ECO 375–Homework 1

University of Toronto Mississauga

Due: 17 October, 2023, 5 PM

ANSWER KEY

Theoretical Problems

1. Suppose $Y = X^2 + 3$. For this problem, you may use the fact that, if $W \sim N(0,1)$, then $E[W^3] = 0$ and $E[W^4] = 3$. Recall that $N(\mu, \sigma^2)$ indicates a variance of σ^2.

   (a) Suppose $X \sim N(1, 3)$.

   i. *(points: 5)* What is the variance of Y? [Hint: recall from class how to standardize a normal distribution.]

   Answer

$$
\begin{aligned}
V[Y] &= E[(X^2+3)^2] - E[X^2+3]^2 \\
&= E[X^4 + 6X^2 + 9] - (E[X^2]+3)^2 \\
&= E[X^4] + 6E[X^2] + 9 - (E[X^2]+3)^2
\end{aligned}
$$

   Now, using the fact that $W = \frac{X-\mu}{\sigma}$, with $\mu = 1$ and $\sigma^2 = 3$, we have $X = \mu + \sigma W$, so

$$
\begin{aligned}
E[X^2] &= E[\mu^2 + 2\mu\sigma W + \sigma^2 W^2] \\
&= 1 + 2 \times 1 \times \sqrt{3} \times E[W] + 3 \times E[W^2] \\
&= 1 + 0 + 3 = 4. \\
E[X^4] &= E[\mu^4 + 4\mu^3\sigma W + 6\mu^2\sigma^2 W^2 + 4\mu\sigma^3 W^3 + \sigma^4 W^4] \\
&= 1 + 0 + 6 \times 3 + 0 + 9 \times 3 = 46.
\end{aligned}
$$

   Plugging this in to the formula above,

$$
V[Y] = 46 + 6 \times 4 + 9 - (4+3)^2 = 30.
$$

   ii. *(points: 5)* What is the covariance of X and Y?

   Answer

   For this problem, we will use the properties we found above, along with

$$
\begin{aligned}
E[X^3] &= E[\mu^3 + 3\mu^2\sigma W + 3\mu\sigma^2 W^2 + \sigma^3 W^3] \\
&= 1 + 0 + 3 \times 1 \times 3 \times 1 + 0 = 10.
\end{aligned}
$$

   Now,

$$
\begin{aligned}
C[X,Y] &= E[(X - E[X])(Y - E[Y])] \\
&= E[(X-1)(X^2 + 3 - E[X^2+3])] \\
&= E[(X-1)(X^2 + 3 - 7)] \\
&= E[(X-1)(X^2 - 4)] \\
&= E[X^3 - X^2 - 4X + 4] \\
&= 10 - 4 - 4 \times 1 + 4 = 6.
\end{aligned}
$$

iii. *(points: 2)* In week 6, you will learn that, if we use OLS to estimate the βs in $Y = \beta_0 + \beta_1 X + \epsilon$, then our estimate $\hat{\beta}_1 \xrightarrow{p} \frac{C[X,Y]}{V[X]}$. What is $plim(\hat{\beta}_1)$ given this data? [Hint: this should be a very easy question if you've answered previous parts correctly.]

Answer

Based on the information in the question and my previous answers,

$$\hat{\beta}_1 \xrightarrow{p} \frac{C[X,Y]}{V[X]} = \frac{6}{3} = 2.$$

(b) Now, suppose X takes on only two values: with 50% probability, it is 0, and with 50% probability, it is 2.

i. *(points: 5)* Calculate $E[X]$, $E[X^2]$, $E[X^3]$, and $E[X^4]$.

Answer

Using the definition of expected value for discrete random variables,

$$\begin{aligned}
E[X] &= .5 \times 0 + .5 \times 2 = 1 \\
E[X^2] &= .5 \times 0^2 + .5 \times 2^2 = 2 \\
E[X^3] &= .5 \times 0^3 + .5 \times 2^3 = 4 \\
E[X^4] &= .5 \times 0^4 + .5 \times 2^4 = 8
\end{aligned}$$

ii. *(points: 5)* What is the variance of Y?

