HOMEWORK 3
MATH 463 - SPRING 2017

ALEX THIES

INTRODUCTION. I collaborated with Sienna Allen, Joel Bazzle, Torin Brown, Andy Heeszel, Ashley Ordway, and Seth Temple on this assignment. Due to family and job related time constraints I was unable to complete 4.5.12, 4.5.13, and 4.5.14.

1. PROBLEM 1

1.1. **Assignment.** The goal [of this portion of the assigment] is to understand the relationship of Mortality to the other variables.

Note that NOx and NOxPot are identical, so exclude one (why?)

The variables can be divided into demographic and climate variables. Is there significant evidence that the climate variables should be included in modeling mortality? Explain any test that you perform.

Give a confidence interval for the expected Mortality in Indianapolis. What assumptions are you making to guarantee the confidence level of this interval? Are these assumptions reasonable? Is this interval useful? What is the source of uncertainty, if any, about the mortality rate in Indianapolis?

Give a confidence interval for the coefficient of NOx. Does the length of this interval depend on which other variables you include in the model? Discuss.

Suppose all variables are included except NOxPot, and you want to test that the coefficients of NOx and Education are both zero. Estimate the power of the appropriate test of this hypothesis, when the coefficients are 1 and -10, respectively.

If the power is low, discuss why.

1.2. **Report.**

1.2.1. *F-Test on climate variables as a small model.* Let the climate variables be $\beta_{r+1}, \beta_{r+2}, \ldots, \beta_k$. We ignore one of NOx and NOxPot because they are linearly dependent, and thus do not belong in the same linear span. Note that we also ignore one of the rows of data[1] because it contains some `NULL` entries. We test the null hypothesis that $\beta_{r+1} = \beta_{r+2} = \cdots = \beta_k = 0$, by perform an F test using the following statistic, and level $\alpha = 0.05$,

$$F = \frac{(RSS_0 - RSS_1)/(k-r)}{RSS_1/(n-k)}$$

Note that in order for this test to work, we must have the underlying assumption that the residuals are normally distributed.[2] To carry out this test in R, we fit two models `f1` and `f2`, and use `anova` to compute the F-statistic; Table 1 illustrates our

[1] Fort Worth.

[2] This is true by the Central Limit Theorem.

result. We see that $F = 7.53$, which we compare with the critical value $f^* = 0.3$. Observe that $0.3 < 7.53$, or that our p-value of $p = 5.97 \times 10^{-6}$ is considerably less than any level α, either way we have sufficient evidence to reject the null hypothesis that the climate variables should not be included in the model.

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	51	118228.93				
2	44	53806.71	7	64422.21	7.53	0.0000

TABLE 1. F Test of climate variables

1.2.2. *Confidence Interval for $Mort_{Indy}$.* We compute a confidence interval in R for $\mathbb{E}(Y_{\text{Mort}_{\text{Indy}}})$,

```
CI.indyMort <- subset(predict(f1,interval = 'confidence'),
                      row.names(
                          predict(f1, interval = 'confidence'))
                      == 'Indianapolis, IN')
```

Thus, we have the confidence interval $(941.5, 980.57)$. In order to compute this interval we rely on the OLS assumptions, namely that the residuals are centerred about 0, independent from the realized values of $\boldsymbol{X}$, and normally distributed. This interval could be used to test whether or not the OLS assumptions are appropriate. We could determine the percentage of actual mortality rates which lie in their respective confidence intervals, with a high percentage indicating that the design matrix is well suited for our experiment, on the other hand a low percentage indicating that the design matrix needs some work because it is not accurately modelling the data. The first thing that we ought to check is whether or not the residuals are actually normally distributed, if they are not, then the critical value which is used to construct the confidence interval would be different. Any uncertainty in the measurements are either a result of statistical noise, or poor choices in variables.

1.2.3. *Confidence Interval for NOx.* We compute a 95% confidence interval for the coefficient β_{NOx} using a t-statistic in R,

```
alph1 <- 0.05
NOxtstar1 <- abs(qt(alph1,summary(f1)$df[2]))
NOxtstat1 <- subset(summary(f1)$coef,
                    row.names(summary(f1)$coef) == 'NOx')[1]
NOxSE1 <- subset(summary(f1)$coef,
                 row.names(summary(f1)$coef) == 'NOx')[2]
CI1.NOxlow <- NOxtstat1 - NOxtstar1*NOxSE1
CI1.NOxhigh <- NOxtstat1 + NOxtstar1*NOxSE1
```

Thus, we have the confidence interval,

$$(-0.36, 2.71)\,.$$

If we exclude some variables[3] we have the following confidence interval for NOx,

$$(-0.07, 4.02)\,.$$

[3]We exclude all non-climate variables, because we already have that model as an object in our R chunks.

