

Homework 3A

STAT 242: Intermediate Statistics

Don't forget to load any packages using `library()` that you may need for this homework!!

Problem 1: Adapted from *The Statistical Sleuth*, Problem 5.25

The R code below reads in data with annual incomes as of 2005 for a random sample of 2584 Americans who were selected for the National Longitudinal Survey of Youth in 1979 and who had paying jobs in 2005. The data set also includes a code for the number of years of education that each individual had completed by 2006: <12, 12, 13-15, 16, or >16.

```
library(Sleuth3)
income <- ex0525

income$Educ <- as.factor(income$Educ)
```

- (a) Make a suitable plot of the data, showing the distribution of incomes in 2005 (Income2005) separately for each level of education (Educ). Based on your data, how should you proceed with your analysis? (3 points)
- (b) Do the data provide evidence that at least one of the five groups has a different mean (of the square root of) income than the other groups? Conduct a relevant hypothesis test, stating our hypotheses with clearly defined group means and written sentences explaining what each hypothesis means in context. Be sure to interpret your results in context! (6 points)
- (c) Do the data provide evidence that there is a difference in mean incomes for people with a high school education or less (12, <12) and people with at least some undergraduate education (13-15, 16, >16)? Conduct a relevant hypothesis test, clearly stating your hypotheses in terms of equations involving some of the group means and written sentences explaining what each of the hypotheses means in context. Be sure to interpret your results in context! (5 points)
- (d) Given that you conducted two hypothesis tests total, do you need to adjust for multiple comparisons? Justify your answer. (Hint, look at the class notes on this.) (1 point)