**1. Introduction**

Are women smarter than men? If you happen to be male and want to keep breathing, the answer is a definite yes. However, multitudes of studies have explored and compared intelligence levels of the two sexes in previous years utilizing various methods. The results have been largely ineffective or outright disregarded by subject matter experts.

Throughout the decades, six dominant positions have been taken on the difference in intelligence levels of males and females [1]:

1. Intelligence cannot be accurately measured.
2. There are no differences at all because no theories support a difference.
3. There are no mean differences between the sexes but there ARE differences at the extremes.
4. Evolutionary theory suggests a range of differences in intellectual abilities that make up overall intelligence.
5. Sex differences that emerge are not real due to societally learned traits
6. There are real differences between sexes with males having a four to eight-point test score advantage after the age of 15.

Anecdotal and scientific evidence exists showing stark discrepancies in the learning habits and maturity rates of males and females. Young girls are thought to mature and learn language faster than young boys. Boys tend to excel at mathematical reasoning and learn self confidence and self worth while girls are more proficient at linguistics, have better memory and place more value in group interactions than boys. These differences are thought to be maintained or enhanced as age increases.

For this study, we intend to investigate the latter: linguistic and mathematical proficiency differences between male and female youths utilizing multivariate analysis of variance techniques.

**2. Experimental Design**

In 1979, 12,686 compensated individuals participated in the National Longitudinal Survey of Youth (NLSY). The NLSY is a general survey of youth behavioral and educational characteristics. All participants in the NLSY survey were born between 1957 and 1964 and were non-institutionalized civilian or military status. All 50 states in the United States were represented. [2]

Of these general survey participants, 11,914 subjects completed the Armed Services Vocational Aptitude Battery (ASVAB) test. The ASVAB consists of a battery of 10 tests that measure knowledge and skill across ten academic and a few professional areas.

The results of the survey we’re going to investigate include intelligence test scores across four categories: arithmetic knowledge, mathematical knowledge, word knowledge and paragraph comprehension. Because no randomization was applied, we cannot associate causality with the results of subsequent analysis. However, an association can be found if evidence is shown to support one. Inferences would also not pertain to present day, as the data set is quite old.

**3.** **Exploratory Data Analysis**

One of the major issues with the data lies in age difference. To address this issue, each variable has been normalized for the age of the respondent. For instance, if a 17-year-old respondent scores lower on the ASVAB than a 21-year-old, the score is not likely to deviate as largely thanks to data normalization. Researchers used item response theory (IRT) and resulting theta scores to normalize the raw data scores. Therefore, in our analysis, we will be working with normalized responses only, controlling for age differences.

Full exposition to explanatory and response variables is required to understand model set up and additional analysis. Data were obtained via The Statistical Sleuth data sets [3] and will be presented in order of response and factor. All subjects received scores on each of the variables presented, thus no missing data is present.

**3.1 Response Variables**

|  |  |  |
| --- | --- | --- |
| **Variable 1** | **Levels** | **Description** |
| Arith | Quantitative | Arithmetic Reasoning Score |
| Respondents were tested on their ability to use basic arithmetic to solve word problems. The total number of observations is 2584. The range of the normalized Arith variable is 30 (from 0 to 30) with a non-normal, multi-modal distribution that is skewed slightly left, as expected with test scores. The mean of arith is 18.5. Due to sample size, univariate normality is not much of a concern (as it will not be with other variables). However, further assumption checks for MANOVA will be executed later. | | |
| **Variable 2** | **Levels** | **Description** |
| Math | Quantitative | Mathematics Knowledge Score |
| Respondents were tested on early high school mathematics knowledge including algebra and geometry. The range of Math is 25 (from 0 to 25) with a multi-modal distribution that is skewed more moderately left. The mean of Math is 14.2.  ../../../../../Screen%20Shot%202015-11-05%20at%207.21.43%20AM.png  ../../../../../Screen%20Shot%202015-11-05%20at%207.12.16%20AM.png../../../../../Screen%20Shot%202015-11-05%20at%207.12.26%20AM.png | | |

