SMML Project 1

John Kubale

9/3/2024

- Note: For each question, include R code and output *pertinent* to your answers.

1. Use `hprice` in `faraway` package.

- The data includes 324 observations coming from 36 US metropolitan statistical areas (MSAs) over 9 years from 1986-1994: $36 \times 9 = 324$
- Assume that the MSAs in the data are a simple random sample of the population of MSAs in the US. See https://www2.census.gov/geo/maps/metroarea/us_wall/Mar2020/CBSA_WallMap_Mar2020.pdf for MSAs.
- Refer to the R manusal for `faraway` for the background information about this dataset as well as variable definitions.
- The housing sale price is the outcome variable of interest. Because the data set has a natural log transformed price variable, `narsp`, we recode this to create `homeprice` by transforming `narsp` back to the dollar unit for an easier interpretation as follows:

```
hprice$homeprice<-exp(hprice$narsp)*1000
```

1) What are the mean and the variance of `homeprice`? What do they mean?

2) Construct a 95% confidence interval of the average `homeprice`. What does the confidence interval imply?

3) Estimate the average `homeprice` by whether the MAS was adjacent to a coastline, noted in `ajwtr`, and the standard errors.

4) Test the difference in `homeprice` between coastline MSAs and non-coastline MSAs. Clearly state the formula for the hypothesis, the test method and your rationale for selecting the method. What do you conclude about the hypothesis?

5) Estimate the Pearson correlation coefficient between `homeprice` and per capita income of the MSA of a given year, noted in `ypc`.

6) Test whether the correlation coefficient between `homeprice` and `ypc` is 0 or not.

Clearly state the hypothesis including the formula. What do you conclude?
7) Can you say that per capita income has an effect on the home sales price using the results from #6)? Why or why not?

8) Test the normality of `homeprice`. Would this test result change your responses to #1) to 7)? Why or why not?