Daniel Lesh
ECON 3313
2/11/2021

Homework #1

Q.1 $n = 100$, $\bar{x} = 2.6$, $\sigma = 0.4$

1) Hypothesis Test @ 5% significance level:
- $H_0: m_u = 3$; $H_1: m_u \neq 3$
- t-test

$$t = \frac{\bar{x} - mu}{SD/\sqrt{n}} \rightarrow t = \frac{2.6 - 3}{.4/\sqrt{100}} \rightarrow t = -10$$

- Reject H_0 if $|t| > t_{\frac{\alpha}{2}, n-2}$; .05 significance level:

$$|-10| > t_{\frac{.05}{2}, 100-1} \longrightarrow 10 > 1.96$$

- In this case, $10 > 1.96$, so we would reject H_0. In other words, we fail to find the evidence that the mean weight of patties is equal to 3 oz.

2) Generate a 95% CI:
- $CI = \bar{x} \pm z \frac{SD}{\sqrt{n}} \rightarrow CI = 2.6 \pm 1.96\left(\frac{0.4}{\sqrt{100}}\right)$

$CI = 2.6 \pm 0.0784$

$\boxed{CI = (2.5216, 2.6784)}$

3) Yes, there is significant evidence for concern that the mean weight of the patties falls below 3 oz. When generating the 95% CI, the sample mean weight falls betwen 2.52 oz and 2.68 oz. This falls below the desired 3oz weight of patties. In conclusion, the CI provides evidence for concern that the mean weight of patties falls below 3oz.

Q.2 $\widehat{Earn} = 232 + 4.2 \cdot Age$, $R^2 = 0.15$, $SER = 287.21$

1) $\bar{y} = \hat{\beta}_0 + \hat{\beta}_1(\bar{x}) \rightarrow \bar{y} = 232 + 4.2(x)$
- $y = 232 + 4.2(18)$; $y = 232 + 4.2(32)$

$\boxed{y = \$307.60 \text{ for 18 yr old}}$ $\boxed{y = \$366.40 \text{ for 32 yr old}}$

2) $415.96 = 232 + 4.2x$
$-232 \quad -232$

$\frac{183.96}{4.2} = \frac{4.2x}{4.2}$

$43.8 \approx 44 = x \rightarrow \boxed{44 \text{ years of age}}$

$\widehat{Earn} = 232 + 4.2(46)$

$\widehat{Earn} = \$425.20$

$\boxed{\$425.20 \text{ weekly earnings}}$ @ 46 years of age.

3) Age should matter in the determination of earnings for a variety of reasons. Age is important in determining one's prime money-making years. For example, someone who is 35 years old is more likely to make more money than someone who is 18 years old. Those that near retirement (65+) are more likely to make less money than someone younger than 55 years old.
The results do not necessarily suggest that there is a guarantee for earnings to rise for everyone as they become older, as age only accounts for $R^2 = .15$. This suggests that age is not explaining much variation in weekly earnings. Weekly earnings most likely depends on a variety of factors including age, showing why age may not account for most of the variation in weekly earnings in the regression.

4) $\bar{x} = 34$ years old ; $\widehat{Earn} = 232 + 4.2(x)$

$\widehat{Earn} = 232 + 4.2(34)$

$\boxed{= \$374.80} \rightarrow$ avg. weekly income in sample

Q.3 $\widehat{Cost} = 7{,}311.17 + 3{,}985.20 \times Rep - .20 \times Size + 8{,}406.79 \times Dpriv$
$- 416.38 \times Dlibart - 2{,}376.51 \times Dreligion$, $R^2 = 0.72$, $SER = 3{,}773.35$

1) For every 1 unit increase in Reputation, it raises the tuition cost by roughly $3,985.20.

For every 1 unit increase in Size, it decreases the tuition cost by roughly $0.20.

Since Dpriv is a binary variable, there is a $8,406.79 added cost to tuition if the school is private.

Since DLibart is a binary variable, there is a -$416.38 decrease in cost to tuition if it is a liberal arts school.

Since Dreligion is a binary variable, there is a -$2,376.51 decrease in cost to tuition if it is a school w/ religious affiliation.