|  |  |  |
| --- | --- | --- |
| **Variable 3** | **Levels** | **Description** |
| Parag | Quantitative | Paragraph Comprehension Score |
| Paragraph comprehension questions consisted of SAT-like reasoning questions, testing the respondents on their ability to fully grasp the points of the word problem. The range of the normalized Parag variable is 15 (from 0 to 15) with a heavily left skewed distribution and a mean of 11.2.  ../../../../../Screen%20Shot%202015-11-05%20at%207.19.21%20AM.png  ../../../../../Screen%20Shot%202015-11-05%20at%207.08.52%20AM.png../../../../../Screen%20Shot%202015-11-05%20at%207.09.53%20AM.png | | |
| **Variable 4** | **Levels** | **Description** |
| Word | Quantitative | Word Knowledge / English Comprehension |
| Word knowledge questions tested respondents’ ability to appropriately find errors in and complete an English sentence. The range of the normalized Word variable is 35 (from 0 to 35). The distribution of word is left skewed with a bi-modal distribution and a mean of 26.6. | | |

**3.2 The Factor**

Our analytical method is a one-way multivariate analysis of variance (MANOVA). This means that there is one factor we’re interested in and we do not consider any blocking variables. Exposition to the factor follows:

|  |  |  |
| --- | --- | --- |
| **Variable 9** | **Levels** | **Description** |
| Gender | 2 | The sex of the subject |
| 1278 females and 1306 males make up the 2,584 participants in our unbalanced data set.  **1:** Male **2:** Female  Initial means analysis shows the most significant differences in arithmetic and math, with males having higher means than females (Appendix Fig.1). Females outperform males in paragraph comprehension by almost 1.5 points. Full hypothesis testing is required to confirm means differences. | | |

**3.3 Additional Assumption Checking**

Multivariate analysis of variance considers four critical assumptions as the foundation of accurate analysis:

1. The data from group has common mean vector
2. The data from all groups have a common variance co-variance matrix Σ
3. The subjects are independently sampled
4. The data are multivariate normally distributed

Because sampling methods are inclusive of youth born between 1957 and 1964 and we are interested in analyzing the population of youth in the 1979 timeframe, we do not assume that subpopulations have different mean vectors and assumption one is met.

A Bartlett’s test for homogeneity of variance co-variance matrices between groups shows that the groups do not have common variance co-variance matrices at an alpha level of 0.05 (p = 0.0086). Additional log, inverse and square root transformations were attempted to meet MANOVA variance co-variance assumptions, however, transformations proved to worsen the lack of commonality between variance co-variance matrices (Appendix Fig. 2). Statistical literature indicates that alpha levels for the box test extension (Bartlett’s test) should be considered at an alpha of 0.001[4] and univariate non-normality, which we have seen, may provide inaccurate results for the test of homogeneity of variance co-variance matrices [5]. Therefore, using these assumptions, it is appropriate to move forward with MANOVA. However, to err on the side of being conservative, we will use pillai’s trace as our test statistic [5].

We assume independence in the data. Sampling methods do not cluster observations and data is not serially correlated. Indeed, serial correlation is not an issue in the study because our data is not time series related. Further, special care was taken to ensure samples varied widely across states, academic institutions and other youth programs.

Multivariate normality is not an issue given the sample size of 2,584, with over 1,000 observations for each level of the gender factor. We noticed bi-modal and multi-modal distributions in the individual variable analysis, which is often appropriate for diagnosing multivariate normality. However, given tenants of the central limit theorem and large sample sizes, normality is not an issue. This lack of overall normality could affect the test for homogeneity previously mentioned.

**4.** **Multivariate Analysis of Variance**

Multivariate analysis of variance or MANOVA allows multiple measures to be considered when analyzing data for one or two factors. In our case, we are interested in testing the assumption that males tend to perform more effectively at arithmetic and mathematics while females perform more effectively at linguistic skills, such as paragraph comprehension and word knowledge. MANOVA is critical in this case because single univariate tests would fail to capture the natural correlation between the variables we are interested in analyzing. Correlation is obvious in the pearson’s r table given below and in regression plots in figures 3 and 4 in the appendix:

|  |  |
| --- | --- |
| **Correlation Summary** | **Summary Table** |
| * Moderate to strong correlation is present across all variables, especially between math and arithmetic (r= 0.82 p = <.0001) and parag and word (r = 0.77 p = <.001) * Because of these correlations, MANOVA is favored over single, univariate ANOVAs to capture correlation between variables * MANOVA performs more favorably than ANOVA when correlation is present | ../../../../../Screen%20Shot%202015-11-05%20at%203.30.22%20PM.png |

