Homework #3

3.3, 3.5, 3.8, 3.18, 3.26, 3.28, 3.30

Hayden Atchley

27 September 2022

3.3

If the crew were aware of whether seeding was conducted, they may have—consciously or not—changed their behaviors in a way that would influence the results.

3.5

Though there are several samples of dioxin levels in the Vietnam veterans that are substantially higher that in other veterans, these are all extreme outliers. Additionally, there are over 6 times as many observations in the former group, so the outliers represent an even smaller proportion of this group. Further analysis is required to determine if the difference is stastically significant.

3.8

The observations are not independent, but that is by design. The paired t-test is used in this case, since each observation is paired with another, related observation.

3.18

a

The difference in means is *not* resistant, as a dramatic change in one observation changes this statistic considerably.

b

The average of the ranks for the trauma group minus the average of the ranks for the non-trauma group *is* resistant, as a dramatic change in one observation would not change the rankings, and therefore the difference of their averages, very much if at all.

3.26

The log-transformed data from the Agent Orange study (adding 0.1 to all values since the dataset contains zeros) gives:
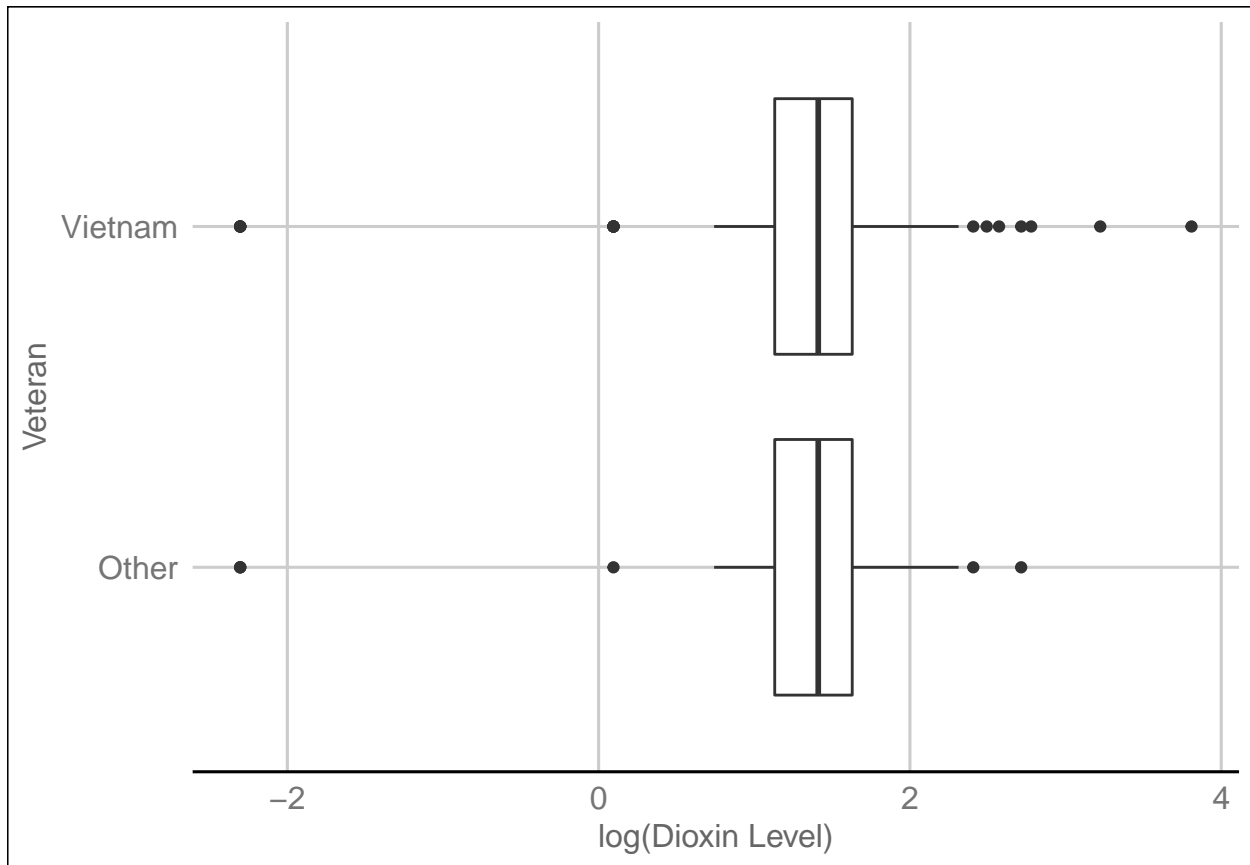
Running a t-test on these transformed samples gives a p-value of 0.343 and a 95% confidence interval of $-0.084 < \mu_2 - \mu_1 < 0.241$ for the log-transformed data. Back-transforming gives us that the median dioxin of the Vietnam veterans is between $e^{-0.084} = 0.919$ and $e^{0.241} = 1.273$ times as much as that of the other veterans.

