Spreadsheets in Education (eJSiE)

Volume 7 | Issue 3

Article 2

June 2014

Parametric Statistics and the General Linear Model

John A. Rochowicz Jr Alvernia University, john.rochowicz@alvernia.edu

Follow this and additional works at: http://epublications.bond.edu.au/ejsie



This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License.

Recommended Citation

Rochowicz, John A. Jr (2014) Parametric Statistics and the General Linear Model, *Spreadsheets in Education (eJSiE)*: Vol. 7: Iss. 3, Article 2. Available at: http://epublications.bond.edu.au/ejsie/vol7/iss3/2

This Regular Article is brought to you by the Bond Business School at ePublications@bond. It has been accepted for inclusion in Spreadsheets in Education (eJSiE) by an authorized administrator of ePublications@bond. For more information, please contact Bond University's Repository Coordinator.

Parametric Statistics and the General Linear Model

Abstract

Too many students acquire statistical knowledge and techniques independent of each other. The purpose of this paper is to illustrate the many connections mathematically between parametric statistics and the General Linear Model. With these various connections students will see that parametric statistical analyses are essentially one technique the General Linear Model

Parametric inferential statistics include t-tests of hypothesis, Analysis of Variance (ANOVA), correlation and regression. The ability of analyzing data in a variety of ways and using the General Liner Model (GLM) as a mathematical tool that can be applied for making these inferences are described. This paper shows connections between t-tests, ANOVA's, correlations, and regression. The capabilities of EXCEL to do mathematics, apply the Data Analysis Toolpak, find p-values for these tests and perform matrix algebra operations are demonstrated. Numerous ways of achieving the same results are displayed.

Keywords

The General Linear Model, Mathematical Statistics, Matrix Algebra, Inferential Statistics and Spreadsheets

Distribution License

This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License.

Parametric Statistics and the General Linear Model

John A Rochowicz Jr

Professor of Mathematics

Alvernia University

Abstract

Too many students acquire statistical knowledge and techniques independent of each other. The purpose of this paper is to illustrate the many connections mathematically between parametric statistics and the General Linear Model. With these various connections students will see that parametric statistical analyses are essentially one technique the General Linear Model

Parametric inferential statistics include t-tests of hypothesis, Analysis of Variance (ANOVA), correlation and regression. The ability of analyzing data in a variety of ways and using the General Liner Model (GLM) as a mathematical tool that can be applied for making these inferences are described. This paper shows connections between t-tests, ANOVA's, correlations, and regression. The capabilities of EXCEL to do mathematics, apply the Data Analysis Toolpak, find p-values for these tests and perform matrix algebra operations are demonstrated. Numerous ways of achieving the same results are displayed.

Keywords: The General Linear Model, Mathematical Statistics, Matrix Algebra, Inferential Statistics, and Spreadsheets

Introduction

Many students as well as researchers study and apply inferential statistics with no understanding that certain procedures are mathematically the same process. Connections and generalizations can be made as more statistical concepts are studied, compared and analyzed. The application of various statistical tests of hypothesis to inferences can be performed in a number of different ways. But all these tests: the t-test of hypothesis, the one way Analysis of Variance (ANOVA), correlation and regression analysis can all be analyzed applying one procedure called the General Linear Model (GLM). The General Linear Model enables the student or researcher to unify a number of inferential tests of hypothesis with one technique.

Students study many undergraduate and graduate statistics courses without the understanding that all parametric tests of hypothesis: t-tests, ANOVA's, correlations and regression are part of a more general process. When students mention they never know what test to use, they lack the ability to recognize that a specific inference can be conducted in various ways and that the General Linear Model can be applied for all of them.

The aim of this paper is to illustrate these various connections and how to apply the GLM to the analysis of a variety of statistical techniques. From many classroom experiences and textbooks in the social sciences, students learn parametric tests of hypothesis independent of each other. The unifying features and power of the GLM is invaluable to the student and the researcher.

Any statistical analysis involving linear regression, correlation, ANOVA's and ttests of hypothesis can be easily conducted in EXCEL through classical calculations, the Data Analysis Toolpak and Matrix Algebra. An example illustrating these approaches follows. Although there are limitations to using EXCEL for some statistical applications, this software can illustrate with mathematics all the necessary computations for successful analysis.

Parametric Statistics

A statistic is a numerical quantity that comes from a random sample of the population. The goal of inferential statistics is to infer, predict or estimate the population parameter from these statistics. Various statistics for inference include among others, the mean, median, mode, range, variance, standard deviation, correlation coefficient and the coefficients beta weights for linear

regression. In order to make inferences about a population parameter a researcher applies tests of hypothesis. There are many inferential tests of hypothesis. The most common inferential tests of hypothesis include: t-Tests of Hypothesis, One Way ANOVA's, Correlation Analysis, and Linear Regression.

In order to present valid results and findings, assumptions must be considered when conducting parametric tests of hypothesis. For purposes of this paper the assumptions for the example provided will be assumed verified.

Tests of Hypothesis

In conducting inferential statistics the following steps are usually followed: a) State the null and alternative hypothesis b) Calculate the test statistic. Is there enough evidence to reject null hypothesis and determine that the differences do not depend on chance. c) Set the significance level. d) Determine the observed significance level p-value. e) Decide to reject or accept the null hypothesis.

Data Analysis Toolpak is an add-in for EXCEL that can conduct many types of hypothesis tests. Illustrations of the Toolpak are provided for t-tests of hypothesis, one way ANOVA's, Correlation and Regression. In order to use Toolpak it must be installed. The techniques presented here through EXCEL are suitable for illustrating their relationships to the GLM.

Matrix Algebra for Tests of Hypothesis

If the researcher or learner uses the GLM to conduct inferential statistics knowledge about Linear Matrix Algebra is necessary. Linear Algebra [2] is a mathematical field that allows the researcher to analyze many variables using matrices. Applying the GLM process to inferential statistics is accomplished by using Linear Algebra. The main operations for matrices in the application of the GLM include the abilities to multiply, invert and transpose matrices. These abilities can be performed in EXCEL.

The steps for multiplying two matrices in EXCEL are found at <u>http://office.microsoft.com/en-us/excel-help/mmult-function-HP010342697.aspx</u> The syntax for multiplying 2 matrices is "=MMULT(array1,array2)" In order to get the completed calculated product do not use enter but "ctrl+shift+enter.

The steps for inverting a matrix are at <u>http://office.microsoft.com/en-us/excel-help/minverse-HP005209179.aspx</u>. The syntax for inverting a matrix is

"=MINVERSE(array)". In order to invert a matrix it must be square and to calculate the entire matrix array use "ctrl+shift+enter".

The steps for transposing a matrix are at <u>http://office.microsoft.com/en-us/excel-help/transpose-function-HP010069834.aspx</u> The syntax for transposing a matrix is "TRANSPOSE(array)". In order to get the correct transposed matrix again use "ctrl+shift+enter". Recall that the transpose of a matrix interchanges rows and columns in a matrix.

Various tests of hypothesis can be conducted with the General Linear Model using Matrix Algebra.

EXCEL, Statistics and Matrix Algebra

In Excel all computing formulas can be constructed and applied. EXCEL can calculate all test statistics, t, F, r, and beta coefficients for regression. With Excel p-values for all tests of hypothesis can also be found and from them decisions to reject or accept the null hypothesis are made. The availability of the EXCEL add-in Data Analysis Toolpak also allows the user to conduct: two independent samples t-tests, one-way ANOVA's, correlation, regression without any actual calculations using the classical formulas. The techniques are automated and emphasis is placed on understanding and interpretation. The capabilities of EXCEL to perform matrix operations enhance one's learning experiences. All parametric statistics can be analyzed with the general linear model by way of matrix algebra. The techniques necessary and available in EXCEL for performing matrix operations for statistical analyses are: matrix multiplication, matrix inversion, and matrix transposition.

The Example for Analysis

The following grades on a certain academic standardized calculus II test were collected for 2 groups of students:

Grou	ıp A								
56	65	69	78	72	82	81	87	89	81
Grou	ıp B								
78	88	89	91	84	87	92	94	92	90

Group A represents students taking calculus II right after high school, ages 18 to 25 years old while Group B represents students ages 26 to 30. Is there a significant difference in the mean grades for the 2 groups of students? Is there a relationship between the two types of calculus II students? Test at alpha = 0.05 This example is analyzed with the following techniques: a) t-test; b) 1way ANOVA; c) Correlation; d) Linear Regression; e) Data Analysis Toolpak; f) Matrix Algebra/Linear Models

Inferential Technique 1: t-Tests of Hypothesis

The t-test of hypothesis is an inferential test of hypothesis that determines whether there is a difference in the means of two independent or dependent sample means. The question of interest is: Is there a difference in the mean test grades for the two groups of students on a calculus II exam?

The t-test of hypothesis can be conducted by classical calculations. If the variances are known the t statistic is found by applying:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Note that \bar{x}_1 and \bar{x}_2 are the sample means, n_1 and n_2 are the sample sizes and σ_1^2 and σ_2^2 are the variances for the populations.

