Take Home Exam 1

BUS 735: Business Decision Making and Research

Instructor: James M. Murray, Ph.D.

Fall 2017

Directions

Type up your answers to the questions that follow and upload your submission to the appropriate D2L dropbox folder. Please include the following in every answer:

- R code you used.

- R output.

- Description of the answer to the question.

Reporting only the code and output without a descriptive answer to the questions will not earn any points. Submit your work in a word processing document or PDF document.

Data

The dataset `cps2017.RData` includes the variables related employment, education, and health insurance for 1,207 workers in 2017. Download and load the data set with the following command:

```
load(url("http://murraylax.org/datasets/cps2017.RData"))
```

1. Level of Education: Ordinal scale, variable name: `edu`

```
class(df$edu)
```

```
## [1] "ordered" "factor"
```

```
levels(df$edu)
```

```
## [1] "Less than high school"          "High school degree or equivalent"
## [3] "Some College"                   "Four Year College Degree"
## [5] "Advanced College Degree"
```

2. Classification of worker: Nominal scale, type of employer, variable name: `class`

```
class(df$class)
```

```
## [1] "factor"
```

```
levels(df$class)
```

```
## [1] "Employed by Federal Government"
## [2] "Employed by private company"
## [3] "Employed by State/Local Government"
```

3. Usual weekly hours: Ratio scale, number of hours usually worked per week, variable name: `UsualWeeklyHours`

4. Dummy variable for paid by the hour: Binary variable, equal to 1 if worker is paid by hour, 0 otherwise, variable name: `PaidByHour`

5. Union coverage: Nominal scale, description of union membership and coverage, variable name: `Union`

```
class(df$Union)
```

```
## [1] "factor"
```

```
levels(df$Union)
```

```
## [1] "No union coverage"              "Union coverage but not a member"
## [3] "Union member"
```

6. Union membership: Binary variable, equal to 1 if a member of a union, 0 otherwise, variable name: `UnionBin`

7. Weekly earnings: Ratio scale, usual weekly earnings from wages and salary in dollars, variable name: `WeeklyEarnings`

8. Usual hourly earnings: Ratio scale, usual hourly earnings from wages and salary in dollars, variable name: `UsualHourEarnings`

9. Health insurance coverage: Nominal scale, description if and how person is covered by health insurance, variable name: `HealthInsurance`

```
class(df$HealthInsurance)
```

```
## [1] "factor"
```

```
levels(df$HealthInsurance)
```

```
## [1] "Employer Insurance"      "Medicaid"
## [3] "Medicare"                "Military Health Insurance"
## [5] "Not covered by insurance"  "Private Insurance"
```

Problems

1. Answer the following questions investigating the relationship between usual hourly earnings and usual number of hours per week?

   a. Illustrate the relationship with a scatter plot with a best-fit linear function. Put usual weekly hours on the horizontal axis and usual hourly earnings on the vertical axis. To make the best use of the space in the scatter plot, zoom in on the area where usual weekly hours is between 0 and 80 and usual hourly earnings is between \$0 and \$100.

   b. Assuming a linear relationship is valid, test the hypothesis that usual weekly hours and usual hourly earnings are correlated.

   c. Assuming a monotonic relationship other than a linear relationship is valid, test the hypothesis that usual weekly hours and usual hourly earnings are correlated.

2. Answer the following questions regarding a linear regression that predicts usual hourly earnings based on usual weekly hours.

   a. Estimate and report the linear regression line.

   b. What is the predicted impact on usual hourly earnings for working one additional hour per week?

   c. Test the hypothesis that hours per week has a positive effect on hourly wage.

3. Show a bar plot with means and error bars for confidence intervals (using the normal distribution) for the usual hourly earnings for different classifications of workers. What classification that has the highest mean usual hourly earnings. Use the visual to determine if this mean is statistically significantly different from the others. Explain your answer.

4. Is there evidence of a relationship between level of education and classification of workers? Test the appropriate hypothesis.

5. Show a bar plot with mean usual hourly earnings broken down by education and classification. Are there some classifications with larger differences in pay due to education. Describe the differences.

6. Is there evidence that the median usual hourly earnings is different for union workers versus non-union workers? Test the appropriate hypothesis.

7. Construct and interpret 95% confidence intervals for the medians for each union workers and non-union workers. Which types of workers have a higher median hourly earnings.

8. Is there a relationship between classification of worker and health insurance coverage? Test the appropriate hypothesis. If there is statistical evidence, describe the differences in health insurance coverage by classifications of workers.