

HOMEWORK 2

MATH 463 - SPRING 2017

ALEX THIES

INTRODUCTION. I collaborated with Joel Bazzle, Torin Brown, Ashley Ordway, and Seth Temple on this assignment. The histograms, and scatter plots which are requested in Problems 1 and 2 are bunched together at the end of assignment; trying to make them appear in a neat manner throughout the document proved difficult.

1. A REGRESSION MODEL

1.1. **Assignment.** In this part, you will replicate Yule's regression equation for the metropolises unions, 1871-81. See Chapter 1 of Freedman (2009) for a discussion. Fix the design matrix \mathbf{X} at the values reported in Table 3 there, available at

<http://pages.uoregon.edu/dlevin/DATA/yule.txt>

(Subtract 100 from each entry to get the percent changes.) Yule assumed

$$\Delta \text{Paup}_i = a + b \cdot \Delta \text{Out}_i + c \cdot \Delta \text{Old}_i + d \cdot \Delta \text{Pop}_i + \varepsilon_i$$

for 32 metropolitan unions i . For now, suppose the errors ε_i are I.I.D., with mean 0 and variance σ^2 .

- (a) Estimate a , b , c , d and σ^2 .
- (b) Compute the SEs.
- (c) Are these SEs exact, or approximate?
- (d) Plot the residuals against the fitted values. (This is often a useful diagnostic: If you see a pattern, something is wrong with the model. You can also plot residuals against other variables, or time, or ...)

1.2. **Solution.** We estimate the coefficients, standard deviation, standard errors, and residuals of Yule's linear model with R. In Figure 1, we have the summary of the linear model, showing the values of the estimates and standard errors for the coefficients a , b , c , and d . Note that the standard errors are approximations

Coefficients	Estimate	SE	t value	$\Pr(> t)$
a	12.88356	10.36722	1.243	0.224
b	0.75209	0.13499	5.572	5.83e-06 ***
c	0.05560	0.22336	0.249	0.805
d	-0.31074	0.06685	-4.648	7.25e-05 ***

Residual standard error: 9.547 on 28 degrees of freedom.
Multiple R-squared: 0.6972, Adjusted R-squared: 0.6647.
F-statistic: 21.49 on 3 and 28 DF, p -value: 2.001e-07

FIGURE 1. Summary of Linear Model

because they are constructed from the estimated coefficients, which are themselves approximations. Furthermore, from the summary we see that the residual standard error is $\sigma = 9.547$, thus $\sigma^2 = 91.1452$. Figure 2 is the plot of residuals, e_i , against the fitted values \hat{y}_i . There is no discernible pattern, the points appear to have low correlation.

2. THE t -TEST

2.1. Assignment. Make a t -test of the null hypothesis that $b = 0$. What do you conclude? If you were arguing with Yule, would you want to take the position that $b = 0$ and he was fooled by chance variation? In this part, you will do a simulation to investigate the distribution of $t = \hat{b}/\text{SE}$, under the null hypothesis that $b = 0$.

- (a) Set the parameters in Yule's equation as follows: $a = -40$, $b = 0$, $c = 0.2$, $d = -0.3$, $\sigma = 15$. Fix the design matrix \mathbf{X} as in Part 1. Generate 32 $N(0, \sigma^2)$ errors and plug them into the equation

$$\Delta \text{Paup}_i = -40 + 0 \cdot \Delta \text{Out}_i + 0.2 \times \Delta \text{Old}_i - 0.3 \times \Delta \text{Pop}_i + \varepsilon_i,$$

to get simulated values for ΔPaup_i for $i = 1, 2, \dots, 32$.

- (b) Regress the simulated ΔPaup_i on ΔOut , ΔPop and ΔOld . Calculate \hat{b} , SE, and t .
- (c) Repeat (b) and (c) 1000 times.
- (d) Plot a histogram for the 1000 \hat{b} 's, a scatter diagram for the 1000 pairs $(\hat{b}, \hat{\sigma})$ and a histogram for the 1000 t 's.
- (e) What is the theoretical distribution of \hat{b} ? of $\hat{\sigma}^2$? of t ? How close is the theoretical distribution of t to normal?
- (f) Calculate the mean and SD of the 1000 \hat{b} 's. How does the mean compare to the true b ("True" in the simulation.) How does the SD compare to the true SD for \hat{b} ?
- (g) Would it matter if you set the parameters differently? For instance, you could try $a = 10$, $b = 0$, $c = 0.1$, $d = -0.5$ and $\sigma = 25$. What if $b = 0.5$? What if $\varepsilon_i \sim \sigma \cdot (\chi_5^2 - 5) / \sqrt{10}$? The simulation in this exercise is for the level of the test. How would you do a simulation to get the power of the test?

