169 61 4 60 1
11 eo73 38 H NA 69 6 108 1
12 eo78 33 H NA 64 9 180 1
  Cooper_1 Cooper_2 Hora_Inicio Tipo_suelo Frec_Media Distancia Minutos
1 3090 3288 M Otro 156.1500 12242.42 56
2 1668 2019 M Otro NA 3227.75 30
3 3010 3160 T Asfalto 171.1800 8261.54 39
4 3000 3100 T Asfalto 138.1740 12182.60 58
5 2450 2550 T Otro 137.2000 11133.33 64
6 3200 3400 T Otro 128.0700 11800.00 65
7 2500 2600 T Otro 149.8500 NA 49
8 2130 2300 T Asfalto 149.2200 4173.33 33
9 1920 2210 T Asfalto 157.4500 4453.85 31
10 3138 3211 T Otro 148.0000 12100.00 58
11 2458 2497 M Asfalto 132.1284 10477.90 69
12 2414 2275 M Asfalto 140.7598 11743.60 75
  m_minuto VO2Max mejora experimento_r lev_mod13 cook_mod13
1 255.0500 44.88 198 Respiración 0.08919597 5.035709e-02
2 107.5900 17.08 351 Control 0.17620494 2.445499e-01
3 211.8300 42.31 150 Respiración 0.10828122 1.828785e-02
4 210.0400 38.04 100 Respiración 0.28965866 4.587641e-02
5 173.9600 34.26 100 Control 0.13990683 1.296143e-04
6 181.5400 36.47 200 Respiración 0.09087886 2.631942e-02
7 NA NA 100 Control 0.24025797 2.616109e-01
8 126.4600 22.22 170 Control 0.38480100 1.266589e-07
9 143.6700 25.92 290 Control 0.69227090 1.781114e+00
10 206.9756 42.81 73 Control 0.09266954 1.801806e-02
11 181.5600 35.76 39 Control 0.15812808 1.912796e-03
12 156.5300 30.34 -139 Control 0.53774602 2.384262e-01

```

6.5.3 Resultados Experimentos (Tratamiento de missings)

```

Recodificación experimento
> mejora<-Cooper_2-Cooper_1
> corredores<- data.frame(corredores,mejora)
> experimento_r<-rep(NA, length(experimento))
> experimento_r<-factor(experimento_r, levels=c("Control", "Respiración"))
> experimento_r[experimento==1]<-"Control"
> experimento_r[experimento==3]<-"Respiración"
> corredores<- data.frame(corredores,experimento_r)

```

6.5.3.1 Análisis descriptivo

```

> summary(corredores)
  corredor      Edad      Sexo      Altura      Peso
6 :1  Min. :11.00  H:9  Min. :148.0  Min. :41.00
7 :1  1st Qu.:34.25  M:1  1st Qu.:169.0  1st Qu.:61.25

```

```

4      :1  Median :38.00          Median :176.5  Median :66.50
9      :1  Mean   :37.40          Mean   :173.4  Mean   :67.30
10     :1  3rd Qu.:44.75          3rd Qu.:182.5 3rd Qu.:70.00
1      :1  Max.   :50.00          Max.   :188.0  Max.   :99.00
(Other) :4
H_entr_semana t_experiencia  experimento  Cooper_1  Cooper_2
Min.   :2.0  Min.   : 1.0  Min.   :1.0  Min.   :1920  Min.   :2210
1st Qu.:2.5  1st Qu.: 42.0 1st Qu.:1.0  1st Qu.:2423 1st Qu.:2349
Median :6.5  Median : 60.0  Median :1.0  Median :2729  Median :2825
Mean   :5.3  Mean   : 68.7  Mean   :1.8  Mean  :2681  Mean  :2799
3rd Qu.:7.0  3rd Qu.: 96.0 3rd Qu.:3.0  3rd Qu.:3070 3rd Qu.:3198
Max.   :9.0  Max.   :180.0  Max.   :3.0  Max.   :3200  Max.   :3400

```

```

Hora_Inicio  Tipo_suelo  Frec_Media  Distancia  Minutos
M:3          Asfalto:6  Min.   :128.1  Min.   : 4173  Min.   :31.00
T:7          Otro   :4  1st Qu.:137.4  1st Qu.: 8816  1st Qu.:43.25
              Median :144.4  Median :11438  Median :58.00
              Mean   :145.8  Mean   : 9857  Mean   :54.80
              3rd Qu.:154.4  3rd Qu.:12025 3rd Qu.:64.75
              Max.   :171.2  Max.   :12242  Max.   :75.00

```

```

      m_minuto      VO2Max      mejora      experimento_r
Min.   :126.5  Min.   :33.77  Min.   : -139.00  Control   :6
1st Qu.:160.9  1st Qu.:46.23  1st Qu.: 79.75  Respiración:4
Median :181.6  Median :49.95  Median : 125.00
Mean   :184.8  Mean   :48.66  Mean   : 118.10
3rd Qu.:209.3  3rd Qu.:55.09  3rd Qu.: 191.00
Max.   :255.1  Max.   :56.89  Max.   : 290.00

```

```

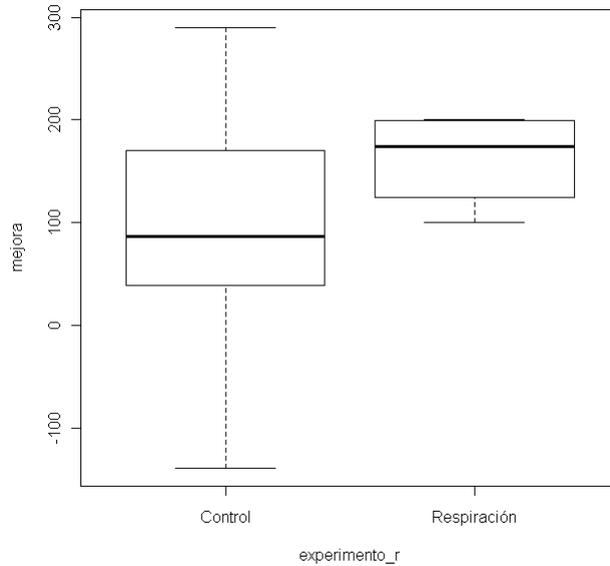
> cor(data.frame(mejora,t_experiencia,Edad,Altura,Peso,H_entr_semana, Frec_Media,
+ VO2Max,m_minuto),use="pairwise.complete.obs")

```

```

      mejora t_experiencia      Edad      Altura      Peso
mejora      1.00000000  -0.86778136 -0.07993886 -0.20636373  0.30012441
t_experiencia -0.86778136  1.00000000  0.30090832  0.53004850  0.02818626
Edad          -0.07993886  0.30090832  1.00000000  0.46292998  0.38135416
Altura        -0.20636373  0.53004850  0.46292998  1.00000000  0.82839180
Peso          0.30012441  0.02818626  0.38135416  0.82839180  1.00000000
H_entr_semana -0.58431219  0.82075270  0.44120216  0.28446783 -0.08192931
Frec_Media    0.35844097 -0.49551270 -0.33167549 -0.06232505  0.11788107
VO2Max        -0.18424368  0.38592766  0.66755034  0.35588641  0.06917480
m_minuto      0.06015627  0.16836614  0.46252141  0.24512395  0.07706568
H_entr_semana  mejora      Frec_Media      VO2Max      m_minuto
mejora      -0.58431219  0.35844097 -0.18424368  0.06015627
t_experiencia  0.82075270 -0.49551270  0.38592766  0.16836614
Edad          0.44120216 -0.33167549  0.66755034  0.46252141
Altura        0.28446783 -0.06232505  0.35588641  0.24512395
Peso          -0.08192931  0.11788107  0.06917480  0.07706568
H_entr_semana 1.00000000 -0.66801421  0.42065273  0.27529807
Frec_Media    -0.66801421  1.00000000 -0.02266483  0.20741958
VO2Max        0.42065273 -0.02266483  1.00000000  0.92456546
m_minuto      0.27529807  0.20741958  0.92456546  1.00000000

```



```
> oneway.test(mejora ~ experimento_r, data=corredores)

One-way analysis of means (not assuming equal variances)

data:  mejora and experimento_r
F = 1.3536, num df = 1.000, denom df = 6.493, p-value = 0.2856
```

Estudio de normalidad

```
> shapiro.test(mejora)

Shapiro-Wilk normality test

data:  mejora
W = 0.9326, p-value = 0.4742

> shapiro.test(mejora[experimento_r=="Control"])

Shapiro-Wilk normality test

data:  mejora[experimento_r == "Control"]
W = 0.9744, p-value = 0.9208

> shapiro.test(mejora[experimento_r=="Respiración"])

Shapiro-Wilk normality test

data:  mejora[experimento_r == "Respiración"]
W = 0.8746, p-value = 0.3162
```

6.5.3.2 Búsqueda modelo

```
> step1_mejora<-step(corredores.mejora,~
experimento_r+Altura+Frec_Media+Edad+Peso+H_entr_semana+
+ +m_minuto+t_experiencia+Tipo_suelo,data=corredores)
Start: AIC=96.04
mejora ~ 1
```

	Df	Sum of Sq	RSS	AIC
+ t_experiencia	1	91419	29980	84.057
+ H_entr_semana	1	41448	79951	93.866
<none>			121399	96.043
+ Frec_Media	1	15597	105802	96.667
+ experimento_r	1	12848	108551	96.924
+ Peso	1	10935	110464	97.099
+ Altura	1	5170	116229	97.607

```
+ Tipo_suelo      1      4051 117348 97.703
+ Edad            1       776 120623 97.978
+ m_minuto        1       439 120960 98.006
```

Step: AIC=84.06
mejora ~ t_experiencia

	Df	Sum of Sq	RSS	AIC
+ experimento_r	1	13174	16806	80.269
+ Peso	1	12800	17180	80.489
+ Altura	1	10858	19122	81.560
+ H_entr_semana	1	6087	23893	83.788
<none>			29980	84.057
+ m_minuto	1	5315	24665	84.105
+ Edad	1	4382	25598	84.477
+ Frec_Media	1	824	29156	85.778
+ Tipo_suelo	1	506	29474	85.887
- t_experiencia	1	91419	121399	96.043

Step: AIC=80.27
mejora ~ t_experiencia + experimento_r

	Df	Sum of Sq	RSS	AIC
+ Peso	1	8176	8630	75.604
+ Altura	1	4833	11974	78.879
<none>			16806	80.269
+ Frec_Media	1	2716	14090	80.507
+ H_entr_semana	1	2536	14271	80.634
+ Edad	1	1290	15517	81.471
+ m_minuto	1	74	16732	82.225
+ Tipo_suelo	1	10	16796	82.263
- experimento_r	1	13174	29980	84.057
- t_experiencia	1	91744	108551	96.924

Step: AIC=75.6
mejora ~ t_experiencia + experimento_r + Peso

	Df	Sum of Sq	RSS	AIC
+ H_entr_semana	1	5740	2890	66.664
+ Altura	1	4537	4093	70.145
+ Frec_Media	1	3953	4677	71.478
<none>			8630	75.604
+ Tipo_suelo	1	1409	7221	75.821
+ Edad	1	18	8612	77.583
+ m_minuto	1	1	8629	77.603
- Peso	1	8176	16806	80.269
- experimento_r	1	8550	17180	80.489
- t_experiencia	1	93203	101833	98.285

Step: AIC=66.66
mejora ~ t_experiencia + experimento_r + Peso + H_entr_semana

	Df	Sum of Sq	RSS	AIC
+ Edad	1	1042	1848	64.192
+ Tipo_suelo	1	671	2219	66.023
<none>			2890	66.664
+ Frec_Media	1	466	2424	66.907
+ Altura	1	243	2647	67.785
+ m_minuto	1	19	2871	68.599
- experimento_r	1	4180	7070	73.610
- H_entr_semana	1	5740	8630	75.604
- Peso	1	11381	14271	80.634
- t_experiencia	1	52763	55653	94.243

Step: AIC=64.19
mejora ~ t_experiencia + experimento_r + Peso + H_entr_semana + Edad

	Df	Sum of Sq	RSS	AIC
+ Frec_Media	1	673	1175	61.663
<none>			1848	64.192
+ m_minuto	1	119	1729	65.528
+ Tipo_suelo	1	19	1829	66.087
+ Altura	1	18	1830	66.094

```
- Edad 1 1042 2890 66.664
- experimento_r 1 4493 6340 74.521
- H_entr_semana 1 6764 8612 77.583
- Peso 1 11962 13810 82.306
- t_experiencia 1 53596 55444 96.205
```

Step: AIC=61.66

mejora ~ t_experiencia + experimento_r + Peso + H_entr_semana +
Edad + Frec_Media

	Df	Sum of Sq	RSS	AIC
+ m_minuto	1	918	257	48.473
<none>			1175	61.663
+ Altura	1	17	1158	63.516
+ Tipo_suelo	1	12	1163	63.559
- Frec_Media	1	673	1848	64.192
- Edad	1	1249	2424	66.907
- H_entr_semana	1	3124	4299	72.635
- experimento_r	1	5151	6326	76.499
- Peso	1	12042	13217	83.867
- t_experiencia	1	49169	50344	97.240

Step: AIC=48.47

mejora ~ t_experiencia + experimento_r + Peso + H_entr_semana +
Edad + Frec_Media + m_minuto

	Df	Sum of Sq	RSS	AIC
+ Tipo_suelo	1	188	69	37.348
+ Altura	1	99	158	45.621
<none>			257	48.473
- m_minuto	1	918	1175	61.663
- Frec_Media	1	1472	1729	65.528
- experimento_r	1	1477	1734	65.557
- Edad	1	2135	2392	68.773
- H_entr_semana	1	2356	2614	69.659
- Peso	1	12834	13092	85.771
- t_experiencia	1	49874	50131	99.198

Step: AIC=37.35

mejora ~ t_experiencia + experimento_r + Peso + H_entr_semana +
Edad + Frec_Media + m_minuto + Tipo_suelo

	Df	Sum of Sq	RSS	AIC
+ Altura	1	69.2	0.0	-Inf
<none>			69.2	37
- Tipo_suelo	1	188.0	257.2	48
- experimento_r	1	794.7	863.9	61
- Edad	1	878.3	947.5	62
- m_minuto	1	1093.5	1162.7	64
- Frec_Media	1	1495.5	1564.7	67
- H_entr_semana	1	1948.8	2018.0	69
- Peso	1	7250.2	7319.4	82
- t_experiencia	1	17904.2	17973.4	91

Step: AIC=-Inf

mejora ~ t_experiencia + experimento_r + Peso + H_entr_semana +
Edad + Frec_Media + m_minuto + Tipo_suelo + Altura

	Df	Sum of Sq	RSS	AIC
<none>			0.0	-Inf
- Altura	1	69.2	69.2	37
- Tipo_suelo	1	158.3	158.3	46
- H_entr_semana	1	759.6	759.6	61
- experimento_r	1	765.3	765.3	61
- Edad	1	815.5	815.5	62
- Peso	1	888.0	888.0	63
- m_minuto	1	1138.4	1138.4	65
- Frec_Media	1	1544.9	1544.9	68
- t_experiencia	1	3178.4	3178.4	76

Mensajes de aviso perdidos

attempting model selection on an essentially perfect fit is nonsense
> summary(step1_mejora)

Call:

```
lm(formula = mejora ~ t_experiencia + experimento_r + Peso +
  H_entr_semana + Edad + Frec_Media + m_minuto + Tipo_suelo +
  Altura, data = corredores)
```

Residuals:

ALL 10 residuals are 0: no residual degrees of freedom!

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	453.4561	NA	NA	NA
t_experiencia	-2.5269	NA	NA	NA
experimento_rRespiración	36.9724	NA	NA	NA
Peso	3.8795	NA	NA	NA
H_entr_semana	16.7713	NA	NA	NA
Edad	-1.6968	NA	NA	NA
Frec_Media	-1.7911	NA	NA	NA
m_minuto	0.6545	NA	NA	NA
Tipo_sueloOtro	-22.4561	NA	NA	NA
Altura	-1.8097	NA	NA	NA

Varianza incremental:

```
> mejora.lm00<-lm(mejora~1,data=corredores)
> mejora.lm0<-lm(mejora~Altura,data=corredores)
> mejora.lm1<-lm(mejora~Frec_Media,data=corredores)
> mejora.lm2<-lm(mejora~Edad,data=corredores)
> mejora.lm3<-lm(mejora~ VO2Max,data=corredores)
> mejora.lm4<-lm(mejora~Peso,data=corredores)
> mejora.lm5<-lm(mejora~H_entr_semana,data=corredores)
> mejora.lm6<-lm(mejora~t_experiencia,data=corredores)
> mejora.lm7<-lm(mejora~experimento_r,data=corredores)
> mejora.lm8<-lm(mejora~Tipo_suelo,data=corredores)
> mejora.lm9<-lm(mejora~m_minuto,data=corredores)
> anova(mejora.lm00,mejora.lm0)
Analysis of Variance Table
```

```
Model 1: mejora ~ 1
Model 2: mejora ~ Altura
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1      9 121399
2      8 116229  1    5169.9 0.3558 0.5673
> anova(mejora.lm00,mejora.lm1)
Analysis of Variance Table
```

```
Model 1: mejora ~ 1
Model 2: mejora ~ Frec_Media
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1      9 121399
2      8 105802  1    15597 1.1794 0.3091
> anova(mejora.lm00,mejora.lm2)
Analysis of Variance Table
```

```
Model 1: mejora ~ 1
Model 2: mejora ~ Edad
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1      9 121399
2      8 120623  1     775.77 0.0515 0.8262
> anova(mejora.lm00,mejora.lm3)
Analysis of Variance Table
```

```
Model 1: mejora ~ 1
Model 2: mejora ~ VO2Max
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1      9 121399
2      8 117278  1     4121 0.2811 0.6104
> anova(mejora.lm00,mejora.lm4)
Analysis of Variance Table
```

```
Model 1: mejora ~ 1
Model 2: mejora ~ Peso
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1      9 121399
2      8 110464  1    10935 0.7919 0.3995
> anova(mejora.lm00,mejora.lm5)
```

Analysis of Variance Table

```

Model 1: mejora ~ 1
Model 2: mejora ~ H_entr_semana
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1      9 121399
2      8  79951  1    41448 4.1474 0.07609 .

```

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> anova(mejora.lm00,mejora.lm6)
Analysis of Variance Table

```

```

Model 1: mejora ~ 1
Model 2: mejora ~ t_experiencia
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1      9 121399
2      8  29980  1    91419 24.395 0.001136 **

```

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> anova(mejora.lm00,mejora.lm7)
Analysis of Variance Table

```

```

Model 1: mejora ~ 1
Model 2: mejora ~ experimento_r
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1      9 121399
2      8 108551  1    12848 0.9469  0.359
> anova(mejora.lm00,mejora.lm8)
Analysis of Variance Table

```

```

Model 1: mejora ~ 1
Model 2: mejora ~ Tipo_suelo
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1      9 121399
2      8 117348  1    4050.8 0.2762 0.6135

```

```

> mejora.lm10<-lm(mejora~t_experiencia+Edad,data=corredores)
> mejora.lm11<-lm(mejora~t_experiencia+VO2Max,data=corredores)
> mejora.lm13<-lm(mejora~t_experiencia+Peso,data=corredores)
> mejora.lm13b<-lm(mejora~t_experiencia+Altura,data=corredores)
> mejora.lm14<-lm(mejora~t_experiencia+experimento_r,data=corredores)
> mejora.lm15<-lm(mejora~t_experiencia+Tipo_suelo,data=corredores)
> mejora.lm16<-lm(mejora~t_experiencia+m_minuto,data=corredores)
> anova(mejora.lm6,mejora.lm10)
Analysis of Variance Table

```

```

Model 1: mejora ~ t_experiencia
Model 2: mejora ~ t_experiencia + Edad
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1      8  29980
2      7 25598  1    4382 1.1983 0.3099
> anova(mejora.lm6,mejora.lm11)
Analysis of Variance Table

```

```

Model 1: mejora ~ t_experiencia
Model 2: mejora ~ t_experiencia + VO2Max
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1      8  29980
2      7 26742  1    3237.7 0.8475 0.3879
> anova(mejora.lm6,mejora.lm13)
Analysis of Variance Table

```

```

Model 1: mejora ~ t_experiencia
Model 2: mejora ~ t_experiencia + Peso
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1      8  29980
2      7 17180  1    12800 5.2154 0.05633 .

```

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> anova(mejora.lm6,mejora.lm13b)
Analysis of Variance Table

```

```

Model 1: mejora ~ t_experiencia
Model 2: mejora ~ t_experiencia + Altura

```

```

  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1      8 29980
2      7 19122  1    10858 3.975 0.08641 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> anova(mejora.lm6,mejora.lm14)
Analysis of Variance Table

Model 1: mejora ~ t_experiencia
Model 2: mejora ~ t_experiencia + experimento_r
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1      8 29980
2      7 16806  1    13174 5.4869 0.05166 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> anova(mejora.lm6,mejora.lm15)
Analysis of Variance Table

Model 1: mejora ~ t_experiencia
Model 2: mejora ~ t_experiencia + Tipo_suelo
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1      8 29980
2      7 29474  1    506.2 0.1202  0.739
> anova(mejora.lm6,mejora.lm16)
Analysis of Variance Table

Model 1: mejora ~ t_experiencia
Model 2: mejora ~ t_experiencia + m_minuto
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1      8 29980
2      7 24665  1    5315.4 1.5086  0.259

> summary (mejora.lm14)

Call:
lm(formula = mejora ~ t_experiencia + experimento_r, data = corredores)

Residuals:
    Min       1Q   Median       3Q      Max
-72.867 -29.177   7.444  22.617  76.666

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    215.1789    28.5990   7.524 0.000135 ***
t_experiencia    -1.8445     0.2984  -6.182 0.000453 ***
experimento_rRespiración  74.0889    31.6292   2.342 0.051660 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 49 on 7 degrees of freedom
Multiple R-squared:  0.8616,    Adjusted R-squared:  0.822
F-statistic: 21.78 on 2 and 7 DF,  p-value: 0.0009872

