June 2013

Methodology for Evaluating Statistical Equivalence in Face Recognition Using Live Subjects with Dissimilar Skin Tones

Rigoberto Chinchilla
Eastern Illinois University, rchinchilla@eiu.edu

120th ASEE Annual Conference & Exposition
FRANKLY, WE DO GIVE A D*MN
June 23-26, 2013

Paper ID #6155

Methodology for Evaluating Statistical Equivalence in Face Recognition Using Live Subjects with Dissimilar Skin Tones

Dr. Rigoberto Chinchilla, Eastern Illinois University

Rigoberto Chinchilla, PhD in Integrated Engineering, Ohio University, is an Associate Professor of Applied Engineering and Technology at Eastern Illinois University (EIU) since 2004. His teaching and research interests include Quality Design, Biometric and Computer Security, Clean Technologies, Automation and Technology-Ethics. Dr. Chinchilla has been a Fulbright and a United Nations scholar, serves in numerous departmental and university committees at EIU and has been awarded several research grants in his career. Dr. Chinchilla can be reached at rchinchilla@eiu.edu.

Mr. Bryan G. Baker, Eastern Illinois University

Methodology for Evaluating Statistical Equivalence in Face Recognition Using Live Subjects with Dissimilar Skin Tones

Abstract

The general purpose of this study is to propose a methodology that can be employed in the application of facial recognition systems (FRS) to determine if a statistically significant difference exists in a facial recognition system's ability to match two dissimilar skin tone populations to their enrolled images. A particular objective is to test the face recognition system's ability to recognize dark or light skin tone subjects. In addition to the direct comparison of results from two different populations, this study uses a Box Behnken Design to examine four factors commonly effecting facial recognition systems. Four factors were tested, the horizontal angle of the camera viewing the subject, both horizontally to the left and right; the vertical angle, both above and below the subject's line of sight, ;the distance the subjects are from the camera, and the intensity of the illumination on the subject. Experimentation was approached from the assumption that subjects are cooperative, following guidelines for proper enrollment and submission for matching. The experimentation of the four factors was conducted using two sets of three subjects. One set was dark skin tone males, and the second set was light skin tone males. The results of the study showed a significance statistical difference at $p = 0.05$ level between the two skin tones, with greater difficulty identifying the light skin tone test subjects than those with dark skin tone.

Introduction

Facial recognition, in particular among biometrics, has held the hope and promise of the ability to accurately and efficiently screen the masses and alert authorities when a person on a watch list appears. The reality however, is that the human ability to recognize people we have seen or match them to their photographic image is an extremely difficult ability to artificially recreate. The purpose of this study is to determine if the proposed methodology can be employed in the application of facial recognition systems to determine if a statistical significant difference exists in a facial recognition system's ability to match two dissimilar skin tone populations to their enrolled images. If the matching scores of different races can be shown to have a statistical equivalence, then a system will much more likely be viewed as having achieved an effective "fairness."

However, if a system cannot be counted on to accurately match persons of any race or skin tone to their enrolled image with a statistical equivalence between the varying races or skin tones, then the usefulness of a system will be called into doubt and rightfully criticized as biased. An inaccurate system will result in less cooperation, and frustration by the users due to false negatives. It could also be anticipated that government and industry alike would fail to adopt the technology if there is an unacceptable rate of false positives.