Observe that this CI is about longer that the previous one, and centered about a different μ. In class it was mentioned that the big model is better than the small one because it not only considers the possibilities of the small model, but all others. I suspect that this is the reason why our CI gets larger, that is, the degree of confidence with which we assert that $\beta_{NOx} = \mu$ is less under the small model than the large one.

1.2.4. *Power test.* In order to compute the power of the test of the alternative hypothesis that $\beta_{\text{NOx}} = 1$ and $\beta_{\text{Ed}} = -10$ at level $\alpha = 0.05$. Note that the values for β_{NOx} and β_{Ed} are approximately their estimates under the OLS model. We must first compute the non-centrality parameter δ, we do this in R,

```
    NOxperp <- residuals(lm(NOx~.-Mortality-Education,
                            data = X))
    Edperp <- residuals(lm(Education~.-Mortality-NOx,
                           data = X))
    de <- ((Ed1^2)*sum(Edperp^2) + (NOx1^2)*sum(NOxperp^2)
           + (2*NOx1*Ed1)*sum(NOxperp*Edperp))/summary(f1)$sigma^2
```

Now we can compute the critical value, and power.

```
    Fstar2 <- qf(1-alph2, 2, anova1$Res.Df[2])
    power <- 1-pf(Fstar2, 2, anova1$Res.Df[2], ncp = de)
    power
## [1] 0.24
```

We see that the power is 0.24, which is quite low. Given that we used roughly our estimates under the large OLS model for this test, the low power indicates that we may not be including enough relevant (orthogonal) variables in the model.

2. Book Problems

Problem 4.5.2. In the OLS regression model, do the residuals always have mean 0? Discuss briefly.

Solution. True, observe that the standard model $\boldsymbol{Y} = \boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{\varepsilon}$, thus we can define the errors as $\boldsymbol{\varepsilon} = \boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}$. Recall that $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$, after substitution we have $\boldsymbol{\varepsilon} = \boldsymbol{Y} - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$. Let $\boldsymbol{H} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$, so $\boldsymbol{\varepsilon} = \boldsymbol{Y} - \boldsymbol{H}\boldsymbol{Y} = (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y}$. Observe that $\mathbf{1}^T\boldsymbol{\varepsilon} = \sum_{i=0}^{n} \boldsymbol{\varepsilon}_i = \langle \mathbf{1}, (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y}\rangle$. Note that the constant vector is in the linear span of $\boldsymbol{X}$, and that $(I - \boldsymbol{H})\boldsymbol{Y}$ is in the linear span orthogonal to $\boldsymbol{X}$, thus, $\sum_{i=0}^{n} \boldsymbol{\varepsilon} = \langle \mathbf{1}, (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y}\rangle = 0$. Therefore, $\sum_{i=0}^{n} \boldsymbol{\varepsilon}_i/n = 0$, which we aimed to show. □

Problem 4.5.3. True or false, and explain. If, after conditioning on $\boldsymbol{X}$, the disturbance terms in a regression equation are correlated with each other across subjects, then

(a) The OLS estimates are likely to be biased.
(b) The estimated standard errors are likely to be biased.

Solution.

(a) This is false by Theorem 2 on page 43 of the text.

(b) This is true. Suppose that there is correlation in the residuals, i.e., $y_i = x_i^T + \varepsilon_i$ and that $\text{Var}(\varepsilon) = s$. The OLS estimates take the form

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}, \\ &= \boldsymbol{\beta} + (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\end{aligned}$$

The variance of the estimates is

$$\begin{aligned}\text{Var}(\hat{\boldsymbol{\beta}}) &= \mathbb{E}[(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}], \\ &= (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{s}^T\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}.\end{aligned}$$

Because $\boldsymbol{s} \neq \sigma^2\boldsymbol{I}$ we see that the standard errors are biased in this case.