```

6.5.3.3 Estudio de residuos

```

> lm6_student=rstudent(mejora.lm6)
> lm6_student
      1          2          3          4          5          6          7
1.1189138 -0.4737254  1.4962748 -0.5627056  1.1605637 -1.4384475  0.8917723
      8          9         10
-1.0624926 -0.1123270 -1.3479941

> lm14_student<-rstudent(mejora.lm14)
> lm14_student
      1          2          3          4          5          6
0.43061220 -2.19428591  0.78724294 -0.09359263  0.47660500 -1.04116443
      7          8          9         10
2.58237992 -0.67782152  0.50434663 -0.71537308

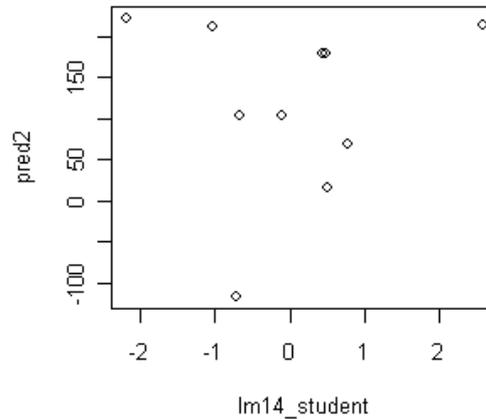
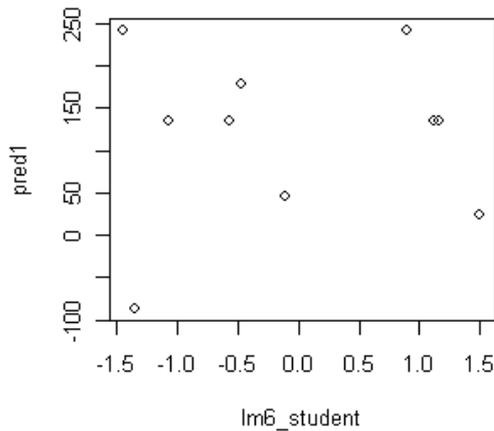
```

```

> corredores
  corredor Edad Sexo Altura Peso H_entr_semana t_experiencia experimento
1     Javi  34   H   169   62           7           60           3
2     Mario  38   H   178   70           2           36           3
3     Xavi  45   H   188   83           7          120           3
4     Bea  50   M   160   54           7           60           1
5     Paco  46   H   175   70           7           60           3
6 Alejandro 11   H   148   41           2            2           1
7     Antonio 35   H   183   99           2            1           1
8     Pepe  44   H   169   61           4           60           1
9     eo73  38   H   181   69           6          108           1
10    eo78  33   H   183   64           9          180           1

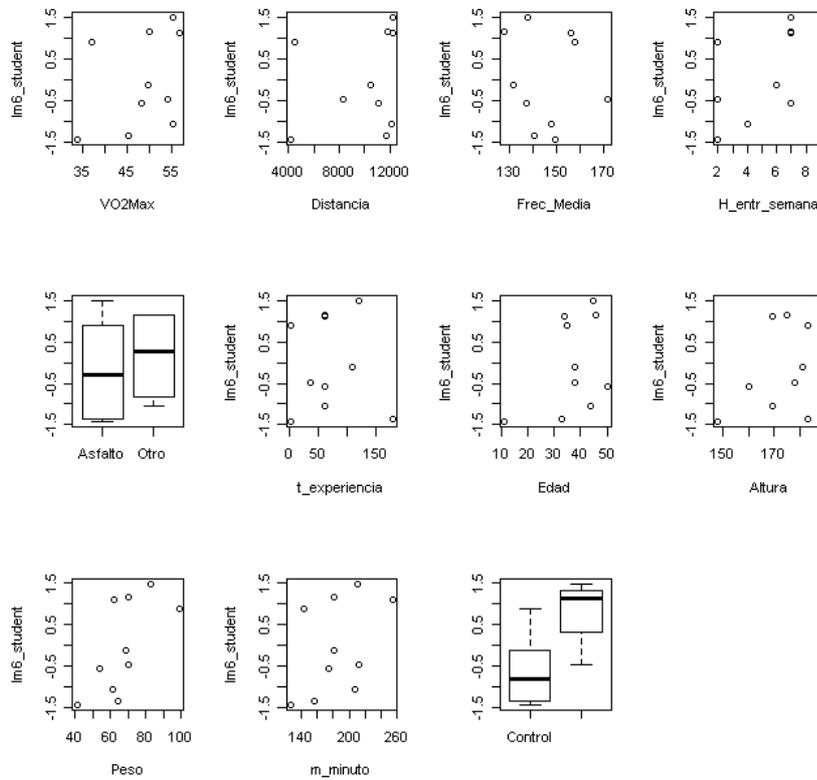
  Cooper_1 Cooper_2 Hora_Inicio Tipo_suelo Frec_Media Distancia Minutos
1      3090      3288           M      Otro    156.1500  12242.42     56
2      3010      3160           T  Asfalto    171.1800   8261.54     39
3      3000      3100           T  Asfalto    138.1740  12182.60     58
4      2450      2550           T      Otro    137.2000  11133.33     64
5      3200      3400           T      Otro    128.0700  11800.00     65
6      2130      2300           T  Asfalto    149.2200   4173.33     33
7      1920      2210           T  Asfalto    157.4500   4453.85     31
8      3138      3211           T      Otro    148.0000  12100.00     58
9      2458      2497           M  Asfalto    132.1284  10477.90     69
10     2414      2275           M  Asfalto    140.7598  11743.60     75

  m_minuto VO2Max mejora experimento_r   pred1   pred2
1  255.0500  56.89   198  Respiración 134.11813 178.60015
2  211.8300  54.23   150  Respiración 178.30607 222.86721
3  210.0400  55.50   100  Respiración 23.64828  67.93249
4  173.9600  48.46   100    Control 134.11813 104.51125
5  181.5400  50.10   200  Respiración 134.11813 178.60015
6  126.4600  33.77   170    Control 240.90565 211.48998
7  143.6700  36.99   290    Control 242.74682 213.33445
8  206.9756  55.37    73    Control 134.11813 104.51125
9  181.5600  49.81    39    Control  45.74225  15.97713
10 156.5300  45.48  -139    Control -86.82157 -116.82406
  
```



```

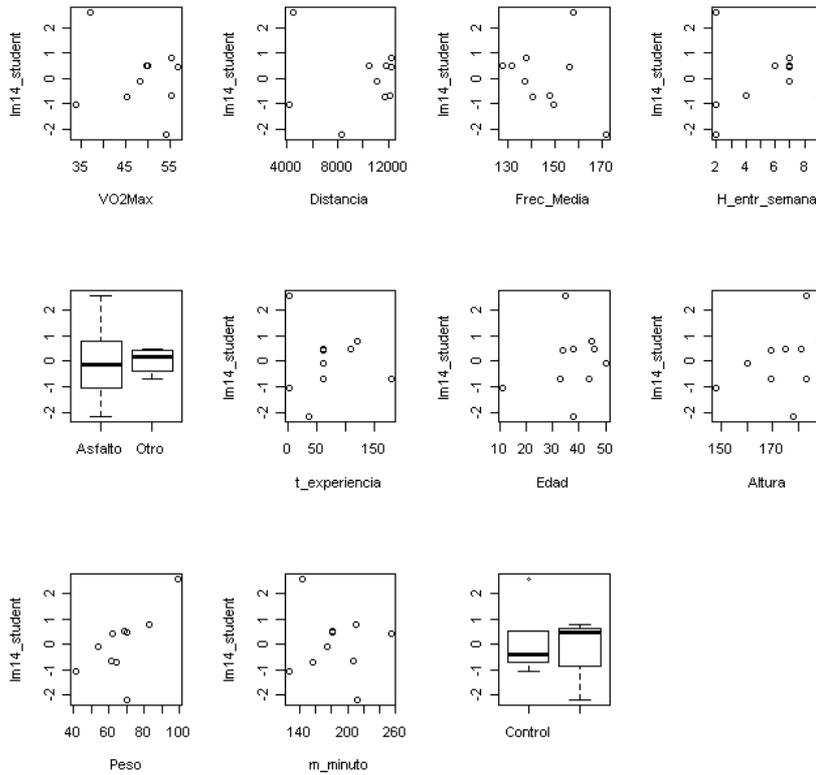
> par(mfrow=c(3,4))
> plot(VO2Max,lm6_student)
> plot(Distancia,lm6_student)
> plot(Frec_Media,lm6_student)
> plot(H_entr_semana,lm6_student)
> plot(Tipo_suelo,lm6_student)
> plot(t_experiencia,lm6_student)
> plot(Edad,lm6_student)
> plot(Sexo,lm6_student)
> plot(Altura,lm6_student)
> plot(Peso,lm6_student)
> plot(m_minuto,lm6_student)
> plot(experimento_r,lm6_student)
  
```



```

par(mfrow=c(3,4))
plot(VO2Max,lm14_student)
plot(Distancia,lm14_student)
plot(Frec_Media,lm14_student)
plot(H_entr_semana,lm14_student)
plot(Tipo_suelo,lm14_student)
plot(t_experiencia,lm14_student)
plot(Edad,lm14_student)
plot(Altura,lm14_student)
plot(Peso,lm14_student)
plot(m_minuto,lm14_student)
plot(experimento_r,lm14_student)

```



```
> dwtest(mejora.lm6)
```

Durbin-Watson test

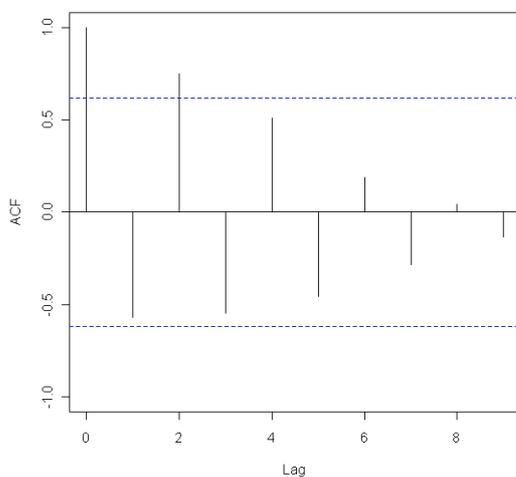
```
data: mejora.lm6
DW = 3.0385, p-value = 0.9554
alternative hypothesis: true autocorrelation is greater than 0
```

```
> dwtest(mejora.lm14)
```

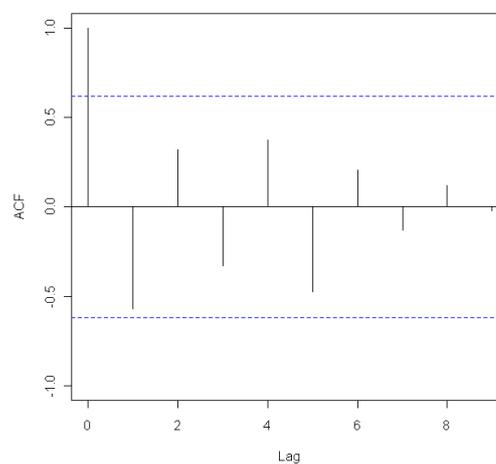
Durbin-Watson test

```
data: mejora.lm14
DW = 3.3421, p-value = 0.9942
```

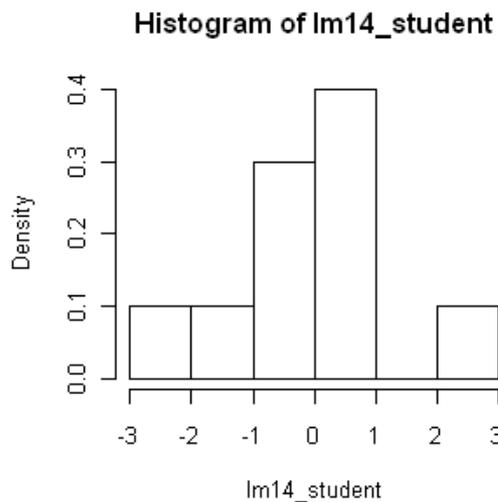
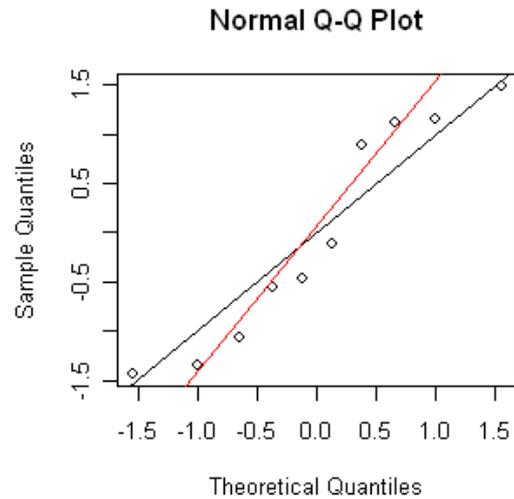
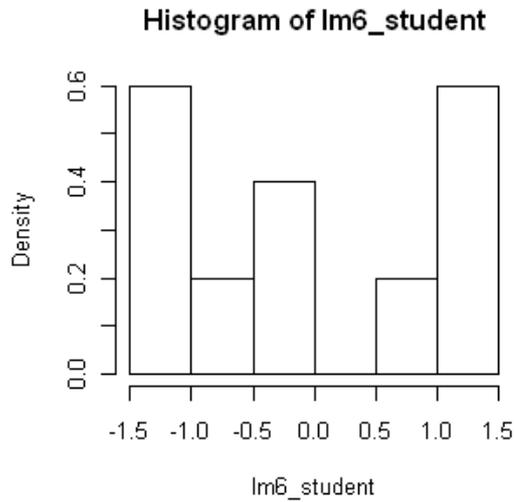
Series lm6_student



Series lm14_student



```
> par(mfrow=c(2,2))
> hist(lm6_student,freq=F)
> qqnorm(lm6_student)
> abline(0,1)
> qqline(lm6_student,col=2)
> hist(lm14_student,freq=F)
> qqnorm(lm14_student)
> abline(0,1)
> qqline(lm14_student,col=2)
```



6.5.3.4 Detección valores influyentes

```
> lev_lm14<-hatvalues(mejora.lm14)
> anclatge_sosp<-which(lev_lm14>2*3/length(mejora))
> anclatge_sosp
10
10
> corredores<-data.frame(corredores,lev_lm14)
> corredores
  corredor Edad Sexo Altura Peso H_entr_semana t_experiencia experimento
1     Javi  34   H   169   62           7           60           3
2     Mario  38   H   178   70           2           36           3
3     Xavi  45   H   188   83           7          120           3
4     Bea   50   M   160   54           7           60           1
5     Paco  46   H   175   70           7           60           3
6 Alejandro 11   H   148   41           2            2           1
7 Antonio  35   H   183   99           2            1           1
8     Pepe  44   H   169   61           4           60           1
```

```

9      eo73  38  H   181  69          6          108          1
10     eo78  33  H   183  64          9          180          1
  Cooper_1 Cooper_2 Hora_Inicio Tipo_suelo Frec_Media Distancia Minutos
1      3090   3288          M      Otro   156.1500  12242.42    56
2      3010   3160          T      Asfalto 171.1800   8261.54    39
3      3000   3100          T      Asfalto 138.1740  12182.60   58
4      2450   2550          T      Otro   137.2000  11133.33   64
5      3200   3400          T      Otro   128.0700  11800.00   65
6      2130   2300          T      Asfalto 149.2200   4173.33   33
7      1920   2210          T      Asfalto 157.4500   4453.85   31
8      3138   3211          T      Otro   148.0000  12100.00   58
9      2458   2497          M      Asfalto 132.1284  10477.90   69
10     2414   2275          M      Asfalto 140.7598  11743.60   75
  m_minuto VO2Max mejora experimento_r   pred1   pred2 lev_lm14
1  255.0500  56.89   198  Respiración 134.11813  178.60015  0.2530036
2  211.8300  54.23   150  Respiración 178.30607  222.86721  0.2903819
3  210.0400  55.50   100  Respiración 23.64828   67.93249  0.3464494
4  173.9600  48.46   100      Control 134.11813  104.51125  0.1693458
5  181.5400  50.10   200  Respiración 134.11813  178.60015  0.2530036
6  126.4600  33.77   170      Control 240.90565  211.48998  0.3306511
7  143.6700  36.99   290      Control 242.74682  213.33445  0.3356200
8  206.9756  55.37    73      Control 134.11813  104.51125  0.1693458
9  181.5600  49.81    39      Control 45.74225   15.97713  0.2245233
10 156.5300  45.48  -139      Control -86.82157 -116.82406 0.6276753

> cook_lm14=cooks.distance(mejora_lm14)
>
> distcook_sosp<-which(cook_lm14>4/(length(mejora)-3))
> cooksosp<-corredores[distcook_sosp,]
> cooksosp
  corredor Edad Sexo Altura Peso H_entr_semana t_experiencia experimento
7  Antonio  35  H   183  99          2          1          1
  Cooper_1 Cooper_2 Hora_Inicio Tipo_suelo Frec_Media Distancia Minutos
7      1920   2210          T      Asfalto   157.45   4453.85    31
  m_minuto VO2Max mejora experimento_r   pred1   pred2 lev_lm14
7  143.67  36.99   290      Control 242.7468 213.3344 0.33562
> corredores<-data.frame(corredores,cook_lm14)
> corredores
  corredor Edad Sexo Altura Peso H_entr_semana t_experiencia experimento
1      Javi  34  H   169  62          7          60          3
2      Mario  38  H   178  70          2          36          3
3      Xavi  45  H   188  83          7         120          3
4      Bea  50  M   160  54          7          60          1
5      Paco  46  H   175  70          7          60          3
6  Alejandro 11  H   148  41          2          2          1
7      Antonio 35  H   183  99          2          1          1
8      Pepe  44  H   169  61          4          60          1
9      eo73  38  H   181  69          6         108          1
10     eo78  33  H   183  64          9         180          1
  Cooper_1 Cooper_2 Hora_Inicio Tipo_suelo Frec_Media Distancia Minutos
1      3090   3288          M      Otro   156.1500  12242.42    56
2      3010   3160          T      Asfalto 171.1800   8261.54    39
3      3000   3100          T      Asfalto 138.1740  12182.60   58
4      2450   2550          T      Otro   137.2000  11133.33   64
5      3200   3400          T      Otro   128.0700  11800.00   65
6      2130   2300          T      Asfalto 149.2200   4173.33   33
7      1920   2210          T      Asfalto 157.4500   4453.85   31
8      3138   3211          T      Otro   148.0000  12100.00   58
9      2458   2497          M      Asfalto 132.1284  10477.90   69
10     2414   2275          M      Asfalto 140.7598  11743.60   75
  m_minuto VO2Max mejora experimento_r   pred1   pred2 lev_lm14
1  255.0500  56.89   198  Respiración 134.11813  178.60015  0.2530036
2  211.8300  54.23   150  Respiración 178.30607  222.86721  0.2903819
3  210.0400  55.50   100  Respiración 23.64828   67.93249  0.3464494
4  173.9600  48.46   100      Control 134.11813  104.51125  0.1693458
5  181.5400  50.10   200  Respiración 134.11813  178.60015  0.2530036
6  126.4600  33.77   170      Control 240.90565  211.48998  0.3306511
7  143.6700  36.99   290      Control 242.74682  213.33445  0.3356200
8  206.9756  55.37    73      Control 134.11813  104.51125  0.1693458
9  181.5600  49.81    39      Control 45.74225   15.97713  0.2245233
10 156.5300  45.48  -139      Control -86.82157 -116.82406 0.6276753
  cook_lm14
1  0.0236912512

```

```

2 0.4250950290
3 0.1158012663
4 0.0006934728
5 0.0288278800
6 0.1763814255
7 0.6204643568
8 0.0338350412
9 0.0274754330
10 0.3091411661