Review of Literature

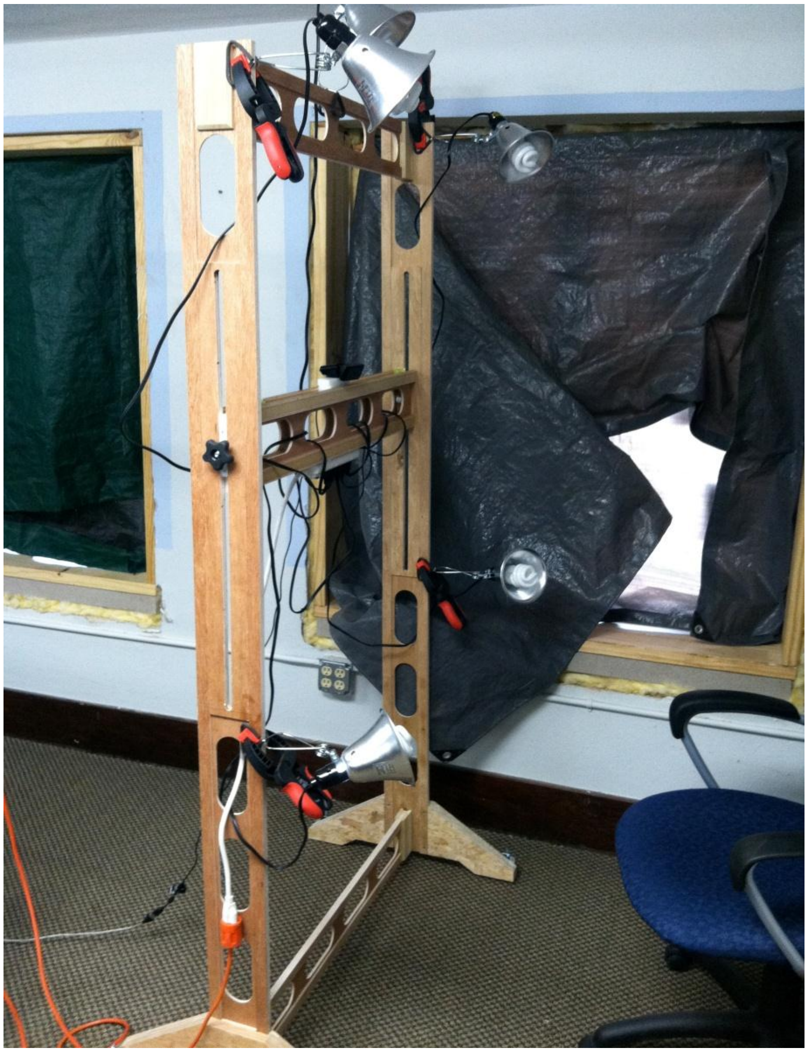
The change in appearance of someone's face often becomes much larger than the difference between two different faces under the same illumination[1]. The accuracy of face recognition degrades quickly when the illumination is dim or when the face is not uniformly illuminated[2]. A person's appearance will typically change dramatically if the intensity of light reflected from the face is changed[1]. Add to this a change in the direction the illumination is originating from, resulting in shading and shadows being created, the angle of view a camera has on the subjects face, and a person can become unrecognizable to a facial recognition system.[1]

Hiremath and Prabhakar [2] noted that there are also variances in how light reflects from human faces depending on the color, or skin tone of people from different races and ethnic groups. According to Gao, Shan, Chai & Fu[3], evaluation of both algorithms and commercially available facial recognition systems has shown that the performance of most systems degrades significantly when there are variations in both the illumination and the pose of the subject. Beveridge, et al[4], conducted a study using images of 1,072 subjects (two of each) from the FERET data set. The population consisted of white, Asian, African-American, or other. Beveridge, et al. reported that older subjects were more easily recognized than younger subjects, facial recognition works best if a person wears glasses and, "white subjects are harder to recognize than Asian, African-American or other subjects, even when the system is trained with racially balanced data sets."[4]

Participants Materials and Procedures

The research was conducted using six male subjects. Three males with dark skin tone and three males with light skin tone were chosen for their contrasting skin tones. These skin tones were selected for the purpose of determining if skin tone plays a significant, measurable factor in the ability of facial recognition systems to correctly identify a subject at varying combinations of illumination levels, angles and distances. The camera angle in relation to the test subject's face was altered in both the horizontal and vertical planes (15 degrees). The distance the camera was placed from the person's face was also varied at three different distances. Additionally, the illumination intensity was varied between three settings. A framework was constructed as a platform on which both the lights and the camera where mounted. The mounting location of the camera on the framework was constructed in such a manner that it was possible to adjust the location of the camera in both the horizontal and vertical plane. Additionally, the camera was mounted on a base that pivoted so the camera could be directed at a test subject when moved horizontally to the left and right of a test subject, or vertically above or below a test subject's line of sight. The framework with the camera mount made possible a consistently repeatable setup for each test subject.

