STOR 664 Homework 1, Fall 2023

Richard L. Smith

August 24, 2023

Due: Thursday August 31. Hand in on gradescope no later than 11:59 pm, August 31, 2023.

These assignments use the NEW versions of the Amherst and Mount Airy/Charleston datasets, available on

http://rls.sites.oasis.unc.edu/faculty/rs/source/Data/amh1.csv

and

http://rls.sites.oasis.unc.edu/faculty/rs/source/Data/mta1.csv

Theory Part. Well, not really "theory" in the sense or proving new mathematical results, but I'd like you to work through these parts directly, evaluating $\bar{x}$, $\bar{y}$, $\sum(y_i-\bar{y})^2$, $\sum(x_i-\bar{x})(y_i-\bar{y})$, etc. in order to calculate $\hat{\beta}_0$, $\hat{\beta}_1$, the residual variance s^2, etc., using the direct formulas given in the text rather than by using the regression commands in R or any other language (it's fine if you use R or Excel as a calculator to evaluate these numerical quantities quickly).

In each case assume the model

$$y_i \quad = \quad \beta_0 + \beta_1(x_i - \bar{x}) + \epsilon_i$$

where the ϵ_i satisfy either of the assumptions (a) or (b) in the text.

1. For the Amherst data, find (a) $\hat{\beta}_0$ and $\hat{\beta}_1$, (b) the residual variance estimator s^2, (c) the standard errors of $\hat{\beta}_0$ and $\hat{\beta}_1$, (d) formally test $H_0:\ \beta_1 = 0$ against the alternative $H_1:\ \beta_1 \neq 0$. Assume the x variable is Year, the y variable is annual mean temperature in Amherst, and that the standard assumptions are correct.

2. For the Mount Airy/Charleston data, find (a) $\hat{\beta}_0$ and $\hat{\beta}_1$, (b) the residual variance estimator s^2, (c) the standard errors of $\hat{\beta}_0$ and $\hat{\beta}_1$, (d) formally test $H_0:\ \beta_1 = 0$ against the alternative $H_1:\ \beta_1 \neq 0$. Assume the x variable is summer mean temperature in Mount Airy, the y variable is summer mean temperature in Charleston, and that the standard assumptions are correct.

   R Part. Now repeat the same calculations in R, using the `lm` command. Typical commands will be of the form

```
lm(y~x)
summary(lm(y~x))
# or
summary(lm(y~x1+x2)) # if there are two covariates x1 and x2
```

after defining the variables x and y.

3. For the Amherst data, verify the results of part 1 and extend also to the model in which Year^2 is used as a covariate as well as Year. (Hint: you may want to rescale first, e.g. replace Year by Year/1000.) Draw a scatterplot of temperature against year, and show both fitted curves on the plot (you may find the `lm(...)$fitted` command useful for this). What do you conclude about the possibility that the trend may be nonlinear?

4. For the Mount Airy/Charleston data, verify the results of part 2. Also consider (separately for Mount Airy and Charleston) whether there appears to be a time trend against Year, and state your conclusions from that.