Figure 1 illustrates EXCEL cell formulas for this t-test of hypothesis. In figure 2 the EXCEL results are shown. The t-statistic is -3.47804 and the p-value is 0.002684. In order to find the p-value use the command TDIST with a positive t-value.

XI	E) ~ (2 ~	📓 🗠	17							ZZ Fo	r Submission !	Show Cells I	Format - Micr	osoft Excel	non-commer	cial use	-	_	_	_					×
File	Home	Ins	ert Pag	ge Layout	Formulas	Data	Review	View	Developer	Add-	Ins										-			۵ 🕜 🗆	p 🛛
Ê	🔏 Cut		Calibri	-	11 · A	- _^ =	= =	≫	Wrap Text		General	Ŧ			Normal	Bad		Good		-	* 🔳	Σ AutoSum	* 2 7	A	
Past	Format	Painter	BI	<u>u</u> • ⊞	- 👌-	<u>A</u> - ≣	= =	使使 🔤	Merge & Cer	ter *	\$~%,	00. 00. 0.4 00.	Conditional Formatting *	Format as Table *	Neutral	Calc	ulation	Check Ce	ell 🚽	Insert	Delete Format	Clear *	Sort &	Find & Select *	
	Clipboard	6		Font		G.		Alignment		G	Number	r G				Styles					Cells		Éditing		
	B2		• (=	∫x t-	Test By Fo	rmula																			~
	А		В	С	D	E	F	G	н	1	J	K	L	M	N	0	Р	Q	R	S	Т	U	v	W	-
1																									Ē
2		- 6	-Test By	ormula																					
3				Ī																					
4 A	۱		3																						
5																									
6		56	78																						
7		65	88					_																	
8		69	89			n =		10																	
9		78	91				(447.0	471//0007	10 4 81 4 04 - 0																
10		72	84			t =	=(A1/-8	\$17)/(SQR1	(0.1*(A21+B	21)))															
11		82	8/			a contrar	TOUCT	100(510)	0.01																
12		01	52			p-value	=10151(ABS(F10),1	10,2)																
14		84	94																						
15		81	90																						
16																									=
17 =	AVERAGE(A	6:A15	AVERAG	E(B6:B15)																					
18																									
19 =	VAR(A6:A15) :	VAR(B6:	B15)																					
20																									
21																									
22																									
23																									
24																									
25																									
26																									
27																									
28																									
29																									
30																									
31																									
14 4	► H t-Test	By Fo	mula 🦯	t-Test and	Pooled Var	ance / t	Test By Te	oolpak / A	NOVA By For	mula /	ANOVA By	Toolpak /	Regression E	By Formula	Rear	0									•
Read	y 🛄 🗌																					10	0% -		÷
				W			1	7	2		-				-		6.00				-	-		3:08	PM
				1.1			1997 V	-																3/12/	2014

Figure 1: t-Test of Hypothesis with EXCEL Formulas

X	5	* (°	× 🧾 🗠	-		-	-			-		-		ZZ F	or Subn	nission -	Microsoft E	cel non	-comme	ercial use			-				-	-	- 0	×
F	ile	Home	e Inse	rt Page	e Layout	Formulas	Data	Re	view	View	De	/eloper	Add-I	ns														G	ם 🕜 د	# XX
ľ	*	Cut		Calibri	Ŧ	11 × A*	=	=	= *	×*	📑 Wra	ip Text	•	Seneral		٣			2 🖪	Iormal	Bad		Good	-	1	*	Σ AutoSum	2	A	
Pa	te 🦪	Forma	t Painter	BI	u • 🖽	• 🙆 • 🛓	<u>∖</u> =	= :	= 1	e ejz	- An Mer	ge & Cent	er *	\$ - 9	, ,	00. 0.* 0.4 00.	Condition	al For ar as Ta	nat N	leutral	Calcula	ation	Check Cel		Insert D	elete Format	Clear *	Sort & Filter *	Find & Select *	
	Clipb	oard	5		Font		6		A	lignme	nt		G	N	lumber	G					Styles					Tells	Ec	liting		
	E	32	•	- (°	<i>f</i> _x t-1	Test By Forr	nula																							~
- 24	A		В	С	D	E	F		G		Н	- I		J	K		L I	М	N	0	Р	Q	R	S	т	U	v	w	х	
1																														
2		t-	-Test By	ormula		t-Test of	Hypothe	esis, \	Variano	ces Kn	own																			
3																														
4	A	в																												
6		56	78																											
7		65	88																											
8		69	89			n =		10																						
9		78	91																											
10		72	84			t =	-3.478	804																						
11		82	87					_																						
12		81	92			p-value	0.0020	684																						
13		87	94																											
15		81	90																											
16																														=
17		76	88.5																											
18																														
19	107.3	333 :	21.83333																											
20																														
21																														
22																														
25																														
25																														
26																														
27																														
28																														
29																														
30																														
31																														-
14 4	F H	t-Tes	st By For	mula / t	-test and I	ooled Variar	nce / t-	-test I	By Tool	pak 🦯	ANOV	A By Form	ula /	ANOV	A By To	olpak /	Regression	By For	mula 📿	Correl 4										
Rea	dy 🙎																										100	6 🖂 🚽		÷
6		0			V (1	1	e																-	× 🦉	3:09	PM
		-																											5/12/	2014

Figure 2: t-Test of Hypothesis with EXCEL Results

Inferential Technique 2: t-Test of Hypothesis, Variances Unknown

The t-test of hypothesis for two groups can be conducted if the variances are unknown by pooling the variances and calculating

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2},$$

where n_1 and n_2 are the sample sizes and s_1^2 and s_2^2 are the sample variances for each group. The t statistic is then calculated by applying:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Figure 3 illustrates the EXCEL cell formulas for t-Test of Hypothesis, Variances Unknown. Figure 4 displays the results of this t-test. The t-statistic is -3.47804 and the p-value is 0.002684.



Figure 3: t-Test of Hypothesis with EXCEL Formulas, Variances Unknown

XI	30		۵ 🛅	=				-			ZZ F	For Submission	- Microso	ft Excel non-	commer	cial use	_		-	_	_	_			- 0	
F	ile	Home	Inser	t Page	Layout	Formulas	Data P	eview Vi	ew De	veloper Ad	id-Ins													6	ວ 🕜 🗆	₽ X
ľ	X	ut		Calibri	* 1	1 * A* A*	- = =	= >>-	⊒ i Wra	ap Text	General		-	57	No	rmal	Bad		Good		3 - 3	×	Σ AutoSum	· 剂	an i	
Pas	te 🛷 F	opy * ormat i	Painter	BI	u • ⊞ •	<u>ð - A</u> -		三 使 9	Me	rge & Center +	\$ - 9	% •	Condi	tional Form	at Ne	utral	Calculat	tion	Check Cell		Insert De	lete Format	Clear *	Sort &	Find &	
	Clipbo	ard	G		Font		6	Align	ment	6	N	lumber	G	ung · as rab	IC .	9	Styles				c	ells	G2	diting	select -	
_	B	,		(=	fx t-Te	est of Hypoth	hesis																	-		~
	^	-	D	6	0	6	c	G	u		1	V	1	54	N	0	D	0	P	c	т		M	M	×	
1	A		0	C	U			0			,	ĸ		IVI	IN .	0	F	ų	R	5		0		vv	~	-
2		t-T	est of I	ypothesis	s	Population	n Variance	s Unknown																		
3																										
4	Α	В																								_
5																										
6		56	78		n1 =n2 =	10																				
/		65	88		-02	64 50333																				
0		70	07		52	04.30335																				
10		70	84		t=	-3 47804																				
11		82	87		-	5.47004																				
12		81	92		p-value =	0.002684																				
13		87	94																							
14		89	92																							
15		81	90																							_
16																										=
17		76	88.5																							
18																										
19	107.33	33 21	1.83333																							
20																										
22																										
23																										
24																										
25																										
26																										
27																										
28																										
29																										
30																										
31																										
5Z 14 4	F H	t-Test	t By Forn	nula t-t	est and Po	oled Varian	ce t-tes	By Toolpak	ANOV	A By Formula	ANOV	A By Toolpak	Regres	sion By Form	ula /	Correl 4										
Rea	dy 🛅																						10	% ⊖		÷
6			X		VI 🧧				6		e - 1				-	-		1	1.1				-		3:10	PM
		_						<u> </u>	\sim																3/12/	2014