```

6.5.4 Resultados juego de datos cursas

6.5.4.1 Análisis descriptivo

```

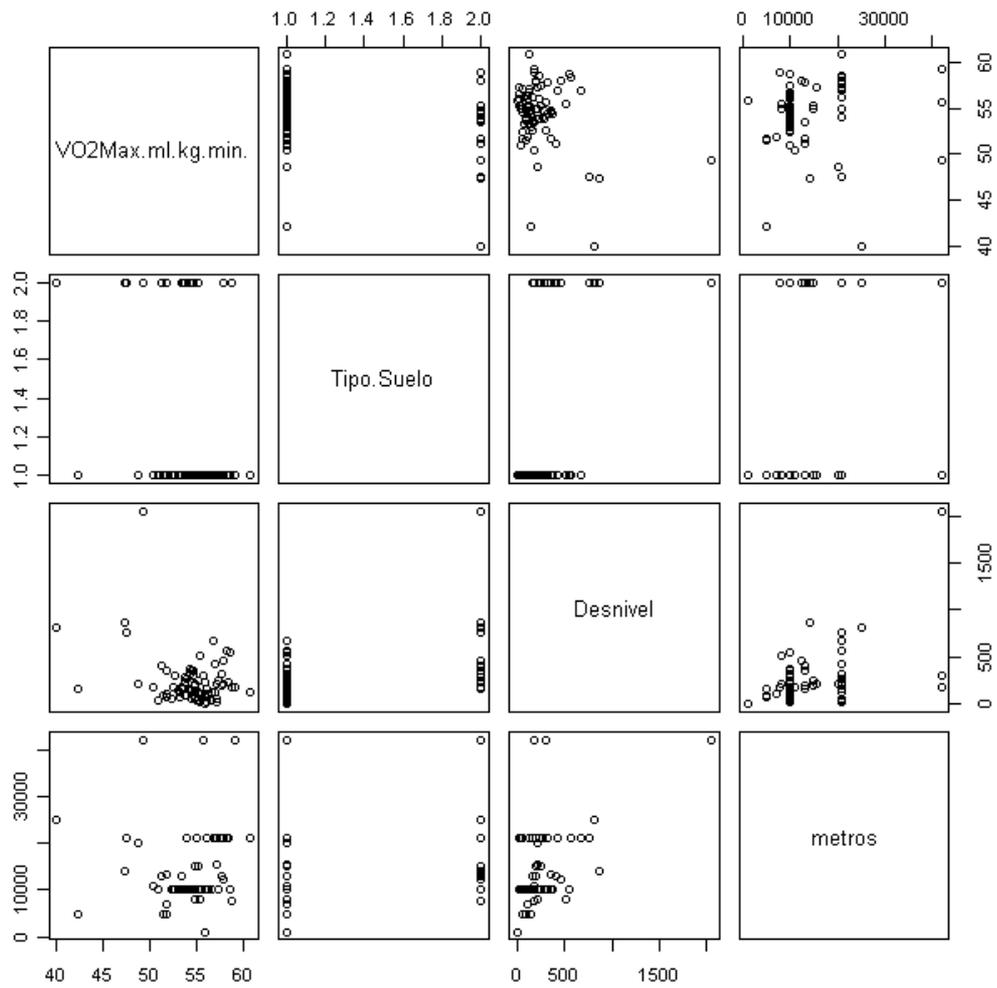
> summary(cooper_curses)

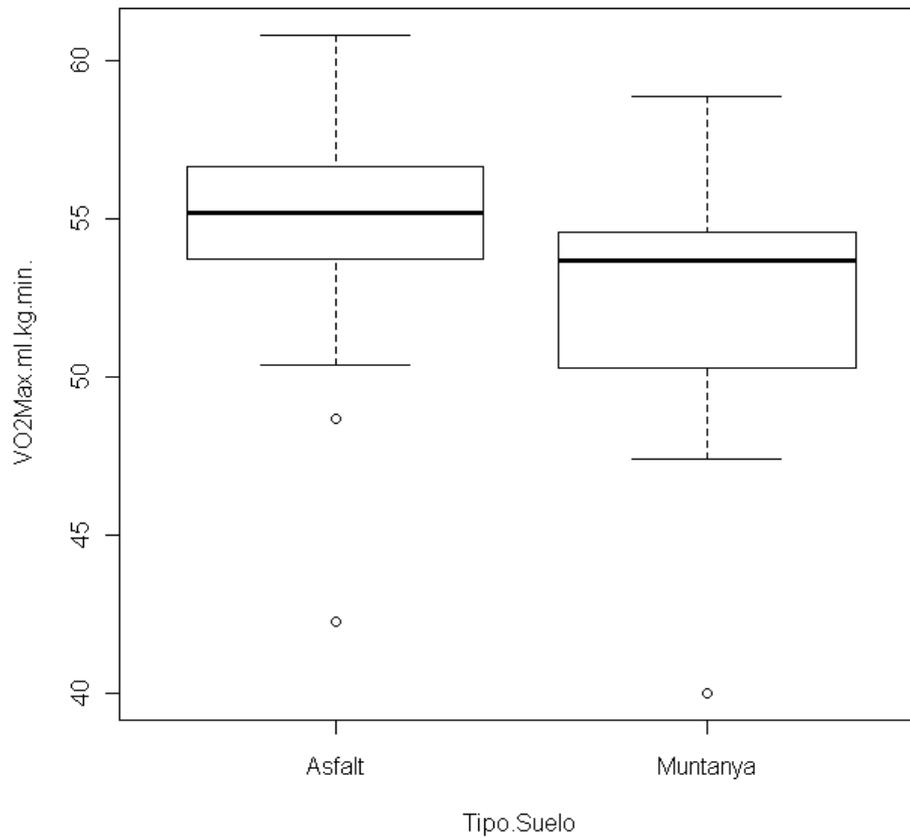
```

curso	metros	segundos
10 k Nou Barris : 1	Min. : 1000	Min. : 204.1
10 km Bellavista : 1	1st Qu.:10000	1st Qu.: 2825.9
10 km Collbatonina-olesa: 1	Median :10000	Median : 2948.8
10 km de Pineda de Mar : 1	Mean :13546	Mean : 4130.1
10 km Pont de Vilomara : 1	3rd Qu.:15450	3rd Qu.: 4940.4
10 km Roda de Ter : 1	Max. :42000	Max. :17267.9
(Other) :72		

Km_hora	VO2Max.ml.kg.min.	Cooper	Tipo.Suelo
Min. : 6.839	Min. :40.01	Min. :1454	Asfalt :63
1st Qu.:12.096	1st Qu.:53.47	1st Qu.:2474	Muntanya:15
Median :12.507	Median :54.86	Median :2553	
Mean :12.316	Mean :54.42	Mean :2526	
3rd Qu.:12.885	3rd Qu.:56.31	3rd Qu.:2638	
Max. :17.643	Max. :60.81	Max. :3589	

Desnivel	Dificultad
Min. : 0.00	Moderat :35
1st Qu.: 95.75	Facil :22
Median : 164.50	0 :12
Mean : 237.71	Dificil : 4
3rd Qu.: 288.75	A ull : 1
Max. :2051.00	inf 2008: 1
	(Other) : 3



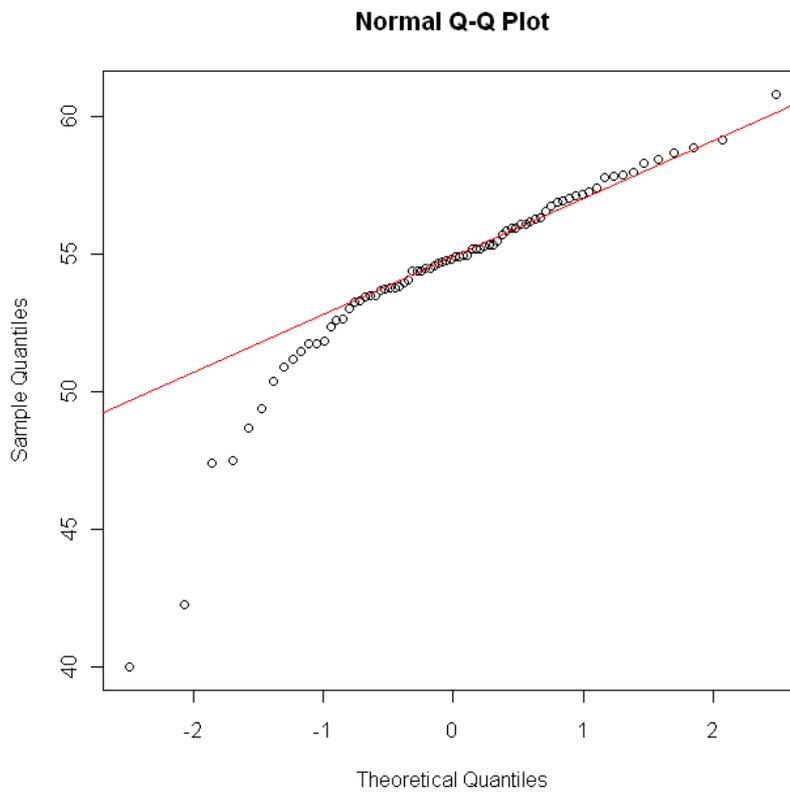


```
> t.test(VO2Max.ml.kg.min. ~ Tipo.Suelo, data=cooper_curses)

Welch Two Sample t-test

data: VO2Max.ml.kg.min. by Tipo.Suelo
t = 2.1674, df = 16.367, p-value = 0.04528
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.06483038 5.40892640
sample estimates:
 mean in group Asfalt mean in group Muntanya
      54.94757          52.21069
```

6.5.4.2 Estudio de normalidad de los datos



```
> VO2Max_teo<-rnorm(78,mean(VO2Max.ml.kg.min.),sd(VO2Max.ml.kg.min.))
> ks.test(VO2Max.ml.kg.min., VO2Max_teo)
```

Two-sample Kolmogorov-Smirnov test

```
data: VO2Max.ml.kg.min. and VO2Max_teo
D = 0.1026, p-value = 0.81
alternative hypothesis: two-sided
```

El resultado nos induce a no rechazar la hipótesis de normalidad.

6.5.4.3 Búsqueda del modelo

```
> cooper_curses.lm<-lm(VO2Max.ml.kg.min.~1, data=cooper_curses)
> step_cooper_curses <-step(cooper_curses.lm,~ metros*Desnivel*Tipo.Suelo)
Start: AIC=190.35
VO2Max.ml.kg.min. ~ 1
```

	Df	Sum of Sq	RSS	AIC
+ Tipo.Suelo	1	90.750	781.77	183.78
+ Desnivel	1	86.049	786.48	184.25
<none>			872.52	190.34
+ metros	1	4.356	868.17	191.96

```
Step: AIC=183.78
VO2Max.ml.kg.min. ~ Tipo.Suelo
```

	Df	Sum of Sq	RSS	AIC
+ Desnivel	1	28.414	753.36	182.89
<none>			781.77	183.78
+ metros	1	8.533	773.24	184.92
- Tipo.Suelo	1	90.750	872.52	190.34

```
Step: AIC=182.89
VO2Max.ml.kg.min. ~ Tipo.Suelo + Desnivel
```

```

              Df Sum of Sq   RSS   AIC
+ Desnivel:Tipo.Suelo  1    76.430 676.93 176.55
+ metros              1    44.682 708.68 180.12
<none>                753.36 182.89
- Desnivel            1    28.414 781.77 183.78
- Tipo.Suelo         1    33.115 786.48 184.25

Step: AIC=176.55
VO2Max.ml.kg.min. ~ Tipo.Suelo + Desnivel + Tipo.Suelo:Desnivel

              Df Sum of Sq   RSS   AIC
+ metros              1    51.651 625.28 172.36
<none>                676.93 176.55
- Tipo.Suelo:Desnivel  1    76.430 753.36 182.89

Step: AIC=172.36
VO2Max.ml.kg.min. ~ Tipo.Suelo + Desnivel + metros + Tipo.Suelo:Desnivel

              Df Sum of Sq   RSS   AIC
+ metros:Desnivel     1    44.244 581.03 168.63
+ metros:Tipo.Suelo   1    40.176 585.10 169.18
<none>                625.28 172.36
- metros              1    51.651 676.93 176.55
- Tipo.Suelo:Desnivel  1    83.399 708.68 180.12

Step: AIC=168.63
VO2Max.ml.kg.min. ~ Tipo.Suelo + Desnivel + metros + Tipo.Suelo:Desnivel +
  Desnivel:metros

              Df Sum of Sq   RSS   AIC
+ metros:Tipo.Suelo   1    84.419 496.62 158.39
<none>                581.03 168.63
- Desnivel:metros     1    44.244 625.28 172.36
- Tipo.Suelo:Desnivel  1   122.552 703.59 181.56

Step: AIC=158.39
VO2Max.ml.kg.min. ~ Tipo.Suelo + Desnivel + metros + Tipo.Suelo:Desnivel +
  Desnivel:metros + Tipo.Suelo:metros

              Df Sum of Sq   RSS   AIC
+ metros:Desnivel:Tipo.Suelo  1    32.507 464.11 155.11
- Tipo.Suelo:Desnivel         1     6.814 503.43 157.45
<none>                        496.62 158.39
- Tipo.Suelo:metros           1    84.419 581.03 168.63
- Desnivel:metros             1    88.487 585.10 169.18

Step: AIC=155.11
VO2Max.ml.kg.min. ~ Tipo.Suelo + Desnivel + metros + Tipo.Suelo:Desnivel +
  Desnivel:metros + Tipo.Suelo:metros + Tipo.Suelo:Desnivel:metros

              Df Sum of Sq   RSS   AIC
<none>                464.11 155.11
- Tipo.Suelo:Desnivel:metros  1    32.507 496.62 158.39
> summary(step_cooper_curses)

Call:
lm(formula = VO2Max.ml.kg.min. ~ Tipo.Suelo + Desnivel + metros +
    Tipo.Suelo:Desnivel + Desnivel:metros + Tipo.Suelo:metros +
    Tipo.Suelo:Desnivel:metros, data = cooper_curses)

Residuals:
    Min       1Q   Median       3Q      Max
-11.3773  -1.2362   0.1597   1.4897   5.8677

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.187e+01  1.186e+00  43.750 < 2e-16 ***
Tipo.SueloMuntanya  1.485e+01  3.049e+00   4.869 6.70e-06 ***
Desnivel      6.232e-03  5.778e-03   1.079  0.2845
metros       2.061e-04  8.647e-05   2.384  0.0198 *
Tipo.SueloMuntanya:Desnivel -1.911e-02  8.143e-03  -2.347  0.0217 *
Desnivel:metros -2.851e-07  3.637e-07  -0.784  0.4357
Tipo.SueloMuntanya:metros -1.178e-03  2.838e-04  -4.152 9.16e-05 ***
Tipo.SueloMuntanya:Desnivel:metros  8.605e-07  3.886e-07   2.214  0.0301 *

```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

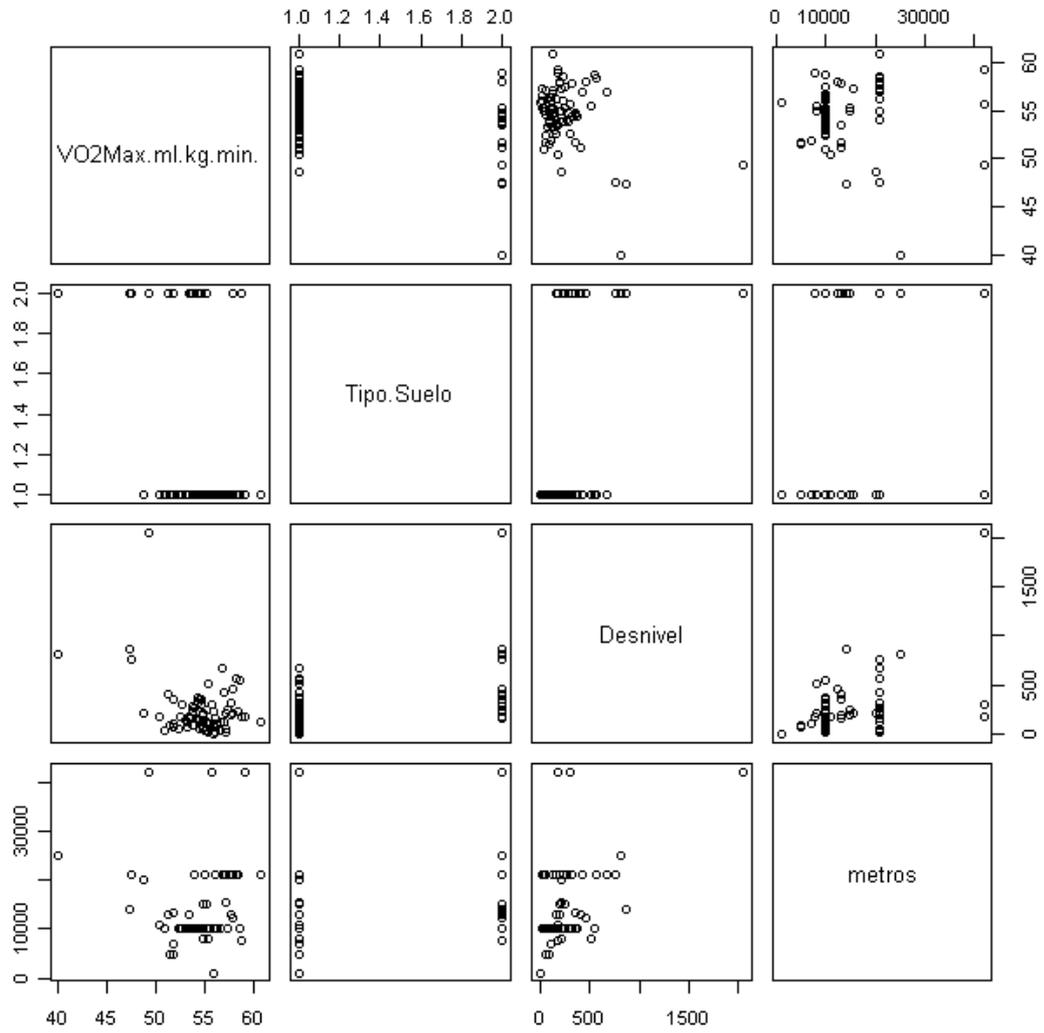
```
Residual standard error: 2.575 on 70 degrees of freedom
Multiple R-squared:  0.4681,    Adjusted R-squared:  0.4149
F-statistic:  8.8 on 7 and 70 DF,  p-value: 1.014e-07
```

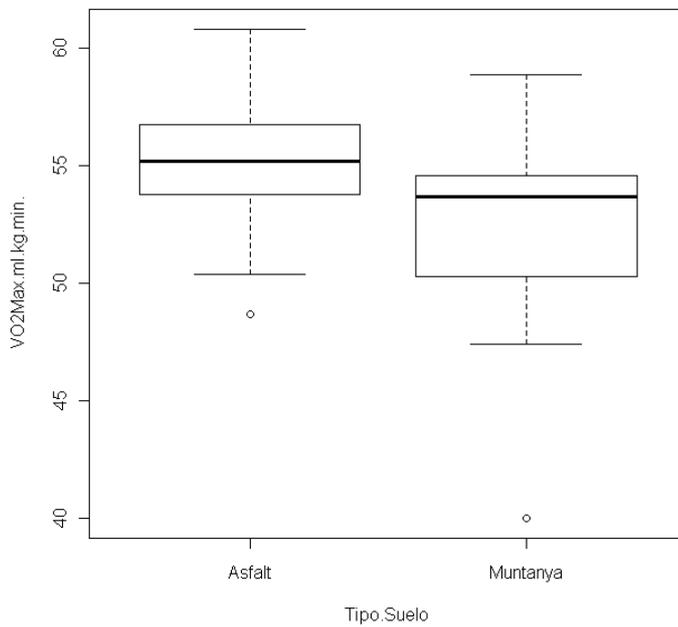
6.5.5 Resultados juego de datos cursas sin la carrera de la maquinista 5 km

6.5.5.1 Análisis descriptivo

```
> summary(cooper_curses2)
      cursa      metros      segundos
10 k Nou Barris      : 1  Min.      : 1000  Min.      : 204.1
10 km Bellavista      : 1  1st Qu.:10000 1st Qu.: 2827.8
10 km Collbatonina-olesa: 1  Median :10000 Median : 2951.2
10 km de Pineda de Mar : 1  Mean   :13657 Mean   : 4159.7
10 km Pont de Vilomara : 1  3rd Qu.:15600 3rd Qu.: 5114.0
10 km Roda de Ter      : 1  Max.   :42000 Max.   :17267.9
(Other)                :71
      Km_hora      VO2Max.ml.kg.min.      Cooper      Tipo.Suelo
Min.      : 6.839  Min.      :40.01  Min.      :1454  Asphalt :62
1st Qu.:12.136  1st Qu.:53.48  1st Qu.:2482  Muntanya:15
Median :12.515  Median :54.92  Median :2553
Mean   :12.350  Mean   :54.58  Mean   :2532
3rd Qu.:12.903  3rd Qu.:56.32  3rd Qu.:2640
Max.   :17.643  Max.   :60.81  Max.   :3589

      Desnivel      Dificultad
Min.      : 0.0  Moderat :35
1st Qu.: 95.0  Facil   :21
Median : 168.0  0       :12
Mean   : 238.8  Dificil : 4
3rd Qu.: 292.0  A ull   : 1
Max.   :2051.0  inf 2008: 1
(Other) : 3
```

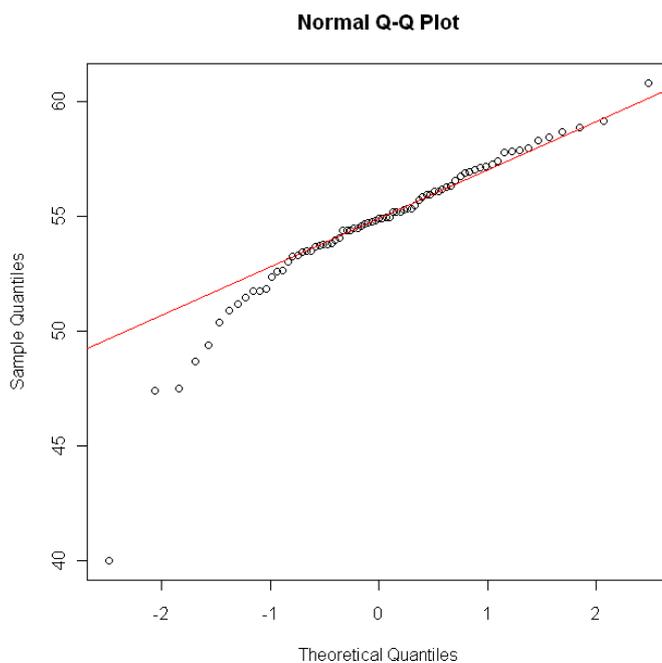




```
> oneway.test(VO2Max.ml.kg.min. ~ Tipo.Suelo, data=cooper_curses2)
One-way analysis of means (not assuming equal variances)

data: VO2Max.ml.kg.min. and Tipo.Suelo
F = 5.5644, num df = 1.000, denom df = 15.582, p-value = 0.03175
```

6.5.5.2 Estudio de normalidad de los datos



```
> ks.test(VO2Max.ml.kg.min., VO2Max_teo2)
Two-sample Kolmogorov-Smirnov test
```

```
data: VO2Max.ml.kg.min. and VO2Max_teo2
D = 0.1558, p-value = 0.3086
alternative hypothesis: two-sided
```

