

A group of researchers interested in astrology had come to believe that the earning power of an individual was greatly influenced by their astrological sign. The researchers obtained data from 1856 individuals who responded to the 2008 General Social Survey (GSS) and proceeded to examine the effect of astrological sign on socioeconomic status. The GSS contains a variable, named Socioeconomic Index (SEI), which converts a respondent's job category into a number representing a rough estimate of their general socioeconomic status at the time of the interview (GSS, Davis & Smith, 1972-2008). In addition, a respondent's zodiac sign was collected by the GSS, which was used in the analysis.

To determine if there were mean differences in SEI across the different signs of the zodiac, the researchers conducted a one-way ANOVA using SEI as the dependent variable and zodiac sign as the independent variable. Prior to running the analysis, a boxplot of SEI by zodiac sign was investigated and can be found in Figure 1. Upon visual inspection, the boxplot suggested that there did not seem to be a real pattern of mean differences, and that the data was void of many outliers. Following the plot, the results of the analysis revealed a non-significant (at  $\alpha = 0.05$ ) difference between all the means across all zodiac signs ( $F_{11, 1844} = 1.225$ ,  $p = 0.285$ ). Furthermore, the effect size for the analysis was very small ( $0.001$  – calculations shown in Appendix A), indicating there was very little difference between the mean SEI of the astrological signs.

Although the ANOVA model was not statistically significant, the researchers still wanted to check the assumptions of the analysis to ensure the p-value was accurate. The ANOVA model assumes: 1) that observations are independent, 2) that error terms are normally distributed with a zero mean and 3) homogeneous variance. Because of the structure of the data, no tests could be conducted to investigate the first assumption of independence (which could be investigated in, say, nested samples). For the second assumption, of normality, the residuals from the analysis were saved and plotted using the Q-Q plot feature in SPSS. If the residuals were normally distributed, the points would fall exactly on the diagonal line in the plot. From visual inspection, the points seemed to deviate from the line in large quantities near the lower end of the residual range. For further inspection, Figure 3 displays the residuals plotted using a histogram. The assumption of normality is likely violated. The third assumption, homogeneity of error variance was tested using Levene's test (at  $\alpha = 0.05$ ) which was not significant ( $F_{11, 1844} = 1.727$ ,  $p = 0.062$ ). Based on the results of the test, the researchers concluded that the homogeneity of variance assumption likely holds.

Because one of the assumptions was violated (normality of error terms) additional investigation was merited. The researchers noted that the distribution of residuals (Figure 3) was bi-modal, making it likely that transformations of the data would not yield normality. Therefore the researchers decided to use a more robust test of the means – the Brown-Forsythe test. As with the previous analysis, the test revealed no significant (at  $\alpha = 0.05$ ) differences between the means ( $BF = 1.223$ ,  $p = 0.266$ ).

Although shaken by the results of the one-way ANOVA, the researchers pressed forward with a pre-specified contrast comparing the means of the earth signs (Taurus, Virgo, and Capricorn) to the means of the water signs (Cancer, Scorpio, and Pisces). The researchers believed the earth signs would have higher SEI than the water signs, particularly due to the fact that a not-so-well-known assistant professor was a Scorpio and thereby brought the SEI down with his salary. The researchers formed a contrast testing this hypothesis, weighting the earth signs with coefficients of  $1/3$  and the water signs with coefficients of  $-1/3$ . All other signs received a zero for their coefficients. The contrast was not significant at  $\alpha = 0.05$  ( $t_{1844} = 0.32$ ,  $p = .980$ ), revealing that the earth signs likely did not have a higher SEI than the water signs.

Undeterred, the researchers ordered one final analysis comparing the mean SEI for all pairs of signs. As this was a post-hoc analysis with multiple statistical tests, the chance of committing a Type-I error was fairly high. Therefore, the researchers used Tukey's method of Type-I error control setting the family-wise Type-I error rate to 0.05. Of the 66 paired comparisons made, not a single one was significantly different.

Having tried repeatedly to find a significant difference in SEI between signs of the zodiac, the researchers finally realized the errors of their ways. Steadfast in their devotion to astrology, they could not believe their result. They decided SEI was an inaccurate indicator of a person's true socioeconomic status. They then sought out to find better data because, in their words, "everyone knows the zodiac predicts how much a person will earn."