Figure 4: t-Test of Hypothesis with EXCEL Results, Variances Unknown

X 🚽	1) ~ (a 📓 🗸	-	-		-		-			Z	Z For Sub	mission -	Microsoft Excel	I non-con	nmercial use	-		-	-		-			. 0	x
File	Ho	me Inse	rt Pa	ge Layout	Formulas	Data	Review	View	Dev	eloper .	Add-Ins													۵	0 -	₽ 83
Ê	🔏 Cut		Calibri	* 1	ı · A /	. = =	-	8	📑 Wra	p Text	Gene	ral	¥	50		Normal	Bad		Good		3 - 7	× 📰	Σ AutoSum	· 🖅 🕯	n.	
Paste	Cop	y *	BI	<u>u</u> • = •	<u> - A</u>	· = =	= +	z (z	Mer	ge & Center	- 5 -	%,	00. 0.* 0.4 00.	Conditional	Format	Neutral	Calcula	ition	Check Cell		Insert Del	ete Format	Fill *	Sort & Fi	nd &	
	Clipboard	l G		Font		6		Alianme	nt		6	Number	5	Formatting *	as Table *		Styles				Ce	lls 🗸	Clear +	Filter * Se diting	lect *	
	B2	-	(=	fx t-te:	st By Toolp	ak																				~
	A	В	С	D	E	F	G		н	1	J	K		L M		N O	Р	Q	R	S	Т	U	v	W	х	
1																										
2		t-test By 1	oolpak																							
3																										_
4 A		В		t-Test: Tw	o-Sample	Assuming	Equal V	ariance	s																	
5							-																			
6	56	78			Variable 1	Variable 2																				
-	65	88		Mean	/6	31 93335																				
0	70	03		Obconut	107.5555	21.00000																				
10	70	94		Realed V	64 59222																					
11	82	87		Hypothes	04.30333																					
12	81	92		df	18																					
13	87	94		t Stat	-3,47804																					
14	89	92		P(T<=t) or	0.001342																					
15	81	90		t Critical o	1.734064																					
16				P(T<=t) tv	0.002684																					=
17				t Critical t	2.100922																					
18																										
19																										
20																										
21																										_
22																										
23																										
24																										
25																										
20																										
28																										
29																										
30																										
31																										
32	N AS	Tort By Erro	nub /	t tort and De	olod Varies -	. t.t	t Du T-	alaak	ANON	Dy Correct		WA DU T	anhak	Pographian P	· Formale	Correl										
Ready	- C-I	rest by roll	muld /	cross and Po	oveu variano	e t-Ces	C DY 10	орак /	ANOVA	s by romful		JVA DY T	ларак /	regression B)	r ronndia	Constant					- eff	6	用 [1] 100	% (-)	-0	-
				wl 🙆				7	1		-								1.1		-				3:12	м
-								<u> </u>	U																3/12/2	014

Figure 5: t-Test of Hypothesis with Data Analysis Toolpak, Variances Unknown

The results are same regardless of the technique. And since the p-value is less than or equal to alpha of 0.05 there is a significant difference between the mean grades for the different age groups. Using the Data Analysis Toolpak also displays the same t-statistics and p-value.

Inferential Technique 3: One-Way ANOVA

In a one-way ANOVA random samples are selected from 2 or more than 2 groups and the goal is to determine if there is a difference in population means for these groups. Consider the same example for the two groups of students. The ANOVA can be calculated [1] or analyzed by the Data Analysis Toolpak.

The calculations for the one-way ANOVA involve:

- a) Find the sums of each variable: $\sum x_1$, $\sum x_2$.
- b) Calculate the squared sums $(\sum x_1)^2$ and $(\sum x_2)^2$.
- c) Calculate the means of each group.

The ANOVA process breaks into sections called the sum of squares between groups (SSB), the sum of squares within (error) groups (SSW) and the sums of squares for the total variance (SST). The formulas are:

SST = $\sum X^2 - \frac{(\sum X)^2}{N}$ where X represents all data elements and N represents the total number of observations. The degrees of freedom are N-1.

 $SSB = \frac{(\sum x_1)^2}{n_1} + \frac{(\sum x_2)^2}{n_2} + \dots + \frac{(\sum x_n)^2}{n_n} - \frac{(\sum X)^2}{N}$ where X represents all data elements and N is the total number of observations. The degrees of freedom are number of groups minus 1.

SSW = SST-SSB

The mean squares are measures of SS terms divided by the degrees of freedom for each section of the ANOVA. MST is the variance of all the data elements in the example and all other values are calculated with the following formulas:

MSW =
$$\frac{SSW}{dfw}$$
; MSB = $\frac{SSB}{dfb}$; and F = $\frac{MSB}{MSE}$

All these formulas are implemented in Figure 6. The degrees of freedom for numerator are the number of groups minus 1 (cell 15) and the degrees of freedom for the denominator are number of items minus 2 or 18 (cell I16). The total degrees of freedom are 20 minus 1 or 19 (cell I17).

The F statistic is 12.09677 and the p-value is 0.002684. In order to find the p-values with the one-way ANOVA the "=FDIST()" command must be used. The

command uses the degrees of freedom for the numerator and the degrees of freedom for the denominator. In this example the value in cell K15 with 1 and 18 degrees of freedom respectively are used. Figure 7 displays the EXCEL results from these formulas and Figure 8 displays the results with the Data Analysis Toolpak.

X 🛃	12) ~ (21	× 🧾 🌣	[v	-	-	-	-	_	ZZ For Sub	mission Show	Cells Format	 Microsoft Ex 	cel non-	commercial us	ie -	-	-	-		-	_			×
File	Home	Inse	rt Page	Layout	Formulas	Data	Review	View Develope	er Add-Ins				_						m		E trace	-		⊕) XX
	A Cut		Calibri	* 1	1 * A* A	. = =	: 😑 📎	Wrap Text	Gene	ral	· ·	55	No	rmal	Bad	C	Good			P 💷	Z Autosu	Ż	n i	
Paste	Copy Copy	t Dainster	BI	I • 🖽 •	🔕 - <u>A</u>	. = =	:= #	🛱 🔜 Merge &	Center * \$ *	% , 5	Cond	tional Forma	t Ne	utral	Calcula	tion	Check Cell		Insert	Delete Forma		Sort &	Find &	
Ť	Jinboard	t Painter		Font			Δlie	apment	G.	Number	Forma	tting * as Table	2 4	Sh	vies				Ť	Cells	Z Clear +	Filter * 1	elect *	
	82		(=	fr ANC	VA By Form	nula									,									~
	A	в	C	D	E	F	G	н	1	J	K	L	м	N	0	Р	Q	R	S	6 Т	U	v	w	
1																								- î
2	A	NOVA B	Formula																					
3																								
4	×	1	x2	x1^2	x2^2																			
5	A		В				SSTotal =		=SUM(D18:E1	L8) minus	=-SUM(B18	8:C18)^2/20=	I5-K5			no of iter	ms -2	df =		18				
6																								_
7		56	78	3136	6084		Sstbetwe	en =	=B18^2/10	plus	=C18^2/10	=-(B18+C18	^2/20	= 17+K7-L7		2 groups-	1	df =		1				
8		65	88	4225	7744																			
9		69	89	4761	7921		Sswithin	-	=M5-N7															_
10		78	91	6084	8281																			
11		72	84	5184	7056																			
12		82	87	6724	7569		ANOVA 1	able:	10															
13		81	92	6561	8464		Source	55	ar	MS	F	р												
10		07	01	7303	9464		Rotwoon	- 17+87 17		1_015/015	-115/116	-EDIST(V15	1 10)											
15		91	94	6561	\$100		Within	-M5-N7		19 - 416/116	-)13/)10	-FDI31(K13,	1,10]											=
17		01		0001	0100		Total	-SUM(H15-H16)		19 -H17/I17														_
18	-	SUM(87	=SUM(C7	=SUM(D7	=SUM(E7:E	16)	Total	-5011(1125.1120)		15 -1127/127														
19			sources	Sources	oonqene																			
20																								
21								r^2 =	SSB/SST	=H15/H1	7													
22																								
23																								
24																								
25																								
26																								
27																								
28																								
29																								
30																								
31																								
32	N A To	at Du Care	anda / a	Test and De	alad Masima	- /+	ab Du Taala			OVA Du Taala	ale Dama	anian Du Carro		Demil 4									_	*
Ready	- t-re	sc by POII		resultitu Pu	voleu valiano	e Z trie	sc by Toop	ANOVA BY		OVA BY TOOL	an / Regre	sson by Point		Negio 4	_	_	_	_			岡岡町 1	00% (-0-	- U
				7 6			-			1. A.				-					-	-			3:13 F	M
<u> </u>			2 💾				u 🖵 💭					-											3/12/2	014