**4.1 Overall Hypothesis and Initial Model Fit**

Utilizing the data on 2,584 participants in ASVAB test, we consider the four quantitative variables previously mentioned on one factor (gender) with two levels (male, female) to test the null hypothesis:

**for the entire set of response variables within each group (arith, parag, word, math)**

While forming a formal alternative hypothesis:

**for at least one and at least one variable where i and j represent male and female groups.**

Our initial investigation is general, meaning we wish to understand if there is any difference in scores across variables and groups in our data set. We will consider pillai’s trace as our primary test statistic. Should the overall MANOVA prove significant, contrasts will be performed to formally test differences between genders. Further, estimates formed on each variable and the differences between males and females in our data set.

MANOVA indicates significance across all test statistics (Fig 5). Our chosen test statistics, pillai’s trace, is very significant (p <0.0001) Therefore, there are differences in scores of at least one academic variable between males and females. Indeed, a profile plot shows, at least visually, that males tend to perform better on arithmetic and math knowledge while females perform better on paragraph comprehension:

|  |  |
| --- | --- |
| **Summary** | **Profile Plot** |
| Profile plot showing male means (blue) and female means (black).  Visual evidence indicates males perform better on arithmetic and math while females perform better on paragraph comprehension.  Word knowledge seems to be about equal. Formal contrasts are necessary. | ../../../../../Screen%20Shot%202015-11-05%20at%204.19.51%20PM.png |

**4.2 Univariate Analyses**

Univariate analyses of variance were carried out for each of the four academic variables. These analyses allow us to identify significant differences in gender *across each variable*, with the caveat that our identified alpha level needs to be adjusted for the experimentwise error rate.

Each univariate analysis formally tests the null hypothesis:

Across each variable: arith, word, parag, math. The alternative hypothesis can be found as:

Across each variable.

In this instance, we are not dealing with vectors of variables for each survey participant, rather, we are investigating a standard univariate ANOVA, looking for group differences for each *individual* variable.

Given an alpha level of 0.05, we control the experimentwise error rate by dividing alpha by the number of variables we are testing, such that our adjusted alpha = 0.05 / 4. Our univariate alpha level is thus 0.0125 instead of 0.05. Results for each variable can be found in the table below:

|  |  |  |
| --- | --- | --- |
| **Response Variable** | **p value** | **Significant** |
| Arith | <0.0001 | Yes |
| Word | 0.9364 | No |
| Parag | <0.0001 | Yes |
| Math | 0.0023 | Yes |

Results show that means for arithmetic, paragraph comprehension and mathematics knowledge differ among genders.

**4.3 Individual Contrasts and Estimates**

Univariate analysis, in this case, is adequate in identifying significance in group mean differences because there are only two groups to test. However, we’d like to execute pre-planned contrasts and estimates of linear combinations with a formal hypothesis and contrast coefficients of:

**with**

Because our contrast is the same as the test for no overall effect, pillai’s trace and all other test statistics are significant at p < 0.0001, indicating there is indeed a difference among the academic variables between males and females. Further, all individual contrasts except for word knowledge are statistically significant, as indicated by the univariate analyses previously.

Now that we’ve formally tested variable hypotheses and confirmed differences in arithmetic, math, and paragraph comprehension, estimates and confidence intervals can be ascertained. The formula for the bonferroni corrected confidence interval for estimate of the difference in arithmetic between genders can be found below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Academic Variable** | **Difference** | **Standard Error** | **Multiplier\*SE Half Width** | **Confidence Interval** |
| Arith | 2.04 | 0.28 | 0.77 | [1.27, 2.81] |
| Math | 0.75 | 0.24 | 0.66 | [0.09, 1.41] |
| Word | -0.02 | 0.28 | 0.77 | [-0.79, +0.75] |
| Parag | -0.57 | 0.12 | 0.33 | [-0.90, -0.24] |

Estimates show males tend to perform better on math (+0.75) and arithmetic (+2.04) based questions, while females perform better on paragraph comprehension (+0.57). There is no difference in word knowledge between male youths and female youths in our study. As you can see, the bonferroni adjusted confidence interval crosses zero and we have confirmed no statistical significance for differences in word knowledge.