6.5.5.3 Búsqueda del modelo

```
> cor(VO2Max.ml.kg.min.,Desnivel)
[1] -0.3618876
> cor(VO2Max.ml.kg.min.,metros)
[1] 0.01890079

> cooper_curses.lm2<-lm(VO2Max.ml.kg.min.~1, data=cooper_curses2)
> step_cooper_curses2 <-step(cooper_curses.lm2,~ metros*Desnivel*Tipo.Suelo)
Start: AIC=174.39
VO2Max.ml.kg.min. ~ 1

              Df Sum of Sq    RSS    AIC
+ Tipo.Suelo  1   104.518 617.91 164.35
+ Desnivel   1    94.611 627.82 165.58
<none>                                722.43 174.39
+ metros     1     0.258 722.17 176.36

Step: AIC=164.36
VO2Max.ml.kg.min. ~ Tipo.Suelo

              Df Sum of Sq    RSS    AIC
+ Desnivel   1    29.894 588.02 162.54
<none>                                617.91 164.35
+ metros     1     1.769 616.14 166.13
- Tipo.Suelo  1   104.518 722.43 174.39

Step: AIC=162.54
VO2Max.ml.kg.min. ~ Tipo.Suelo + Desnivel

              Df Sum of Sq    RSS    AIC
+ Desnivel:Tipo.Suelo  1    72.599 515.42 154.39
+ metros              1    24.319 563.70 161.28
<none>                  1    588.02 162.54
- Desnivel              1    29.894 617.91 164.35
- Tipo.Suelo            1    39.800 627.82 165.58

Step: AIC=154.39
VO2Max.ml.kg.min. ~ Tipo.Suelo + Desnivel + Tipo.Suelo:Desnivel

              Df Sum of Sq    RSS    AIC
+ metros              1    29.652 485.76 151.83
<none>                  1    515.42 154.39
- Tipo.Suelo:Desnivel  1    72.599 588.02 162.54

Step: AIC=151.83
VO2Max.ml.kg.min. ~ Tipo.Suelo + Desnivel + metros + Tipo.Suelo:Desnivel

              Df Sum of Sq    RSS    AIC
+ metros:Desnivel     1    47.068 438.70 145.98
+ metros:Tipo.Suelo   1    36.182 449.58 147.87
<none>                  1    485.76 151.83
- metros              1    29.652 515.42 154.39
- Tipo.Suelo:Desnivel  1    77.932 563.70 161.28

Step: AIC=145.98
VO2Max.ml.kg.min. ~ Tipo.Suelo + Desnivel + metros + Tipo.Suelo:Desnivel +
Desnivel:metros

              Df Sum of Sq    RSS    AIC
+ metros:Tipo.Suelo   1    79.513 359.18 132.58
<none>                  1    438.70 145.98
- Desnivel:metros     1    47.068 485.76 151.83
- Tipo.Suelo:Desnivel  1   121.437 560.13 162.80

Step: AIC=132.58
VO2Max.ml.kg.min. ~ Tipo.Suelo + Desnivel + metros + Tipo.Suelo:Desnivel +
Desnivel:metros + Tipo.Suelo:metros
```

```

              Df Sum of Sq  RSS   AIC
+ metros:Desnivel:Tipo.Suelo  1    29.527 329.66 127.98
- Tipo.Suelo:Desnivel         1     7.531 366.71 132.18
<none>                        359.18 132.58
- Tipo.Suelo:metros           1    79.513 438.70 145.98
- Desnivel:metros             1    90.398 449.58 147.87

Step:  AIC=127.98
VO2Max.ml.kg.min. ~ Tipo.Suelo + Desnivel + metros + Tipo.Suelo:Desnivel +
  Desnivel:metros + Tipo.Suelo:metros + Tipo.Suelo:Desnivel:metros

              Df Sum of Sq  RSS   AIC
<none>                        329.66 127.98
- Tipo.Suelo:Desnivel:metros  1    29.527 359.18 132.58
> summary(step_cooper_curses2)

Call:
lm(formula = VO2Max.ml.kg.min. ~ Tipo.Suelo + Desnivel + metros +
    Tipo.Suelo:Desnivel + Desnivel:metros + Tipo.Suelo:metros +
    Tipo.Suelo:Desnivel:metros, data = cooper_curses2)

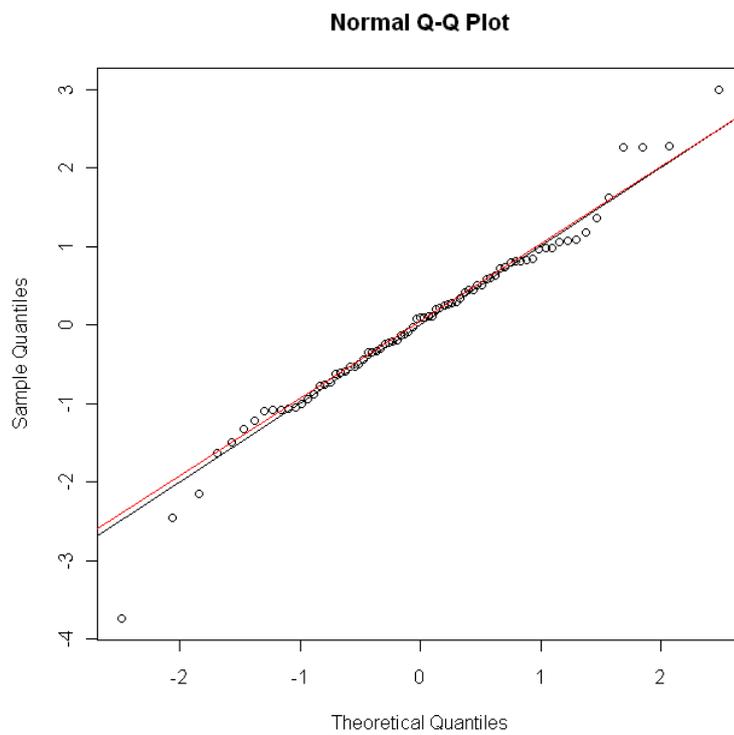
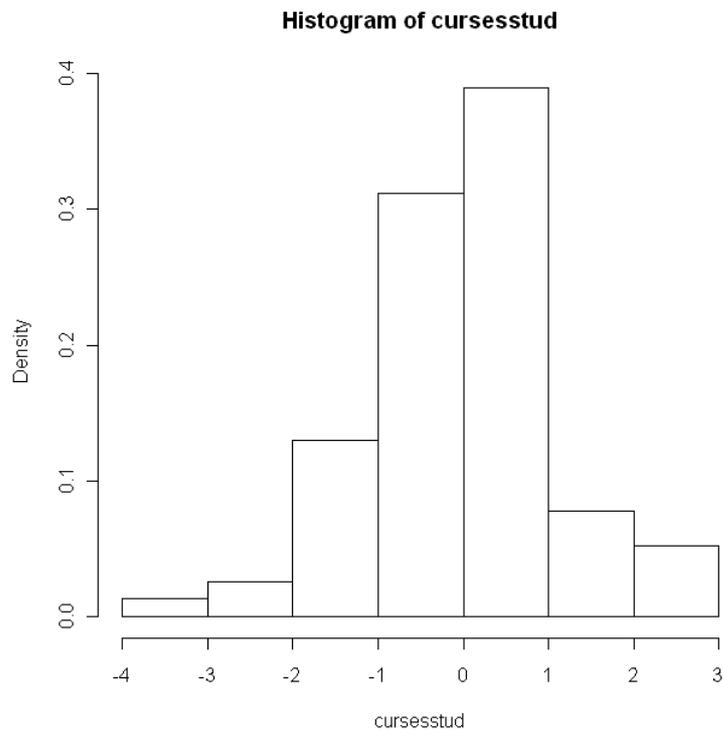
Residuals:
    Min       1Q   Median       3Q      Max
-7.3798 -1.3365  0.2148  1.3559  5.8677

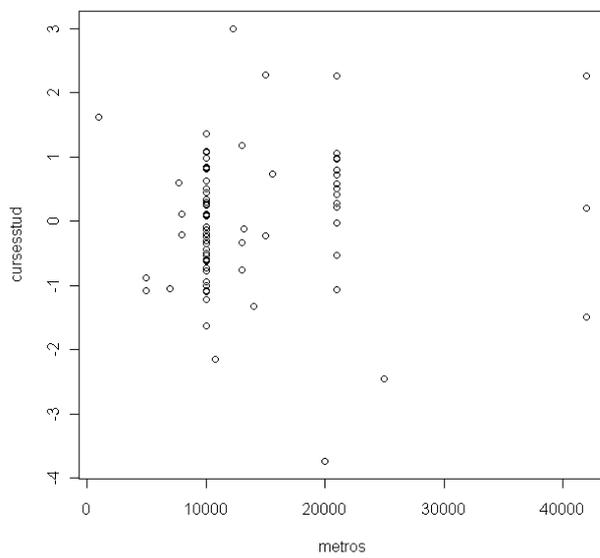
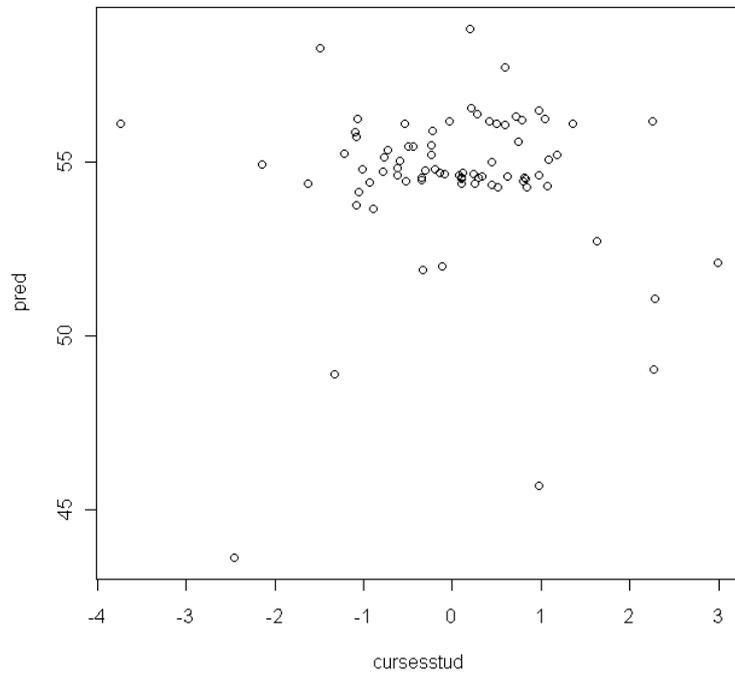
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    5.253e+01  1.014e+00  51.801 < 2e-16 ***
Tipo.SueloMuntanya  1.419e+01  2.592e+00   5.475 6.63e-07 ***
Desnivel        5.857e-03  4.905e-03   1.194  0.23651
metros         1.676e-04  7.376e-05   2.273  0.02616 *
Tipo.SueloMuntanya:Desnivel -1.874e-02  6.913e-03  -2.711  0.00846 **
Desnivel:metros -2.449e-07  3.088e-07  -0.793  0.43037
Tipo.SueloMuntanya:metros -1.140e-03  2.411e-04  -4.729 1.16e-05 ***
Tipo.SueloMuntanya:Desnivel:metros  8.203e-07  3.300e-07   2.486  0.01534 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

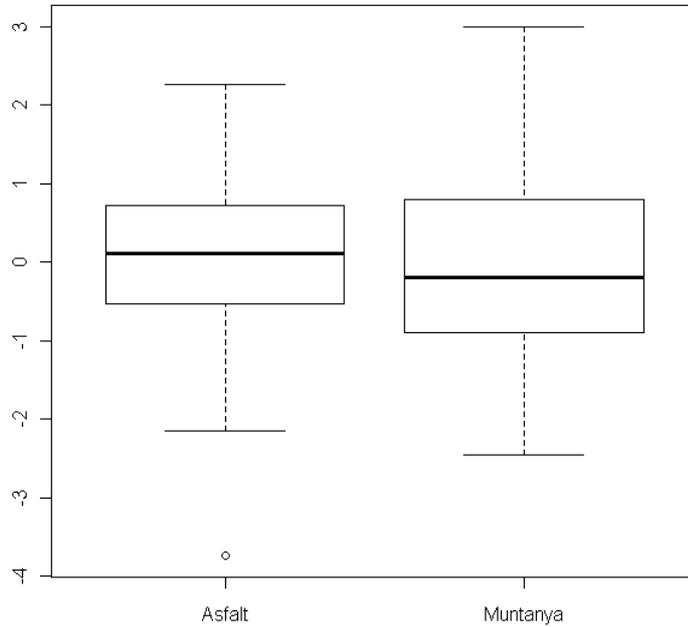
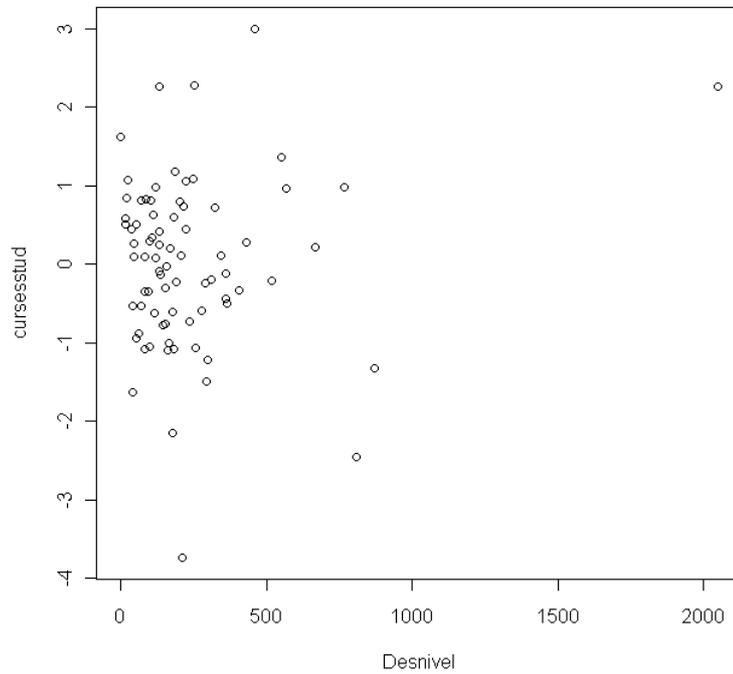
Residual standard error: 2.186 on 69 degrees of freedom
Multiple R-squared:  0.5437,    Adjusted R-squared:  0.4974
F-statistic: 11.74 on 7 and 69 DF,  p-value: 9.67e-10

```

6.5.5.4 Estudio de residuos







```
> dwtest(cooper_curses.lm2)
```

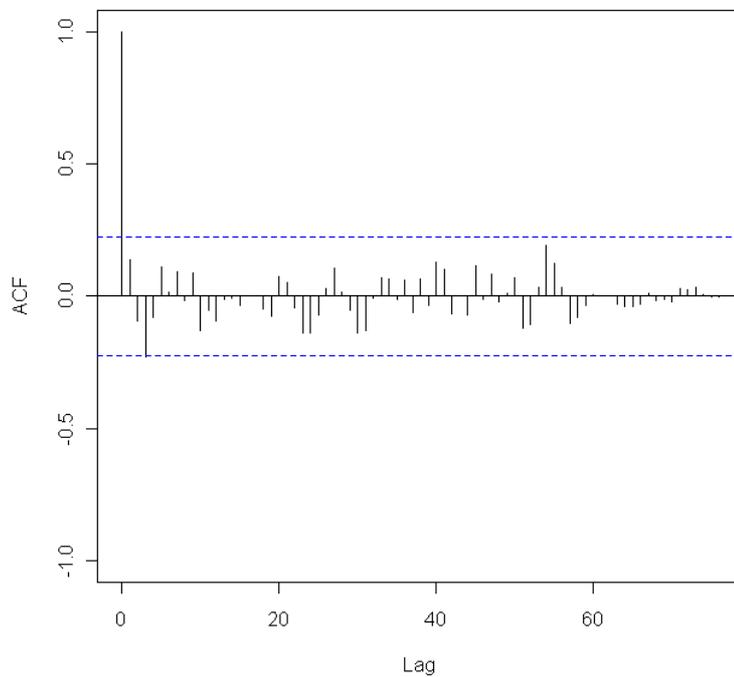
```
Durbin-Watson test
```

```
data: cooper_curses.lm2
```

```
DW = 1.6561, p-value = 0.05404
```

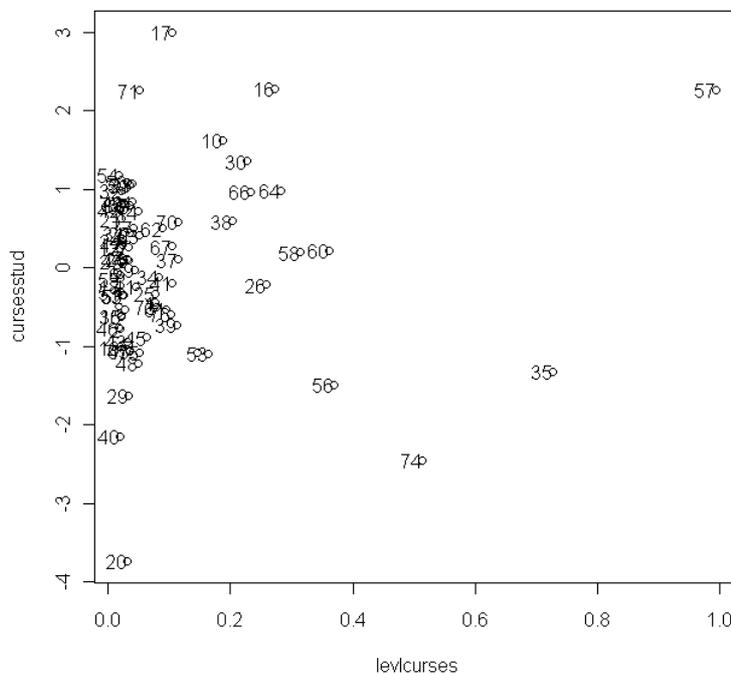
```
alternative hypothesis: true autocorrelation is greater than 0
```

Series cursesstud

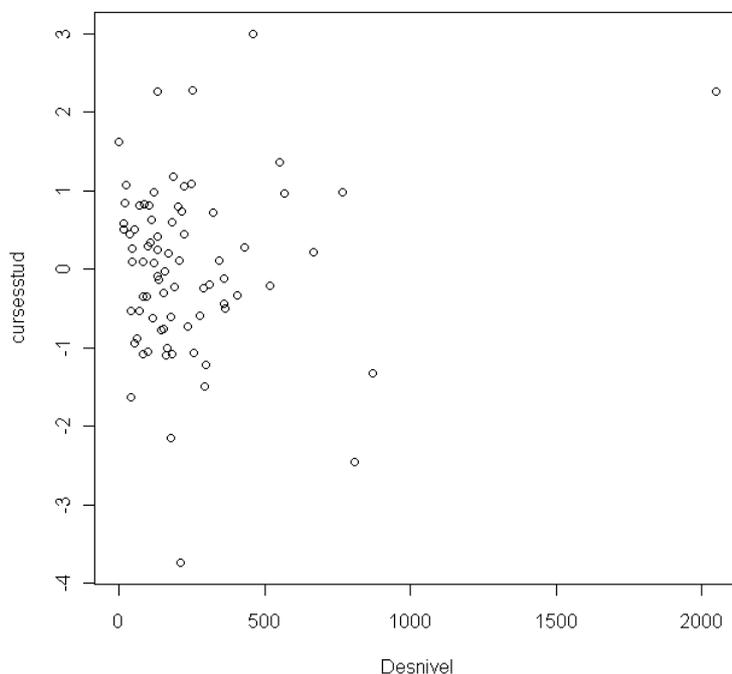


6.5.5.5 Detección de valores influyentes a priori y a posteriori

```
plot(levlcourses,cursesstud)
text(x=levlcourses,y=cursesstud,labels=row.names(cooper_courses2), adj=1)
```



```
plot(cookcourses,cursesstud)
text(x=cookcourses,y=cursesstud,labels=row.names(cooper_courses2), adj=1)
```

6.5.6 Resultados juego de datos cursas sin las carrera de Behovia y selva marítima

6.5.6.1 Análisis descriptivo

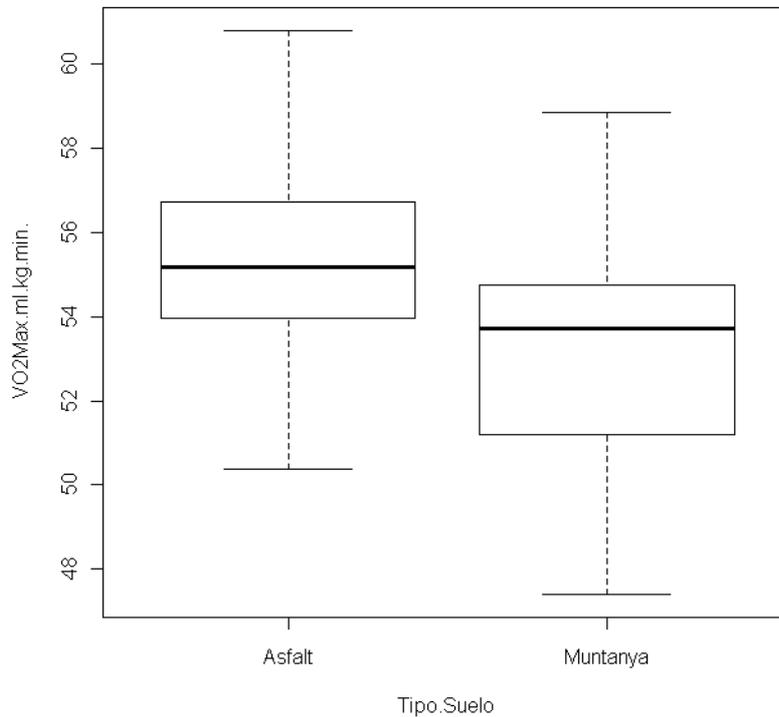
```
> summary(vo2max_sinbehobia)
```

curso	metros	segundos
10 k Nou Barris : 1	Min. : 1000	Min. : 204.1
10 km Bellavista : 1	1st Qu.:10000	1st Qu.: 2826.5
10 km Collbatonina-olesa: 1	Median :10000	Median : 2946.4
10 km de Pineda de Mar : 1	Mean :13421	Mean : 3998.8
10 km Pont de Vilomara : 1	3rd Qu.:15000	3rd Qu.: 4405.5
10 km Roda de Ter : 1	Max. :42000	Max. :17267.9
(Other) :69		