Figure 6: ANOVA with EXCEL Formulas

X 🚽	17 - (1	× 🧾 🙃	7	-	-	-		-	-	ZZ	For Submiss	ion - M	licrosoft Excel r	non-com	mercial use	e	-	-	-	-	-	-			- 0	×
File	Hom	ie Inse	rt Page	e Layout	Formulas	Data	Review	View D	eveloper	Add-Ins														G) 🕜 🗆	p 🛛
Ê	🔏 Cut		Calibri	* 1	1 · A A	= =		₽w	ap Text	Gener	əl	٣	55		Normal		Bad		Good	-	1	7	Σ AutoSum	` 打	an a	
Paste	J Form	at Painter	BI	u • 🖽 •	🄕 - <u>A</u> -	. = =	: = 12	律 🗷 M	erge & Center	- S -	% , 🐮	00. 0 0.¢ 0	Conditional I	Format	Neutral		Calculat	tion	Check Cel		Insert D	elete Format		Sort &	Find &	
	lipboard	G		Font		6	Ali	gnment		6	Number	G	ronnatting u.			S	ityles				(Tells		diting		
	B2		. (=	∫∗ ANO	OVA By Form	nula																				¥
	A	в	с	D	E	F	G	н	1	J	K	L	M	N		0	Р	Q	R	s	т	U	v	W	х	
1																										
2	1	ANOVA B	Formula																							
3																										
4)	a	x2	x1^2	x2^2																					
5	/	Α.	В				SSTotal =		137245	minus	135301.3		1943.7	5												
6																			no. of gro	ups -1						
7		56	7	8 3136	5 6084		SSBetwe	en =	57760	plus	78322.5	13530	01.3	78	1.25		df =		1							
8		65	8	8 4225	5 7744																					
9		69	8	9 4761	L 7921		SSWithin	=	1162.5										no of iten	ns -2						
10		78	9	1 6084	1 8281												dt =	1	В							
11		/2	8	4 5184	1 /056																					
12		82	8	/ 6/24	1 7569		C		-16			-														
13		81	9.	2 0000	L 8404		source	55	ai	IVI5	r	p														
19		89	9	 7303 7921 	8464		Retween	781.2	. 1	781.25	12 09677	0.002	684													
16		81	9	6561	8100		Within	1162	18	64 58333	12.05077	0.002	.004													=
17					0100		Total	1943.7	19	102.3026																
18		760	88	5 58726	5 78519																					
19																										
20																										
21							R^2 =		SSB/SST	0.401929)															
22																										
23									r =	0.633979	9															
24																										
25																										
26																										
27																										
28																										
29																										
30																										_
31																										
32	H t-T	est Ry For	mula / t	-test and Po	oled Variance	a / t-te	st By Toolo		A By Formi	a ANO	VA By Toolo	ak / I	Regression By I	Formula	Corre											•
Ready	1							Allor	, / 0////						, conq							1	10	9% 🗩	-0-	+
-				V (1	I E		<	-						1.6		1.0	1	-	-	-		3:15	PM (2014

🗶 🖬 🤊	· (' · 🍯	∞ ⊽		-	-		-	_	-	ZZ	For Submiss	ion -	Microsoft Excel n	n-com	mercial use	-	-	-	-	_				- 0	x
File	Home I	nsert	Page La	ayout	Formulas	Data R	leview 1	view De	veloper	Add-Ins														≏ 🕜 🗆	⊕ X3
🖹 👗	Cut	Calibi	1	* 11	ı · A A	. = =	- *-	📑 Wr	ap Text	Genera		¥		4	Normal	Bad		Good		3-3	×	Σ AutoSu	" 🖅	A	
Paste	Сору т	в	τυ	• m •	3 - A	· = =	= #	E Me Me	rge & Center		% , *	00.0	Conditional F	ormat	Neutral	Calcu	lation	Check Cel		Insert De	lete Format	J Fill ▼	Sort &	Find &	
• 👽	Format Painte	er	-	Font			Alic	unment			lumber		Formatting * as	Table *		Chular					elle	⊘ Clear *	Filter *	Select *	
Chipbh	12	- (n		fx ANC	VA By Tool	lpak	005	minerite			tumber					styles					cus	1	conting		~
A	В	0		D	E	F	G	н	1	J	к		L M	N	I 0	Р	Q	R	S	T	U	v	w	Х	
1																									- î
2	ANOVA	B Tool	pak																						
3		Ī																							
4	A	В																							
5			-																						
6		56	78		Anova: Sin	ngle Factor																			
/		65	88		CURANANDY	,																			
0		70	01		SUMMARY	Count	6	4	Mariana																
9		78	91		Groups	Count 10	50m	Average	107.2222																
10		92	0%		Column 2	10	700	00 5	21 02222																
12		81	97		Column 2	10	005	00.J	21.03333																
13		87	94																						
14		89	92		ANOVA																				
15		81	90	Sou	rce of Varia	SS	df	MS	F	P-value	F crit														
16					Between	781.25	1	781.25	12.09677	0.002684	4.413873														=
17					Within Gr	1162.5	18	64.58333																	
18																									
19					Total	1943.75	19																		
20																									
21																									
22																									
23																									
24																									
25																									
26																									
27																									+
20																									
30																									
31																									
32																									-
14 F FI	t-Test By F	Formula	t-te	est and Po	oled Variance	e 🦯 t-test	t By Toolpa	k 🦯 ANOV	A By Formul	a ANOV	A By Toolp	ak /	Regression By F	ormula	Corre	ſ					-	_	-		١
Ready 🔮					_				_			_		_	_								00% 😑 —	-0-	÷
3		X	W					i e		S										-	-	-		3:14 3/12/2	PM 2014

Figure 7: ANOVA with EXCEL Results

Figure 8: ANOVA with Data Analysis Toolpak

Again the results indicate that since the p-value is less than or equal to alpha of 0.05 there is a significant difference between the mean grades for the two groups. The p-value is the same as it was for the t-test.

The one-way analysis of variance or F-test of hypothesis is a test of hypothesis for determining if there is a difference in means between two or more groups. The researcher can accomplish the same results for an independent samples t-test as a one–way ANOVA as shown above. Note that $t^2 = F$

Inferential Technique 3: Linear Correlation and Regression

Consider the same example and analyze the relationship with the Pearson correlation. There are connections between t and r.

The t-statistic and r Pearson correlation are related by:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

Solving for r gives:

$$r = \frac{t}{\sqrt{n-2+t^2}}$$

The r^2 (coefficient of determination) measures how data fit a linear relationship and is also determined by the ANOVA calculations above by:

$$r^2 = \frac{SSB}{SST}$$

The example can also be analyzed by applying linear correlation or regression. In order to do so the data has to be coded. Let x be 1 if the score is a member of the first group and a 0 if the score is in the second group. See Figure 9 for coding.

Considering the same example above, a researcher wants to determine if there is linear relationship between the two variables x and y. The goal is to determine if there is a correlation and or if there is a linear relationship. The approach is usually conducted by finding the correlation coefficient and the least squares linear equation or line of best fit [3].

The results are found by using the formulas of statistics [3]. In EXCEL these computations can be easily determined. The correlation is found by the calculating formulas: $\sum x$, $\sum y$, $\sum xy$, $\sum x^2$, $\sum y^2$,

In order to determine the strength of linearity the researcher calculates the Pearson correlation coefficient. This value is found by calculating

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

where the *terms* S_{xy} S_{xx} and S_{yy} are:

a)
$$S_{xx} = \sum x^2 - \frac{(\sum x)^2}{n}$$

b)
$$S_{yy} = \sum y^2 - \frac{(\sum y)^2}{n}$$

c)
$$S_{xy} = \sum xy - \frac{\sum x \sum y}{n}$$

d)
$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

e) The beta coefficients are $b_0 = \frac{1}{n} (\sum y - b_1 \sum x)$ and $b_1 = \frac{S_{xy}}{S_{xx}}$

Figure 9 displays the EXCEL formulas for these calculations. Figure 10 displays the EXCEL results for Linear Correlation and Regression. Figure 11 shows the correlation results with Data Analysis Toolpak and figure 12 displays the linear regression analysis with the Toolpak.

X 🖬 🤊	• (* • 🗐 🗠	 -	-	-	-	-	-	ZZ	For Sub	mission Sho	w Cells F	ormat - Mic	rosoft Exce	non-co	ommercia	luse	-	-	-	_	_			- 0	×
File	Home Inse	t Pa	ige Layout	Formulas	Data Review	v Vie	w De	veloper A	dd-Ins															≏ 🕜 🗆	9 23
<u> </u>	Cut	Calibri	Ŧ	11 · A A	= = =	\$\$/~~	🔐 Wra	ip Text	Gene	ral	Ŧ	50		Norm	nal	Bad		Good			>	Σ AutoS	um * 赶	A	
Paste 2	Copy *	BI	u • ==	• <u>ð</u> • <u>A</u>		ŧ₽ ŧ₽	- Mer	ge & Center *	s -	· % ,	00. 0.	Conditiona	Format	Neut	tral	Calcul	ation	Check Ce		Insert I	Delete Format	Fill *	Sort 8	Find &	
Clipb	ioard G		Font		3	Alignm	nent	5		Number	G	Formatting	as lable *			Styles				÷	Cells	CZ Clear	Editing	Select *	
6	82 -	(=	∫x R€	gression By F	ormula	-																			~
A	В		С	D	E		F	G	Н	1	J	К	L		М	N	0	Р	Q	R	S	т	U	v	-
1																									Ē
2	Regression	n By For	rmula																						
3																									_
4																									
5	x	1 Y	50	xy	X^2	1	2126		- 20																
7		1	50		55	1	4225	n	= 20																
8		1	69		59	1	4761	Sx	v	=D27-(1/	20)*B27	*C27													
9		1	78		78	1	6084			(,-	,														
10		1	72		72	1	5184	Sx	x	=E27-(1/2	0)*B27	^2													
11		1	82		32	1	6724				Ľ.														
12		1	81		31	1	6561	Sy	Y	=F27-(1/2	20)*(C2	7^2)													
13		1	87	, i i i i i i i i i i i i i i i i i i i	37	1	7569																		
14		1	89		59	1	/921	b3		=18/110															
15		1	81	1	31	1	6561																		
16		0	78	3	0	0	6084	bC)	=(1/20)*(C27-I14	*B27)													
17		0	88		0	0	7744	co	rrelati	on															
18		0	89		0	0	7921	r		=18/(SQR	T(110*1	(2))													
19		0	91		0	0	8281																		
20		0	84		0	0	7056			1140700	07/00 0	N (CODT)													
21		0	8/		0	0	7509	τ:		=(118-5Q	KT(20-2))/(SQRT(1	(18~2))												
22		0	94		0	0	9926																		
2.5		0	92		0	0	8464																		
25		0	90		0	0	8100																		
26					-	-																			
27	=SUM(B6:	325) =	SUM(C6:C2	5 =SUM(D6:D2	5) =SUM(E6:E2	5) =SUM	(F6:F25)																		
28																									
29																									
30																									
31																									
32																									-
14 4 F FI	/ t-Test By To	olpak	ANOVA By	/ Formula 📈 A	NOVA By Toolpa	ik Reg	gression	By Formula	Regr	ession By To	olpak	<u>/ GLM / 😪</u>													
Ready		1 11				-			_	-	_	_	_	_	-	_	_	_	_				100% -	0	+
			W	9			e																	 3:18 3/12/ 	PM 2014
																								5/12/	-