The facial recognition software used was VeriLook 5.0[5], a commercially available program. The recommendations from VeriLook[5], in the use of the software were followed. The recommended minimal camera resolution was 640 x 480 pixels for face enrollment and recognition. The camera resolution for this study was set at 1280 X 720 pixels. VeriLook's documentation called for the use of "several images during enrollment" to increase the “recognition quality and reliability”. For this study, the default setting in the software of five enrollment images was increased to ten.



The VeriLook[5], recommendation was followed by limiting the angle of the camera’s view to the test subject at ± 15 degrees for both the head pitch (vertical angle) and head yaw (horizontal angle). The 15 degrees angle for capturing an image of the test subjects at the varying distances were calculated and marked on the adjustable framework that supported the camera (Table 1). This made it possible during the experiment for the camera to be moved to the various horizontal and vertical positions with accuracy, thus ensuring the repeatability of the experiment for each test subject. In order to build the Box-Behnken matrix; fifteen degrees to the left or below the subject’s line of sight, from the viewpoint of the operator facing the test subject, was represented as -1. Fifteen degrees to the right or above the subject’s line of sight was represented as 1. When the camera was directly in the subject’s line of sight, the setting was represented as 0 in the matrix, (Appendix A).

The minimum and maximum effective distances the subjects were seated from the camera were determined by first enrolling the subject at the distance of 24 inches. The enrollment was done with the camera level with the person's eyes. Following the enrollment, a series of identifying photos were taken with the VeriLook software and scores recorded. Numerous recognition photographs were taken at varying distances ranging from 16 inches to 42 inches. Multiple photos were taken at different distances and the average score at various distances were calculated. It was determined that reliable matching to the enrolled image occurred at distances between 45.72 cm (18 in) and 91.44 cm (36 in). These were chosen as the minimum and maximum effective distances, with 68.58 cm (27 in) being an equal distance between the two. These distances provided the values for the Box-Behnken matrix: 45.72 cm = -1; 68.58 cm = 0; and 91.44 cm = +1.

Table 1: Angle Offset Calculations

Offset Calculated for 15-Degree Angle		
Distance	Offset	Angle
45.72 cm	12.065 cm	15
68.58 cm	18.41 cm	15
91.44 cm	24.44 cm	15

The illumination recommendations from VeriLook[5], were for, "equal distribution on each side of the face and from top to bottom with no significant shadows with the face region." Additionally, VeriLook recommended, "Avoid glares on face skin or glasses that are produced by some types of illumination." Illumination was from five (5) dimmable compact fluorescent bulbs, connected through a compatible dimmer switch. The lights were mounted with one directly in front of the seating location of the test subjects, and approximately two feet higher than the subject being photographed. Two more were mounted at the same height, but 45.72 cm (18 in) to the left and right of center line. The remaining two lights were mounted approximately 45.72 cm (18 in) below the test subjects' line of sight and 45.72 cm (18 in) to the left and right of center line. All of the lights were directed at the seating position for the test subjects. This arrangement of the light sources fully illuminated each test subject's face without creating any shadows across the face. The lighting intensities were set using a Sekonic L-308s light meter.

The highest illumination level (640 Lux) was selected from the research done by Harris[6], in which various facilities such as airports and offices were visited and illumination levels measured. 640 Lux was the average level of illumination in the various facilities examined. The next illumination level was set at (320 Lux), 50% of the highest. The third and lowest illumination level was set at (160 Lux), 50% of the mid setting. These illumination levels provided the values for the Box-Behnken matrix, (Table 2). The backdrop was 18% gray. An 18% gray backdrop is an established best practice set by the National Institute of Science and Technology for use by law enforcement in the capture of mug shots; "The subject whose image is being captured shall be positioned in front of a background which is 18% gray with a plain smooth flat surface" (NIST, 1997). A "Kodak" neutral gray card was used to verify the 18% gray reflectance. The test subjects were all males, three with a dark skin tone and three with a light skin tone. Strikingly opposite skin tones were chosen for the purpose of highlighting any differences that might exist in a systems ability to match an enrolled image to a matching image.