2.2. Solution. I attempted to make a function which would perform our simulations and output our results and figures, given the coefficients, standard deviation, and desired number of simulations. I failed. Instead, I have a block of R code that sort of works like a function, with which I performed all computations asked in this problem. In essence, we have four simulations, each with slightly different parameters. Parts (a) - (d) are performed in R, Figures 3, 4, and 5 are the requested histograms of \hat{b} and t , and the scatter plot of the points $(\hat{b}, \hat{\sigma}^2)$, respectively.

A summary of the computations referenced below may be found in Figure 6.

For (e) we find that the theoretical distribution for \hat{b} is normal with $\mu = 0$ and $\sigma = 0.21$. The parameters for the theoretical distribution are from an assumption and computation, respectively. Under H_0 we assume that $\mu = 0$, and we compute the 'true' standard deviation for \hat{b} thus $\text{Var}(\hat{b}) = \sqrt{\sigma^2(\mathbf{X}^T \mathbf{X})^{-1}}$, where \mathbf{X} is the design matrix. Furthermore, we find that $t \sim t_{28}$. Lastly, as we discussed during class on April 26, we find that upon properly scaling $\hat{\sigma}^2$ that $\hat{\sigma}^2 \sim \chi_{28}^2$.

For part (f) we compute $\mu_{\hat{b}} = -5.232548 \times 10^{-5}$ and $\sigma_{\hat{b}} = 0.2149544$. Note that these values are very close to their ‘true’ counterparts under H_0 .

For part (g) we perform another simulation, with different coefficients a , c , and d , as well as a larger standard deviation for the error terms. While the changes in coefficients have an effect on the values for the various estimates and estimated parameters than we compute, but we maintain the result that \hat{b} is well approximated by a Normal distribution, with true mean and variance very close to the sample mean and variance, as before. If we instead leave the coefficients and error variance as they were in the initial simulation, but test an alternative hypothesis that $b \neq 0$ ¹ we find that while the true and sample means are similar as before, the true and sample variances differ a great deal.² If we again revert to the conditions of our initial simulation, but operate under the assumption that $\varepsilon_i \sim \sigma \cdot (\chi_5^2 - 5) / \sqrt{10}$, instead of our running assumption that $\varepsilon_i \sim N(0, \sigma^2)$, we find that not much changes from the initial simulation, aside from the fact that the sample variance for SE is marginally greater than when $\varepsilon_i \sim N(0, \sigma^2)$.

If we were instead attempting to gain insight into the power of the test which we are performing, rather than the level, we would set a value³ for b under H_1 and hold α fixed. After running our simulation we would determine the power by computing the ratio of the number of times we correctly reject H_1 with the total number of iterations of our simulation.

I suspect that the point of this assignment, is that while a least squares model as used in 1.2 is the best way to approximate a curve to some data, the least-squares model is not well suited to making qualitative inferences about causality. Put another way, while a least-squares curve is the best approximation of a cloud of points in \mathbb{R}^n , manipulating the model will not necessarily yield predictive information, because a good approximation does not necessarily imply that the model ‘knows’ about which variables have a causal effect on the result.

3. BALANCE SCALE

3.1. Assignment. A two balance scale reports the difference between the weights of the right and left plates, plus a random measurement error.

Suppose you have 4 objects whose weights you wish to estimate with the scale, and are allowed 12 measurements. One approach is to measure each weight alone 3 times. (How would you then estimate the four weights with this information?) Is there a better way to use the 12 allowed measurements?