Figure 9: Correlation and Linear Regression with EXCEL Formulas

	• (= • (= ~	14								221	For Submissi	on - N	Aicrosoft Excel r	ion-com	imercial u	ise	_		-	_	_	_				_
File	Home Inse	rt Page	e Layout	Formulas	E	Data R	eview \	/iew De	eveloper A	id-Ins														6	ے 🕜 د	₽ X
🖹 🕷	Cut	Calibri	*	11 · A	Ă	= =	- *-	₩	ap Text	General	1	¥		4	Norma	al	Bad		Good		3- 3	×	Σ AutoSum	· 🗛	A	
Paste	Copy *	B /	u • m	- 3	A -	= =			rae & Center *	\$ - 0	× •	.00	Conditional I	Format	Neutra	al	Calcula	tion	Check Cell		Insert D	elete Format	🗿 Fill 👻	Sort &	Find &	
	Format Painter				-		-						Formatting * a:	Table *						×.				Filter *	Select *	
Clipbi	oard 🚱		Font		- Gi		Alig	nment		N	lumber	Gr				51	tyles				(ells		diting		_
8	12 .	. (*	J≈ Re	egression E	By Fo	rmula																				×
A 1	В	С	D	E		F	G	н	1	J.	K	L	M	1	4	0	Р	Q	R	S	Т	U	V	W	x	-
1																										
2	Regressio	h By Form	nula																							
3																										
4																										
5	x	y e	e xy	x''2	- Y'	2126		n - 20																		
7	1		о с	50 65	1	4335		11 = 20																		
·	1	6	۰ ۵	69	1	4225		Swir	62.5																	
9	1	7	8	78	1	6084		JAY	-02.5																	
10	1	7	2	70	1	5184		Svv	5																	
11	1	8	2	82	1	6724		JAN .																		
12	1	8	1	81	1	6561		Svv	1943.75																	
13	1	8	7	87	1	7569		-11																		
14	1	8	9	89	1	/921		b1	-12.5																	
15	1	8	1	81	1	6561																				
16	0	7	8	0	0	6084		b0	88.5																	=
17	0	8	8	0	0	7744		correlatio	in																	
18	0	8	9	0	0	7921		r	-0.63398																	
19	0	9	1	0	0	8281																				_
20	0	8	4	0	0	7056																				
21	0	8	7	0	0	7569		t =	-3.47804																	
22	0	9	2	0	0	8464																				
23	0	9	4	0	0	8836																				
24	0	9.	2	0	0	9100																				
25		5	0	0	0	8100																				
27	10	164	5 7	60	10	137245																				+1
28	10	104			10	137243																				
29																										
30																										
31																										
32																										-
нары	/ t-test By To	olpak /	ANOVA By	Formula	ANG	OVA By To	oolpak R	egression	By Formula	Correlat	tion By Tool	pak	Regression E	By Toolp	ak 🖉 1	In[] 4										
Ready 2				_		-	_	_	_	_	_		_	_	_		_	_	_	_			≝CI 10	»s (=)	0	+
6			<u>v</u>	3				ı e																	3:18 3/12/	PM 2014

Figure 10: Correlation and Linear Regression with EXCEL Results

X 🚽	17 • (*	۵ 🗎 ۱	-	-	-	-		-		-	-	ZZ For S	ubmission -	Microsoft Exce	l non-cor	nmercial use	-	-	-	-		_			- 0	×
File	Hom	ne Inse	rt Paç	ge Layout	Formulas	Data F	leview	Vier	w De	/eloper	Add-Ins													6	· 🕜 🗆	÷ 🕄
1	🔏 Cut		Calibri	× 1	1 × Δ* Δ	- = _		æ	⇒ wr	in Text	Ge	neral	Ŧ			Normal	Bad		Good		3- 3	× 🕋	Σ AutoSun	· A	40.	
Paste	🐚 Сору	(*	n r	n - 1 m -								- 9/	.0 .00	Conditional	Format	Noutral	Calcul	ation	Chock Coll	-	Insert De	lete Format	🛃 Fill 🔻	Zu Sort &	Find &	
*	🛷 Form	at Painter	D 1	Q . ⊞ .	· · A	. = =	-	sin sin	rea mer	ge oc cente		- 70	.000	Formatting *	as Table •	iveutiai	Calcul	ation	Check Cell			· ·	Clear *	Filter *	Select *	
	Clipboard	Gi.	6	Font	- Inter Dur	G		Alignm	ient		6	Numb	er G				Styles				0	ells		Editing		
	BZ		0	J _x Cor	relation By	гоограк											-	-	-	-	_					×
1	A	В	C	D	E	F	G	1	н	_	J		к	L M		N O	P	Q	R	5		U	v	W	X	-
2	E.	Correlatio	By Too	lnak																						
3		Conclutio	,	- part																						
4																										
5	,	x	v																							
6		1	5	56		Column 1	Colur	nn 2																		
7		1		55	Column 1	1																				
8		1		59	Column 2	-0.63398		1																		
9		1	1	78																						
10		1	3	72																						
11		1	8	82																						
12		1	8	81																						
13		1	8	87																						_
14		1	8	89																						_
15		1	8	81																						
16		0		/8																						_
10		0		58																						
10		0		59 51																						
20		0		24																						
20		0		87																						
22		0		92																						
23		0		94																						
24		0		92																						
25		0	9	90																						
26																										
27																										
28																										
29																										
30																										
31																										
32	M /+	tort By Tr	olook	ANOVA BUS	Formula A	NOVA By T	oolook	Po	grassian I	by Formula	Com	abtion	By Toolpak	Pagrarcia	Pr Tool	nak Inii 4									_	× 1
Ready		resuby 10	πηρακ ζ	AnovA by I		NOVA BY 1	σοφακ	. X Re	yression i	sy roiffuid	, con	ciacion	ву тобран	. regression	107 100	par (110 4						1	用口川 10	0% (_0_	-
				w 2				5	6		10						- I				-				3:19	PM
-						- H		-	<u> </u>																3/12/	2014