Figure 1

Boxplot of Socioeconomic Index by Zodiac Sign

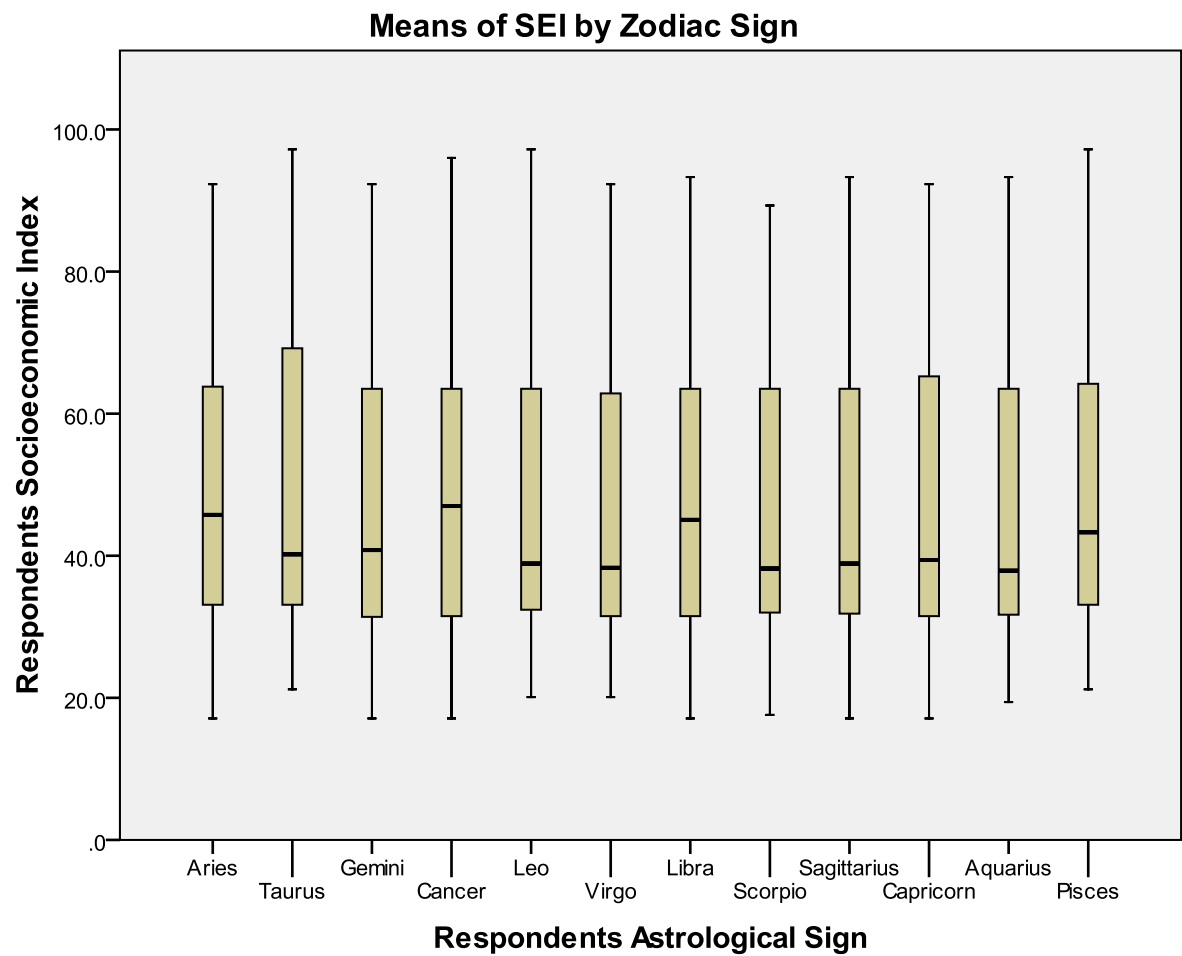


Figure 2

Q-Q Plot of Residuals from Omnibus ANOVA Model

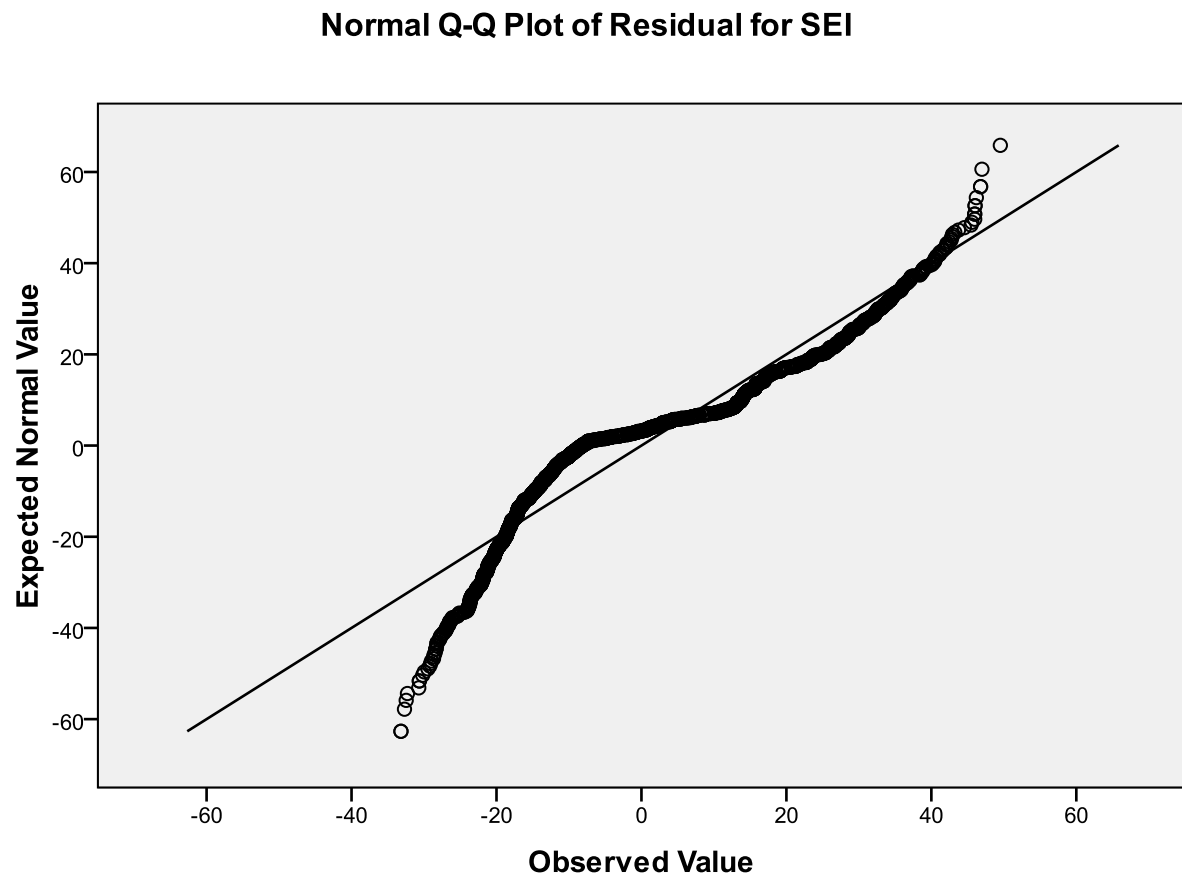


Figure 3

Histogram of Residuals with Normal Curve Superimposed

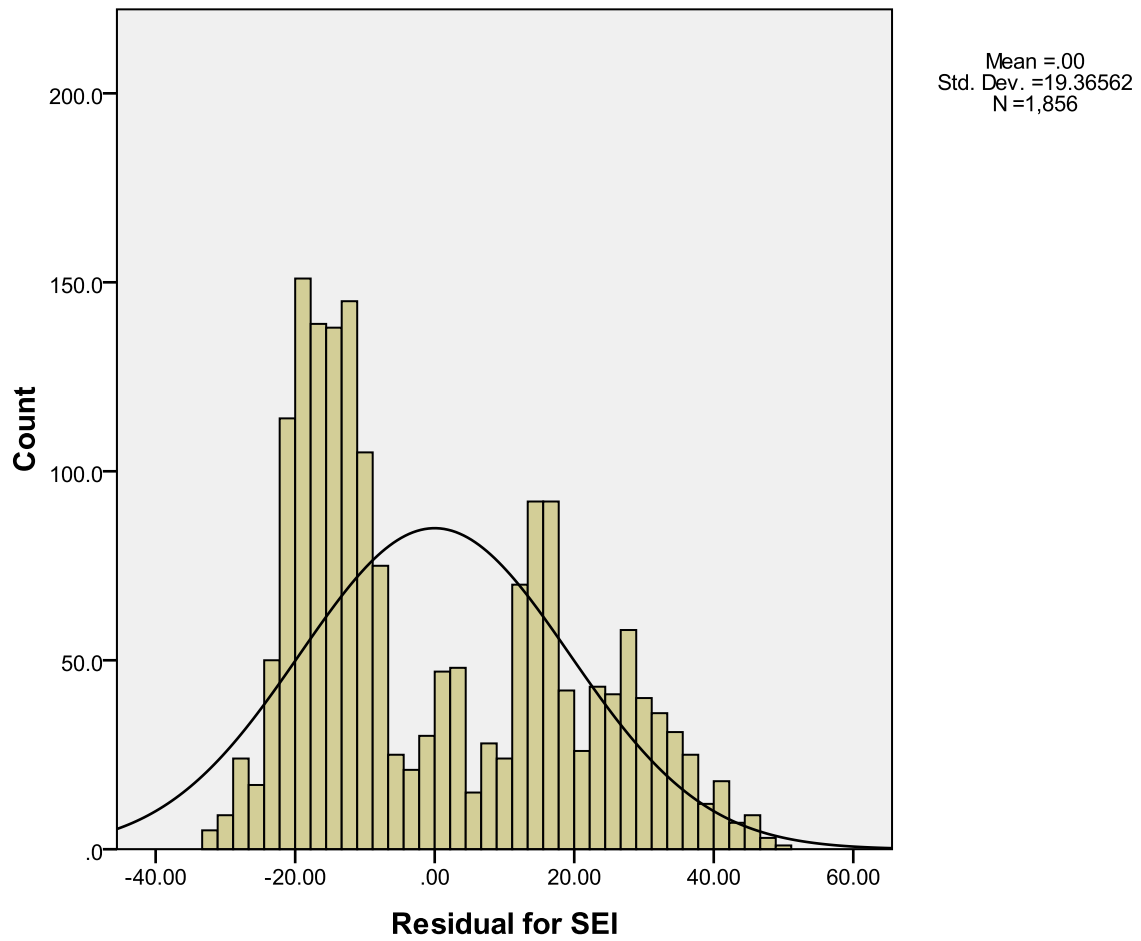


Table 1

Descriptive Statistics of Socioeconomic Status by Zodiac Sign

<b>Sign</b>	<b>N</b>	<b>Mean</b>	<b>Standard Deviation</b>
Aries	160	50.283	19.385
Taurus	143	51.172	21.211