Answer

Using the formula we calculated above,

$$\begin{aligned}
V[Y] &= E[X^4] + 6E[X^2] + 9 - (E[X^2] + 3)^2 \\
&= 8 + 6 \times 2 + 9 - (2+3)^2 = 4.
\end{aligned}$$

iii. *(points: 5)* What is the covariance of X and Y?

Answer

$$\begin{aligned}
C[X,Y] &= E[(X - E[X])(Y - E[Y])] \\
&= E[(X-1)(X^2 + 3 - E[X^2 + 3])] \\
&= E[(X-1)(X^2 + 3 - 5)] \\
&= E[X^3 - X^2 - 2X + 2] \\
&= 4 - 2 - 2 + 2 = 2.
\end{aligned}$$

iv. *(points: 2)* What is the $plim(\hat{\beta}_1)$ given this data?

Answer

Using the information we got above,

$$\hat{\beta}_1 \xrightarrow{p} \frac{C[X,Y]}{V[X]} = \frac{2}{1} = 2.$$

You might notice (though you don't have to mention this to get full credit) that this is the same value we got using the normal distribution above. Why is this? Note that β_1 measures $\frac{dY}{dX}$; in this case, $\frac{dY}{dX} = 2X$, a linear function. When you use OLS to estimate the slope of a non-linear function (such as $X^2 + 3$), it effectively measures the average slope over the values of the independent variable(s). Recall that $E[g(X)] = g(E[X])$ if $g(X)$ is linear, so the average of the slope measures the slope of the average in this case; since both distributions of X have the same expected value, the β_1 we estimate will be the same. (This is not a formal proof - it's here instead to give you some intuition for what OLS is doing.)

2. A company wants to improve their online advertising. They therefore develop two advertising campaigns, called A and B, which will be shown to users of a social media company, Y. Their goal is to determine which campaign is better at getting users to click the ad and buy something from their website. Note: to receive full credit for this question, please reference terms and ideas discussed in class in each part of this question. For example, if an idea is bad because it violates an assumption, explicitly state the assumption (e.g. SLR.3) and explain why it is violated.

   (a) *(points: 6)* To test which campaign works better, the company shows campaign A to older users of Y and B to younger users. They then measure whether each person who views the ad clicks it and buys something from their website. They define Y_i as the amount of money person i spends, and D_i as a dummy variable that takes a value of 1 if person i saw campaign A, 0 if person i saw campaign B. They then use OLS to estimate $Y_i = \beta_0 + \beta_1 D_i + u_i$. Will their estimate of β_1 be an unbiased estimate of the effect of A vs B? Explain why or why not.

   Answer
   The data they are using violates SLR.4: that $E[u|D] = 0$. (You could state instead that it violates MLR.4.) In particular, older people might be more or less likely to buy from the site than younger people, depending on what they sell. If that's true, high values of D_i (for older people) would be associated with high or low values of u_i, violating the assumption. You could also explain that this will mean that the result does not tell us the causal effect of the campaign.

   (b) *(points: 6)* Now, to test which campaign works better, for one year, the company randomly chooses 3 days a week to show campaign A, and 3 days a week to show campaign B; one day a week, they will use neither campaign. (For example, they might randomly choose Mondays, Tuesdays, and Fridays for campaign A; and Wednesdays, Thursdays, and Saturdays for campaign B. As in the previous problem, they define Y_i as the amount of money person i spends, and D_i as a dummy variable that takes a value of 1 if person i saw campaign A, 0 if person i saw campaign B. They then use OLS to estimate $Y_i = \beta_0 + \beta_1 D_i + u_i$. Will their estimate of β_1 be an unbiased estimate of the effect of A vs B? Explain why or why not.

   Answer
   This violates SLR.2 (or, equivalently, MLR.2): This is not a random sample of the population because errors are correlated. For example, Saturdays might happen to be the busiest shopping day of the week. In that case, the error term would be high for all Saturdays, which are all either treated or untreated.

   (c) *(points: 6)* The company decides to hire you to investigate this question. How would you design a study that would best estimate the effect of A vs B? Explain each step you would take, and why it would work.

