**Lab 4: One and two sample difference tests; ANOVA; Goodness of fit tests**

Geog 4300/6300-Shannon

Fall 2016

Value: 20 points

1. ***(4 points)*** You’re interested in knowing how uniform the income distribution is across states for people responding to the Current Population Survey data you used in Lab 3 (find the data in that folder). Using that data, look at three indicators of normality for the average weekly income (EARNWEEK) for **each of** **two states of your choosing**. Exclude any responses listing “9999.99” as the value.
   1. a Shapiro-Wilk normality test
   2. a Q-Q plot with a line showing the normal distribution
   3. a histogram

Copy and paste the script you use and the output from R. Interpret what they tell us about the normality of that variable for each state.

1. ***(4 points)*** Based on the results of question one, test whether the incomes of survey respondents in these two states were equal or different. Provide the following information with your answer:
2. ***Appropriate*** measures of central tendency and distribution for both states (e.g., mean, median, range, IQR, st. dev.--whatever fits the data based on question 1)
3. H0 and HA
4. The appropriate test
5. The results of your test
6. Your substantive interpretation of what you found, including whether you reject the null hypothesis.
7. ***(4 points)*** The number of minor, moderate, and severe storms were measured for five regions over ten years. The results are listed below. Using a chi-square test, evaluate whether there are any significant differences in the distribution of storm severity across regions. Include your null & alternative hypothesis, your work/R commands, your results, and your interpretation of these results.

|  |  |  |  |
| --- | --- | --- | --- |
| Region | Minor | Moderate | Severe |
| 1 | 8 | 10 | 11 |
| 2 | 10 | 11 | 11 |
| 3 | 13 | 18 | 15 |
| 4 | 11 | 17 | 12 |
| 5 | 10 | 8 | 9 |

4. ***(8 points)*** The table below provides snowfall amounts from 12 reporting stations in southern Illinois for the winters of 1995, 2000, and 2005. Conduct an ANOVA analysis on this data (including post-hoc tests) to determine if and how snowfall may have changed during this time with ***90% confidence.***

You should be able to copy and paste this table into Excel and then import it into R if you would like. It is also available for download in the lab folder.

Report the following:

* Null and alternate hypothesis
* Critical F statistic
* Results of your ANOVA analysis
* Results from post-hoc testing

Summary and interpretation of these results (about 1 paragraph.)

|  |  |  |
| --- | --- | --- |
| Station | Snow | Year |
| 1 | 9.8 | Y1995 |
| 2 | 8.1 | Y1995 |
| 3 | 7.7 | Y1995 |
| 4 | 7.1 | Y1995 |
| 5 | 8.6 | Y1995 |
| 6 | 7.9 | Y1995 |
| 7 | 8.1 | Y1995 |
| 8 | 9.5 | Y1995 |
| 9 | 8 | Y1995 |
| 10 | 8.7 | Y1995 |
| 11 | 9.2 | Y1995 |
| 12 | 7.3 | Y1995 |
| 1 | 8.8 | Y2000 |
| 2 | 6.8 | Y2000 |
| 3 | 7.6 | Y2000 |
| 4 | 6.4 | Y2000 |
| 5 | 7.8 | Y2000 |
| 6 | 7.2 | Y2000 |
| 7 | 8.5 | Y2000 |
| 8 | 9 | Y2000 |
| 9 | 7.1 | Y2000 |
| 10 | 8.8 | Y2000 |
| 11 | 8.9 | Y2000 |
| 12 | 6.4 | Y2000 |
| 1 | 8.7 | Y2005 |
| 2 | 6.5 | Y2005 |
| 3 | 7.7 | Y2005 |
| 4 | 6.5 | Y2005 |
| 5 | 5.9 | Y2005 |
| 6 | 6.9 | Y2005 |
| 7 | 8.2 | Y2005 |
| 8 | 7.5 | Y2005 |
| 9 | 6.5 | Y2005 |
| 10 | 8.3 | Y2005 |
| 11 | 5.2 | Y2005 |
| 12 | 6.5 | Y2005 |