Gemini	154	47.747	18.644
Cancer	155	49.794	19.479
Leo	149	47.685	18.991
Virgo	192	46.320	18.654
Libra	166	49.563	19.898
Scorpio	141	46.579	18.660
Sagittarius	136	47.825	18.741
Capricorn	143	49.387	20.539
Aquarius	151	46.321	18.769
Pisces	166	50.408	20.052
<b>Total</b>	<b>1856</b>	<b>48.576</b>	<b>19.436</b>

## References

Davis, J. A., & Smith, T. W. (1972-2008). *General social surveys*. Chicago: NORC.

Keppel, G., & Wickens, T. (2004). *Design and analysis: a researcher's handbook*. New York: Prentice.

## Appendix A

### Calculation of Effect Size

#### Overall ANOVA

The effect size for the overall ANOVA uses the omega-squared provided by formula 8.11 of Keppel and Wickens (2004, p. 164):

$$\hat{\omega}^2 = \frac{SS_A - (a - 1)MS_{S/A}}{SS_{total} + MS_{S/A}} = \frac{5,081.581 - (12 - 1) * 377.264}{700,757.049 + 377.264} = 0.001$$

Alternatively, we could use the partial eta-squared effect size:

$$\eta_{<A>}^2 = \frac{SS_A}{SS_A + SS_{S/A}} = \frac{5,081.581}{5,081.851 + 695,675.468} = 0.007$$