Suppose that

$$x_{i,j} = \begin{cases} +1 & \text{if weight } j \text{ is included on the right plate in the } i\text{-th measurement} \\ -1 & \text{if weight } j \text{ is included on the left plate in the } i\text{-th measurement} \end{cases}$$

Then the model we are investigating is $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where \mathbf{Y} is the vector of the 12 scale readings, $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4)'$ is the vector of the true weight of the four objects, and $\boldsymbol{\varepsilon}$ is the vector of 12 measurement errors. The vector equation $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ is equivalent to the 12 individual equations.

$$Y_i = \beta_1 x_{i,1} + \beta_2 x_{i,2} + \beta_3 x_{i,3} + \beta_4 x_{i,4} + \varepsilon_i, \quad i = 1, 2, \dots, 12.$$

¹in this case we pick $b = 0.5$

²My assumption is that this provides evidence against the alternative hypothesis, but I’m not very sure at this point.

³Say, $b = 0.5$, for instance.

For example, if in the first measurement we put weight 1 on the right plate and weight 2 on the left, then the first reading of the scale is

$$Y_1 = \beta_1 - \beta_2 + \varepsilon_1.$$

Find a design matrix \mathbf{X} that does a better job of estimating β than the design matrix corresponding to measuring each weight 3 times alone. (What is the former matrix?) Discuss the choice of design matrix.

3.2. Solution. The somewhat naïve way to measure our objects is provided by the following design matrix \mathbf{X}_0 ,

$$\mathbf{X}_0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The nice thing about \mathbf{X}_0 is that its column vectors are mutually orthogonal, the bad thing is that its variance is larger than it could be, let's compute it,⁴

$$\text{Var}(\mathbf{X}_0) = (1/3) \cdot \mathbf{I}_4$$

So, we're trying to make a design matrix \mathbf{X}_1 , which has variance less than 1/3. After some trial and error on the blackboard, we come upon the following design matrix \mathbf{X}_1 .

$$\mathbf{X}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & 1 \\ 0 & -1 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & -1 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

We compute $\text{Var}(\mathbf{X}_1)$,

$$\text{Var}(\mathbf{X}_1) = \begin{bmatrix} 0.2 & 0 & 0 & 0 \\ 0 & 0.2 & 0 & 0 \\ 0 & 0 & 0.3 & 0 \\ 0 & 0 & 0 & 0.14 \end{bmatrix}$$

⁴With R, of course.

Observe that $\text{Var}(\mathbf{X}_1)_{i,i} \leq \text{Var}(\mathbf{X}_0)_{i,i}$, for $i = 1, 2, 3, 4$, so we've achieved our goal. Now let's go through how we constructed \mathbf{X}_1 , and what it can tell us about an optimal experiment. The first four rows of \mathbf{X}_1 are each ones in order to get a base reading for the mass of each object. After that, we went row by row, modifying two entries at a time in an effort to maintain the mutual orthogonality of the column vectors. Surprisingly, this only took about ten minutes. The interesting aspect of what we found given this approach, is that placing two objects on the same side of a two balance scale is an effective way of estimating their respective masses, given an initial reading. The big picture, I suppose, is that upon recording an initial measurement for each object, a more efficient way to measure their respective masses is to compare them to one another, rather than to compare them against nothing.

