Homework 5

Shun-Lung Chang, Dilip Hiremath

1. First of all, load the data frame Wage from the library ISLR. You start out with a close look at wage differences between the two health levels.

```
data(Wage, package = "ISLR")
```

(a) (1.5 points) Compute mean and standard deviation of wage for each health level separately. Summarize the result in an English sentence.

(b) (1 point) Compute the standard errors for the mean wages in the two groups.

2. (2.5 points) Create a plot showing the mean wages for the two groups and corresponding error bars, i.e. add lines of length one standard error of the mean to both sides of the mean.

3. (2.5 points) Using an appropriate statistical procedure, test whether average wage is the same for workers with health level "1. at most Good" and workers with health level "2. at least Very Good". Formulate the null and alternative hypothesis and report the results in an English sentence refering to the relevant numbers.

4. Plot a box plot of the workers raw wage (variable wage using the education level (variable education) as grouping.

(a) (half a point) Are half of the wages for workers who have less than a high school degree below the first quartile of the wage for workers with some college degree?

(b) (half a point) Do half of the workers with a HS degree have higher wages than three quarters of the high school dropouts?

(c) (half a point) The minimum wage of workers with advanced degree is larger than the median wage of high school dropouts?

(d) (half a point) The interquartile range differs substantially between all groups.

(e) (half a point) Spread as measured by the length of the whiskers differs substantially between all groups.

5. You want to assess the wage difference between educational groups. Before you run the appropriate statistical test, you check some of the assumptions for ANOVA. In particular, you assess homoscedasticity.

(a) (1.5 points) Looking at the boxplot in Question ?? does homoscedasticity hold for the five groups? Give reasons for your answer!

(b) (1 point) Select a suitable variance test to check on this. Does the test confirm homoscedasticity?

6. Now, you assess the wage difference between educational groups using a statistical test.

(a) (1 point) Using an appropriate statistical test check whether wages are equal across education groups. Report the result in a complete English sentence including the relevant numbers!

(b) (1 point) From the ANOVA table derive the total sum of squares for wages and compare this result with the variance of wage when multiplied by 2999.

(c) (half a point) Which proportion of total variation in wages is due to the group differences in education?

7. Having found an overall difference, you now want to use a post-hoc test with Holm correction, to asses which marital status groups do actually differ
significantly in wages?