Figure 11: Correlation with Data Analysis Toolpak

New Net Republic New	- 1	(* • 📄 🗠	[v				-			ZZ F	or Submiss	ion - Mic	crosoft Excel no	on-com	mercial use	-	-	-		_	_		6	- 0	-
Corport Determine If I i i i i i i i i i i i i i i i i i i	le H	ome Inse	rt Page	e Layout	Formulas	Data P	Review \	riew De	veloper	Add-Ins														ລ 🕜 🗆	, é
Corp. Part of the second of the	🛯 🔏 Cu	t	Calibri	× 1	11 · A /	· = =	= >>-	⊒ i Wra	ap Text	General		¥	50	1	Normal	Bad		Good		3- 3	× 📰	Σ AutoSum	- 27	æ	
Uncer Print Augenet Number Other Differ Cols Eating B C D E F G H 1 X M N D P Q R S T U V W X Unrear Regression By Toolpak Integration Strutter Integrater Integration Strutter	te JFo	py * rmat Painter	BI	u • = •	• <u>&</u> • <u>A</u>	· = =	≣ (# (🖉 🔤 Me	rge & Center	· • •	6 , 5	0 .00 C	conditional Fo	ormat	Neutral	Calcula	ition	Check Cel		Insert De	elete Format	Fill * Clear *	Sort &	Find &	
12 • (m) <i>X</i> Linear Regression By Toolpak <i>x x</i>	Clipboar	rd G		Font		G.	Alig	nment		G N	umber	G				Styles				0	ells		Editing		
A B C D E F G H I J K L M N O P Q R S T U V W/ X Linear Reference B C D E F G H I J K L M N O P Q R S T U V W/ X Linear Reference SUMMARY OUTPUT S6 1 Algenesity Explored <	B2		• (=	∫x Line	ear Regress	ion By Tool	pak																		
v x y x y x Si 1 Agression By Toolpak y x Si 1 Agression Statistics Gi 1 Agression Statistics Gi 1 Agression Statistics Gi 1 Multiple* (0.338703) Size 1 Size 1 Adjusted (0.388708) Size 1 Adjusted (0.388708) Size 1 Adjusted (0.388708) Size 1 NOVA Regression 1 112.5 (9.45331 Montard Size 12.5 (9.53331 Size 13 Size 13 Alotect 11 13 Size 13 Size 13 Size 13 Size 13 Size 12.5 (9.537) Size 12.5 (9.537) </td <td>А</td> <td>В</td> <td>С</td> <td>D</td> <td>E</td> <td>F</td> <td>G</td> <td>н</td> <td>I</td> <td>J</td> <td>к</td> <td>L</td> <td>м</td> <td>N</td> <td>0</td> <td>Р</td> <td>Q</td> <td>R</td> <td>S</td> <td>т</td> <td>U</td> <td>V</td> <td>W</td> <td>X</td> <td></td>	А	В	С	D	E	F	G	н	I	J	к	L	м	N	0	Р	Q	R	S	т	U	V	W	X	
Unear Regression By Toolpak v x SUMMARY OUTPUT v x 9 X SUMMARY OUTPUT v v x v 66 1 Regression Statistics v v x v 73 1 R Square 0.41929 v v v v v 72 1 Alguited 10.36576 v <																									
y X SUMMARY OUTPUT 65 1 Regression Statistics 66 1 Regression Statistics 71 1 R Square 0.401329 72 1 Aljuited 10.358703 81 1 Observati 82 1 Standard 18.036376 83 1 Observati 81 1 Observati 83 1 Regression 1 71 Akvia 84 0 75 1 83 1 74 1 75 1 84 0 76 1 77 1 84 0 70 1 71 1 72 1 73 7 74 1 75 1 76 1 77 1 78 1 78		Linear Re	ression B	y Toolpak																					
y x SUMMARY OUTPUT Image: solution status 69 1 Regression Status Image: solution status 69 1 Multiple 6.033979 Image: solution status Image: solution status 78 1 R-Square 0.013939 Image: solution status Image: solution status 78 1 Adjusted 1.0368703 Image: solution status Image: solution status 81 1 Sundard 1.606976 Image: solution status Image: solution status 81 1 Observati 20 Image: solution status Image: solution status 81 1 Observati 20 Image: solution status Image: solution status 81 1 Observati 20 Image: solution status Image: solution status 81 1 Image: solution status 10.000564 Image: solution status Image: solution status 81 1 Image: solution status 10.000564 Image: solution status Image: solution status 91 0 Image: solution status 10.000564 <t< td=""><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></t<>																									
y x SUMMARY OUT/01 65 1 Regression Statistics 69 1 Multiple F. 6.53376 78 1 R.Square 0.401329 72 1 A.Quittel 0.368703 82 1 Standard 18.085703 82 1 Standard 18.085703 81 1 Observati 89 1 ANOVA Breestool 81 1 0 0 72 0 73 0 74 18 10 19 10 19 11 1162.5 12 12.05677 13 1162.5 14 1162.5 15 1162.5 16 1161.05 18 1162.5 19 0 10 1041.75 18 1162.5 19 0 10 10.002664																									
30 1 Regression Statistics 69 1 Multiple 7 0.53379 78 1 RSquere 0.41929 72 1 Adjustel 0.368703 82 1 Standard 18.038768 81 1 Observati 89 1 ANOVA 89 1 ANOVA 81 1 Observati 82 0 Regression 1 781.25 78 0 Regression 1 781.25 88 0 Residual 11 102.56 (-53333) 89 0 Total 19 1943.75 91 0 Coefficientandard fr. 15tot P-value lower 550/pper 95.0% 87 0 Intercept 18.10807 9.38313 92 0 Xvarable -12.5 3.53376 -3.47804 0.002684 -20.0507 -4.34934 92 0 Xvarable -12.5 3.593976 -3.47804 0.002684 -20.0507 -4.34934 92 0 Xvarable -12.5 3.593976 -3.47804 0.002684 -20.0507 -4.34934 92 0 Varable -		Y CC	x		SUMMAR	YOUIPUI																			
00 1 Multiple F, 053979 72 1 R Square 0.4(1929 72 1 Alguited 1.38793 82 1 Standard 18.06376 82 1 Standard 18.06376 81 1 Observati 20 87 1 ANOVA. 81 1 Observati 20 83 1 ALOVA. 84 0 Regression 172.52 12.06777 85 0 Regression 172.52 12.05777 84 0 Coefficientandard Er 15tot 91 0 Coefficientandard Er 15tot 92 0 Variable -12.5 3.593976 93 0 Variable -12.5 3.593976 -3.47804 0.002644 90 0 Variable -12.5 3.593976 -3.47804 0.002644 -20.0507 -4.94934 91 0 Variable -12.5 3.593976 -3.47804 0.002644 -20.0507 -4.94934 92 0 0 0		00																							
09 1 Midliple 7 063379 78 1 Alguste 0103876 82 1 Standard 18 03876 81 1 Observati 83 1 Regression 1 71 ANOVA 84 0 Coefficientindiard in 15205 75 0 Regression 1 781.25 84 0 Coefficientindiard in 1510t Puolue lower 55.00 per 95.0% 87 0 Intercept 85.10687 93.8313 90 0 Variable -12.5 3.533376 -3.47804 0.002684 -20.0507 4.94934 92 0 XVariable -12.5 3.53376 -3.47804 0.002684 -20.0507 4.94934 92 0 XVariable -12.5 3.539376 -3.47804 0.002684 -20.05		65	-		Regression	n Statistics																			
78 1 Alguited 10.38576 72 1 Alguited 10.38576 82 1 Standard 10.08376 81 1 Observati 89 1 ANOVA 81 1 Observati 80 1 73L25 78 0 Regression 1 1.75.12 1.20570 1.75.12 1.20570 0.002644 81 1 Observati 80 Residual 18 10 Total 1 11 1.15.2 5.45333 91 O Coefficientsondard Erit 91 0 Coefficientsondard Erit 1.5007 92 0 XVariable -1.25 3.539376 92 0 XVariable -1.25 3.539376 91 0 Coefficientsondard Erit 1.44934 -20.0507 92 0 XVariable -1.25 3.539376 -3.47804 0.002644 -20.0507 92 0 0 0 0 0<		09		1	Multiple I	0.033979																			
1 Substrate 10.330/03 82 1 Standard 16.303/03 81 1 Observati 20 87 1 ANCVA 89 1 ANCVA 81 1 Observati 20 87 1 Regression 1 78.10,5 12.5 12.05077 88 0 Regression 1 78.10,5 12.5 12.05077 89 0 Total 19 19.10,2 5.05333 90 0 Coefficientandard fr. 15tat P-value lower 55:0/pper 95:0/w r5.05/pper 95:0/w 84 0 Coefficientandard fr. 15tat P-value lower 95:0/pper 95:0/w r5.05/pper 95:0/w 87 0 Intercept 85.2 5.24125 34.3435 92 0 Xivatable -12.5 3.533376 -3.47804 0.002684 -20.0507 4.34934 -20.0507 4.34934 92 0 Xivatable -12.5 3.539376 -3.47804 0.002684 -20.0507 4.34934 -20.0507 4.34934 -0.0507 4.34934 -0.0507 4.34934 -0.0507 4.34934 -0.0507 4.34934 -		70		1	Adjusted	0.401929																			
az 1 Semial 1 Aussis 81 1 Observati 20 87 1 ANOVA 81 1 Off SS 781.25 83 1 Regressio 1 78 Regressio 1 781.25 781.25 84 0 Regressio 1 781.25 12.05677 84 0 Coefficientandard Er 1 194.25 4.5333 91 0 Coefficientandard Er 1 Stat Puole Lower 950/pper 95.0% 87 0 Intercept 85.2 2.941.25 M.82435 S.726-18 83.16687 93.83913 83.16687 93.83913 91 0 Coefficientandard Er 1 Stat Puole Lower 950/pper 95.0% 1		12		1	Aujusteu	0.306703																			
83 1 Observation 000 89 1 ANOVA f 55 MS F gnfpconce 81 1 Regression 1 781.25 12.05077 0.0026544 88 0 Residual 11.102.5 65.0333.3 1 1 84 0 Coefficientandard Fr 15.001 P-volue Iower 55.0% pper 95.0% 87 0 Intercept 85.25.2412.53 At3455 57.251 81.0627 93.8313 92 0 X.Variable -12.5 3.593976 -3.47804 0.002684 -20.0507 -4.94934 92 0 X.Variable -12.5 3.593976 -3.47804 0.002684 -20.0507 -4.94934 92 0 0 0 0.002684 -20.0507 -4.94934 -20.0507 -4.94934 92 0 0 0 0.002684 -20.0507 -4.94934 -0.0507 -4.94934 90 0 0 0 0.002684 -20.0507 -4.94934 -0.0507 -4.94934		02		1	Obconut	1 8.030370																			
89 1 ANOVA Image: Single		01		1	Observati	20																			
81 1 Regressio 1 781.25 78		89		1	ΔΝΟΥΔ																				
78 0 Regression 1 781.25 7		81		1	A1101A	df	22	MS	E	anificance															
88 0 Residual 18 1122.5 64.58333 89 1 Total 19 1243.75 91 0 Coefficientsindrard Err. t Stot P-volue lower 95:0%/per 95:0%/per 95:0%/per 95:0% 81 0 Coefficientsindrard Err. t Stot P-volue lower 95:0%/per 95:0%/per 95:0%/per 95:0% 87 0 Intercept 88:5 S-22:18 S3:16067 93:8313 92 0 X Xarable -12:5 3:593976 -3:47804 0.002644 -20:0507 -4:54934 94 0 0 0 0 0 0 0 0 92 0 90 0 0 0 0 0 0 0 90 0		78		n	Regressio	1	781.25	781.25	12 09677	0.002684															
89 0 Total 19 1943.75 91 0 Coefficientzndard Err. (t Stat P-value lower 95%/pper 95.0% 84 0 Coefficientzndard Err. (t Stat P-value lower 95%/pper 95.0% 87 0 Intercept 85.5 2.541.325 34.8745 5.72E-18 83.1667 93.83913 83.1667 93.83913 92 0 X/variable -12.5 3.539376 -3.47804 0.00264 -20.0507 4.34934 -20.0507 4.34934 94 0 92 0 90 0 90 90 90 0 90 90 90 0 91 91 91 91 91 91 91 91 91 91 91 91 91 92 92 92 92 92 92 92 91 91 91 91 91 91 91 91 91 91 91 91 91 91 92 92 92 92 92 91 91 91 91 91 91 91 91 <td></td> <td>88</td> <td></td> <td>n</td> <td>Residual</td> <td>18</td> <td>1162.5</td> <td>64.58333</td> <td>121050777</td> <td>01002001</td> <td></td>		88		n	Residual	18	1162.5	64.58333	121050777	01002001															
91 0 Coefficientsindard Erit 1 Stat P-value Lower 35%Upper 95.0% 84 0 Coefficientsindard Erit 1 Stat P-value Lower 35%Upper 95.0% 87 0 Intercept 88.25 244125 M.82435 S.72E-18 83.10667 93.8913 92 0 XVariable -12.5 3.593976 -3.47804 0.002664 -20.0507 4.54934 94 0 90 0 0 0.002664 -20.0507 4.54934 -20.0507 4.54934 92 0 0 0.002664 -20.0507 4.54934 -20.0507 4.54934 92 0 0 0 0.002664 -20.0507 4.54934 -20.0507 4.54934 90 0 0 0 0.002664 -20.0507 4.54934 -20.0507 4.54934 91 0 0 0 0.002664 -20.0507 4.54934 -20.0507 4.54934 92 0 0 0 0.002664 -20.0507 4.54934 -20.0507 4.54934 93 0 0 0 0		89		- D	Total	19	1943.75																		
84 0 Coefficientsindard Erit 15tat P-value Lower 35%Upper 95%ower 95.0% 87 0 Interrept 88.5 2.54122 3.42845 5.72E-18 83.10667 93.83913 83.10667 93.83913 92 0 X Variable -12.5 3.593976 -3.47804 0.002684 -20.0507 4.54934 20.0507 4.54934 94.0 94 0 92 0 92 0 99.0 0 99.0 0 99.0 0		91	(- D																					
87 0 Intercept 86.5 2.54123 34.84245 5.72E-18 83.16087 93.8913 83.16087 93.8913 92 0 X.Variable -12.5 3.533376 -3.47804 0.002564 -20.0507 4.54934 -20.0507 4.54934 94 0 92 0 </td <td></td> <td>84</td> <td></td> <td>- 0</td> <td></td> <td>Coefficient</td> <td>andard Fr</td> <td>t Stat</td> <td>P-value</td> <td>ower 95%</td> <td>Inner 95%</td> <td>ower 95</td> <td>.09 pper 95.0</td> <td>36</td> <td></td>		84		- 0		Coefficient	andard Fr	t Stat	P-value	ower 95%	Inner 95%	ower 95	.09 pper 95.0	36											
92 0 XVariable -12.5 3.593976 -3.47804 0.002684 -20.0507 -4.54934 -20.0507 -4.54934 92 0 90 0 9 0 9 0 9 0 9 0 9 0 9 0 9 0 9 0		87		0	Intercept	88.5	2.541325	34.82435	5.72E-18	83.16087	93.83913	83.160	87 93.83913												
94 0 92 0 90 0 90 0 8 / ttest By Toolpak / AlkOVA By Formula / AlkOVA By Toolpak / Regression By Toolpak / h) 4		92	(D	X Variable	-12.5	3.593976	-3.47804	0.002684	-20.0507	-4.94934	-20.05	07 -4.94934												
92 0 90 0 N Ctest By Toolpak / AHOVA By Formula / AHOVA By Toolpak / Regression By Toolpak / In (III 100 0 -		94		D																					
90 0 91 Ctest By Toobak / ANOVA By Formula / ANOVA By Toobak / Regression By Toobak / bill III 1005 0 1005 0		92	(D																					
N / Ltest By Toolpak / AHOVA By Formula / AHOVA By Toolpak / Regression By Toolpak / Ini (III III) 100%		90	(D																					
Ctest By Toolpak / ANOVA By Formula / ANOVA By Toolpak / Regression By Toolpak / Iri)																									
N / L'Etett By Toolpak / AHOVA By Formula / AHOVA By Toolpak / Regression By Toolpak / In 4 III III 100% O																									
M / ttest By Toolpak / ANOVA By Formula / ANOVA By Toolpak / Regression By Toolpak / Pri 4 III III III III III III III III III																									
Ctest By Toolpak / AHOVA By Formula / AHOVA By Toolpak / Regression By Toolpak / bil 4																									
N / L'test By Toolpak / ANOVA By Formula / ANOVA By Toolpak / Regression By Toolpak / In) 4 III III 100% O																									
M / ttest By Toolpak / ANOVA By Formula / ANOVA By Toolpak / Regression By Formula / Correlation By Toolpak / Pri 4 III III III III III III III III III																									
	H Z	t-test By To	nolnak /	ANOVA By I	Formula /	ANOVA By T	oolpak	Regression	Ry Formula	Correlat	ion By Too	lnak I	Regression R	v Tool	nak In 4										
														,								田田 10	0% (
													-		-		-	-						2.10	ē