□

Problem 4.5.5. You are using OLS to fit a regression equation. True or false, and explain:

(a) If you exclude a variable from the equation, but the excluded variable is orthogonal to the other variables in the equation, you wont bias the estimated coefficients of the remaining variables.
(b) If you exclude a variable from the equation, and the excluded variable isnt orthogonal to the other variables, your estimates are going to be biased.
(c) If you put an extra variable into the equation, you won't bias the estimated coefficients – as long as the error term remains independent of the explanatory variables.
(d) If you put an extra variable into the equation, you are likely to bias the estimated coefficients – if the error term is dependent on that extra variable.

Solution.

(a) This is true, and follows from the lecture notes. Let $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ and $\theta = \boldsymbol{X}\boldsymbol{\beta} = \mathbb{E}[\boldsymbol{Y}]$. Suppose that $\boldsymbol{X} = [x_1, \ldots, x_k]$ and $V = \mathscr{L}\, L[x_1, \ldots, x_k]$. Thus, $\mathbb{E}[\boldsymbol{Y}] \in V$. Additionally, let us suppose that x_k is the orthogonal variable that we decide to exclude. Then, let us call θ_1 be the expectation of the OLS model excluding x_k with $V_0 = \mathscr{L}[x_1, \ldots, x_{k-1}]$ and θ_0 be the expectation of the OLS model without excluding x_k.

$$\begin{aligned}\theta &= \theta_0 + \theta_1 \\ &= (\beta_1 x_1 + \cdots + \beta_{k-1}x_{k-1} + \beta_k x_k).\end{aligned}$$

Notice that,

$$\begin{aligned}\theta_1 &= \theta - \theta_0, \\ &= \theta - \Pi_{v_0}\theta, \\ &= (\beta_1 x_1 + \cdots + \beta_{k-1}x_{k-1} + \beta_k x_k) \\ &\qquad - (\beta_1 x_1 + \ldots \beta_k x_k + \beta_k \Pi_{v_0} x_{k-1} + \Pi_{v_0} x_k), \\ &= \beta_k \Pi_{v_0} x_{k-1} + \Pi_{v_0} x_k.\end{aligned}$$

Because x_k is orthogonal to v_0, then $\theta_1 = \beta_k \Pi_{v_0} x_{k-1} = \boldsymbol{X}\boldsymbol{\beta}$. Thus, if the excluded variable is orthogonal, the estimated coefficients of the remaining variables are not biased.

(b) This is true and follows directly as a consequence of (a). If x_k is not orthogonal to v_0, then $\theta_1 = \beta_k \Pi_{v_0} x_{k-1} + \Pi_{v_0} x_k \neq \boldsymbol{X}\boldsymbol{\beta}$. So, θ_1 is biased in this case.

(c) This is true. Because ε_i is I.I.D., the assumptions of the OLS model hold. Thus, the estimated coefficients will not be biased.
(d) This is true. Because ε_i is not I.I.D. in this case, the assumptions of the OLS model do not hold. So, the estimated coefficients are likely to be biased.

□

Problem 4.5.9. True, or false, and explain:

(a) Collinearity leads to bias in the OLS estimates.
(b) Collinearity leads to bias in the estimated standard errors for the OLS estimates.
(c) Collinearity leads to big standard errors for some estimates.

Solution.

(a) False, the OLS assumes that the residuals are independent from the realized values of $\boldsymbol{X}$, under this assumption the estimates are never biased.
(b) False, see above.
(c) True, collinearity between estimates of coefficients makes the design matrix 'become more singular.' How I interpret that is that the design matrix is a basis for a linear space, therefore its columns are linearly independent. Collinearity makes the columns *almost* linearly dependent (what I've heard referred to as singular). When this happens the determinant for the matrix approaches zero, diving by this number which is approaching zero makes the standard error blow up, in some cases.

□

Problem 4.5.10. Suppose $(X_i, W_i, \varepsilon_i)$ are I.I.D. as triplets across subjects $i = 1, \ldots, n$, where n is large; $\mathbb{E}(X_i) = \mathbb{E}(W_i) = \mathbb{E}(\varepsilon_i) = 0$, and ε_i is independent of (X_i, W_i). Happily, X_i and W_i have positive variance; they are not perfectly correlated. The response variable Y_i is in truth this:

$$Y_i = aX_i + bW_i + \varepsilon_i.$$

We can recover a and b, up to random error, by running a regression of Y_i on X_i and W_i. No intercept is needed. Why not? What happens if X_i and W_i are perfectly correlated (as random variables)?

