Public Policy 529
Fall 2023 Problem Set #10

Francisco Brady

2023-12-06

Due on Wednesday, December 6th

1. In the regression analysis below, the dependent variable is the number of minutes it takes for an ambulance to arrive when called to an emergency. The independent variable is the distance in miles from the location where the ambulance is waiting to the location of the emergency. You have data on n=52 cases in which an ambulance was called to an emergency.

$$Minutes_i = \beta_0 + \beta_1(Miles_i) + u_i$$

Variable	Coefficient	Standard Error
Intercept	1.2	0.7
Miles	1.8	0.4

Note: $R^2 = .71$

(a) Interpret the estimated coefficient on Miles (β_1). In other words, what does it predict about the relationship between distance and the time it takes for the ambulance to arrive?

A one mile increase is associated with a 1.8 minute increase in the time it takes for an ambulance to arrive.

(b) Are these two coefficients statistically significant? In other words, for each estimated coefficient, can we reject the null hypothesis that the true coefficient is zero? Provide evidence for your answer by finding the t-statistic and p-value for each coefficient. To find the t-statistic for **the intercept**, we divide the coefficient by the standard error of the coefficient:

$$t = \frac{\hat{\beta}_0}{se_{\hat{\beta}_0}}$$
$$t = \frac{1.2}{0.7}$$
$$t = 1.714286$$

At a 95% confidence level with 50 degrees of freedom, we require a critical value of: 2.009. Looking up our observed t-statistic using the same degrees of freedom, we can say that our p-value is between $t_{.05} < p < t_{.10}$. In the case of the intercept, we cannot reject that null hypothesis that the true intercept is not zero.

To find the t-statistic for β_1, we can divide the coefficient by its standard error:

$$t = \frac{\hat{\beta_1}}{se_{\hat{\beta_1}}}$$
$$t = \frac{1.8}{0.4}$$
$$t = 4.5$$

At a 95% confidence level with 50 degrees of freedom, we get a critical value of: 2.009. Looking up our observed t-statistic using the same degrees of freedom, we can say that our p-value is less than $t_{.002}$. For this estimate, we can reject the null hypothesis that the true value of the coefficient is 0.

(c) Suppose there is an emergency 6.2 miles from where the ambulance is waiting. What do we predict is the number of minutes it will take for the ambulance to arrive? Use the model coefficients to calculate:

$$Mins_{6.2} = 1.2 + 1.8(Miles_i)$$
$$Mins_{6.2} = 1.2 + 1.8(6.2)$$
$$Mins_{6.2} = 1.2 + 11.16$$
$$Mins_{6.2} = 12.36$$

2. Use the dataset `StateData_2018` for this question. This dataset contains variables measured at the state level in the United States. You will estimate a bivariate linear regression using the following variables:

- The dependent variable is `gun_deathrate`, which is the number of firearm-related deaths per 100,000 in population in the state. In the sample, this variable ranges from 3.3 to 22.9.
- The independent variable is `lawtotal`, which is the number of laws restricting gun ownership in the state. Specifically, it is a count of different types of possible gun restrictions: requiring a person be 21 years-old to possess a handgun, requiring a person be 21 years-old to purchase a handgun, a ban on "assault-style" weapons, a requirement for universal background checks for gun sales, etc. In the sample, this variable ranges from 1 to 109.

```r
# note: code for question 2a:
model <- lm(gun_deathrate ~ lawtotal, data = state_data)
summary(model)
```

(a) Estimate the model and report the output. See lecture slides or help documents for appropriate commands.

```
##
## Call:
## lm(formula = gun_deathrate ~ lawtotal, data = state_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.4274 -2.1524 -0.2607  2.9935  6.7143
##
```

```
## Coefficients:
##             Estimate Std. Error t value            Pr(>|t|)
## (Intercept) 17.37730    0.70138  24.776 < 0.0000000000000002 ***
## lawtotal    -0.14167    0.01783  -7.945       0.000000000267 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.463 on 48 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.568,  Adjusted R-squared:  0.559
## F-statistic: 63.12 on 1 and 48 DF,  p-value: 0.0000000002669
```

(b) What is the substantive meaning of the estimated intercept (β_0)?
The intercept is 17.377, which means that if the total number of laws regulating gun control is 0, the predicted rate of firearm deaths would be 17.377 per 100,000.

(c) Interpret the substantive meaning of the coefficient on lawtotal (β_1).
An increase of one gun control law is associated with a decrease of 0.14 firearm deaths per 100,000. There is a negative relationship between the dependent and independent variables.

(d) Assess the statistical significance of each of these coefficients at $\alpha = .05$.
The t-statistic for the intercept term is 24.766, and the p-value is 0.0000000000000002, which is lower than .05, so it is statistically significant. The t-statistic for the `lawtotal` variable is 0.000000000267, which is also less than .05, so that coefficient is statistically significant. The critical value for 95% confidence using a more conservative 40 degrees of freedom would be 2.021.