Table 2 Illumination Levels

Illumination Levels	
Lux	Value
160	- 1
320	0
640	1

Following the recommendations of VeriLook[5], a neutral face expression was held by each test subject during the enrollment and identification; eyes open looking straight ahead, no smile, mouth closed. None of the test subjects wore glasses. None had facial hair, and all had short hair affording a clear view of their facial features. The photographs were taken of the test subjects using variable combinations of the twenty-seven runs of the Box-Behnken Design[7]. The experiment was conducted with the three dark toned test subjects and repeated with the three light skin toned test subjects. The total number of photographs taken was eighty-one of the dark skin tone subjects and eighty-one of the light skin tone subjects. The statistical analysis of the scores collected was accomplished using a combination of software and statistical equations. The software used, DOE PRO XL an add-in for Microsoft Excel, is a statistical software program that is customizable for the number of variables, replications, and design types.

Following the initial enrollment of each subject the random sequence of photographs were taken in both the dark skin tone and light skin tone populations. Twenty-seven photographs were taken of each test subject in numerical order from the list of random numbers listed in the Box-Behnken matrix shown in Appendix A. Following each photograph, the matching score generated by the VeriLook software was entered into the corresponding cell in the matrix.

Results

Scores from the Box-Behnken matrix were entered into the DOE PRO statistical analysis software. The results from the dark skin tone subjects (Mean, $\bar{x}$ = 385.65, and Standard deviation, s = 143.18) showed an overall greater ability of the software to identify the subjects over its ability to identify the light skin tone subjects ($\bar{x}$ = 319.259, s = 84.81). Even without the mean score calculated, the eighty-one dark skin tone scores were noticeably higher from a simple visual inspection. Additionally, while the software never failed to identify the dark skin tone subjects throughout twenty-seven runs and three replications of each, it did fail to identify the light skin tone subjects during at least one of the replications across seven of the runs.

There were a total of ten scores of zero from the light skin tone subjects, where the software failed to detect the presence of a face. Interestingly, none of those instances involved the lowest light setting as one of the variables when it would be reasonable to expect that the diminished illumination would adversely affect the software's ability. Nine of the ten replications with scores of zero were with the illumination at the medium level and one was at the highest level. Seven of the ten replications were at the shortest distance from the camera (45.72 cm / 18 in), two at the medium distance (68.58 cm / 27 in), and one at the greatest distance (91.44 cm / 36 in). Six of the ten replications with scores of zero were with the camera positioned to view the test subject fifteen degrees horizontally off center to the left or right. Five of the ten replications with scores of zero, where with the camera positioned to view the test subject fifteen degrees vertically off center above or below the test subject's line of sight. Run fourteen, with the variable combination of L = 0, D = -1, A1 = -1, and A2 = 0, did not result in a score for any of the three replications for the light skin tone subjects. The same run produced the lowest mean score among the twenty-seven runs of the dark skin tone.

For comparison purposes, degrees of freedom were approximated using the following equation[8] where s_1^2 and s_2^2 represent the square roots of the mean of the standard deviations from the dark and light skin tone populations (Appendix D) [8]

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2}$$

n_1 and n_2 represent the number of runs of the dark and the light skin tone test populations.

s_1^2 = 20,500.88 s_2^2 = 7,192.20 df = 42.75 n_1 = 27 n_2 = 27

The mean score of the eighty-one total sample scores from the Dark Skin Tone subjects is represented as $\bar{x}_1 = 385.65$, and the eighty-one total sample scores from the Light Skin Tone subjects is represented as $\bar{x}_2 = 319.25.$ The two-Sample *t** for the confidence interval was determined using the following equation[8]

$$(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

The *t** was calculated at the confidence of 99%, 95% and 90%. The results are as follows: 99% CI [-20.38, 153.16], 95% [2.35, 130.43], and 90% [12.6, 120.18] (Appendix D). At a confidence level of 99% the confidence interval shows the lack of statistical significant difference, supporting the Null Hypothesis. At the 95% Confidence Level, the Confidence Interval shows a statistical significant difference, supporting the Research Hypothesis. At the 90% Confidence Level, the statistical significant difference increased, further supporting the Research Hypothesis.