Solution. If X_i and W_i are not perfectly correlated then let $X_i = [x_1, \ldots, x_k]$ and $W_i = [w_i, \ldots, w_k]$. Thus, we have the design matrix $\mathbf{A}$,

$$\mathbf{A} = \begin{bmatrix} x_1 & w_1 \\ x_2 & w_2 \\ \vdots & \vdots \\ x_k & w_k \end{bmatrix}.$$

In this case, the design matrix does not require the constant vector of 1's because the x_i's and w_i's are independent. Thus, our design matrix has full rank and we can account for all observations. However, if X_i and W_i are correlated, then X_i and W_i are linearly dependent, i.e., $X_i = [x_1, \ldots, x_k]$ and $W_i = [mx_1 + b, \ldots, mx_k + b]$.

So in the case of linear dependence, we have a new design matrix,

$$\mathbf{B} = \begin{bmatrix} x_1 & a_1x_1 + b_1 \\ x_2 & a_2x_2 + b_2 \\ \vdots & \vdots \\ x_k & a_kx_k + b_k \end{bmatrix}.$$

Observe that $\boldsymbol{B}$ will not have full rank, so it will be impossible to estimate our coefficients because we will have infinitely many solutions. □

Problem 4.5.11. (This continues question 10.) Tom elects to run a regression of Y_i on X_i, omitting W_i. He will use the coefficient of X_i to estimate a.

(a) What happens to Tom if X_i and W_i are independent?
(b) What happens to Tom if X_i and W_i are dependent? Hint: see exercise 3B15.

Solution.

(a) If X_i and W_i are independent, then they are almost orthogonal. Thus, the estimated coefficient for X_i will not be affected by omitting W_i, we can regress Y_i onto X_i and the $\hat{\beta}$
(b) According to the text the case of dependence leaves Y subject to omitted-variable bias. This is loosely defined as picking up an effect of the omitted variable W. Ommitted-variable bias has not been discussed in lecture, so I am unable to expound on it more fully.

□

Problem 4.5.12. Suppose $(X_i, \delta_i, \varepsilon_i)$ are I.I.D. as triplets across subjects $i = 1, \ldots, n$, where n is large; and X_i, δ_i, ε_i are mutually independent. Furthermore, $\mathbb{E}(X_i) = \mathbb{E}(\delta_i) = \mathbb{E}(\varepsilon_i) = 0$ while $\mathbb{E}(X_i^2) = \mathbb{E}(\delta_i^2) = 1$ and $\mathbb{E}(\varepsilon_i^2) = \sigma^2 > 0$. The response variable Y_i is in truth this:

$$Y_i = aX_i + \varepsilon_i.$$

We can recover a, up to random error, by running a regression of Y_i on X_i. No intercept is needed. Why not?

Solution. □

Problem 4.5.13. (Continues question 12.) Let c, d, e be real numbers and let $W_i = cX_i + d\delta_i + e\varepsilon_i$. Dick elects to run a regression of Y_i on X_i and W_i, again without an intercept. Dick will use the coefficient of X_i in his regression to estimate a. If $e = 0$, Dick still gets a, up to random error – as long as $d \neq 0$. Why? And whats wrong with $d = 0$?

Solution. □

Problem 4.5.14. (Continues questions 12 and 13.) Suppose, however, that $e \neq 0$. Then Dick has a problem. To see the problem more clearly, assume that n is large. Let $Q = XW$ be the design matrix, i.e., the first column is the X_i and the second column is the W_i . Show that

$$Q^TQ/n = \begin{pmatrix} \mathbb{E}(X_i^2) & \mathbb{E}(X_iW_i) \\ \mathbb{E}(X_iW_i) & \mathbb{E}(W_i^2) \end{pmatrix}, \quad Q^TY/n = \begin{pmatrix} \mathbb{E}(X_iY_i) \\ \mathbb{E}(W_iY_i) \end{pmatrix}$$

(a) Suppose $a = c = d = e = 1$. What will Dick estimate for the coefficient of X_i in his regression?

(b) Suppose $a = c = d = 1$ and $e = -1$. What will Dick estimate for the coefficient of X_i in his regression?
(c) A textbook on regression advises that, when in doubt, put more explanatory variables into the equation, rather than fewer. What do you think?

Solution. □

E-mail address: athies@uoregon.edu