   Answer
   The best option is a randomized controlled trial: randomly assign some people to get campaign A, and some people to get campaign B. Then as in previous questions, define Y_i as the amount of money person i spends, and D_i as a dummy variable that takes a value of 1 if person i saw campaign A, 0 if person i saw campaign B. They then use OLS to estimate $Y_i = \beta_0 + \beta_1 D_i + u_i$. This will satisfy SLR.2 because each observation is independent of others. It satisfies SLR.4 because the independent variable is randomly assigned, so those with high values of D_i are no more or less likely to have high values

of u_i. You could also note, though you don't have to for full credit, that this will satisfy SLR.1 (because there are only two values for the independent variable, the conditional expectation must be linear); it will satisfy SLR.3 (there will be two values if we have enough observations; and it will satisfy SLR.5 (since the dependent variable is randomly assigned, the variance of u will be unrelated to it).

3. A research team analyzes a data set and finds the following. Variable X takes on each integer value between 0 and 10, inclusive. (That is, it takes a value of 0, 1, 2, 3, ... , 9, 10.) There are 8 observations for each of these values (that is, 88 total observations). For half of observations where $X = c$, $Y = c$; for the other half, $Y = c + 1$. For example, when $X = 3$, there are four observations where $Y = 3$ and four observations where $Y = 4$. The researchers all want to estimate β_1 in the model $Y = \beta_0 + \beta_1 X + u$.

   (a) *(points: 6)* Researcher A only uses 24 observations: those where $X = 4$, $X = 5$, or $X = 6$. Calculate SST_x, SST_{xy}, and the OLS estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ researcher A will estimate.

   Answer

   Each of these answer requires calculating a sum. To make this simpler, I will calculate it by multiplying the number of observations with a value times that value. For example, to calculate the sum of the Xs, I note that the value 4 appears 8 times, 5 appears 8 times, and 6 appears 8 times, so I multiply 4 by 8, plus 5 by 8, plus 6 by 8. Thus we first calculate:

$$\begin{aligned} \bar{X} &= \frac{1}{24}(8 \times 4 + 8 \times 5 + 8 \times 6) = 5 \\ \bar{Y} &= \frac{1}{24}(4 \times 4 + 4 \times 5 + 4 \times 5 + 4 \times 6 + 4 \times 6 + 4 \times 7) = 5.5 \end{aligned}$$

   Using this, we have

$$\begin{aligned} SST_x &= \sum_{i=1}^{n}(x_i - \bar{x})^2 \\ &= 8 \times (4-5)^2 + 8 \times (5-5)^2 + 8 \times (6-5)^2 \\ &= 16. \end{aligned}$$

   and

$$\begin{aligned} SST_{xy} &= \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \\ &= 4(4-5)(4-5.5) + 4(4-5)(5-5.5) + 4(5-5)(5-5.5) + 4(5-5)(6-5.5) \\ &\quad +4(6-5)(6-5.5) + 4(6-5)(7-5.5) \\ &= 16. \end{aligned}$$

   Now, $\hat{\beta} = \frac{SST_{xy}}{SST_x} = \frac{16}{16} = 1$.

   We can calculate $\hat{\beta}_0$ using the formula from the notes:

$$\begin{aligned} \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} \\ &= 5.5 - 1 \times 5 = 0.5. \end{aligned}$$

   (b) *(points: 6)* Researcher B also only uses 24 observations: those where $X = 0$, $X = 5$, or $X = 10$. Calculate SST_x that researcher B will estimate. In order to not waste your time, I will not have you calculate SST_{xy}, $\hat{\beta}_0$, or $\hat{\beta}_1$. However, I hope you can see that $\hat{\beta}_0$ and $\hat{\beta}_1$ are the same for researchers A and B. If not, come to office hours and I will explain!

Answer
We can calculate:

$$\bar{X} = \frac{1}{24}(8 \times 0 + 8 \times 5 + 8 \times 10) = 5$$

So

$$\begin{aligned} SST_x &= \sum_{i=1}^{n}(x_i - \bar{x})^2 \\ &= 8 \times (0-5)^2 + 8 \times (5-5)^2 + 8 \times (10-5)^2 \\ &= 400. \end{aligned}$$

(c) *(points: 6)* Calculate the estimated standard error of $\hat{\beta}_1$ for each researcher. [Hint: you may use the fact that both researchers estimate the same value for $\hat{\beta}_0$ and $\hat{\beta}_1$.]