The two-sample *t* significance test was calculated using the following equation [8]

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

The results from the two-sample “t” significance showed that the test was not statistically significant at the specified level, $t = 2.71$, $df = 42.75$, $p > 0.005$ for a 99% confidence level (Appendix D). However, it was significant at the specified level, $t = 2.07$, $df = 42.75$, $p < 0.025$ for a 95% confidence level, as well as at the 90% confidence level, $t = 1.68$, $df = 42.75$, $p < 0.05$. These results confirm the statistical significant differences found with the confidence intervals of 95% and 90%.

Summary

Our study shows that traditional-designed experiments, like the Box-Behnken design, coupled with extreme discipline when conducting the experiment and a strong background and knowledge of the face recognition software used, can be used to detect if biases related to skin tone are present in the system. A statistically significant difference to recognize dark skin tone persons and light skin tone persons was found to exist between the ability of the facial recognition system at the 90% Confidence Level; $p = 0.05$ and at the 95% Confidence Level; $p = 0.025$,. The facial recognition system had greater difficulty recognizing the light skin tone test subjects. This was especially evident in the existence of seven runs, or seven combinations of variables with the light skin tone test subjects, which resulted in a total of ten replications that returned a score of zero when the system failed to detect the presents of a human face. Even when the runs that returned a score of zero were excluded, reducing the number of runs to twenty, calculating the t^* showed a statistical significance at the 90% confidence level, $p = 0.05$.

References

[1] Belhumeur, P. N. (2006). Retrieved 03/27/11, www1.cs.columbia.edu/~belhumeur/journal/face- challenge.pdf Ongoing Challenges in Face Recognition. *Department of Computer Science.* Columbia University. New York, New York.

[2] Hiremath, P. S., & Prabhakar, C. J. (2008). Symbolic factorial discriminant analysis for illumination invariant face recognition. International Journal of Pattern Recognition and Artificial Intelligence, vol. 22, no. 3 (2008) pg. 371-387. doi:10.1142/S021800140800634X

[3] Chai, X., Gao, W., Fu, X., & Shan, S. (2003). Virtual face image generation for illumination and pose insensitive face recognition. Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03). 2003 IEEE International Conference on Issue Date: 6-10 April 2003. p. IV - 776-9 vol.4. doi:10.1109/ICME.2003.1221270

[4] Beveridge, J. R., Bolme, D., Draper, B. A., & Givens, G. (2003). A statistical assessment of subject factors in the PCA recognition of human faces. *Computer Vision and Pattern Recognition Workshop, 2003. CVPRW '03.* *d*oi:10.1109/CVPRW.2003.10088.

[5] VeriLook (2011). Retrieved 05/12/11, http://www.neurotechnology.com. Biometric and Artificial Intelligence Technologies.

[6] Harris, H. (2010). *Facial recognition system screening evaluation methodology for complexion biases* (Unpublished master’s thesis). Eastern Illinois University, Charleston IL.

[7] Schmidt, S. R., & Launsby, R.G. (1994). Understanding industrial designed experiments, Box-Behnken Designs, pg 3-31, 4th Edition

[8] McCabe, G. P., Moore, D. S. (1989), Introduction to the practice of statistics, Comparing Two Means, pg 541-549, 3rd Edition NIST, (1997). Retrieved 06/23/11, https://mail.co.coles.il.us/exchweb/bin/redir.asp?URL=http://www.itl.nist.gov/iad/vip/face/bpr_mug3.html

Appendix A: Sample of Randomization Form

R1, R2 and R3 representing the three test subjects with the random order of testing numbered 1 thru 81.

Run	L	D	A1	A2	R1	R2	R3
1	-1	-1	0	0	30	38	39
2	-1	+1	0	0	45	36	31
3	+1	-1	0	0	5	37	1
4	+1	+1	0	0	70	52	55
5	0	0	-1	-1	16	32	3
6	0	0	-1	+1	12	72	42
7	0	0	+1	-1	65	62	4
8	0	0	+1	+1	19	78	56
9	0	0	0	0	81	26	22
10	-1	0	0	-1	60	6	73
11	-1	0	0	+1	14	18	44
12	+1	0	0	-1	58	51	17
13	+1	0	0	+1	41	25	21
14	0	-1	-1	0	67	74	29
15	0	-1	+1	0	49	71	64
16	0	+1	-1	0	40	8	50
17	0	+1	+1	0	9	47	54
18	0	0	0	0	69	15	57
19	-1	0	-1	0	66	13	28
20	-1	0	+1	0	79	63	10
21	+1	0	-1	0	33	2	76
22	+1	0	+1	0	20	46	43
23	0	-1	0	-1	75	80	7
24	0	-1	0	+1	48	61	77
25	0	+1	0	-1	68	24	11
26	0	+1	0	+1	23	34	35
27	0	0	0	0	53	59	27