(e) Suppose a state has 20 laws that involve a restriction on gun ownership. What do we predict would be the level of firearm-related deaths in that state?
We can use the model coefficients to calculate:

$$\begin{aligned} gundeaths &= 17.377 - 0.142(lawtotal) \\ gundeaths &= 17.377 - 0.142(20) \\ gundeaths &= 17.377 - 2.84 \\ gundeaths &= 14.537 \end{aligned}$$

(f) Suppose this state were to pass five additional laws restricting gun ownership. What is the change in the predicted level of firearm-related deaths?

$$\begin{aligned} gundeaths_{25} &= 17.377 - 0.142(25) \\ gundeaths_{25} &= 17.377 - 3.55 \\ gundeaths_{25} &= 13.827 \\ gundeaths_{20} &= 14.537 \\ gundeaths_{25-20} &= 13.827 - 14.537 \\ gundeaths_{25-20} &= -0.71 \end{aligned}$$

(g) Interpret the R^2 for this regression.
The R^2 value is 0.559, which means that the `lawtotal` variable explains around 55% of the variance of the dependent variable.

(h) Interpret the standard error of the estimate from this regression. In other words, what does this number mean substantively?
The residual standard error of the model is 3.463, which means that on average the model is off by 3.463 in predicting the rate of gun deaths per 100,000.

3. Continue with the `StateData_2018` for this question. You will estimate a bivariate regression using the following variables:

- The dependent variable is `pctui`, which is the percentage of a state's population that does not have health insurance. In the sample, this variable ranges from 3.3 to 19.9.
- The independent variable is `MedicaidExp`, which is a dichotomous (0,1) indicator of whether a state chose to expand Medicaid eligibility as part of the Affordable Care Act (ACA). The variable is coded as 1 if the state expanded Medicaid and 0 otherwise. As of 2018, 32 states, plus the District of Columbia, had expanded Medicaid eligibility, while 18 states did not. Since the independent variable is dichotomous, a bivariate regression is essentially a difference of means test.

```r
# note: code for question 3a
model <- lm(pctui ~ MedicaidExp, data = state_data)
summary(model)
```

(a) Estimate the model and report the output.

```
##
## Call:
## lm(formula = pctui ~ MedicaidExp, data = state_data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -6.100 -1.589 -0.200  1.611  7.300
##
## Coefficients:
##             Estimate Std. Error t value            Pr(>|t|)
## (Intercept)  12.6000     0.6551  19.235 < 0.0000000000000002 ***
## MedicaidExp  -4.7606     0.8143  -5.846         0.000000404 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.779 on 49 degrees of freedom
## Multiple R-squared:  0.4109, Adjusted R-squared:  0.3989
## F-statistic: 34.18 on 1 and 49 DF,  p-value: 0.0000004038
```

(b) What is the substantive meaning of the estimated intercept (β_0)?
The intercept is 12.6. The interpretation is that in states that did **not** choose to expand Medicaid, on average the percentage of the population that is uninsured is 12.6%.

(c) Interpret the substantive meaning of the coefficient on MedicaidExp (β_1).
The decision to expand Medicaid is associated with a **decrease** of 4.67 on average in the percentage of the state population that is uninsured. The relationship is negative.

(d) What do we predict is the mean level of pctui in a state that has expanded Medicaid as part of the ACA?

$$pctui = 12.6 - 4.7606(\text{MedicaidExp})$$
$$pctui = 12.6 - 4.7606(1)$$
$$pctui = 7.8394$$

```r
# note: code for question 3e
t.test(pctui ~ MedicaidExp, data = state_data) #, var.equal = TRUE)
```

(e) Using your software, perform a t-test for the difference of means. Identify on the output the items that match up with your regression output. Note: the regression method for a difference of means test is equivalent to using the equal variance assumption in a two-sample t-test, so the relevant t-statistics will generally differ somewhat unless you apply the equal variance assumption in the t-test (which I would not recommend in this case).

```
##
##  Welch Two Sample t-test
##
## data:  pctui by MedicaidExp
## t = 5.5289, df = 29.9, p-value = 0.000005292
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  3.001867 6.519345
## sample estimates:
## mean in group 0 mean in group 1
##       12.600000        7.839394
```

- The mean in group 0 matches the intercept estimate
- The mean in group 1 matches the predicted value of the mean level of `pctui` in a state that has expanded Medicaid.
- The t-statistic (5.5289) on `MedicaidExp` is slightly larger than the t-statistic in the regression output (-5.846)
- The difference between the two means is the coefficient estimate on `MedicaidExp`: 12.600000 - 7.839394 = 4.760606
- The p-value is 0.000005292, compared to 0.000000404 in the regression.