Figure 12: Linear Regression with Data Analysis Toolpak

The Pearson correlation is -0.63398. The values for the coefficients b₀ and b₁ are respectively 88.5 and -12.5. The t-statistic is calculated from the r shown above and is -3.47804. The p-value is the same as before 0.002684.

Inferential Technique 4: The GLM

The General Linear Model is the method that enables any researcher or learner to conduct parametric inferences by applying matrix algebra. In constructing these matrices the researcher sees the unifying of mathematics behind all parametric tests of hypothesis: t-tests, 1way ANOVAS, correlations and linear regression analysis. As shown above all the different tests of hypothesis presented yield the same results. Through the use of matrix algebra all these calculations are presented again. The value of GLM is illustrated on the same scores for the different age groups.

Matrix Algebra and the General Linear Model (GLM)

In studying statistics, researchers are interested in determining relationships between usually many variables, the design variables (independent) x's and the response variables (dependent) y's. Since this paper addresses linear relationships among numerous variables the use of matrix notation is appropriate. Matrices are tables of numbers or variables that allow the user to indicate how a statistical design can be analyzed

This equation Y = Xb + e represents the basic experimental designs described above, where Y is the matrix of response or dependent variables and X is the matrix of design or independent variables. The matrix notation is:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} \cdots & X_{1s} \\ \vdots & \vdots & \vdots \\ 1 & X_{n1} \cdots & X_{ns} \end{bmatrix} \begin{bmatrix} b_0 \\ \vdots \\ b_s \end{bmatrix} + \begin{bmatrix} e_0 \\ \vdots \\ e_n \end{bmatrix}$$

Consider the expression Y = Xb where **e** is negligible. The b's are the beta weights or coefficients of the independent variables x.