L = light; D = distance; A1 = horizontal angle; A2 = vertical angle. Adapted from Schmidt, S. R., & Launsby, R. G. (1994). Understanding industrial designed experiments, Box-Behnken Designs, pg 3-31, 4th Edition

Appendix B: Dark Skin tone

Scores listed under Y1, Y2 and Y3

Factor	A	B	C	D	Dark Skin Scores				
Row #	Light	Distance	Horizontal	Vertical	Y1	Y2	Y3	Mean	SD
1	-1	-1	0	0	466	559	447	490.6667	59.936077
2	-1	1	0	0	496	624	882	667.3333	196.61468
3	1	-1	0	0	742	712	1200	884.6667	273.49832
4	1	1	0	0	518	681	706	635	102.09309
5	0	0	-1	-1	163	181	152	165.3333	14.640128
6	0	0	-1	1	105	120	97	107.3333	11.676187
7	0	0	1	-1	128	114	183	141.6667	36.473735
8	0	0	1	1	76	156	71	101	47.69696
9	0	0	0	0	617	1673	1186	1158.667	528.53035
10	-1	0	0	-1	184	320	235	246.3333	68.704682
11	-1	0	0	1	570	263	171	334.6667	208.93141
12	1	0	0	-1	148	466	255	289.6667	161.80956
13	1	0	0	1	203	386	259	282.6667	93.767443
14	0	-1	-1	0	59	139	72	90	42.930176
15	0	-1	1	0	174	91	50	105	63.174362
16	0	1	-1	0	240	229	121	196.6667	65.759663
17	0	1	1	0	162	247	116	175	66.460515
18	0	0	0	0	973	1952	925	1283.333	579.57945
19	-1	0	-1	0	162	283	159	201.3333	70.741313
20	-1	0	1	0	154	189	280	207.6667	65.041013
21	1	0	-1	0	206	346	187	246.3333	86.835093
22	1	0	1	0	279	256	527	354	150.2631
23	0	-1	0	-1	90	115	258	154.3333	90.643992
24	0	-1	0	1	84	60	73	72.33333	12.013881
25	0	1	0	-1	125	352	233	236.6667	113.54441
26	0	1	0	1	194	179	280	217.6667	54.500765
27	0	0	0	0	1946	748	1408	1367.333	600.03444
							mean	385.65	143.18

Appendix C: Light Skin tone

Scores listed under Y1, Y2 and Y3

Factor	A	B	C	D	Light Skin Scores					
Row #	Light	Distance	Horizontal	Vertical	Y1	Y2	Y3		Mean	SD
1	-1	-1	0	0	356	227	387		323.3333	84.85478
2	-1	1	0	0	359	452	342		384.3333	59.2143
3	1	-1	0	0	615	194	280		363	222.4343
4	1	1	0	0	892	864	828		861.3333	32.08323
5	0	0	-1	-1	109	165	0		91.33333	83.90669
6	0	0	-1	1	183	70	173		142	62.55398
7	0	0	1	-1	157	100	162		139.6667	34.44319
8	0	0	1	1	172	157	204		177.6667	24.00694
9	0	0	0	0	736	1022	741		833	163.6979
10	-1	0	0	-1	410	263	405		359.3333	83.46456
11	-1	0	0	1	186	68	657		303.6667	311.6317
12	1	0	0	-1	386	427	339		384	44.03408
13	1	0	0	1	284	363	294		313.6667	43.0155
14	0	-1	-1	0	0	0	0		0	0
15	0	-1	1	0	142	53	0		65	71.75653
16	0	1	-1	0	150	162	180		164	15.09967
17	0	1	1	0	227	280	201		236	40.26164
18	0	0	0	0	1145	1384	1492		1340.333	177.5735
19	-1	0	-1	0	271	262	232		255	20.42058
20	-1	0	1	0	350	197	237		261.3333	79.34944
21	1	0	-1	0	197	178	311		228.6667	71.93284
22	1	0	1	0	199	213	0		137.3333	119.14
23	0	-1	0	-1	101	0	185		95.33333	92.63009
24	0	-1	0	1	0	126	0		42	72.74613
25	0	1	0	-1	246	208	165		206.3333	40.52571
26	0	1	0	1	204	0	196		133.3333	115.5393
27	0	0	0	0	861	839	637		779	123.4666
								mean	319.25	84.81

