

# Homework 4 Solutions

Statistics 411: Spring 2018, 25 points

1. (12 pts total) Regarding Exercise 23 on page 56:

- (a) (1 pt) Because we are focusing on a single population (states in the US that increased the speed limit in 1996 from 55mph), then of the  $t$ -tools presented as choices, a 1-sample  $t$  tool is appropriate.
- (b) (1 pt) The Appendix presents side-by-side boxplots that compare the percent change in fatalities between the 32 American states that increased their speed limit from 55mph in 1996 to the 18 states that did not. The boxplots show a slight decrease in the median percent change in fatalities in the states that increased the speed limit compared to those that did not. This change looks to be small compared to the spread in the data, so the boxplots fail to suggest there's a real decrease.
- (c) (4 pts total) The 6 steps of the hypothesis test of whether the percent of fatalities increased just among the states that increased the speed limit from 55mph are:

- i. (1 pt) The **statistical hypotheses** regard  $\mu$ , the true mean percent change in fatalities just among the states that increase the speed limit from 55mph:

$$H_0 : \mu = 0$$

$$H_a : \mu > 0$$

- ii. (2 pts) **Assumptions:**

- Other states or other countries may be keen to see how fatalities changed in these 32 states to help decide whether to increase their speed limits. For this population of all states or countries, these 32 states are not a RS: they are all American states and the speed limits were changed in a single year, 1996. One would also expect that rural states like Montana, Idaho, Wyoming, and Washington would have more similar traffic fatalities as opposed to more densely populated states with many urban centers like California, New York and Florida.
- The sample size of  $n = 32 \geq 30$  states is large enough for us to assume that the sample mean has a normal distribution, so the  $t$ -test is appropriate to get a reliable CI even if the data are not normal.

iii, iv. and v (1 pt) The two-sided 90%  $t$ -CI is [-1.9% 2.9%] for the true mean percent change in fatalities just among the states that increased the speed limit from 55mph. See the Appendix for details. Because 0 is in the 90% two-sided CI, we FAIL TO REJECT  $H_0$ .

- (d) (2 pts) **Conclusion:** The evidence fails to suggest that increasing the speed limit is associated with an increase in the percent of traffic fatalities. At 90% confidence, the true percent change in fatalities is between -1.9% and 2.9% after an increase in the speed limit from 55mph.
- (e) (4 pts) **Scope of inference:** Because the 32 states in our data set were all from the United States (and not randomly selected from states or countries around the planet), it is unclear what population of states or countries that the results from these data can be generalized to. These results were from an observational study of the states that chose to increase their speed limits in 1996. Hence these data only failed to show an association (as opposed to causality) between increasing speed limits and an increase in fatalities.

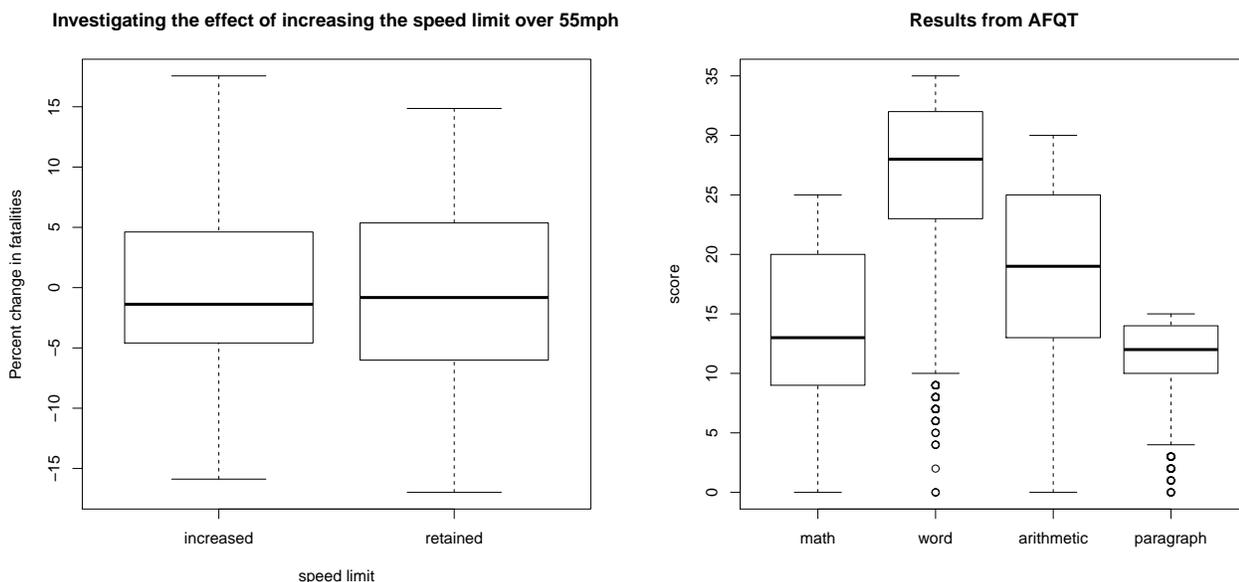
2. (13 pts total) Regarding Exercise 22 on page 55:

- (a) (1 pt) The Appendix presents side-by-side boxplots that compare the 4 component test scores of the AFQT. The boxplots suggest a very large median “word knowledge” score compared to the median “math knowledge” score. In fact, the 25th percentile (i.e., 1st quartile) of the word scores is larger than the 75th percentile (i.e., 3rd quartile) of the math scores!
- (b) (1 pt) To compare the mean “word knowledge” scores to the mean “math knowledge” scores, a  $t$ -test is more appropriate than a  $z$ -test because we have no knowledge regarding the SD of the word scores ( $\sigma_{\text{word}}$ ), the SD of the math scores ( $\sigma_{\text{math}}$ ), or of the SD of the difference ( $\sigma_{\text{word-math}}$ ). We will estimate  $\sigma_{\text{word-math}}$  from the data.
- (c) (1 pt) To compare the mean “word knowledge” scores to the mean “math knowledge” scores, a 2-sample paired  $t$ -test is appropriate because we have (paired) word and math scores for each individual in the data set.
- (d) and (e) (4 pts total) The 6 steps of the hypothesis test of whether young adults’ math knowledge is less than their word knowledge on the average:

- i. (1 pt) The **statistical hypotheses** regard  $\mu_{\text{word}}$  and  $\mu_{\text{math}}$ , the true mean scores for the word and math tests respectively:
- $$H_0 : \mu_{\text{word}} - \mu_{\text{math}} = 0$$
- $$H_a : \mu_{\text{word}} - \mu_{\text{math}} > 0$$
- ii. (2 pts) **Assumptions:**
- The problem states that we have a RS from NLSY of American males and a RS of American females between 16 and 24 years old in 1981 who took the AFQT. Because there is roughly a 50-50 split of males and females in the American population, putting these two RSs together as we do will not over-represent males or females in the sample compared to the population. We will assume, as does the *Sleuth*, that we have a RS of young Americans from NLSY between 16 and 24 years old in 1981.
  - Each individual has word and math scores, so clearly the word scores are not independent of the math scores. However, the differences in word and math scores are independent across the individuals because we have a RS of individuals from NLSY.
  - The sample size of  $n = 2584 \geq 30$  Americans is large enough to support the assumption that the mean difference in math and word test scores is normally distributed, so we can use the  $t$ -test without having to assume normality of the differences in word and math scores.
- iii. iv. and v. (1 pt) The sample mean word score is 12.4 points larger than the mean math score. The test statistic value of -114.7 (see the Appendix for details) means that the mean difference of 12.4 is 114.7 SEs away from the hypothesized difference in  $H_0$   $\mu_{\text{word}} - \mu_{\text{math}} = 0$ . This is a huge disparity between the data and  $H_0$ . The associated  $p$ -value  $< 2.2 \times 10^{-16} < \alpha = 0.01$ , so we REJECT  $H_0$ .
- (f) (2 pts) **Conclusion:** The evidence suggests that young Americans' scores on the mean word score on the AFQT is at least 12 points larger at 99% confidence than the mean math score ( $t = -114.7$ ,  $p$ -value  $< 2.2 \times 10^{-16}$ ).
- (g) (4 pts) **Scope of inference:** As your book suggests, if these young adults are a RS from Americans in NLSY in 1981, then we conclude that, at 99% confidence, the population of all Americans in NLSY in 1981 tend to score at least 12 points higher, on average, on word knowledge compared to math knowledge. This was an observational study, so we make no claims that being a young American in NLSY caused this disparity in test scores. In fact, I would guess that this result for young Americans in NLSY in 1981 is similar to word and math knowledge of the general American population today; however, it would be a stretch to rely only on these data to support that guess.

## Appendix

In this first figure, the boxplot in the pane on the left compares the percent change in fatalities between the American states that increased their speed limit from 55mph in 1996 to the states that did not. The boxplot in the pane on the right compares the scores for the 4 component tests of the military's AFQT taken by the individuals in the sample.



## R-code and R-output

```
#####  
# PROBLEM 1  
library(Sleuth3)  
d=ex0222  
  
# Plot the data  
boxplot(d$Math,d$Word,d$Arith,d$Parag,  
        names=c("math","word","arithmetic","paragraph"),ylab="score",  
        main="Results from AFQT")  
  
# Data analysis  
t.test(s[s$SpeedLimit=="Inc"],$PctChange,conf.level=.9)  
##  
##          One Sample t-test  
##  
##data:  s[s$SpeedLimit == "Inc", ]$PctChange  
##t = 0.345, df = 31, p-value = 0.7324  
##alternative hypothesis: true mean is not equal to 0  
##90 percent confidence interval:  
## -1.932586  2.920086  
##sample estimates:  
##mean of x  
##  0.49375  
  
#####  
# PROBLEM 2  
d=ex0222  
  
# Plot the data  
boxplot(d$Math,d$Word,d$Arith,d$Parag,  
        names=c("math","word","arithmetic","paragraph"),ylab="score",  
        main="Results from AFQT")  
  
# Data analysis  
t.test(d$Math,d$Word,paired=T,alternative="less",conf.level=0.99)  
##  
##          Paired t-test  
##  
##data:  d$Math and d$Word  
##t = -114.687, df = 2583, p-value < 2.2e-16  
##alternative hypothesis: true difference in means is less than 0  
##99 percent confidence interval:  
##      -Inf -12.11056  
##sample estimates:  
##mean of the differences  
##          -12.36146
```