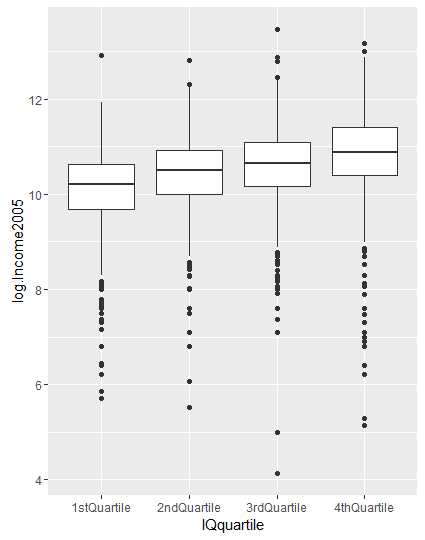
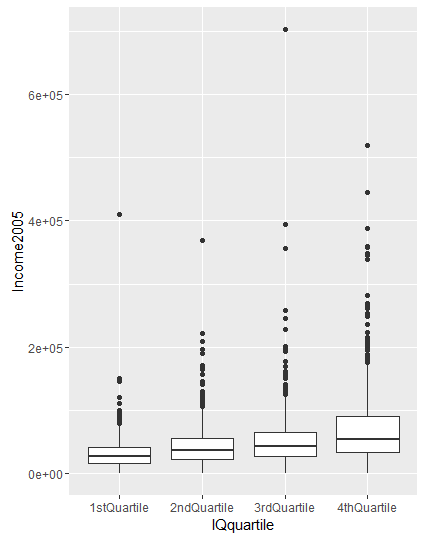
Anthony Le

ST 511 (T 16:00-16:50)

1 November 2016

Homework 5

Question 1: Problem 24

1. 
2. The distributions appear skewed with different spreads (standard deviations) before a log transformation, so it is not justified in making the assumptions that the populations are normal, the variances are equal, and the standard deviations are equal. Thus, a log transformation was used to resolve the skewness to perform a one-way ANOVA on the data since the one-way ANOVA, like the two-sample t-test, assumes the population standard deviations are equal.

> anova(log.ex0524.aov)

Analysis of Variance Table

Response: log.Income2005

Df Sum Sq Mean Sq F value Pr(>F)

IQquartile 3 184.79 61.598 70.165 < 2.2e-16 \*\*\*

Residuals 2580 2264.98 0.878

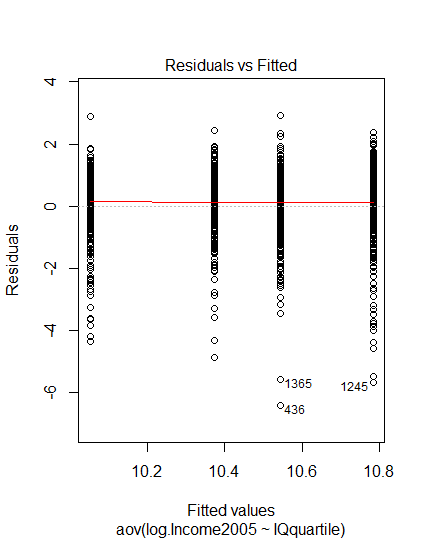
---

Signif. codes:

0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

1. Statistical Conclusion: There is strong evidence that the median annual income in 2005 for those within the highest (4th) quartile of IQ test scores is greater than that of those within the lowest (1st) quartile of IQ test scores (two-sided p-value<2.2e-16, from a two-sample t-test and one-way ANOVA F-test). The median annual income in 2005 for those within the 4th quartile of IQ test scores was estimated to be 0.735 times greater than the median annual income for those within the 1st quartile of IQ test scores. A 95% confidence interval for the difference in annual income in 2005 is 0.632 to 0.837.

Scope of Inference: There is no random sample. The statistical results apply only to Americans who were available for re-interview in 2006 and who had paying jobs in 2005. Because the subjects were not selected randomly from any population, extending this inference to any others with paying jobs in 2005 outside of America is speculative. However, causal conclusion is strong even if it applies only to the recruited subjects. A log transformation on the data was necessary because the distributions appear skewed with different spread, so it is not justified in making the assumptions that the populations are normal, the variances are equal, and the standard deviations are equal.

1. 

Each group appears to have symmetric residuals about 0. The 3rd and 4th quartile groups have a few outliers below 0. The spread of residuals appears to be approximately equal for all four groups where each group has symmetry and normal looking distance around 0 which confirms the equal standard deviation assumption. Note that the data was calculated on the log scale.