- I believe all coefficients have the expected sign except Dlibarts and Dreligion. Usually when universities are recognized as such it drives the price even higher. In this regression, it doesn't, which I find odd.

2) $\widehat{Cost} = 7{,}311.17 + 3{,}985.20(4.5) - .20(1{,}500) + 8{,}406.79(1)$
$- 416.38(1) - 2{,}376.51(0)$

Liberal Arts College Cost = $32,934.98

3) Reputation and Dpriv - statistically significant @ 1%? $\to 2.58$

Reputation $= \dfrac{3{,}985.20 - 0}{664.58} = \boxed{5.997 > 2.58}$

Dpriv $= \dfrac{8{,}406.79}{2{,}154.85} = \boxed{3.901 > 2.58}$

- Both Reputation & Dpriv are statistically significant @ 1%.

4) Construct 99% CI for the coefficient Size:

$CI = -.2 \pm 2.58(.13)$

$= (-.5354, .1354)$

5) $H_0: \beta_{size} = 0$ $\quad$ $H_a: \beta_{size} \neq 0$

$\beta_{plibart} = 0$ $\quad$ $\beta_{plibart} \neq 0$

F-stat = 1.23 → @ 5% sig. level when $q=2$ = 3

$1.23 < 3$ (F-stat)

- cannot reject null hypothesis.

Q.4

1) Log-Linear Model

$\widehat{\ln(Earnings)} = .54 + .083 \times Educ, \quad \bar{R}^2 = .234$

$(.14) \quad (.011)$

- A 1 unit increase in Educ leads to a 8.3% increase in Earnings. to try and test the belief that years of high school education is different from college education, you would need an additional data set w/ values for high school education. Then you would create a new coefficient different from Educ, called HighS. the Educ coefficient can be renamed to Coll. after this, you would re-run the regression and assume these values:

$\widehat{Earnings} = \hat{\beta}_0 + \hat{\beta}_1 \times HighS + \hat{\beta}_2 \times Coll$

2) % change on earnings w/ change to schooling:

$\frac{16-12}{12} \times 100 = 33.33\%$

- a 1% increase in Educ leads to a 4.3% increase in $\widehat{Earnings}$

$4.3\% \cdot 33.33\% = \boxed{113.33\%}$

- The Log-Log model is a better fit because of a higher $\bar{R}^2$ value.

3) $\ln(y+\Delta y) - \ln(y) = \Delta y / y$

$y = \widehat{\text{Earnings}}$

$\Delta \widehat{earnings} / \widehat{earnings} = .033(exper) - .0005(exper)^2$

$- (.033 \cdot exper - .0005(exper)^2)$

$= .0325 - .001$

$= .0315(exper)$

- $H_0: \beta_3 = 0$, $H_a \neq 1$
- Test @ 5% significance level (1.96):

$t = \frac{-.0005 - 0}{.0001} = -5$

$|-5| \approx \boxed{5 > 1.96}$

- Reject Null hypothesis. The population reg. model is not linear.

---

5)

1) $\widehat{Charges} = \underset{(968.63)}{2{,}636.04} + \underset{(22.49)}{255.69}(age) + \underset{(262.16)}{556.53}(children)$, $\bar{R}^2 = 0.0911$

- 95% CI = (10,536.11, 12,303.77)

2) $\widehat{charges} = \underset{(2744.42)}{7{,}899.56} - \underset{(154.83)}{58.27}(age) + \underset{(1.93)}{3.96}(age)^2 + \underset{(274.29)}{723.97}(children)$,

$\bar{R}^2 = .0933$

- 95% CI = (10,243.45, 12,077.53)

3) $\widehat{charges} = \underset{(12{,}896.5)}{26{,}630.1} + \underset{(142.2)}{517.6}(age) - \underset{(5{,}129.7)}{9570.7}\ln(age) + \underset{(274.2)}{707.9}(children)$

$\bar{R}^2 = .0928$

- 95% CI = (-511,633.2, 24537.95)

4) Out of the 3 models, Model #2 is the best for the prediction. Model #2 is the best for prediction because it has the highest $\bar{R}^2$ (.09328). This model accounts for the highest variation when compared to the linear model (Model #1) and linear-log model (Model #3).