Appendix D: Statistical Analysis

Calculation for Degrees of Freedom

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2}$$

$$df = \frac{\left(\frac{20{,}500.88}{27} + \frac{7{,}192.20}{27}\right)^2}{\frac{1}{26}\left(\frac{20{,}500.88}{27}\right)^2 + \frac{1}{26}\left(\frac{7{,}192.20}{27}\right)^2}$$

$$df = 42.75$$

The two-sample "t" Significance Test:

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$t = \frac{(385.65 - 319.259)}{\sqrt{\frac{20{,}500.88}{27} + \frac{7{,}192.20}{27}}}$$

$$t = 2.07$$

The Two-Sample t Confidence Interval

$$(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$(385.65 - 319.259) \pm t^* \sqrt{\frac{20{,}500.88}{27} + \frac{7{,}192.20}{27}}$$

$$66.39 \pm t^* \sqrt{1{,}025.66}$$

The Two-Sample t Confidence Intervals

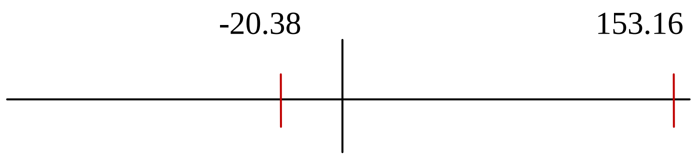
99% Confidence Level

$66.39 \pm 2.71\ (32.02)$

66.39 ± 86.77

[-20.38 → 153.16]

$t = 2.71$



40 *df*: = 2.704

50 *df*: = 2.678

99%: $p = 0.005$

For the 99% confidence level, the results from the two-sample t significance test was <u>not</u> statistically significant at the specified level, $t = 2.71$, $df = 42.75$, $p > 0.005$.

The Two-Sample t Confidence Intervals

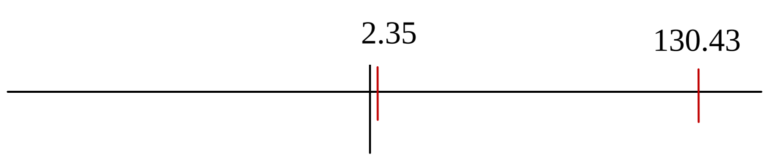
95% Confidence Level

$66.39 \pm 2.0\ (32.02)$

66.39 ± 64.04

[2.35 → 130.43]

$t = 2.0$



40 *df*: = 2.021

50 *df*: = 2.009

95%: $p = 0.025$

For the 95% confidence level, the results from the two-sample t significance test was statistically significant at the specified level, $t = 2.07$, $df = 42.75$, $p < 0.025$.

The Two-Sample t Confidence Intervals

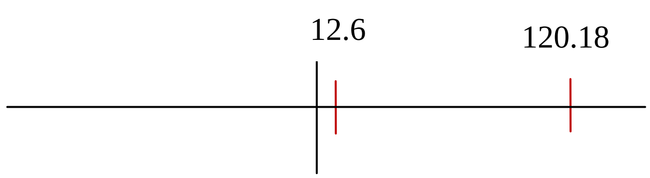
90% Confidence Level

$66.39 \pm 1.68\ (32.02)$

$66.39 \pm 53.79.77$

[12.6 → 120.18]

$t = 1.68$



40 *df*: = 2.021

50 *df*: = 2.009

90%: $p = 0.05$

For the 90% confidence level, the results from the two-sample t significance test was statistically significant at the specified level, $t = 1.68$, $df = 42.75$, $p < 0.05$.