**5.** **Conclusion**

There is evidence to show that male youths perform better at arithmetic reasoning and mathematics knowledge than female youths. Females, on the other hand, tend to perform better at paragraph comprehension than males. This does not mean that males are smarter than females in regard to mathematical knowledge and arithmetic reasoning as our data are obtained from an observational study with no randomization involved. We can infer an association between male youths in the United States and higher arithmetic and mathematics scores and female youths and paragraph comprehension scores. However, as this data was captured in 1979, any inferences to present day should be taken with extreme caution.

Further, a canonical correlation analysis ran on the same data set shows gender explains little of the variance in test scores (squared canonical correlation of 0.08) from a single variate of linear combinations (Appendix Fig. 5). Thus, there are likely other variables that explain how well the different genders perform on each of the academic variables tested. This data could be investigated by including more explanatory variables in a larger canonical correlation study.

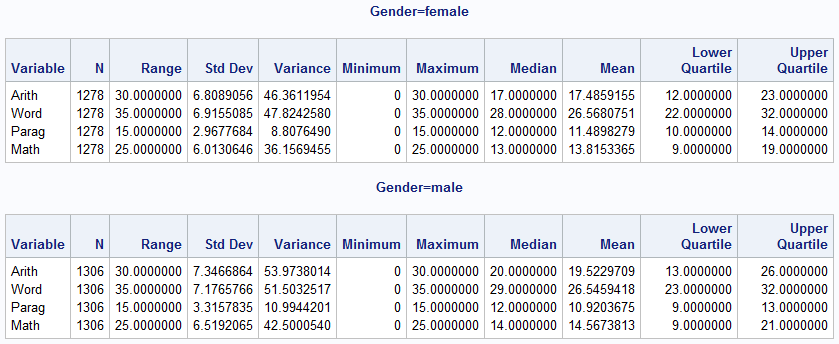
As a final point, this analysis does not encourage men to tell their wives or female counterparts that males are the smarter sex. We are not responsible for any subsequent reaction or injuries (emotional or physical) sustained by males using this analysis as proof of any kind.

**REFERENCES**

1. <https://www.psychologytoday.com/blog/sideways-view/201410/are-men-really-more-intelligent-women>
2. <https://www.nlsinfo.org/content/cohorts/nlsy79/topical-guide/education/aptitude-achievement-intelligence-scores>
3. www.statisticalsleuth.com
4. <https://en.wikiversity.org/wiki/Box%27s_M>
5. <http://www.real-statistics.com/multivariate-statistics/boxs-test-equality-covariance-matrices/boxs-test-basic-concepts/>

**APPENDIX 1 – FIGURES**

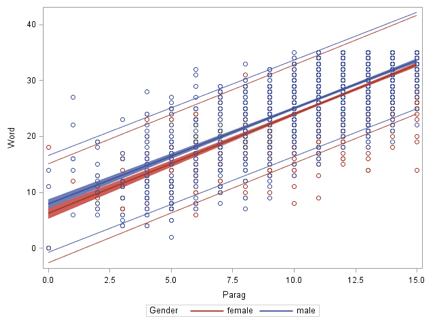
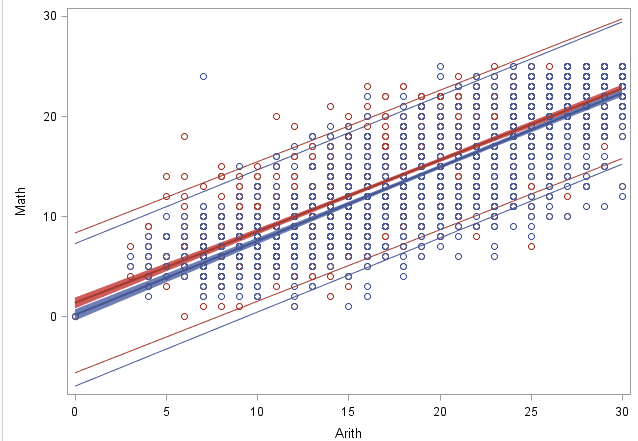
**Figure 1 – Means Tables By Gender**



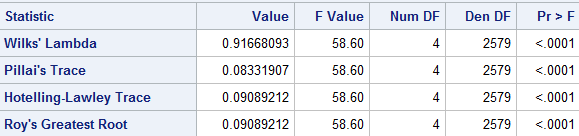
**Figure 2 – Transformations To Correct Variance Co-Variance**

|  |  |  |
| --- | --- | --- |
| **Variable Transformation** | **Test** | **Result (p=)** |
| None | Bartlett’s | 0.0086 |
| Log | Bartlett’s | < 0.0001 |
| Square Root | Bartlett’s | 0.0012 |
| Inverse | Bartlett’s | < 0.0001 |