Answer
To calculate the standard error, we use the formula:

$$\hat{\sigma}^2_{\hat{\beta}_1} = \frac{\hat{\sigma}^2_u}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

where

$$\hat{\sigma}^2_u = \frac{1}{n-2}\sum_{i=1}^{n}\hat{u}_i^2$$

To calculate $\hat{\sigma}^2_u$, note that both researchers estimated $\hat{Y} = 0.5 + X$, so that every observation is exactly 0.5 from that estimated line.That is, $Y = X$ for some observations and $Y = X + 1$ for others, so

$$\hat{u} = Y - \hat{Y} = \begin{cases} X - (0.5 + X) = -0.5 & \text{if } Y = X \\ (X+1) - (0.5 + X) = 0.5 & \text{if } Y = X + 1 \end{cases}$$

Thus $\hat{u}^2 = 0.5^2 = 0.25$ for all 24 observations, so:

$$\hat{\sigma}^2_u = \frac{1}{24-2}24 \times 0.25 = \frac{6}{22} = 0.2727$$

Using this value, and SST_x calculated above, for A:

$$\begin{aligned} \hat{\sigma}^2_{\hat{\beta}_1} &= \frac{0.2727}{16} = 0.01704 \\ \implies SE &= \sqrt{0.01704} = 0.1306 \end{aligned}$$

For B:

$$\begin{aligned} \hat{\sigma}^2_{\hat{\beta}_1} &= \frac{0.2727}{400} = 0.0006818 \\ \implies SE &= \sqrt{0.0006818} = .02611 \end{aligned}$$

You don't have to say this to get full credit, but: note that the standard error is much smaller for B than for A, even though both have the same number of observations with the same errors. This is because the X values are more dispersed; in general, having a more dispersed independent variable leads to more precisely estimated coefficients when using OLS.

Computer Based Problems

Notes:

- If you are using Citrix, remember to save your code and log file on your home computer, not the remote computer!
- Remember to include the Stata code as part of your writeup. (See "Notes" at the start of this assignment.)

4. **Google Trends**. For this problem, you will analyze data set `eco375hw1data2023.dta` using Stata. This data is gathered from Google Trends; see https://trends.google.com/. The data show the frequency of searches for various terms in provinces and territories of Canada over the past five years.[1] Variable `location` presents the province or territory. All other variables present the search interest in the term in the variable's title. You can refer to this value as the "normalized search interest." For this problem, you can pretend that non-economists understand what this means—that is, a non-economist would understand what it means if "normalized search interest in the word 'warm' increasing by 1 unit."

   (a) *(points: 4)* Present a table showing the mean and standard deviation of normalized search interest in each city name in the data set.

   Answer

City	Mean	Standard Deviation
Montreal	13.53846	26.13623
Vancouver	12.46154	26.96793
Toronto	19.53846	24.46976

   Code:

   ```
   sum montreal vancouver toronto
   ```

   (b) *(points: 3)* Use OLS to estimate β_1 in:

   $$\text{rain} = \beta_0 + \beta_1 \text{umbrella} + u$$

   Report the estimate $\hat{\beta}_1$ in words a non-economist could understand. Do *not* use causal language. Should we think of this as an estimate of the causal effect of searches for umbrellas on searches for rain? Why or why not?

   Answer

   Normalized search interest in the word "umbrella" increasing by 1 unit is associated with normalized search interest in "rain" increasing by 0.627965.

   We should not think of this as an estimate of the causal effect because SLR.4 is not satisfied. Many factors that are related to searches for umbrellas might also be related to searches for rain–for example, the extent to which rain is common in that location.

   Code:

[1] Google describes the numbers in this way: "Numbers represent search interest relative to the highest point on the chart for the given region and time. A value of 100 is the peak popularity for the term. A value of 50 means that the term is half as popular. A score of 0 means there was not enough data for this term."

```
reg rain umbrella
```

(c) *(points: 4)* Create a new variable, `aboveavgcold`, that is equal to 1 if the location has normalized search interest above that variable's average, and 0 otherwise. Use a regression to estimate the difference between normalized search interest for pizza when normalized search interest for cold is above or below average. Is this difference significantly different from 0 at the 5% level?

Answer

The difference is 5.6; it is not significantly different from 0 at the 5% level.