Km_hora	VO2Max.ml.kg.min.	Cooper	Tipo.Suelo
Min. : 8.756	Min. :47.40	Min. :1805	Asfalt :61
1st Qu.:12.145	1st Qu.:53.59	1st Qu.:2495	Muntanya:14
Median :12.523	Median :54.92	Median :2558	
Mean :12.455	Mean :54.85	Mean :2552	
3rd Qu.:12.923	3rd Qu.:56.45	3rd Qu.:2642	
Max. :17.643	Max. :60.81	Max. :3589	

Desnivel	Dificultad
Min. : 0.0	Moderat :35
1st Qu.: 90.5	Facil :20
Median : 161.0	0 :12
Mean : 231.6	Dificil : 3
3rd Qu.: 285.5	A ull : 1
Max. :2051.0	inf 2008: 1
	(Other) : 3

```
> plot(VO2Max.ml.kg.min. ~ Tipo.Suelo)
```



```
> oneway.test(VO2Max.ml.kg.min. ~ Tipo.Suelo, data=vo2max_sinbehobia)
```

One-way analysis of means (not assuming equal variances)

```
data: VO2Max.ml.kg.min. and Tipo.Suelo
```

```
F = 5.2786, num df = 1.00, denom df = 15.37, p-value = 0.03602
```

6.5.6.2 Búsqueda del modelo

```
> cor(data.frame(VO2Max.ml.kg.min.,Desnivel,metros))
```

```
      VO2Max.ml.kg.min.  Desnivel  metros
VO2Max.ml.kg.min.      1.0000000 -0.3004549 0.171749
Desnivel                -0.3004549  1.0000000 0.479444
metros                  0.1717490  0.4794440 1.000000
```

*****Búsqueda del modelo que mejor ajuste según criterio de Akaike*****

```
> step_vo2max_sinbehobia <-step(lm0,~ metros*Desnivel*Tipo.Suelo)
```

```
Start: AIC=139.67
```

```
VO2Max.ml.kg.min. ~ 1
```

	Df	Sum of Sq	RSS	AIC
+ Tipo.Suelo	1	53.915	416.25	132.54
+ Desnivel	1	42.444	427.73	134.57
+ metros	1	13.869	456.30	139.43
<none>			470.17	139.67

```
Step: AIC=132.54
```

```
VO2Max.ml.kg.min. ~ Tipo.Suelo
```

	Df	Sum of Sq	RSS	AIC
+ metros	1	16.610	399.64	131.48
+ Desnivel	1	12.772	403.48	132.20
<none>			416.25	132.54
- Tipo.Suelo	1	53.915	470.17	139.67

```
Step: AIC=131.48
```

```
VO2Max.ml.kg.min. ~ Tipo.Suelo + metros
```

	Df	Sum of Sq	RSS	AIC

```

+ metros:Tipo.Suelo 1 94.169 305.48 113.33
+ Desnivel 1 43.729 355.92 124.79
<none> 399.64 131.48
- metros 1 16.610 416.25 132.54
- Tipo.Suelo 1 56.656 456.30 139.43

Step: AIC=113.33
VO2Max.ml.kg.min. ~ Tipo.Suelo + metros + Tipo.Suelo:metros

              Df Sum of Sq  RSS  AIC
<none>                305.48 113.33
+ Desnivel            1  0.937 304.54 115.10
- Tipo.Suelo:metros   1  94.169 399.64 131.48

> summary(step_vo2max_sinbehobia)

Call:
lm(formula = VO2Max.ml.kg.min. ~ Tipo.Suelo + metros + Tipo.Suelo:metros,
    data = vo2max_sinbehobia)

Residuals:
    Min       1Q   Median       3Q      Max
-5.7132 -1.3605  0.0166  1.3504  4.4827

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.332e+01  5.538e-01  96.281 < 2e-16 ***
Tipo.SueloMuntanya  2.720e+00  1.225e+00   2.221 0.029536 *
metros        1.460e-04  3.667e-05   3.980 0.000164 ***
Tipo.SueloMuntanya:metros -3.551e-04  7.591e-05  -4.678 1.34e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.074 on 71 degrees of freedom
Multiple R-squared:  0.3503,    Adjusted R-squared:  0.3228
F-statistic: 12.76 on 3 and 71 DF,  p-value: 9.26e-07

```

Test de varianza incremental.

```

***Método Forward y Backward para considerar modelo factible***
lm1<-lm(VO2Max.ml.kg.min. ~metros)
anova(lm0,lm1)
lm2<-lm(VO2Max.ml.kg.min. ~Desnivel)
anova(lm0,lm2)
lm3<-lm(VO2Max.ml.kg.min. ~Tipo.Suelo)
anova(lm0,lm3)

> lm3<-lm(VO2Max.ml.kg.min. ~Tipo.Suelo)
> anovalm0,lm3)
Analysis of Variance Table

Model 1: VO2Max.ml.kg.min. ~ 1
Model 2: VO2Max.ml.kg.min. ~ Tipo.Suelo
  Res.Df  RSS Df Sum of Sq    F  Pr(>F)
1     74 470.17
2     73 416.25  1    53.915 9.4553 0.002961 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

lm4<-lm(VO2Max.ml.kg.min. ~metros+Tipo.Suelo)
anova(lm3,lm4)
lm5<-lm(VO2Max.ml.kg.min. ~Desnivel+Tipo.Suelo)
anova(lm3,lm5)

> anova(lm3,lm4)
Analysis of Variance Table

Model 1: VO2Max.ml.kg.min. ~ Tipo.Suelo
Model 2: VO2Max.ml.kg.min. ~ metros + Tipo.Suelo
  Res.Df  RSS Df Sum of Sq    F  Pr(>F)
1     73 416.25
2     72 399.64  1    16.61 2.9924 0.08794 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

lm6<-lm(VO2Max.ml.kg.min. ~metros+Tipo.Suelo+Tipo.Suelo:metros)
summary(lm6)
anova(lm4,lm6)
> anova(lm4,lm6)
Analysis of Variance Table

Model 1: VO2Max.ml.kg.min. ~ metros + Tipo.Suelo
Model 2: VO2Max.ml.kg.min. ~ metros + +Tipo.Suelo + Tipo.Suelo:metros
  Res.Df  RSS Df Sum of Sq    F    Pr(>F)
1      72 399.64
2      71 305.48  1    94.169 21.887 1.344e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

lm7<-lm(VO2Max.ml.kg.min. ~metros+Desnivel+Tipo.Suelo+Tipo.Suelo:metros)
summary(lm7)
anova(lm6,lm7)
> anova(lm6,lm7)
Analysis of Variance Table

Model 1: VO2Max.ml.kg.min. ~ metros + +Tipo.Suelo + Tipo.Suelo:metros
Model 2: VO2Max.ml.kg.min. ~ metros + Desnivel + Tipo.Suelo + Tipo.Suelo:metros
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1      71 305.48
2      70 304.54  1    0.93674 0.2153 0.6441

lm8<-lm(VO2Max.ml.kg.min.
~metros+Desnivel+Tipo.Suelo+Tipo.Suelo:metros+Desnivel:Tipo.Suelo)
summary(lm8)
anova(lm7,lm8)
> anova(lm7,lm8)
Analysis of Variance Table
Model 1: VO2Max.ml.kg.min. ~ metros + Desnivel + Tipo.Suelo +
Tipo.Suelo:metros
Model 2: VO2Max.ml.kg.min. ~ metros + Desnivel + Tipo.Suelo +
Desnivel:Tipo.Suelo + Tipo.Suelo:metros +
  Res.Df  RSS Df Sum of Sq    F Pr(>F)
1      70 304.54
2      69 290.79  1    13.746 3.2616 0.07528 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

lm9<-lm(VO2Max.ml.kg.min.
~metros+Desnivel+Tipo.Suelo+Tipo.Suelo:metros+metros:Desnivel)
summary(lm9)
anova(lm7,lm9)
> summary(lm9)

Call:
lm(formula = VO2Max.ml.kg.min. ~ metros + Desnivel + Tipo.Suelo +
  Tipo.Suelo:metros + metros:Desnivel)

Residuals:
    Min       1Q   Median       3Q      Max
-4.9690 -1.0605  0.0119  1.2432  4.8273

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.393e+01  6.499e-01  82.983 < 2e-16 ***
metros       9.154e-05  4.347e-05   2.106 0.038876 *
Desnivel    -3.115e-03  2.459e-03  -1.267 0.209470
Tipo.SueloMuntanya  7.451e+00  2.431e+00  3.065 0.003103 **
metros:Tipo.SueloMuntanya -7.635e-04  2.016e-04  -3.788 0.000321 ***
metros:Desnivel  2.552e-07  1.161e-07   2.198 0.031304 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.031 on 69 degrees of freedom
Multiple R-squared: 0.3947, Adjusted R-squared: 0.3508
F-statistic: 8.997 on 5 and 69 DF, p-value: 1.28e-06

> anova(lm7,lm9)
Analysis of Variance Table

```

```

Model 1: VO2Max.ml.kg.min. ~ metros + Desnivel + Tipo.Suelo + Tipo.Suelo:metros
Model 2: VO2Max.ml.kg.min. ~ metros + Desnivel + Tipo.Suelo + Tipo.Suelo:metros +
  metros:Desnivel
  Res.Df    RSS Df Sum of Sq    F Pr(>F)
1      70 304.54
2      69 284.61  1    19.928 4.8314 0.0313 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 ---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

lm10<-lm(VO2Max.ml.kg.min. ~metros+Desnivel+Tipo.Suelo+Tipo.Suelo:metros+
metros:Desnivel+Tipo.Suelo:Desnivel)
summary(lm10)
anova(lm9,lm10)
> summary(lm10)

Call:
lm(formula = VO2Max.ml.kg.min. ~ metros + Desnivel + Tipo.Suelo +
  Tipo.Suelo:metros + metros:Desnivel + Tipo.Suelo:Desnivel)

Residuals:
    Min       1Q   Median       3Q      Max
-4.4978 -1.3505  0.0348  1.3531  5.4050

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.383e+01  6.346e-01  84.817  <2e-16 ***
metros       7.601e-05  4.294e-05   1.770  0.0812 .
Desnivel    -1.803e-03  2.469e-03  -0.730  0.4678 .
Tipo.SueloMuntanya  6.102e+00  2.447e+00   2.493  0.0151 *

metros:Tipo.SueloMuntanya -4.002e-04  2.577e-04  -1.553  0.1251 .
metros:Desnivel          2.864e-07  1.140e-07   2.513  0.0144 *
Desnivel:Tipo.SueloMuntanya -9.083e-03  4.175e-03  -2.176  0.0331 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.978 on 68 degrees of freedom
Multiple R-squared:  0.4341,    Adjusted R-squared:  0.3841
F-statistic: 8.692 on 6 and 68 DF,  p-value: 5.024e-07

> anova(lm9,lm10)
Analysis of Variance Table

Model 1: VO2Max.ml.kg.min. ~ metros + Desnivel + Tipo.Suelo + Tipo.Suelo:metros +
  metros:Desnivel
Model 2: VO2Max.ml.kg.min. ~ metros + Desnivel + Tipo.Suelo + Tipo.Suelo:metros +
  metros:Desnivel + Tipo.Suelo:Desnivel
  Res.Df    RSS Df Sum of Sq    F Pr(>F)
1      69 284.61
2      68 266.09  1    18.521 4.7331 0.03307 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

lm_completo<-lm(VO2Max.ml.kg.min. ~metros*Desnivel*Tipo.Suelo)
summary(lm_completo)
anova(lm10,lm_completo)
> anova(lm10,lm_completo)
Analysis of Variance Table

Model 1: VO2Max.ml.kg.min. ~ metros + Desnivel + Tipo.Suelo + Tipo.Suelo:metros +
  metros:Desnivel + Tipo.Suelo:Desnivel
Model 2: VO2Max.ml.kg.min. ~ metros * Desnivel * Tipo.Suelo
  Res.Df    RSS Df Sum of Sq    F Pr(>F)
1      68 266.09
2      67 246.64  1    19.452 5.2841 0.02465 *
---
> summary(lm_completo)

Call:
lm(formula = VO2Max.ml.kg.min. ~ metros * Desnivel * Tipo.Suelo)

Residuals:

```

```

      Min      1Q  Median      3Q      Max
-4.6271 -1.3143  0.0996  1.2974  5.6145

```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.235e+01	8.915e-01	58.719	< 2e-16 ***
metros	1.908e-04	6.501e-05	2.934	0.00458 **
Desnivel	6.428e-03	4.308e-03	1.492	0.14036
Tipo.SueloMuntanya	9.735e+00	2.852e+00	3.414	0.00109 **
metros:Desnivel	-2.830e-07	2.712e-07	-1.043	0.30057
metros:Tipo.SueloMuntanya	-6.781e-04	2.777e-04	-2.442	0.01724 *
Desnivel:Tipo.SueloMuntanya	-1.948e-02	6.070e-03	-3.209	0.00205 **
metros:Desnivel:Tipo.SueloMuntanya	6.828e-07	2.970e-07	2.299	0.02465 *

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

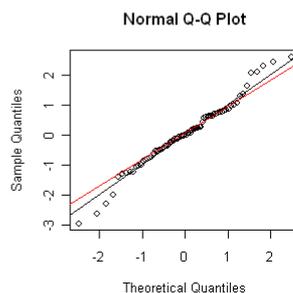
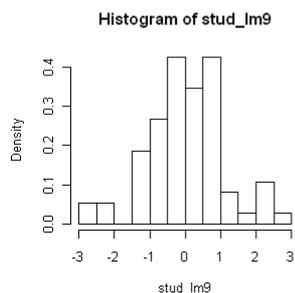
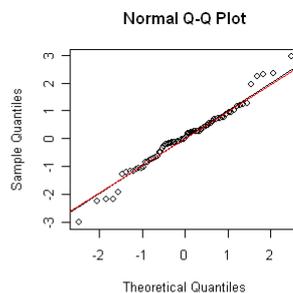
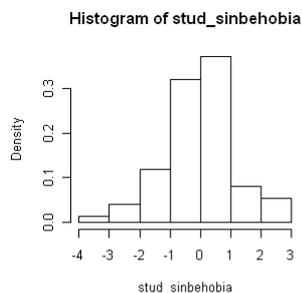
Residual standard error: 1.919 on 67 degrees of freedom
 Multiple R-squared: 0.4754, Adjusted R-squared: 0.4206
 F-statistic: 8.675 on 7 and 67 DF, p-value: 1.539e-07

6.5.6.3 Estudio de residuos

```

par(mfrow=c(2,2))
hist(stud_step,freq=F)
qqnorm(stud_step)
abline(0,1)
qqline(stud_step,col=2)
hist(stud_lm9,freq=F)
qqnorm(stud_lm9)
abline(0,1)
qqline(stud_lm9,col=2)

```



```

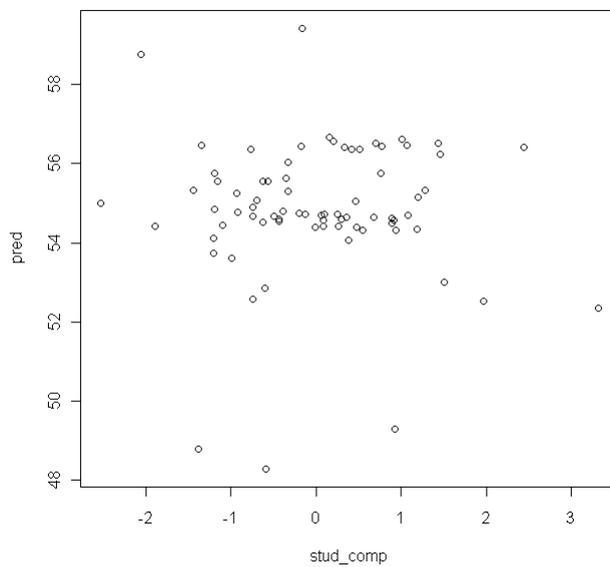
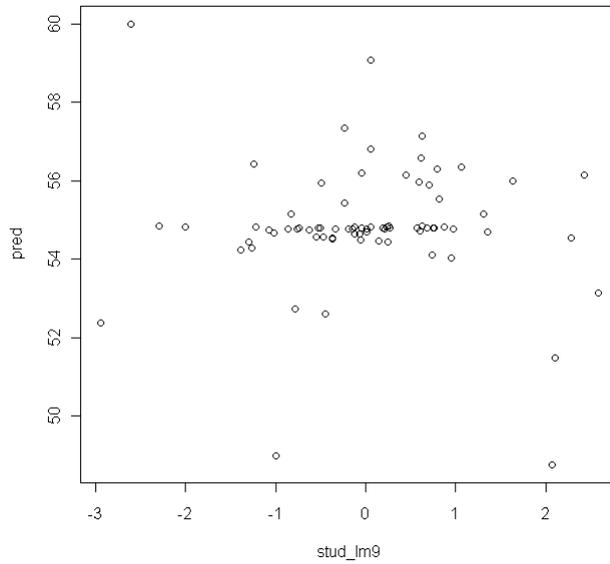
> llista_step<-which(abs(stud_step)>3)
> llista_lm9<-which(abs(stud_comp)>3)
> vo2max_sinbehobia[stud_lm_step_sosp,]
      cursa metros segundos Km_hora
34  Cursa de muntanya Santa Coloma de Queralt 14000 5114.004 9.855291
      VO2Max.ml.kg.min. Cooper Tipo.Suelo Desnivel Dificultad
34      47.40175 2059.688 Muntanya 870 Moderat

> vo2max_sinbehobia[stud_lm9_sosp,]
[1] cursa metros segundos Km_hora
[5] VO2Max.ml.kg.min. Cooper Tipo.Suelo Desnivel
[9] Dificultad
<0 rows> (or 0-length row.names)

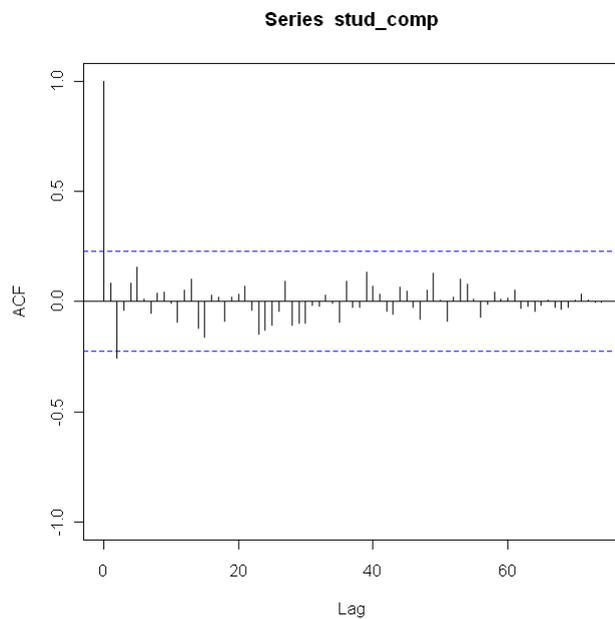
```

6.5.6.3.1 Diagrama bivalente de residuos estandarizados vs previstos

```
noudf<-data.frame(vo2max_sinbehobia)
pred<-predict(lm_completo, newdata=noudf)
cbind(noudf, pred)
plot(stud_comp,pred)
```



6.5.6.3.2 Independencia de los residuos studentizados: Test de Durbin-Watson o ACF (modelo completo)



```
> dwtest(lm_completo)
```