4. HISTOGRAMS AND SCATTER PLOTS

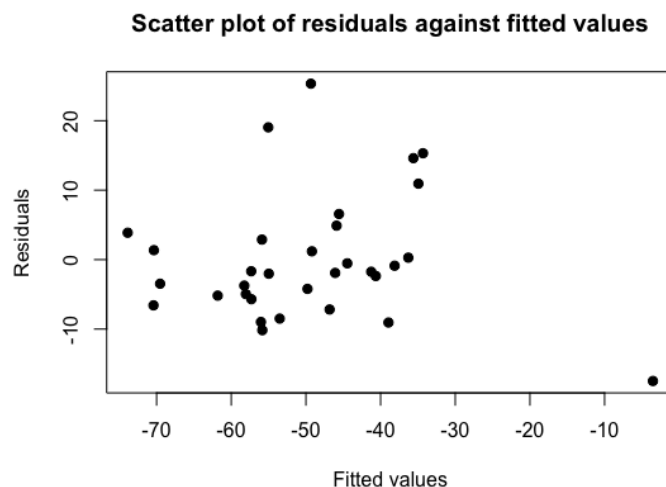


FIGURE 2. Residuals against \hat{y}

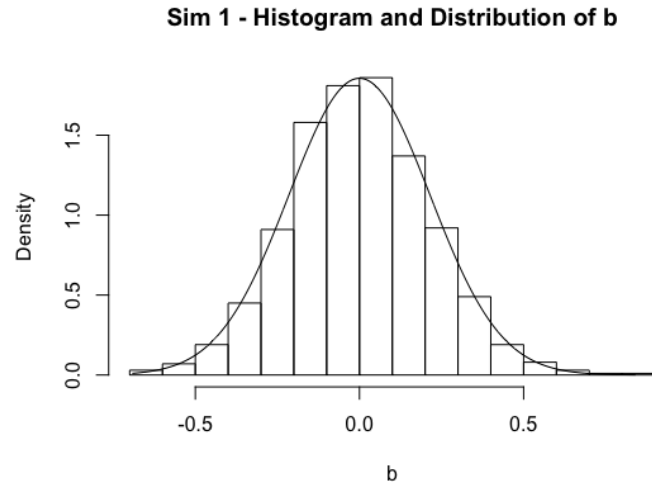


FIGURE 3. Histogram of \hat{b} , with theoretical distribution superimposed

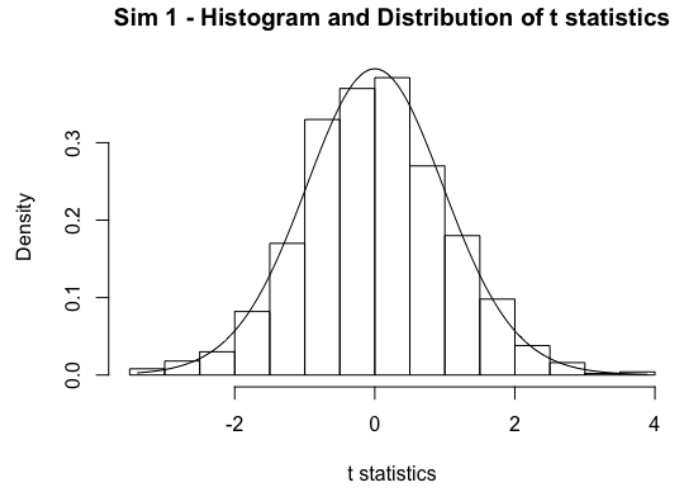
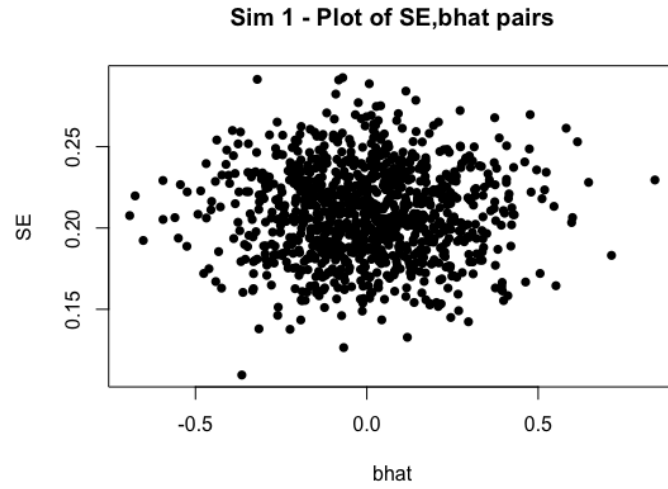


FIGURE 4. Histogram of t , with theoretical distribution superimposed

FIGURE 5. Plot of $(\hat{b}, \hat{\sigma})$.

Sim	$\mu_{\hat{b}}$	$\sigma_{\hat{b}}$	$\mu_{\hat{b} H_0}$	$\sigma_{\hat{b} H_0}$	μ_t	σ_t	μ_{SE}	σ_{SE}
Original	-5.23×10^{-5}	0.215	0	0.212	-6.46×10^{-4}	1.057	0.209	0.0283
Diff Param	-0.00515	0.354	0	0.353	-0.0106	1.045	0.351	0.0468
$H_1 : b = 0.5$	0.4996	0.0143	0.5	0.212	36.228	4.969	0.014	0.0018
New errors	-0.00579	0.204	0	0.212	-0.00947	0.996	0.209	0.0389

FIGURE 6. Summary of Simulations