Code:

```
sum cold
gen aboveavgcold = 0
replace aboveavgcold = 1 if cold > 65.15
reg pizza aboveavgcold
```

(d) *(points: 4)* Use OLS to estimate β_1 in each of the following equations. Report your estimate for each using words a non-economist could understand. Use causal language for this part (even though the results may not actually be causal).

$$\begin{aligned}
\text{poutine} &= \beta_0 + \beta_1 \ln(\text{montreal}) + u \\
\ln(\text{poutine}) &= \beta_0 + \beta_1 \text{montreal} + u \\
\ln(\text{poutine}) &= \beta_0 + \beta_1 \ln(\text{montreal}) + u
\end{aligned}$$

Answer

- A 1% higher normalized search interest in the word "Montreal" causes normalized search interest in the word "poutine" to increase by 0.205 units.
- A 1 unit higher normalized search interest in the word "Montreal" causes normalized search interest in the word "poutine" to increase by 1.58%.
- A 1% higher normalized search interest in the word "Montreal" causes normalized search interest in the word "poutine" to increase by 0.412%.

Code:

```
gen logpoutine = ln(poutine)
gen logmontreal = ln(montreal)
reg poutine logmontreal
reg logpoutine montreal
reg logpoutine logmontreal
```

(e) *(points: 2)* Using the data, briefly present and describe one fact that you think is interesting. Your answer to this question must actually use Stata code and the data to find a fact that was not discussed in a previous question.

Answer

Answers will vary.

5. **Monte Carlo Simulation**. Consider the following model:

$$Y = \beta_0 + \beta_1 X + U$$

where

$$\beta_0 = 2$$
$$\beta_1 = 3$$
$$X \sim N(-1, 1)$$

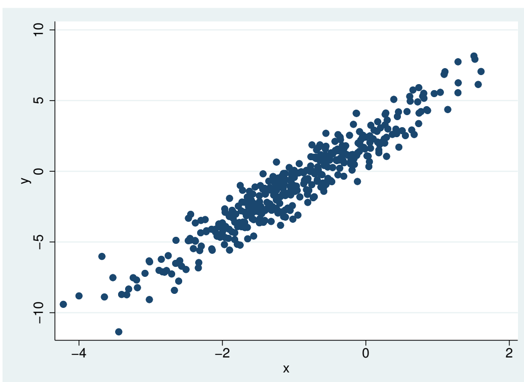
and

$$U \sim N(0, 1),$$

and X and U are independent.

(a) *(points: 6)* Simulate the model once in Stata. Generate a data set $\{y_i, x_i : i = 1, ..., n\}$ with $n = 400$ observations. Show a scatterplot of X and Y. Use OLS to estimate β_0 and β_1; what are your estimates?

Answer



We estimate $\hat{\beta}_0 = 2.126651$ and $\hat{\beta}_1 = 3.068917$.

Code:

```
clear all
set seed 1234
cap program drop regra1
program regra1, rclass
drop _all
set obs 400
gen x = rnormal(-1,1)
gen u = rnormal(0,1)
gen y = 2 + 3* x + u
reg y x
return scalar b0=_b[_cons]
return scalar b1=_b[x]
end
regra1
scatter y x
graph export "graphics/scatter.eps", replace
```

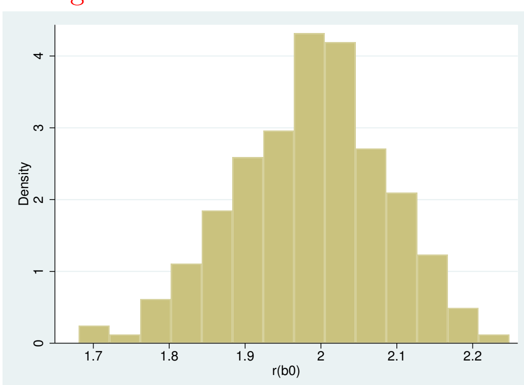
(b) *(points: 6)* Now, simulate this model 200 times in Stata. For each simulation, generate a data set $\{y_i, x_i : i = 1, ..., n\}$ with $n = 200$ observations. Then, for each sample, estimate β_0 and β_1 using OLS and save the results $\hat{\beta}_0$ and $\hat{\beta}_1$.
What are the averages and the standard deviations of $\hat{\beta}_0$ and $\hat{\beta}_1$ across the simulations? *(Note: most likely, the averages you calculate here are closer to the truth than the estimate from one regression in the previous part. This is related to the Law of Large Numbers, which we will cover later in the course.)* Are $\hat{\beta}_0$ and $\hat{\beta}_1$ close to the true β_0 and β_1? Why should we expect that result? Plot the histograms of $\hat{\beta}_0$ and $\hat{\beta}_1$.