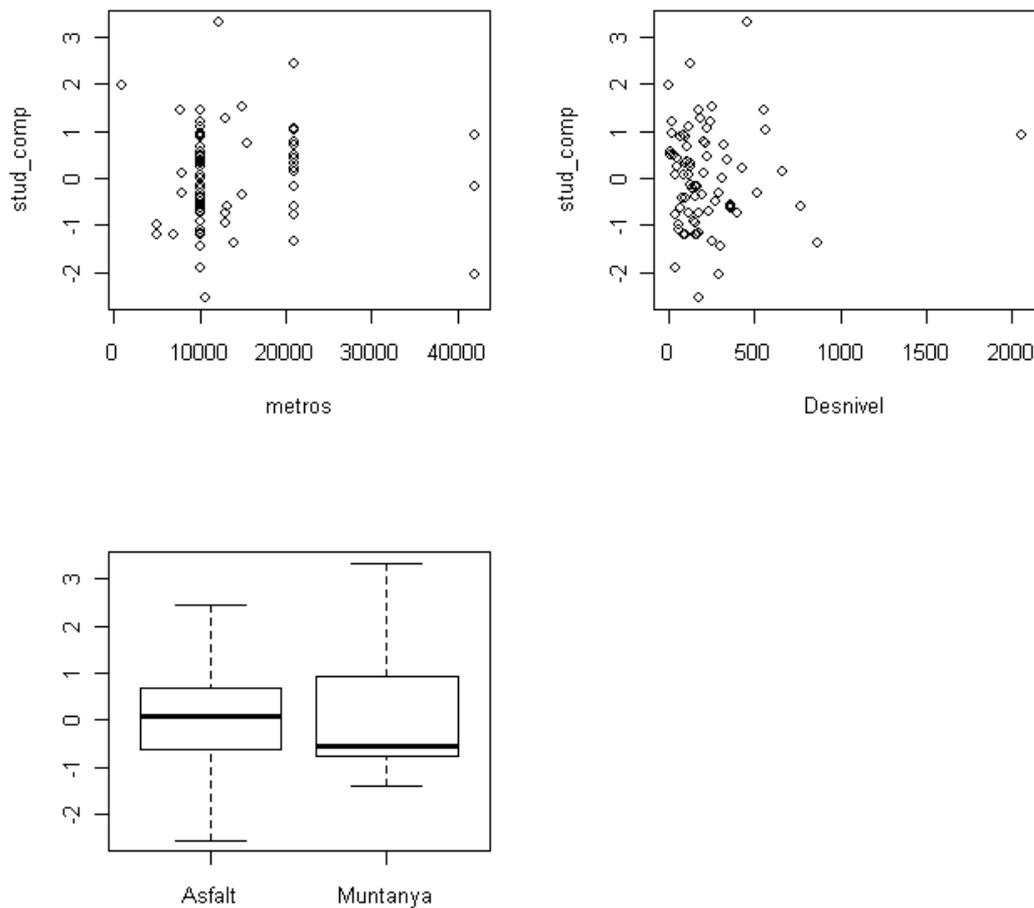
Durbin-Watson test

data: lm_completo

DW = 1.8019, p-value = 0.171

alternative hypothesis: true autocorrelation is greater than 0

6.5.6.3.3 Diagrama bivalente de los residuos estandarizados vs X_i (modelo completo)



6.5.6.3.4 Detección de valores influyentes a priori y a posteriori (modelo completo)

```
> cook_lm9=cooks.distance(lm9)
> cooksosp_lm9
> vo2max_sinbehobia[distcook_sosp,]
      curso metros segundos Km_hora
16 21a Cursa de Fons de Canovelles 15000 4391.645 12.296077
17 26a Pujada i Baixada a Guanta 12300 3302.002 13.410047
29 Cursa de la Sagrera 2011 10000 2599.884 13.846774
34 Cursa de muntanya Santa Coloma de Queralt 14000 5114.004 9.855291
37 Cursa de Santa Cristina d Aro 7700 1926.464 14.389060
55 Marató de Barcelona 42000 13814.180 10.945275
56 Marató de Collserola 42000 17267.933 8.756114
63 Mitja Marató de Collserola 21000 7943.875 9.516766
VO2Max.ml.kg.min. Cooper Tipo.Suelo Desnivel Dificultad stud_lm9
16 55.19843 2524.409 Muntanya 251 Moderat 2.0975943
17 57.95325 2739.287 Muntanya 461 Dificil 2.5878733
29 58.68283 2822.980 Asfalt 553 Moderat 2.2770265
34 47.40175 2059.688 Muntanya 870 Moderat -2.9433074
37 58.86323 2919.790 Muntanya 182 Moderat 1.6306615
55 55.71192 2240.927 Asfalt 293 Moderat -2.6131717
56 49.38874 1805.194 Muntanya 2051 Moderat 2.0631648
```

```
63      47.50632 1948.714  Muntanya      766      Moderat -0.9936232
      lev_lm9      cook_lm9
16 0.20042644  0.17518602
17 0.08676438  0.09795775
29 0.14586401  0.13913422
34 0.23228261  0.39318620
37 0.23273222  0.13127052
55 0.29171482  0.43223114
56 0.97570947  27.21268127
63 0.46116674  0.14085615
```

6.5.6.4 Análisis de los residuos y valores influyentes del modelo resultante del step

```
> oneway.test(VO2Max.ml.kg.min. ~ Tipo.Suelo, data=vo2max_curses)
```

One-way analysis of means (not assuming equal variances)

```
data: VO2Max.ml.kg.min. and Tipo.Suelo
```

```
F = 3.7261, num df = 1.000, denom df = 14.451, p-value = 0.07343
```

```
> lm0<-lm(VO2Max.ml.kg.min.~1, data=vo2max_curses)
```

```
> lm_step <-step(lm0,~ metros*Desnivel*Tipo.Suelo)
```

```
Start: AIC=129.4
```

```
VO2Max.ml.kg.min. ~ 1
```

	Df	Sum of Sq	RSS	AIC
+ Tipo.Suelo	1	32.409	381.51	125.36
+ Desnivel	1	21.640	392.27	127.42
+ metros	1	14.378	399.54	128.78
<none>			413.92	129.40

```
Step: AIC=125.36
```

```
VO2Max.ml.kg.min. ~ Tipo.Suelo
```

	Df	Sum of Sq	RSS	AIC
+ metros	1	16.489	365.02	124.09
<none>			381.51	125.36
+ Desnivel	1	6.330	375.18	126.13
- Tipo.Suelo	1	32.409	413.92	129.40

```
Step: AIC=124.1
```

```
VO2Max.ml.kg.min. ~ Tipo.Suelo + metros
```

	Df	Sum of Sq	RSS	AIC
+ metros:Tipo.Suelo	1	94.694	270.32	103.87
+ Desnivel	1	29.744	335.27	119.81
<none>			365.02	124.09
- metros	1	16.489	381.51	125.36
- Tipo.Suelo	1	34.519	399.54	128.78

```
Step: AIC=103.87
```

```
VO2Max.ml.kg.min. ~ Tipo.Suelo + metros + Tipo.Suelo:metros
```

	Df	Sum of Sq	RSS	AIC
+ Desnivel	1	9.869	260.46	103.12
<none>			270.32	103.87
- Tipo.Suelo:metros	1	94.694	365.02	124.09

```
Step: AIC=103.12
```

```
VO2Max.ml.kg.min. ~ Tipo.Suelo + metros + Desnivel + Tipo.Suelo:metros
```

	Df	Sum of Sq	RSS	AIC
+ metros:Desnivel	1	8.006	252.45	102.81
<none>			260.46	103.12
- Desnivel	1	9.869	270.32	103.87
+ Desnivel:Tipo.Suelo	1	2.367	258.09	104.44
- Tipo.Suelo:metros	1	74.819	335.27	119.81

```
Step: AIC=102.81
```

```
VO2Max.ml.kg.min. ~ Tipo.Suelo + metros + Desnivel + Tipo.Suelo:metros + metros:Desnivel
```

	Df	Sum of Sq	RSS	AIC
--	----	-----------	-----	-----

```

<none>                                252.45 102.81
- metros:Desnivel      1      8.006 260.45 103.12
+ Desnivel:Tipo.Suelo  1      0.059 252.39 104.79
- Tipo.Suelo:metros     1     53.007 305.46 114.91
> summary(lm_step)

Call:
lm(formula = VO2Max.ml.kg.min. ~ Tipo.Suelo + metros + Desnivel +
    Tipo.Suelo:metros + metros:Desnivel, data = vo2max_curses)

Residuals:
    Min       1Q   Median       3Q      Max
-4.5260 -1.1166 -0.0265  1.3304  4.7459

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      5.351e+01  6.325e-01  84.603 < 2e-16 ***
Tipo.SueloMuntanya  7.086e+00  2.309e+00   3.068 0.003089 **
metros           9.963e-05  4.134e-05   2.410 0.018648 *
Desnivel        -2.322e-05  2.558e-03  -0.009 0.992784
Tipo.SueloMuntanya:metros -7.243e-04  1.917e-04  -3.779 0.000334 ***
metros:Desnivel    1.675e-07  1.141e-07   1.469 0.146576
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.927 on 68 degrees of freedom
Multiple R-squared:  0.3901,    Adjusted R-squared:  0.3452
F-statistic: 8.699 on 5 and 68 DF,  p-value: 2.049e-06

```

Forward

```

> summary(lm6)

Call:
lm(formula = VO2Max.ml.kg.min. ~ Tipo.Suelo + metros + Tipo.Suelo:metros)

Residuals:
    Min       1Q   Median       3Q      Max
-4.5771 -1.2449  0.0013  1.3759  4.4201

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      5.332e+01  5.247e-01 101.626 < 2e-16 ***
Tipo.SueloMuntanya  3.173e+00  1.170e+00   2.713 0.00839 **
metros           1.460e-04  3.475e-05   4.201 7.7e-05 ***
Tipo.SueloMuntanya:metros -3.562e-04  7.192e-05  -4.952 4.9e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

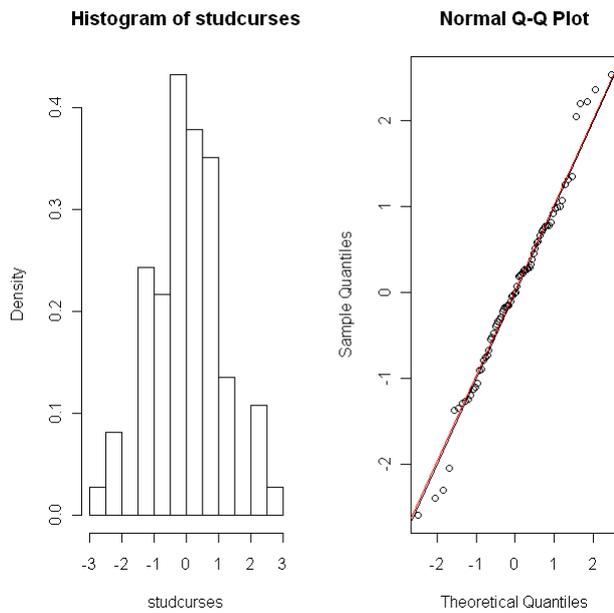
Residual standard error: 1.965 on 70 degrees of freedom
Multiple R-squared:  0.3469,    Adjusted R-squared:  0.3189
F-statistic: 12.39 on 3 and 70 DF,  p-value: 1.362e-06

```

```

*****Linealitat, Homocedasticitat i Independència dels Residus *****
studcurses<-rstudent(lm6)
stud_sosp<-which(abs(studcurses)>3)
vo2max_curses[stud_sosp,]
1] cursa      metros      segundos      Km_hora
[5] VO2Max.ml.kg.min. Cooper      Tipo.Suelo      Desnivel
[9] Dificultad
<0 rows> (or 0-length row.names)

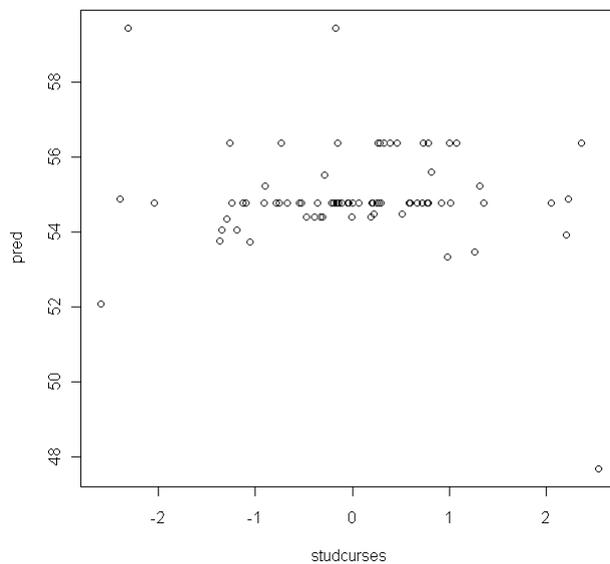
```



6.5.6.4.1 Diagrama bivalente de los residuos estandarizados vs predichos

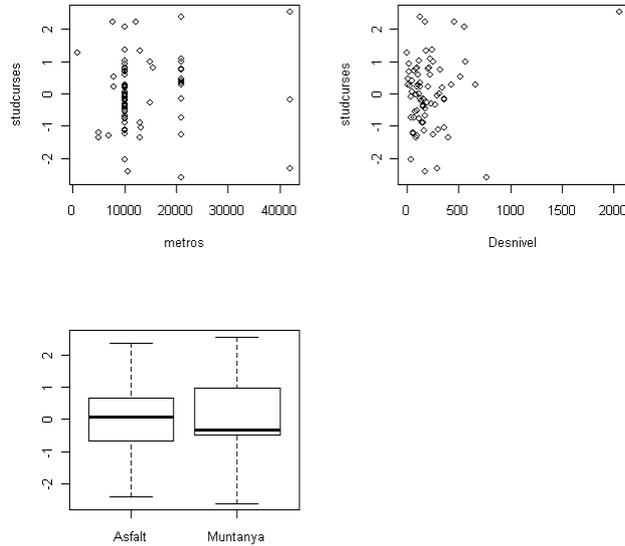
```
noudf<-data.frame(vo2max_curses)
noudf
pred<-predict(lm_completo, newdata=noudf)
cbind(noudf, pred)

plot(studcourses,pred)
```



6.5.6.4.2 Diagrama bivalente de los residuos estandarizados vs regresoras

```
par(mfrow=c(2,2))
plot(metros,studcurses)
plot(Desnivel,studcurses)
plot(Tipo.Suelo,studcurses)
```

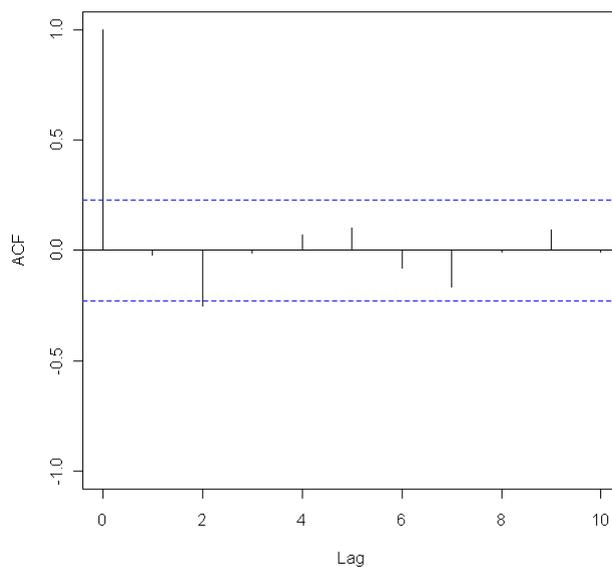


```
> dwtest(lm6)
> dwtest(lm6)
```

Durbin-Watson test

```
data: lm6
DW = 1.9295, p-value = 0.3571
alternative hypothesis: true autocorrelation is greater than 0
Mensajes de aviso perdidos
In dwtest(lm6) :
  exact p value cannot be computed (not in [0,1]), approximate p value will be used
```

Series studcurses



```
anclatge_sosp<-which(levcurses>2*4/length(VO2Max.ml.kg.min.))
> vo2max_curses[anclatge_sosp,]
```

	curso	metros	segundos	Km_hora	VO2Max.ml.kg.min.
36	Cursa de Santa Cristina d Aro	7700	1926.464	14.389060	58.86323
54	Marató de Barcelona	42000	13814.180	10.945275	55.71192
55	Marató de Collserola	42000	17267.933	8.756114	49.38874
56	Marató d Empúries	42000	12452.694	12.141951	59.16840
62	Mitja Marató de Collserola	21000	7943.875	9.516766	47.50632

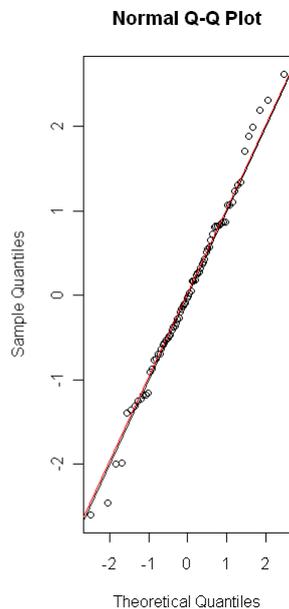
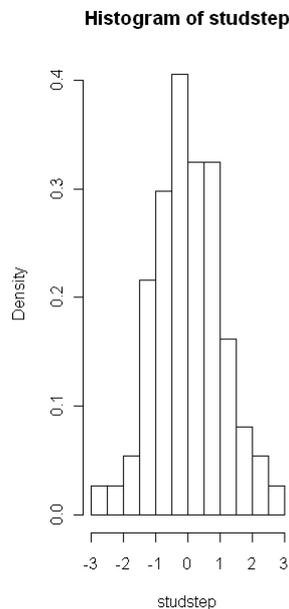
	Cooper	Tipo.Suelo	Desnivel	Dificultad	studcourses	levcourses	cookcourses
36	2919.790	Muntanya	182	Moderat	2.2212723	0.1198979	0.159102031
54	2240.927	Asfalt	293	Moderat	-2.3042402	0.2747505	0.473696392
55	1805.194	Muntanya	2051	Moderat	2.5409587	0.8722751	10.226207685
56	2481.884	Asfalt	171	Moderat	-0.1698658	0.2747505	0.002771217
62	1948.714	Muntanya	766	Moderat	-2.5891693	0.1248355	0.221049598

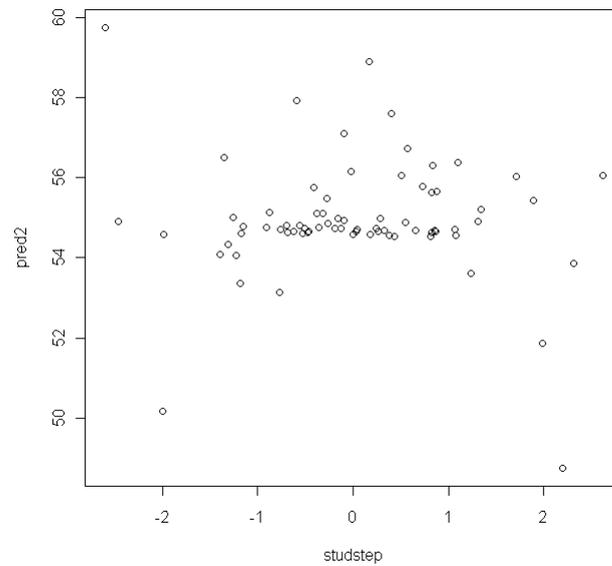
```
> distcook_sosp<-which(cookcourses>4/(length(VO2Max.ml.kg.min.)-8))
> vo2max_curses[distcook_sosp,]
```

	curso	metros	segundos	Km_hora	VO2Max.ml.kg.min.
17	26a Pujada i Baixada a Guanta	12300	3302.002	13.410047	57.95325
36	Cursa de Santa Cristina d Aro	7700	1926.464	14.389060	58.86323
54	Marató de Barcelona	42000	13814.180	10.945275	55.71192
55	Marató de Collserola	42000	17267.933	8.756114	49.38874
62	Mitja Marató de Collserola	21000	7943.875	9.516766	47.50632

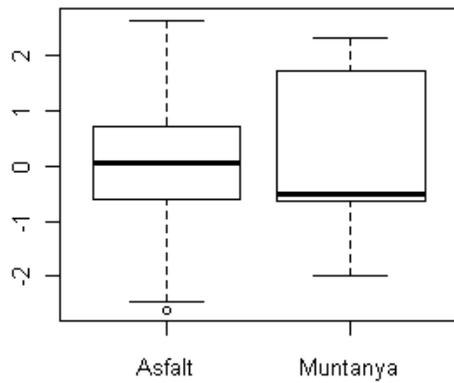
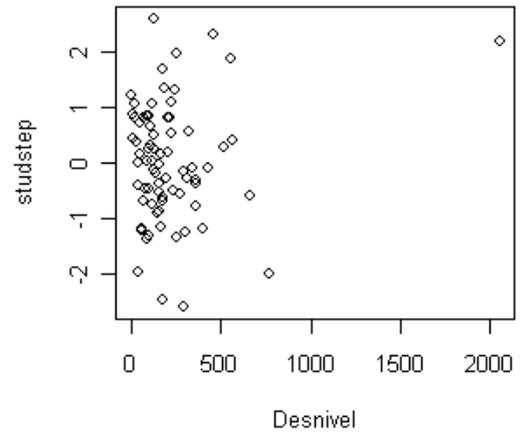
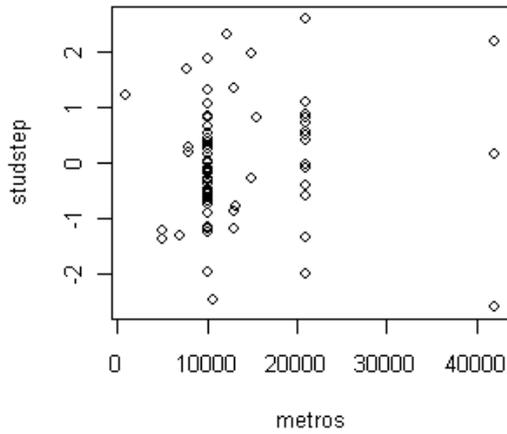
	Cooper	Tipo.Suelo	Desnivel	Dificultad	studcourses	levcourses	cookcourses
17	2739.287	Muntanya	461	Dificil	2.202933	0.08051093	0.1006888
36	2919.790	Muntanya	182	Moderat	2.221272	0.11989791	0.1591020
54	2240.927	Asfalt	293	Moderat	-2.304240	0.27475048	0.4736964
55	1805.194	Muntanya	2051	Moderat	2.540959	0.87227508	10.2262077
62	1948.714	Muntanya	766	Moderat	-2.589169	0.12483551	0.2210496

```
studstep<-rstudent(lm_step)
stud_sosp<-which(abs(studstep)>3)
vo2max_curses[stud_sosp,]
1] curso metros segundos Km_hora
[5] VO2Max.ml.kg.min. Cooper Tipo.Suelo Desnivel
[9] Dificultad
<0 rows> (or 0-length row.names)
```