3.28

A boxplot of the data:
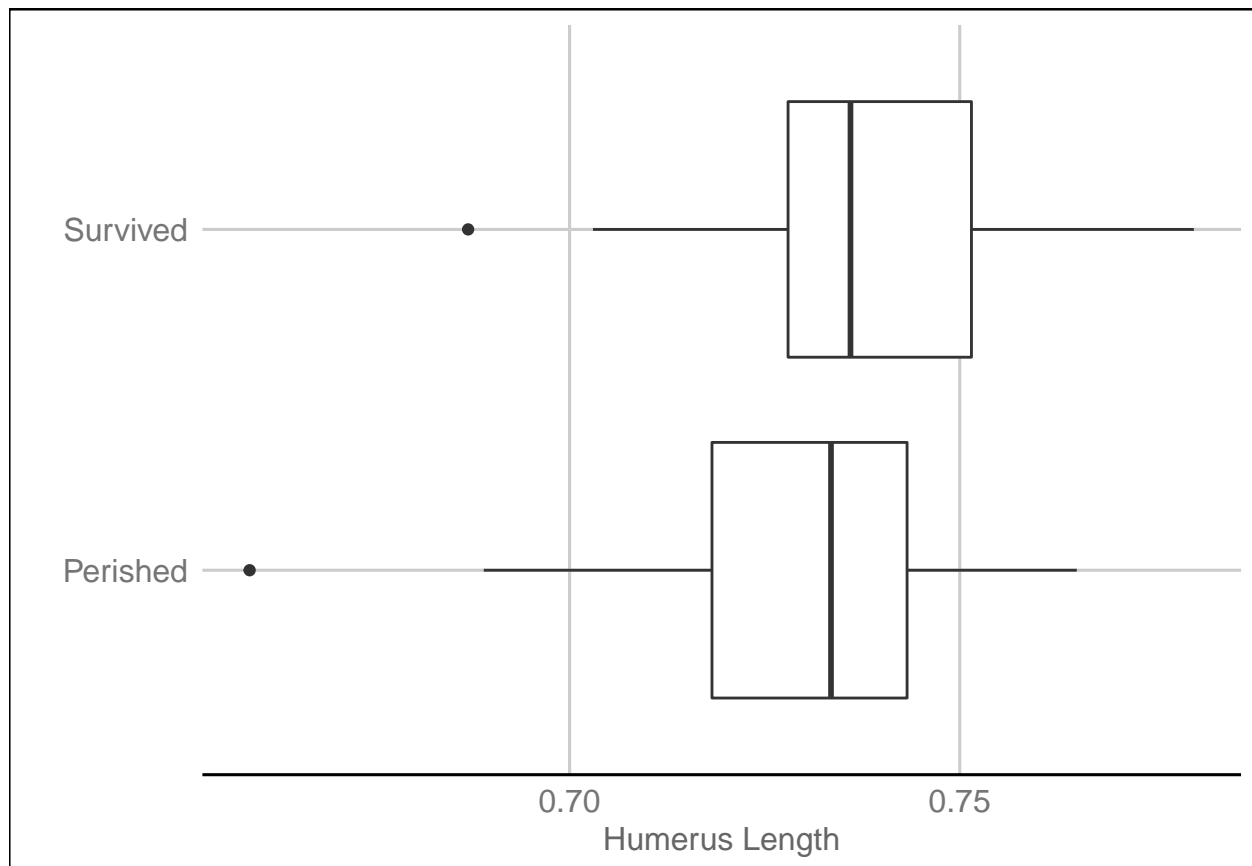
A *p*-value for the difference in humerus lengths with all observations is 0.092, and without the smallest observation in the perished group (0.659), the *p*-value is 0.176. The *p*-value changes a fair bit (by 0.084), but even with all the observations included, there is not strong evidence that the difference in means is not equal to zero. If there were a more drastic difference such that one of these tests *did* show strong evidence of a difference in means, then the removed observation should be double-checked for accuracy.