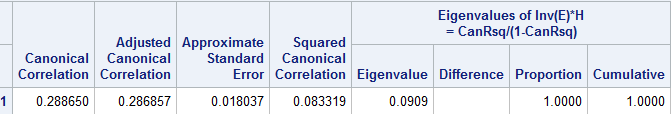
**Figures 3 and 4 – Regression of Math on Arith and Parag and Word**

****

**Figure 5 – Test Statistics for Overall MANOVA**

****

**Figure 6 – Resulting Variate From Canonical Correlation**

****

**APPENDIX 2 – SAS CODE**

/\* import initial data set \*/

**PROC** **IMPORT** out = proj2

DATAFILE = '\\Client\C$\Users\PatrickCoryNichols\Desktop\Data Science\Classes\Stats\Data Sets\sleuth3csv\ex0222.csv'

DBMS = CSV REPLACE;

GETNAMES = YES;

**RUN**;

/\* create second data set with transformed responses \*/

**DATA** project;

SET proj2;

logarith = log(arith);

logword = log(word);

logmath = log(math);

logparag = log(parag);

sqrarith = sqrt(arith);

sqrword = sqrt(word);

sqrmath = sqrt(math);

sqrparag = sqrt(parag);

invarith = **1**/arith;

invword = **1**/word;

invmath = **1**/math;

invparag = **1**/parag;

**RUN**;

/\* basic means analysis \*/

**proc** **means** data = project n range std var min max median mean q1 q3;

var arith word parag math;

**run**;

**proc** **sort** data = project;

by gender;

**run**;

/\* basic means analysis by gender \*/

**proc** **means** data = project n range std var min max median mean q1 q3;

by gender;

var arith word parag math;

**run**;

/\* univariate normality checks \*/

**PROC** **UNIVARIATE** NOPRINT data = project;

VAR arith math parag word;

HISTOGRAM;

QQPLOT;

**RUN**;

/\* bivariate normality checks for ellipses \*/

**PROC** **SGSCATTER**;

MATRIX arith word parag math;

**RUN**;

/\* bartlett's / box test for homogoneity of variance co-variances \*/

/\* this test was repeated 4x for transformed variables \*/

**PROC** **DISCRIM** data = project pool=test;

class gender;

var arith math word parag;

**run**;

/\* correlation table for response variables \*/

**PROC** **CORR**;

VAR arith word parag math;

**RUN**;

/\* scatter grouped by gender for arithmetic and math to confirm visual correlation \*/

**PROC** **SGPLOT**;

scatter x =arith y=math / group = gender;

**run**;

/\* regression line with confidence limits grouped by gender for arithmetic

and math to confirm visual correlation \*/

**PROC** **SGPLOT**;

reg x=arith y=math /

group = gender CLM CLI;

**run**;

/\* regression line with confidence limits grouped by gender for parag

and word to confirm visual correlation \*/

**PROC** **SGPLOT**;

reg x=parag y=word /

group = gender CLM CLI;

**run**;

/\* the following code sets up the profile plot for all four variables \*/

**data** project\_profiles;

set project;

varscore ="arith";amount=arith;output;

varscore="math";amount=math;output;

varscore="word";amount=word;output;

varscore="parag";amount=parag;output;

**run**;

**proc** **sort** data=project\_profiles;

by gender varscore;

**run**;

**proc** **means** data=project\_profiles;

by gender varscore;

var amount;

output out = a mean=mean;

**run**;

**proc** **gplot** data = a;

axis1 length = **3**in;

axis2 length= **4.5**in;

plot mean\*varscore=gender /vaxis=axis1 haxis=axis2;

symbol1 v=J f=special h=**2** l=**1** i=join color=black;

symbol2 v=L f=special h=**2** l=**1** i=join color=black;

**run**;

/\* manova analysis with supporting CCA \*/

**PROC** **GLM** data = project;

CLASS gender;

MODEL arith word parag math = gender;

CONTRAST 'Gender Difference Contrast' gender -**1** **1**;

ESTIMATE 'Gender Difference Estimate' gender -**1** **1**;

lsmeans gender / stderr;

MANOVA h = gender / canonical printe printh;

**RUN**;

**QUIT**;