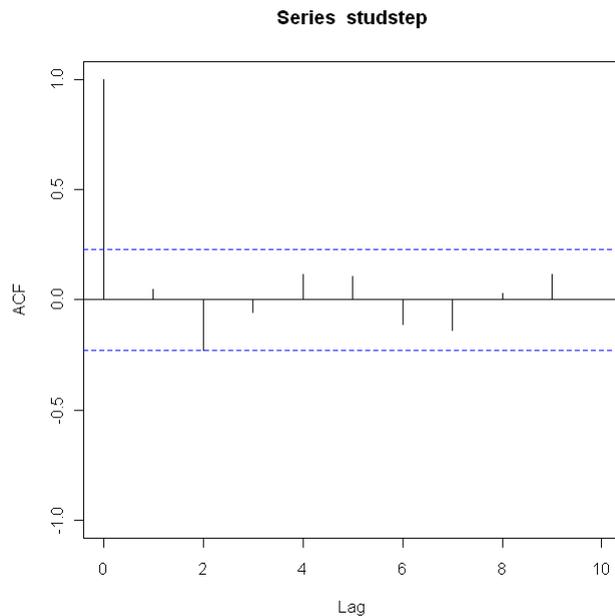
```
> par(mfrow=c(2,2))  
> plot(metros,studstep)  
> plot(Desnivel,studstep)  
> plot(Tipo.Suelo,studstep)
```



```
> dwtest(lm_step)

Durbin-Watson test

data:  lm_step
DW = 1.7824, p-value = 0.1496
alternative hypothesis: true autocorrelation is greater than 0
```



```
> anclatge_step_sosp<-which(levstep>2*6/length(VO2Max.ml.kg.min.))
> vo2max_curses[anclatge_step_sosp,]
```

	curso	metros	segundos	Km_hora
16	21a Cursa de Fons de Canovelles	15000	4391.645	12.296077
25	Cursa de festa major de Terrassa	8000	2175.113	13.240689
29	Cursa de la Sagrera 2011	10000	2599.884	13.846774
36	Cursa de Santa Cristina d Aro	7700	1926.464	14.389060
54	Marató de Barcelona	42000	13814.180	10.945275
55	Marató de Collserola	42000	17267.933	8.756114
56	Marató d Empúries	42000	12452.694	12.141951
58	Mitja Marató Ciutat de Girona	21000	6040.523	12.515472
62	Mitja Marató de Collserola	21000	7943.875	9.516766

	VO2Max.ml.kg.min.	Cooper	Tipo.Suelo	Desnivel	Dificultat	studstep
16	55.19843	2524.409	Muntanya	251	Moderat	1.9882522
25	55.48815	2711.175	Asfalt	520	Facil	0.2792585
29	58.68283	2822.980	Asfalt	553	Moderat	1.8925264
36	58.86323	2919.790	Muntanya	182	Moderat	1.7083939
54	55.71192	2240.927	Asfalt	293	Moderat	-2.5985825
55	49.38874	1805.194	Muntanya	2051	Moderat	2.1950248
56	59.16840	2481.884	Asfalt	171	Moderat	0.1666170
58	56.91158	2553.262	Asfalt	666	Moderat	-0.5920725
62	47.50632	1948.714	Muntanya	766	Moderat	-1.9962919

	levstep	cookstep
16	0.2047865	0.162609716
25	0.1714313	0.002726165
29	0.1705034	0.118213773
36	0.2327490	0.143512989
54	0.2936340	0.431349814
55	0.9757098	30.541599383
56	0.2951263	0.001965343
58	0.2058813	0.015293177
62	0.5037676	0.645926714

```
> distcook_step_sosp<-which(cookstep>4/(length(VO2Max.ml.kg.min.))-8))
> vo2max_curses[distcook_step_sosp,]
```

	curso	metros	segundos	Km_hora	VO2Max.ml.kg.min.
16	21a Cursa de Fons de Canovelles	15000	4391.645	12.296077	55.19843
17	26a Pujada i Baixada a Guanta	12300	3302.002	13.410047	57.95325
29	Cursa de la Sagrera 2011	10000	2599.884	13.846774	58.68283
36	Cursa de Santa Cristina d Aro	7700	1926.464	14.389060	58.86323
54	Marató de Barcelona	42000	13814.180	10.945275	55.71192
55	Marató de Collserola	42000	17267.933	8.756114	49.38874
62	Mitja Marató de Collserola	21000	7943.875	9.516766	47.50632

	Cooper	Tipo.Suelo	Desnivel	Dificultat	studstep	levstep	cookstep
16	2524.409	Muntanya	251	Moderat	1.988252	0.2047865	0.16260972
17	2739.287	Muntanya	461	Dificil	2.317199	0.1032929	0.09686104
29	2822.980	Asfalt	553	Moderat	1.892526	0.1705034	0.11821377

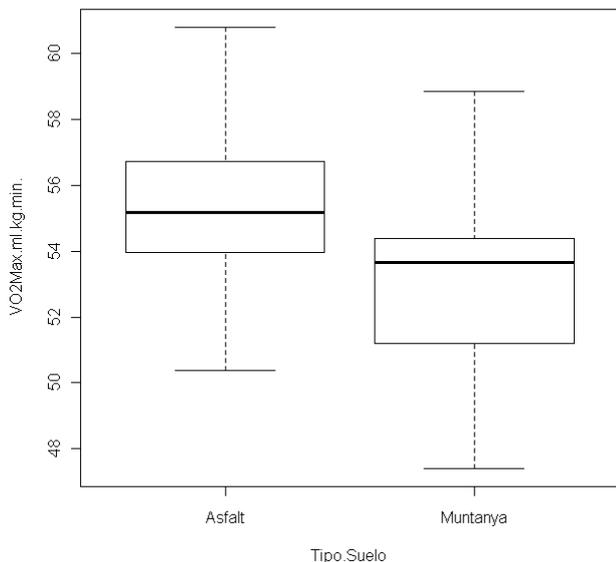
36	2919.790	Muntanya	182	Moderat	1.708394	0.2327490	0.14351299
54	2240.927	Asfalt	293	Moderat	-2.598582	0.2936340	0.43134981
55	1805.194	Muntanya	2051	Moderat	2.195025	0.9757098	30.54159938
62	1948.714	Muntanya	766	Moderat	-1.996292	0.5037676	0.64592671

6.5.7 Resultados juego de datos cursas sin la carrera de Guanta

6.5.7.1 Análisis descriptivo

```
> summary(vo2max_curses)
      cursa      metros      segundos
10 k Nou Barris      : 1  Min.      : 1000  Min.      : 204.1
10 km Bellavista      : 1  1st Qu.:10000  1st Qu.: 2825.9
10 km Collbatonina-olesa: 1  Median :10000  Median : 2944.8
10 km de Pineda de Mar : 1  Mean   :13436  Mean   : 4008.2
10 km Pont de Vilomara : 1  3rd Qu.:15000  3rd Qu.: 4412.5
10 km Roda de Ter      : 1  Max.   :42000  Max.   :17267.9
(Other)                :68
      Km_hora  VO2Max.ml.kg.min.  Cooper  Tipo.Suelo
Min.   : 8.756  Min.   :47.40  Min.   :1805  Asphalt :61
1st Qu.:12.143  1st Qu.:53.55  1st Qu.:2494  Muntanya:13
Median :12.519  Median :54.92  Median :2556
Mean   :12.442  Mean   :54.81  Mean   :2549
3rd Qu.:12.885  3rd Qu.:56.31  3rd Qu.:2638
Max.   :17.643  Max.   :60.81  Max.   :3589

      Desnivel      Dificultad
Min.   : 0.00  Moderat :35
1st Qu.: 88.25  Facil   :20
Median :159.00  0       :12
Mean   :228.54  Dificil : 2
3rd Qu.:273.75  A ull   : 1
Max.   :2051.00  inf 2008:1
(Other) : 3
```



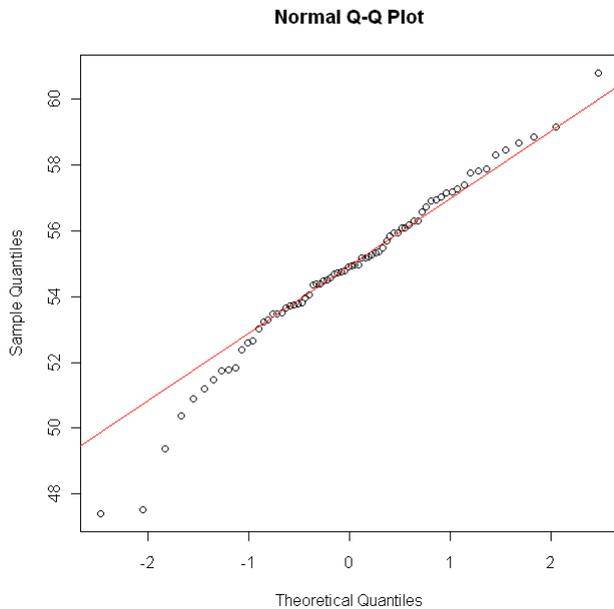
```
> oneway.tes(VO2Max.ml.kg.min. ~ Tipo.Suelo, data=vo2max_curses)
```

One-way analysis of means (not assuming equal variances)

```
data: VO2Max.ml.kg.min. and Tipo.Suelo
F = 7.4733, num df = 1.000, denom df = 14.264, p-value = 0.01594
```

6.5.7.2 Estudio de normalidad de los datos

```
hist(VO2Max.ml.kg.min.,freq=F)
qqnorm(VO2Max.ml.kg.min.)
abline(0,1)
qqline(VO2Max.ml.kg.min.,col=2)
```



```
VO2Max_teo<rnorm(length(VO2Max.ml.kg.min),mean(VO2Max.ml.kg.min.),sd(VO2Max.ml.kg.min.))
ks.test(VO2Max.ml.kg.min., VO2Max_teo)
```

Two-sample Kolmogorov-Smirnov test

```
data: VO2Max.ml.kg.min. and VO2Max_teo
D = 0.0946, p-value = 0.8982
alternative hypothesis: two-sided
```

6.5.7.3 Búsqueda del modelo

```
> cor(data.frame(VO2Max.ml.kg.min.,Desnivel,metros))
                VO2Max.ml.kg.min.  Desnivel  metros
VO2Max.ml.kg.min.  1.0000000 -0.3193506  0.1761236
Desnivel           -0.3193506  1.0000000  0.4835307
metros             0.1761236  0.4835307  1.0000000
```

```
> summary(lm_step)
```

```
Call:
lm(formula = VO2Max.ml.kg.min. ~ Tipo.Suelo + metros + Tipo.Suelo:metros,
    data = vo2max_curses)
```

```
Residuals:
    Min     1Q  Median     3Q     Max
-5.366 -1.278  0.174  1.272  4.836
```

```
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      5.332e+01  5.376e-01  99.193 < 2e-16 ***
Tipo.SueloMuntanya  2.243e+00  1.206e+00   1.859  0.067198 .
metros           1.460e-04  3.560e-05   4.101  0.000109 ***
Tipo.SueloMuntanya:metros -3.459e-04  7.379e-05  -4.688  1.32e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 2.013 on 70 degrees of freedom
Multiple R-squared: 0.3837, Adjusted R-squared: 0.3573
```

```

F-statistic: 14.53 on 3 and 70 DF, p-value: 1.875e-07

> lm_completo<-lm(VO2Max.ml.kg.min. ~metros*Desnivel*Tipo.Suelo)
> summary(lm_completo)

Call:
lm(formula = VO2Max.ml.kg.min. ~ metros * Desnivel * Tipo.Suelo)

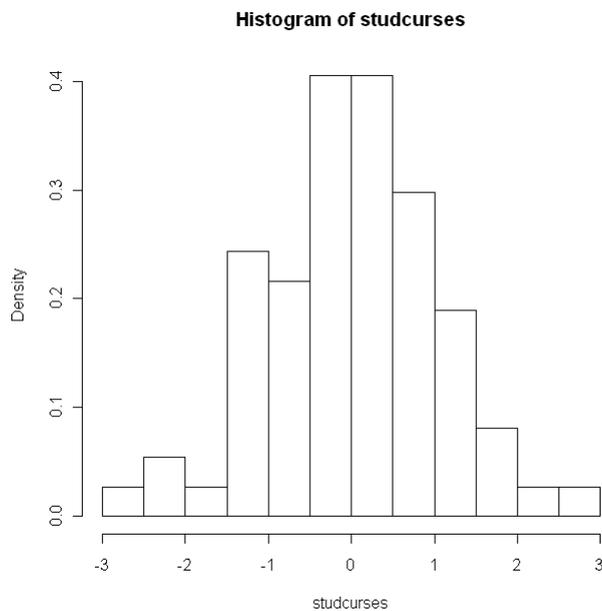
Residuals:
    Min       1Q   Median       3Q      Max
-4.6271 -1.1364  0.0618  1.2789  4.3932

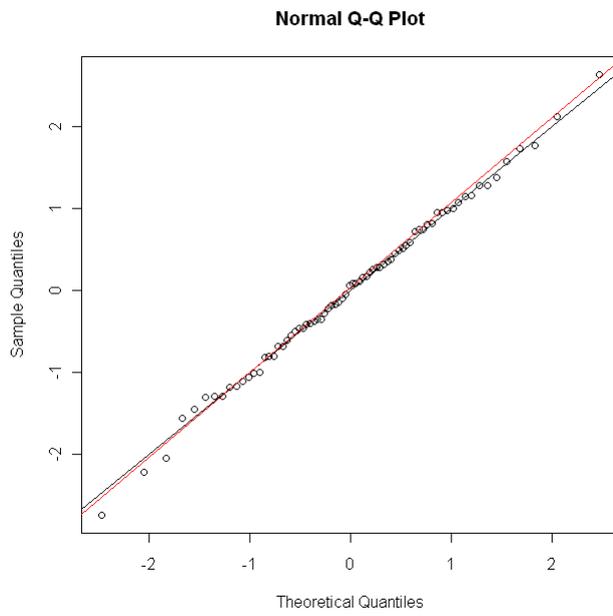
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    5.235e+01  8.314e-01  62.959 < 2e-16 ***
metros         1.908e-04  6.063e-05   3.146 0.002483 **
Desnivel       6.428e-03  4.018e-03   1.600 0.114401
Tipo.SueloMuntanya  9.537e+00  2.660e+00   3.585 0.000640 ***
metros:Desnivel -2.830e-07  2.530e-07  -1.119 0.267360
metros:Tipo.SueloMuntanya -6.449e-04  2.592e-04  -2.489 0.015357 *
Desnivel:Tipo.SueloMuntanya -2.196e-02  5.710e-03  -3.846 0.000273 ***
metros:Desnivel:Tipo.SueloMuntanya  7.290e-07  2.774e-07   2.628 0.010662 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.789 on 66 degrees of freedom
Multiple R-squared:  0.541,    Adjusted R-squared:  0.4923
F-statistic: 11.11 on 7 and 66 DF, p-value: 3.395e-09

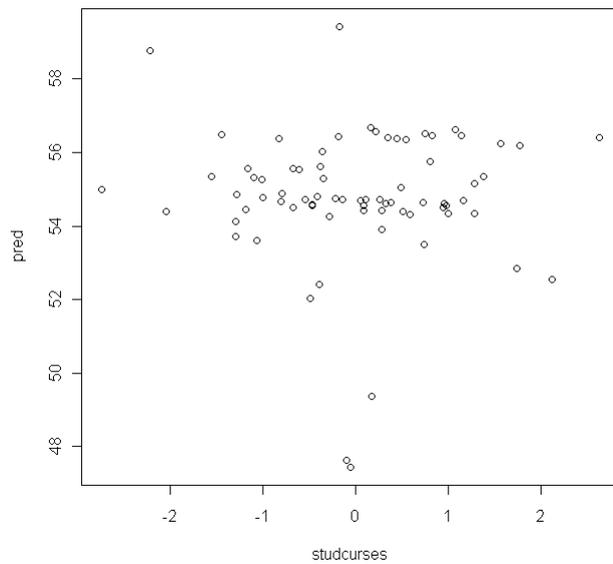
```

6.5.7.4 Estudio de residuos

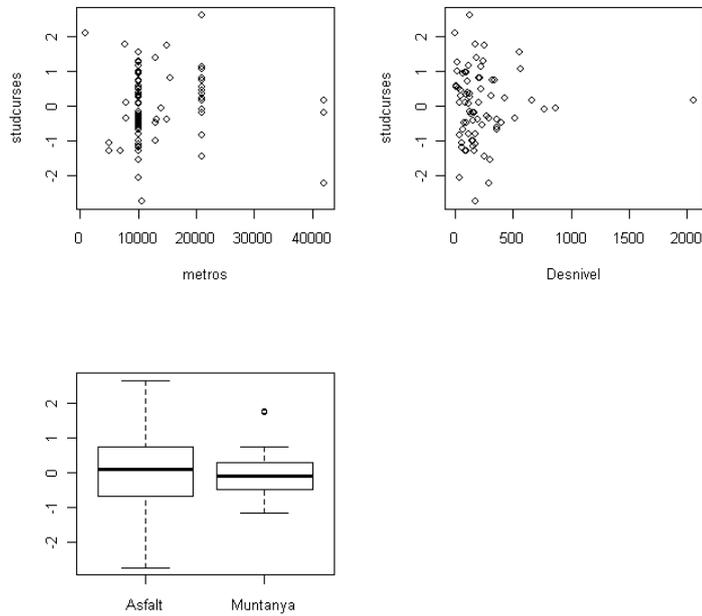




6.5.7.4.1 Diagrama bivalente de residuos studentizados vs previstos



6.5.7.4.2 Diagrama bivalente de los residuos estandarizados vs X_i

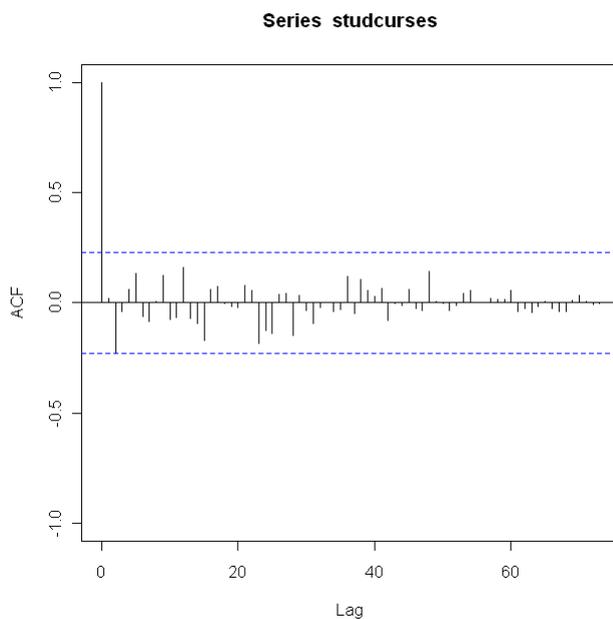


6.5.7.4.3 Independencia de los residuos studentizados: Test de Durbin-Watson o ACF

```
> dwtest(lm_completo)

Durbin-Watson test

data: lm_completo
DW = 1.9185, p-value = 0.3455
alternative hypothesis: true autocorrelation is greater than 0
```



6.5.7.5 Detección de valores influyentes a priori y a posteriori

```

studcourses
  levcourses<-hatvalues(lm_completo)
  cookcourses=cooks.distance(lm_completo)
  vo2max_courses<-data.frame(vo2max_courses,studcourses,levcourses,cookcourses)
  attach(vo2max_courses)
vo2max_courses[stud_sosp,]
 [1] cursa      metros      segundos      Km_hora
 [5] VO2Max.ml.kg.min. Cooper      Tipo.Suelo    Desnivel
 [9] Dificultad  studcourses levcourses    cookcourses
<0 rows> (or 0-length row.names)

> vo2max_courses[anclatge_sosp,]
      cursa metros segundos Km_hora
16      21a Cursa de Fons de Canovelles 15000 4391.645 12.296077
24      Cursa de festa major de Terrassa 8000 2175.113 13.240689
28      Cursa de la Sagrera 2011 10000 2599.884 13.846774
33 Cursa de muntanya Santa Coloma de Queralt 14000 5114.004 9.855291
36      Cursa de Santa Cristina d Aro 7700 1926.464 14.389060
54      Marató de Barcelona 42000 13814.180 10.945275
55      Marató de Collserola 42000 17267.933 8.756114
56      Marató d Empúries 42000 12452.694 12.141951
58      Mitja Marató Ciutat de Girona 21000 6040.523 12.515472
62      Mitja Marató de Collserola 21000 7943.875 9.516766
64      Mitja Marató de l Anoia 21000 5831.977 12.963014
  VO2Max.ml.kg.min. Cooper Tipo.Suelo Desnivel Dificultad studcourses
16      55.19843 2524.409 Muntanya 251 Moderat 1.73563391
24      55.48815 2711.175 Asfalt 520 Facil -0.35452564
28      58.68283 2822.980 Asfalt 553 Moderat 1.56702605
33      47.40175 2059.688 Muntanya 870 Moderat -0.04952041
36      58.86323 2919.790 Muntanya 182 Moderat 1.76698106
54      55.71192 2240.927 Asfalt 293 Moderat -2.21339641
55      49.38874 1805.194 Muntanya 2051 Moderat 0.17097827
56      59.16840 2481.884 Asfalt 171 Moderat -0.17281214
58      56.91158 2553.262 Asfalt 666 Moderat 0.16362749
62      47.50632 1948.714 Muntanya 766 Moderat -0.09555949
64      58.31527 2624.562 Asfalt 570 Dificil 1.07755789
  levcourses cookcourses
16 0.4116703 0.255688194
24 0.2586717 0.005555655
28 0.2268589 0.088122247
33 0.7767818 0.001083084
36 0.2592977 0.132368260
54 0.3730278 0.344027781
55 0.9970694 1.261823566
56 0.3212428 0.001793117
58 0.3613898 0.001922270
62 0.5452625 0.001389543
64 0.2335425 0.044117413

> distcook_sosp<-which(cookcourses>4/(length(VO2Max.ml.kg.min.)-8))
> vo2max_courses[distcook_sosp,]
      cursa metros segundos Km_hora
10 1000 metres Pista Coberta Sabadell 1000 204.050 17.642735
16      21a Cursa de Fons de Canovelles 15000 4391.645 12.296077
28      Cursa de la Sagrera 2011 10000 2599.884 13.846774
36      Cursa de Santa Cristina d Aro 7700 1926.464 14.389060
54      Marató de Barcelona 42000 13814.180 10.945275
55      Marató de Collserola 42000 17267.933 8.756114
  VO2Max.ml.kg.min. Cooper Tipo.Suelo Desnivel Dificultad studcourses
10      55.86202 3589.136 Asfalt 0 0 2.1164574
16      55.19843 2524.409 Muntanya 251 Moderat 1.7356339
28      58.68283 2822.980 Asfalt 553 Moderat 1.5670260
36      58.86323 2919.790 Muntanya 182 Moderat 1.7669811
54      55.71192 2240.927 Asfalt 293 Moderat -2.2133964
55      49.38874 1805.194 Muntanya 2051 Moderat 0.1709783
  levcourses cookcourses
10 0.1887102 0.12371896
16 0.4116703 0.25568819
28 0.2268589 0.08812225
36 0.2592977 0.13236826
54 0.3730278 0.34402778
55 0.9970694 1.26182357