3.30

A boxplot of the data:
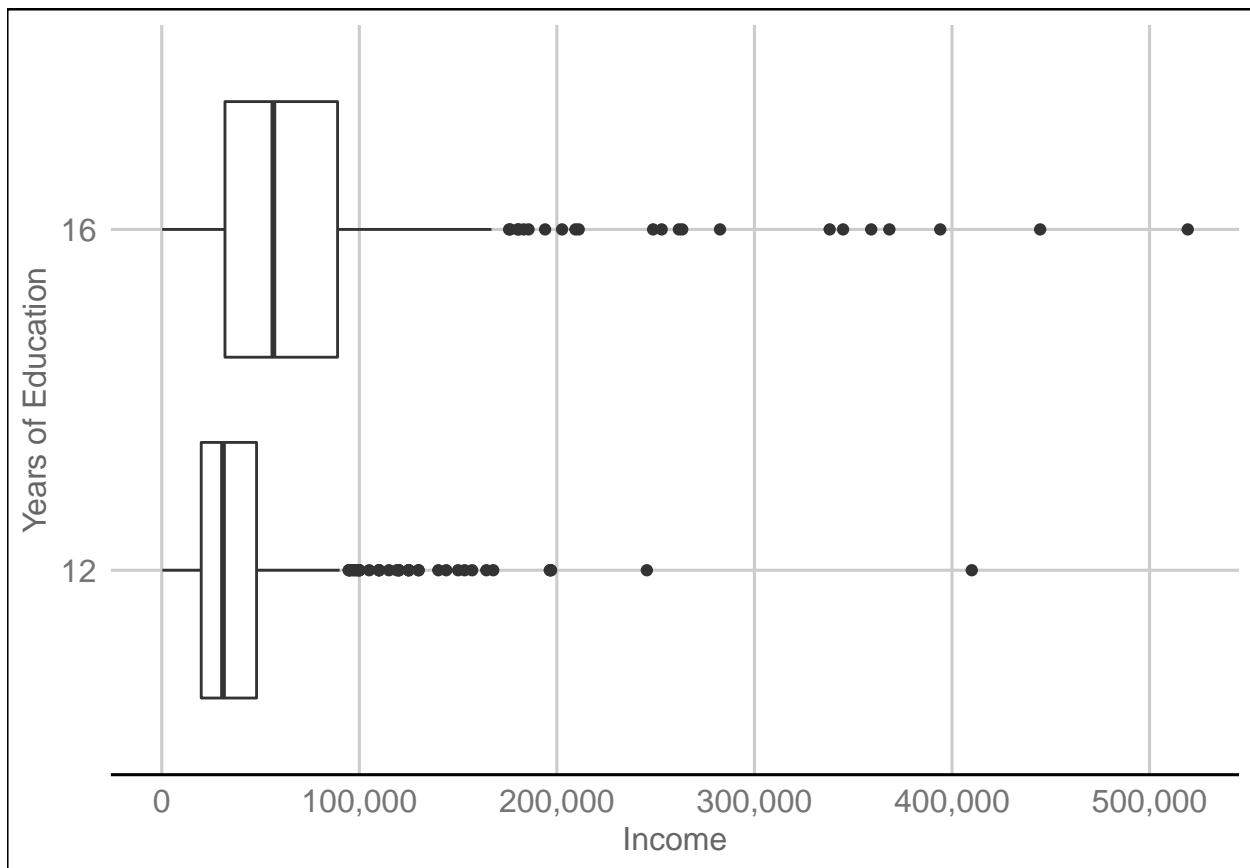
This data is highly skewed in the positive direction, so we look at a log transform as well:
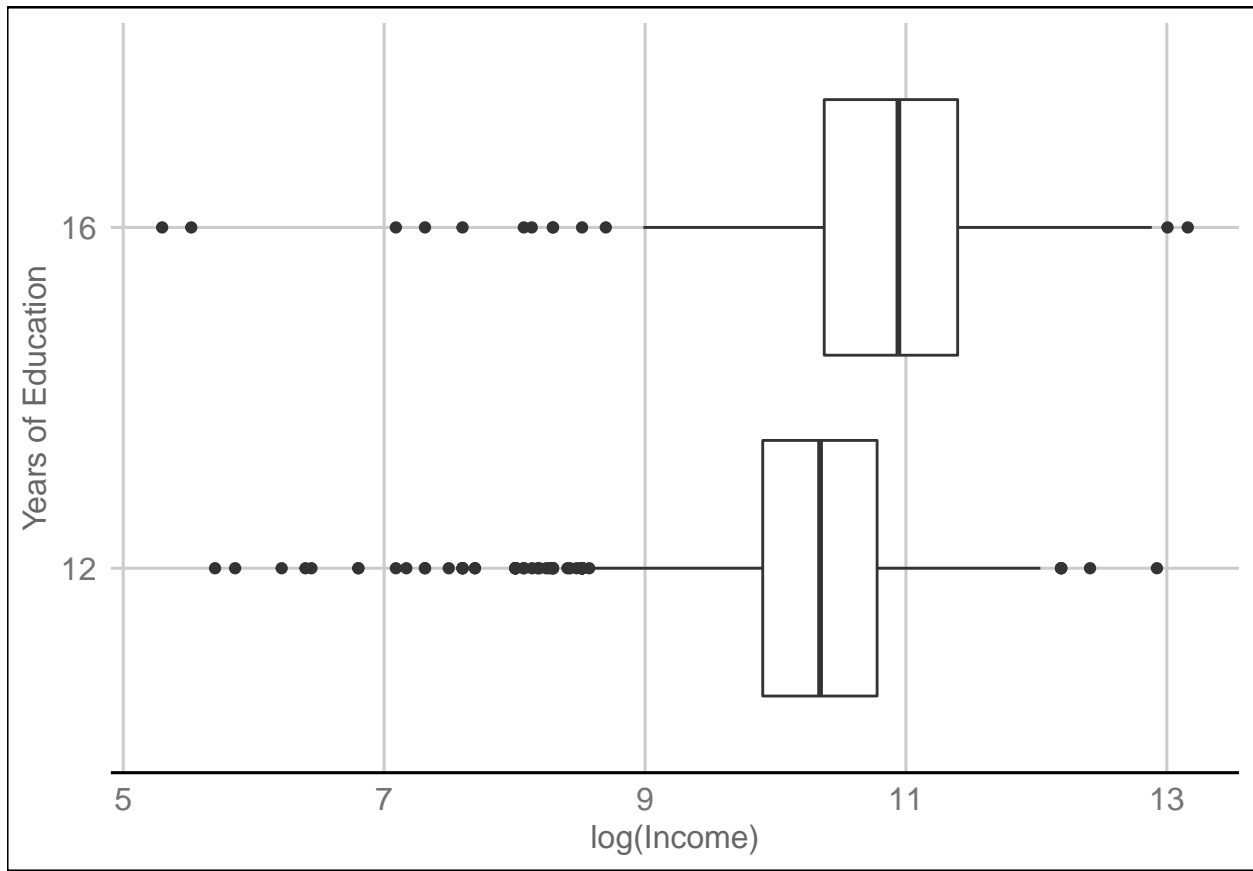
Running a *t*-test on this transformed data gives a *p*-value of 0, with a 95% confidence interval of 0.57 ± 0.107. Restating in terms of the original data, the median income for those with 16 years of education is between $e^{0.463} = 1.59$ and $e^{0.677} = 1.97$ times as much as those with 12 years of education.