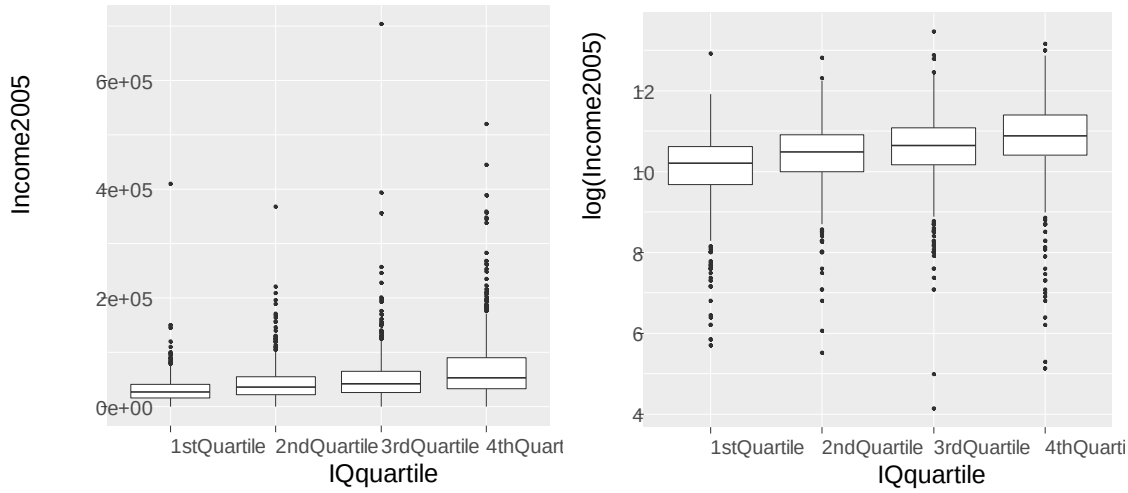
ST 411/511
Homework 5 Answers

1. Problem 24, pages 146-7

(a) Boxplots.




(b)

**Without log transform:**

Analysis of Variance Table

Response: Income2005

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| IQquartile | 3 | 4.9337e+11 | 1.6446e+11 | 82.442 | < 2.2e-16 *** |
| Residuals | 2580 | 5.1466e+12 | 1.9948e+09 | | |

---

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

**With log transform:**

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

(c) **Without log transform:** *Statistical Summary* The estimated difference in mean 2005 income between the 4th and 1st quartiles of IQ test scores is $37,892.55 (95% confidence interval $33,019.51 to $42,765.59).

*Scope of Inference* Causation can't be inferred because IQ score is not randomly assigned. Subjects were a random sample of American men and women, so this is the population to which inference can be made.

Justifications for not using a log transformation include not wanting to complicate the interpretation of the results and doubts about the logged populations being symmetric.

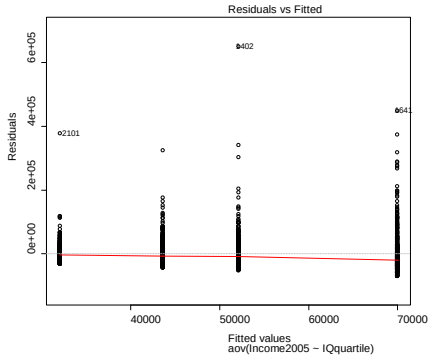
**With log transformation:** *Statistical Summary* The Median 2005 salary for the 4th quartile IQ scorers is estimated to be 108% (or 2.08 times) that of 1st quartile IQ scorers (95% confidence interval for the ratio of medians is 1.88 to 2.31).

*Scope of Inference* Causation can't be inferred because IQ score is not randomly assigned. We don't have information about how the sample subjects were selected, so it's not clear what population we can make inference about.

Justifications for using a log transformation include observing that the sample distributions are right-skewed, and the larger sample standard deviations are associated with larger sample means.

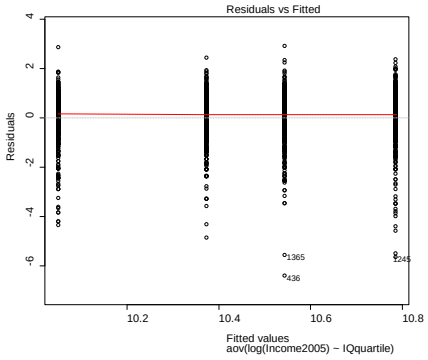
(d)

**Without log transformation:**




Without the log transformation, the residuals do not appear symmetric about 0, and the spread appears to be increasing with fitted value.

**With log transformation:**




With the log transformation, the residuals appear slightly skewed to the left, and the spread looks very similar among the four groups.