```

6.5.8 Resultados juego de datos completo.

6.5.8.1 Cursa de Bombers

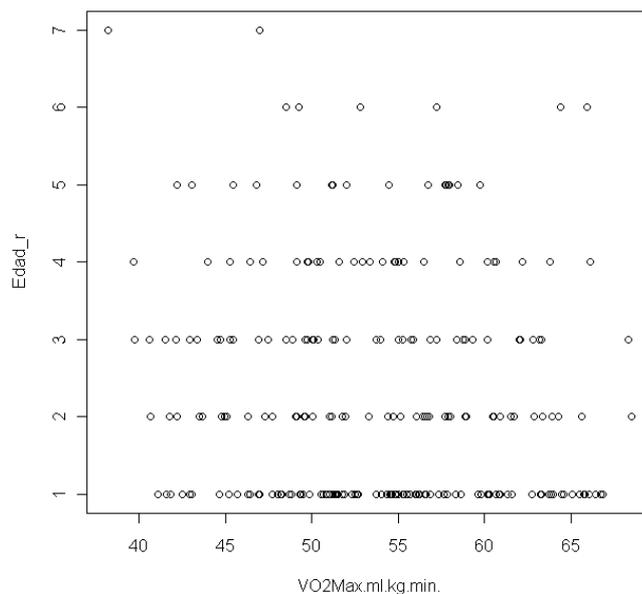
```
> summary(bombers)
```

Lap	Lap.position	Posició	Posició.per.categoria
Km2.5Km5 : 1	Min. :6.673e+05	Min. : 74	Min. : 12
Km2.5Km5Km7.5:246	1st Qu.:1.658e+11	1st Qu.: 1752	1st Qu.: 348
Km5Km7.5 : 2	Median :4.377e+11	Median : 4270	Median :1041
	Mean :2.271e+13	Mean : 5675	Mean :1439
	3rd Qu.:8.080e+11	3rd Qu.: 8510	3rd Qu.:2325
	Max. :1.694e+14	Max. :17657	Max. :5272

Lap.time	Temps.oficial	Personal..Gènere	Uid
08:3417:4427:00: 1	1:01:22: 2	female: 34	Min. : 12.0
08:3617:3126:21: 1	1:04:39: 2	male :215	1st Qu.: 319.0
08:4417:3026:14: 1	38:02 : 2		Median : 647.0
08:4718:1227:39: 1	38:44 : 2		Mean : 758.9
08:4819:0929:46: 1	41:12 : 2		3rd Qu.:1162.0
08:5318:0527:29: 1	43:09 : 2		Max. :2092.0
(Other) :243	(Other):237		

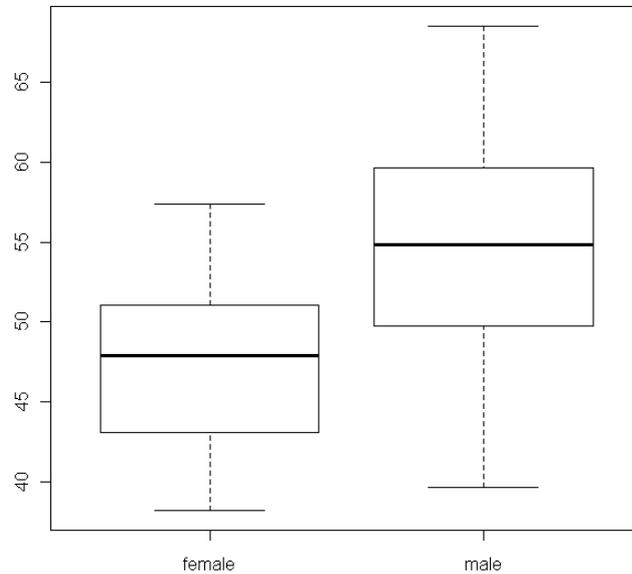
metros	segundos	Km_hora	VO2Max.ml.kg.min.
Min. :10000	Min. :2108	Min. : 7.122	Min. :38.21
1st Qu.:10000	1st Qu.:2607	1st Qu.:10.619	1st Qu.:48.86
Median :10000	Median :2948	Median :12.211	Median :53.70
Mean :10000	Mean :3060	Mean :12.190	Mean :53.64
3rd Qu.:10000	3rd Qu.:3390	3rd Qu.:13.810	3rd Qu.:58.57
Max. :10000	Max. :5055	Max. :17.080	Max. :68.53

Cooper	Edad	Tipo_suelo	Desnivel	Dificultad
Min. :1424	Min. :18.00	Asfalt:249	Min. :56	Facil:249
1st Qu.:2124	1st Qu.:32.00		1st Qu.:56	
Median :2442	Median :37.00		Median :56	
Mean :2438	Mean :38.05		Mean :56	
3rd Qu.:2762	3rd Qu.:44.00		3rd Qu.:56	
Max. :3416	Max. :73.00		Max. :56	



```
> oneway.test(VO2Max.ml.kg.min.~Edad_r, data=bombers)
```

```
One-way analysis of means (not assuming equal variances)
data: VO2Max.ml.kg.min. and Edad_r
F = 1.7526, num df = 6.000, denom df = 12.418, p-value = 0.1896
```



```
> t.test(VO2Max.ml.kg.min.~Personal..Gènere)
```

Welch Two Sample t-test

```
data: VO2Max.ml.kg.min. by Personal..Gènere
t = -7.3795, df = 53.967, p-value = 1e-09
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-8.944642 -5.122741
sample estimates:
mean in group female    mean in group male
      47.56676             54.60045
```

```
step3<step(VO2Max_lm,~Posició*Edad*Posició.per.categoria*Personal..Gènere)
```

```
Step: AIC=217.24
VO2Max.ml.kg.min. ~ Posició + Posició.per.categoria + Personal..Gènere +
  Edad + Posició:Posició.per.categoria + Posició:Personal..Gènere +
  Posició:Edad
```

	Df	Sum of Sq	RSS	AIC
<none>			558.71	217.24
+ Edad:Posició.per.categoria	1	2.58	556.14	218.09
+ Edad:Personal..Gènere	1	2.38	556.33	218.17
+ Posició.per.categoria:Personal..Gènere	1	0.08	558.63	219.20
- Posició:Edad	1	21.97	580.68	224.84
- Posició:Personal..Gènere	1	66.87	625.58	243.39
- Posició:Posició.per.categoria	1	435.43	994.14	358.72

```
> summary(step3)
```

```
Call:
lm(formula = VO2Max.ml.kg.min. ~ Posició + Posició.per.categoria +
  Personal..Gènere + Edad + Posició:Posició.per.categoria +
  Posició:Personal..Gènere + Posició:Edad)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-6.5269 -0.9245 -0.1062  0.9727  5.0266
```

```
Coefficients:
(Intercept)                Estimate Std. Error t value Pr(>|t|)
                6.206e+01  1.002e+00  61.959 < 2e-16 ***
```

```

Posició -1.528e-03 1.382e-04 -11.061 < 2e-16 ***
Posició.per.categoria -2.629e-03 2.291e-04 -11.477 < 2e-16 ***
Personal..Gènerefemale 4.131e+00 7.089e-01 5.827 1.80e-08 ***
Edad -6.599e-02 1.855e-02 -3.557 0.000451 ***
Posició:Posició.per.categoria 2.370e-07 1.729e-08 13.705 < 2e-16 ***
Posició:Personal..Gènerefemale -4.012e-04 7.470e-05 -5.371 1.85e-07 ***
Posició:Edad 8.078e-06 2.624e-06 3.078 0.002323 **
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.523 on 241 degrees of freedom
Multiple R-squared: 0.9517, Adjusted R-squared: 0.9503
F-statistic: 678.1 on 7 and 241 DF, p-value: < 2.2e-16

> summary(step_bombers)

Call:
lm(formula = VO2Max.ml.kg.min. ~ Posició + Posició.per.categoria +
    Personal..Gènere + Edad)

Residuals:
    Min       1Q   Median       3Q      Max
-6.7487 -1.2500 -0.2697  0.9315  6.8049

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  60.6441801  0.6985170  86.818 < 2e-16 ***
Posició      -0.0010611  0.0000643 -16.504 < 2e-16 ***
Posició.per.categoria -0.0012039  0.0002122  -5.673 3.96e-08 ***
Personal..Gènerefemale  2.9340195  0.6625458   4.428 1.43e-05 ***
Edad         -0.0468876  0.0181087  -2.589  0.0102 *
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.03 on 244 degrees of freedom
Multiple R-squared: 0.913, Adjusted R-squared: 0.9116
F-statistic: 640.4 on 4 and 244 DF, p-value: < 2.2e-16

```

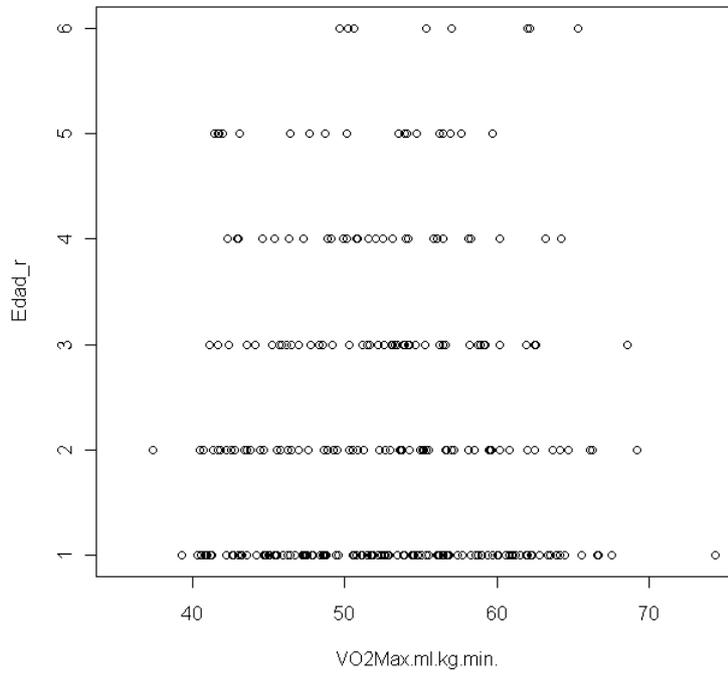
6.5.8.2 Cursa de la Mercè.

```

> summary(merce)
Lap      Lap.position      Posició      Posició.per.categoria      Lap.time
km5:305  Min.       : 3      Min.       : 3      Min.       : 1      21:02 : 3
          1st Qu.: 1072  1st Qu.: 1189  1st Qu.: 731      21:59 : 3
          Median : 2459  Median : 2911  Median : 2127     22:34 : 3
          Mean   : 3503  Mean   : 3837  Mean   : 2744     23:14 : 3
          3rd Qu.: 5531  3rd Qu.: 6410  3rd Qu.: 4323     24:58 : 3
          Max.   : 11178  Max.   : 11146  Max.   : 8337     18:11 : 2
                                     (Other):288
                                     segundos
Temps.oficial Personal..Gènere      Uid      metros
46:53 : 3      female: 42      Min.   : 11      Min.   :10000      Min.   :1895
48:49 : 3      male :263      1st Qu.: 431      1st Qu.:10000      1st Qu.:2704
1:00:03: 2      Median :1009      Median :10000      Median :3054
1:01:21: 2      Mean   :1038      Mean   :10000      Mean   :3188
1:02:00: 2      3rd Qu.:1608      3rd Qu.:10000      3rd Qu.:3653
1:08:50: 2      Max.   :2279      Max.   :10000      Max.   :5843
(Other):291
      Km_hora      VO2Max.ml.kg.min.      Cooper      Edad      Tipo_suelo
Min.   : 6.161      Min.   :35.29      Min.   :1232      Min.   :13.00      Asfalt:305
1st Qu.: 9.856      1st Qu.:46.53      1st Qu.:1971      1st Qu.:31.00
Median :11.789      Median :52.42      Median :2358      Median :36.00
Mean   :11.761      Mean   :52.33      Mean   :2352      Mean   :36.96
3rd Qu.:13.316      3rd Qu.:57.07      3rd Qu.:2663      3rd Qu.:43.00
Max.   :19.002      Max.   :74.38      Max.   :3800      Max.   :73.00

      Desnivel      Dificultad      Edad_r
Min.   :43      Moderat:305      m35   :125
1st Qu.:43
Median :43
Mean   :43
3rd Qu.:43
Max.   :43
                                     m40   : 66
                                     m45   : 49
                                     m50   : 30
                                     m55   : 18
                                     (Other): 9
                                     NA's   : 8

```

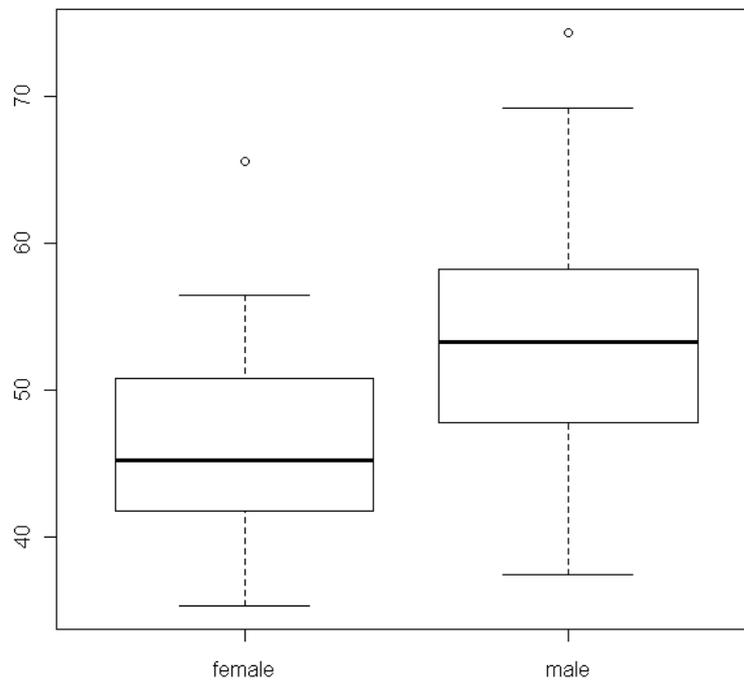


```
> oneway.test(VO2Max.ml.kg.min.~Edad_r, data=merce)
```

One-way analysis of means (not assuming equal variances)

data: VO2Max.ml.kg.min. and Edad_r

F = 1.0984, num df = 5.00, denom df = 49.45, p-value = 0.3733



```
> t.test(VO2Max.ml.kg.min.~Personal..Gènere)

Welch Two Sample t-test

data: VO2Max.ml.kg.min. by Personal..Gènere
t = -6.6678, df = 59.96, p-value = 9.221e-09
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -8.712655 -4.691463
sample estimates:
mean in group female   mean in group male
      46.55358             53.25564

>step_add<step(VO2Max_lm,~Posició+Edad+Posició.per.categoria+
Personal..Gènere)
Step: AIC=565.54
VO2Max.ml.kg.min. ~ Posició + Posició.per.categoria + Personal..Gènere

              Df Sum of Sq   RSS   AIC
<none>                1897.5 565.54
+ Edad                 1     0.51 1897.0 567.46
- Personal..Gènere     1    51.32 1948.9 571.68
- Posició.per.categoria 1   117.92 2015.5 581.93
- Posició              1   398.14 2295.7 621.63

> summary(step_add)

Call:
lm(formula = VO2Max.ml.kg.min. ~ Posició + Posició.per.categoria +
    Personal..Gènere)

Residuals:
    Min       1Q   Median       3Q      Max
-8.0512 -1.0466 -0.1537  1.1572 13.3463

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  57.6925515  1.0976448  52.560 < 2e-16 ***
```

```

Posició                -0.0014356  0.0001807  -7.947  3.86e-14  ***
Posició.per.categoria -0.0009970  0.0002305  -4.325  2.08e-05  ***
Personal..Gènerefemale  3.3449409  1.1723031   2.853  0.00463  **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.511 on 301 degrees of freedom
Multiple R-squared:  0.8761,    Adjusted R-squared:  0.8749
F-statistic: 709.4 on 3 and 301 DF,  p-value: < 2.2e-16

>step_mer<step(VO2Max_lm,~Posició*Edad*Posició.per.categoria*Personal..Gènere)
Step:  AIC=376.44
VO2Max.ml.kg.min. ~ Posició + Posició.per.categoria + Personal..Gènere +
  Posició:Posició.per.categoria + Posició:Personal..Gènere +
  Posició.per.categoria:Personal..Gènere +
  Posició:Posició.per.categoria:Personal..Gènere

              Df Sum of Sq      RSS      AIC
<none>                994.34 376.44
+ Edad                 1     5.261 989.08 376.82
- Posició:Posició.per.categoria:Personal..Gènere  1     55.745 1050.08 391.08

> summary(step_mer)

Call:
lm(formula = VO2Max.ml.kg.min. ~ Posició + Posició.per.categoria +
  Personal..Gènere + Posició:Posició.per.categoria + Posició:Personal..Gènere +
  Posició.per.categoria:Personal..Gènere +
  Posició:Posició.per.categoria:Personal..Gènere)

Residuals:
    Min       1Q   Median       3Q      Max
-6.7173 -0.8462 -0.0799  1.2497 10.0667

Coefficients:
              Estimate Std. Error
(Intercept)  6.379e+01  1.434e+00
Posició      -4.668e-03  8.476e-04
Posició.per.categoria  3.606e-02  1.150e-02
Personal..Gènerefemale  5.388e-01  1.461e+00
Posició:Posició.per.categoria -2.414e-06  7.987e-07
Posició:Personal..Gènerefemale -1.766e-02  2.316e-03
Posició.per.categoria:Personal..Gènerefemale -1.977e-02  1.167e-02
Posició:Posició.per.categoria:Personal..Gènerefemale  3.271e-06  8.017e-07
              t value Pr(>|t|)
(Intercept)  44.475 < 2e-16 ***
Posició     -5.507 7.90e-08 ***
Posició.per.categoria  3.135 0.00189 **
Personal..Gènerefemale  0.369 0.71256
Posició:Posició.per.categoria -3.023 0.00272 **
Posició:Personal..Gènerefemale -7.626 3.29e-13 ***
Posició.per.categoria:Personal..Gènerefemale -1.693 0.09142 .
Posició:Posició.per.categoria:Personal..Gènerefemale  4.080 5.78e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.83 on 297 degrees of freedom
Multiple R-squared:  0.9351,    Adjusted R-squared:  0.9335
F-statistic: 611 on 7 and 297 DF,  p-value: < 2.2e-16

```

7 BIBLIOGRAFIA

- <http://pdf.rincondelvago.com/entrenamiento-y-cualidades-fisicas-basicas.html>
- <http://burgosenrutabtt.es/2010/09/14/un-estandar-para-medir-la-dureza-de-nuestras-rutas-y-carreras/>
- http://www.ibpindex.com/esp/index_e.asp
- http://www.biolaster.com/productos/POWERbreathe/POWERbreathe_versiones/powerbreathe_salud
- <http://math.uprag.edu/clase2samplet.pdf>
- <http://www.udc.es>
- Trabajos de Estadística de Investigación Operativa Vol .32 Núm. 1, 1981, pp. 70 a 93 de Daniel Peña Sánchez de Rivera y Gonzalo Arnáiz Tovar.
- Apuntes de MLGZ de Lidia Montero (UPC)
- <http://www.uoc.edu/in3/emath/docs/Autocorrelacion.pdf>
- http://www.udc.es/dep/mate/estadistica2/sec9_3.html
- Apuntes de Mètodes Estadístics III (UPC)
- <http://www.g-se.com/a/262/la-preparacion-de-los-atletas-para-las-competencias-en-clima-caluroso/>
- http://www.aemet.es/documentos/es/divulgacion/varios/guia_montana.pdf
- http://es.wikipedia.org/wiki/Intensidad_del_campo_gravitatorio
- http://www.saludalia.com/Saludalia/web_saludalia/vivir_sano/doc/ejercicio/doc/adaptacion_mujer_ef.htm
- <http://ca.wikiloc.com/>
- <http://runedia.com>
- <http://corredors.cat>
- <http://jsoleblog.blogspot.com/2010/11/pujada-i-baixada-quanta.html>
- <http://www.itl.nist.gov/div898/handbook/prc/section2/prc213.htm>
- <http://www.atlestisme.com>
- <http://www.marato-cat.org/>