Solving for **b** in the expression using matrix algebra involves the following: First make **X** a square, symmetric matrix by multiplying both sides of the equation by the transpose of **X** or **X**', that is X'Y = X'Xb

As a result a square, symmetric matrix is found. If **X'X** has an inverse, **(X'X)**⁻¹, then multiply both sides by this inverse matrix to get

$(X'X)^{-1} X'Y = (X'X)^{-1} (X'X)b$

Since a matrix multiplied by its inverse is the identity, this product is

(X'X)-1 X'Y=Ib=b

The calculations [4] in parts for each matrix product are shown here:

$$\mathbf{X'X} = \begin{bmatrix} 1 & 1 \cdots & 1 \\ x1 & x2 \cdots & xn \end{bmatrix} \begin{bmatrix} 1 & x1 \\ 1 & x2 \\ \vdots & \vdots \end{bmatrix} = \begin{bmatrix} n & \sum x \\ \sum x & \sum x^2 \end{bmatrix}$$
$$\mathbf{X'Y} = \begin{bmatrix} \sum y \\ \sum xy \end{bmatrix}$$
$$\mathbf{b} = (\mathbf{X'X})^{-1}\mathbf{X'Y}$$
$$\mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}$$

 $SSE = S_{yy} - b_1 S_{xy}$ for Linear Regression

SSE = Y'Y - b'X'Y for GLM

EXCEL Capabilities for the General Linear Model

Since the independent variable X is linear, a column of ones must be entered first. In the second column for X enter 1's if the score is in group 1 and a 0 otherwise. The dependent variable Y scores are entered as a 20 by 1 matrix and the independent variable X is entered as a 20 by 2 matrix. The transpose matrices X' and Y' are found by using "=TRANSPOSE(array)". The transposed matrices are 1 by 20 matrix and 2 by 20 matrix respectively. IN order to multiply matrices they have to be compatible. Multiplying a 20 by 1 matrix by a 1 by 20 produces a 1 by 1 matrix. The details are displayed in the figure. Similarly the products of the required matrices for the GLM are conducted. In order to multiply 2 matrices the command "=MMULT(array1, array2)" is applied. As shown in the figure the inverse of a matrix is also necessary. This is accomplished by using the command "=MINVERSE(array)." The results are displayed in the figure.

EXCEL Cell Formulas for the General Linear Model

The GLM formulas are implemented in EXCEL with the following cell formulas and details:

The transpose of the X matrix, X' is found by using the instructions provided above. The resulting cells display an array with the syntax "{=TRANSPOSE(D2:E21)}".

The transpose of the Y matrix, Y' is found similarly and the resulting array is shown with syntax "{=TRANSPOSE(C2:C21)}".

To find X'X use the multiplication of 2 matrices and get an array syntax "{=MMULT(H2:AA3,D2:E21)}" shown in cells J8 J9 K8 K9.

To find X'Y use the multiplication of 2 matrices and get an array syntax "{=MMULT(H2:AA3,C2:C21)}" in cells P8 and P9.

Finding the X'X inverse matrix applies the technique of finding inverses of matrices defined above. The array syntax "{=MINVERSE(J8:K9)}" is shown in cells J10, J12, K11 and K12.

Finding the Y'Y matrix again applies the multiplication of 2 matrices. The array syntax is "{=MMULT(H6:AA6,C2:C21)}" and shown in cell P11.

The coefficients b_i 's are shown in cells K16 and K17.

The SSE term is displayed in cell O18 and the variance s² term is calculated using SSE/df or SSE/18 and shown in cell I20. The square root of the variance is s and shown in cell I22. With these calculations the t-statistic is found using the formula:

$$t = \frac{b_1}{s\sqrt{\frac{1}{Sxx}}}$$

The excel formula is "=K16/(I22*SQRT(K12))" and is displayed in cell M20. The calculated results are shown in Figure 13. Again, the t statistic is -3.47804 and the p-value is 0.002684.

X 🖌	🚽 🎔 🕫 🖉 🛆 😇 ZZ For Submission With Results - Microsoft Excel non-commercial use											_	_			-	0	x													
File Home Insert Page Layout Formulas Data Review View Developer Add-Ins															۵ (0 - p	P XX														
🗎 🖁 Cu			Calibri	× 1	x 11 x A* *			æ.,	Taking Test			General		Ŧ		- E		Normal			Bad		Good		-	× 🕋	Σ AutoS	Sum + A	a	a.	
-	Copy	*	comort							- wiap lext				0 00	<u></u>	<u>125</u>	ø	Teornal			Dau						🛃 Fill 🔻	Z	U Ur		
* V Forr		at Painter	BI	1 × ⊞ *	<u>∞</u> • <u>A</u>	* = =		te te	He Me	Aerge & Center ×		\$ * 9	6 · 1	000	Form	atting * as	Format is Table *	Neu	trai		Calculati	on	Check Cell	-	insert Di	iete Format	Clear	Fill	er* Sele	u oc ≥ct *	
	Clipboard	G		Font		Gi i		Alignm	nent		G	N	umber	5						Style	es				0	ells		Editing			_
	B2	•	· (=	∫x The	GLM																										*
1	А	В	С	D	E	F	0	;	н	1.1		J	K		L	м	1	N	0		Р	Q	R	S	Т	U	V	W		Х	
1			Y	х																											
2		The GLM	56	1	. 1		X,		1	1		1		1	1	1	1	1		1	1	3	1		L	1	1	1	1	1	
3			65	1	. 1				1	1		1		1	1	1	1	1		1	1	1	. 0	(0	0	0	0	0	0	
4			69	1	. 1		2 by :	20																							
5			/8	1	. 1		RbyC																								
6			/2	1	. 1		Y.		56	65		69	~	5	12	8,	2	81	\$	57	89	81	. 78	81	5 8	9 9	1 8	54	87	92	
-			82	1	1			~	W -			20							why -		1645										
0			01	1	1			^	X =			20	1						A T =		1045										
10			0/	1	1							10	1	,							700										
11			0.0	1	. 1			v	Vinuoro	•		0.1	0.1						viv		127245										
12			70	1				^	Anivers			0.1	-0.					_			137243										
12			99	1	. 0							-0.1	0	2																	
14			89	1	0																										
15			91	1	0			B	eta coeff	icients			88.	5		b'		88.5	-12	.5											
16			84	1	0			X	X invers	e times X*	Ý		-12.	5																	=
17			87	1	0																										
18			92	1	0			S	SE =	Y'Y-b'X'Y	=		13724	5 minu	JS	136082.5	5 =		1162	.5											
19			94	1	. 0																										
20			92	1	0			s'	<u>2 =</u>	64.58333				t =		-3.47804	4														
21			90	1	. 0																										
22								s	=	8.036376																					
23			RbyC	20 by 2																											
24																															
25																															
26																															
27																															
28																															
29																															
30																															
31																															
32		NOVA By 1	Foolpak	Pogrossion	By Formula	/ Corre	btion 0	v Toolo	nk D	agracian P	v Too	look	CLM /	n / -						_									_	_	
Read		NOVA By I		Regression	by rollidia	C Corre	acidii b	y 100p	an / N	egression b	y 100	лрак Д	unt /								_							100% (-		0	
				7 6				-	6				-						-				-		-					3:22 PI	м
-			2 L				H.	<u> </u>																						3/12/20	014

Figure 13: The General Linear Model with EXCEL Results

The goal in using the GLM is to find the beta coefficients that allow the data to fit the best linear model. All the figures in this paper illustrate the relationships between the t statistic, F statistic, correlation coefficient and the coefficients for the regression model.

Conclusions

Linear Models are not usually discussed in elementary statistics textbooks and courses. But there are many parametric statistics techniques that can be performed through the use of matrix algebra. The goal of the General Linear Model is to generate a linear combination of the x (independent) variables for one or more dependent variables y. The concept of linear models can be extended to multivariate where there are many independent and many dependent variables. There are some limitations in using EXCEL for statistical analysis, but the goal of this paper is to illustrate t-tests of hypothesis, one-way ANOVA's, correlation, regression and the relationships that are found using these various parametric statistical methods through matrix algebra. Many students and professors are not aware of the many possible connections.

Students and researchers implement various parametric tests of hypothesis without the knowledge that techniques such as t-tests, ANOVA's, correlation, and regression can all be analyzed through the application of the general linear model. Becoming aware of such possibilities provides the learner with the realization that there is really not one specific test of hypothesis that can answer any specific research question. Any parametric technique can be applied as long as the user can prepare the software for data analysis. When applying the GLM, interpretation and matrix understanding are necessary. Also the student or researcher should recognize the importance of inferential analysis in the determination of relationships that exist. The capabilities of EXCEL to apply computing formulas, use Data Analysis Toolpak, and calculate matrices are invaluable to statistical inference. Also through the application of software and especially spreadsheets concepts can be illustrated in many ways without reliance on calculations.

Learning the importance of the GLM to Parametric statistics becomes evident when students pursue specific situations and approach the situation with different techniques. Emphasis on "statistical literacy" is increased with the application of the GLM. Inferential statistical analysis can be very complex. Technologies relieve the computational burden from the student. Focus is placed on interpretation and design.

Does it really matter which test of hypothesis is used? All the tests of hypothesis discussed can be analyzed in many ways. The general linear model generalizes many concepts into one main technique.

References

[1] Cohen, J. (1968). *Multiple Regression as a General Data-analytic System*. <u>Psychological Bulletin</u>, *70*, 426-443.

[2] Leon, Steven J. () *Linear Algebra, With Applications (6th Edition)*. Prentice Hall Publishing Co. Upper Saddle River, NJ.

[3] Mendenhall III, William, Beaver, Robert J and Beaver, Barbara M. (2013) *Introduction to Probability and Statistics* (14th Edition). Brooks/Cole Cengage Learning, Boston, MA

[4] Wackerly, Dennis D. Mendenhall III, William and Schaeffer, Richard L. (1996) *Mathematical Statistics with Applications* 5th edition. Wadsworth Publishing Co. Belmont, CA.