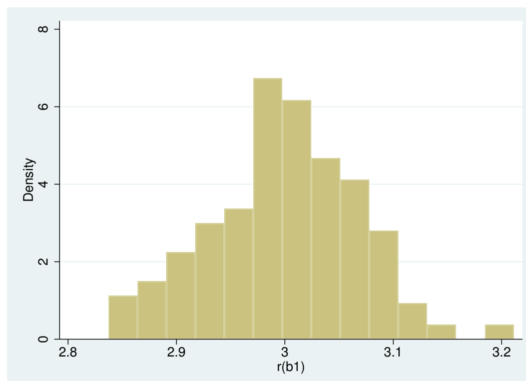
Answer

The averages and the standard deviations of $\hat{\beta}_0$ and $\hat{\beta}_1$ across the 200 simulations that we obtained are:

	Average	*Std.* Deviation
$\hat{\beta}_0$	1.983599	0.1009261
$\hat{\beta}_1$	2.999378	0.0676116

The averages of $\hat{\beta}_0$ and $\hat{\beta}_1$ are close to β_0 and β_1, as expected, since OLS is unbiased (noting that U and X are independent and U has a mean of zero, so $E[U|X] = 0$).
Histograms:





Code:

```
cap program drop regra2
program regra2, rclass
drop _all
set obs 200
gen x = rnormal(-1,1)
gen u = rnormal(0,1)
gen y = 2 + 3* x + u
reg y x
return scalar b0=_b[_cons]
return scalar b1=_b[x]
end
simulate "regra2" b0=r(b0) b1=r(b1), reps(200)
sum b0 b1
histogram b0
graph export "graphics/hist0.eps", replace
histogram b1
graph export "graphics/hist1.eps", replace
```

(c) *(points: 6)* With a bit of algebra, we can see that $X = -\frac{2}{3} - \frac{1}{3}Y - \frac{1}{3}U$. Now, again simulate the same model as above ($Y = \beta_0 + \beta_1 X + U$), 200 times, and for each simulation, generate a data set $\{y_i, x_i : i = 1, ..., n\}$ with $n = 400$ observations. This time, though, you should use OLS to estimate γ_0 and γ_1 in the regression $X = \gamma_0 + \gamma_1 Y + V$. (That is, regression X against Y rather than the other way around.) What are the averages and standard deviations of $\hat{\gamma}_0$ and $\hat{\gamma}_1$, the OLS estimates? Are $\hat{\gamma}_0$ and $\hat{\gamma}_1$ usually close to $-\frac{2}{3}$ and $-\frac{1}{3}$? Why should we expect that result?

Answer

The averages and the standard deviations of $\hat{\gamma}_0$ and $\hat{\gamma}_1$ across the 200 simulations that we obtained are:

	Average	*Std.* Deviation
$\hat{\gamma}_0$	-0.6989677	0.0168848
$\hat{\gamma}_1$	0.2999472	0.0051233

The averages are somewhat close to $-\frac{2}{3}$ and $\frac{1}{3}$. (Full credit for saying it's close or not close as long as your estimates are close to what I found in the table above; I didn't mean this to be a trick question.) We expect that these estimates will not be the same as those values because U and Y are correlated, so $E[U|Y] \neq 0$: we expect U to be higher when Y is higher. (You are not required to calculate $E[U|Y]$.)

By the way, one note you should be able to prove: $R^2 = \hat{\beta}_1 \hat{\gamma}_1$.

Code:

```
cap program drop regra3
program regra3, rclass
drop _all
set obs 400
gen x = rnormal(-1,1)
gen u = rnormal(0,1)
gen y = 2 + 3* x + u
reg x y
return scalar g0=_b[_cons]
return scalar g1=_b[y]
end
simulate "regra3" g0=r(g0) g1=r(g1), reps(200)
sum g0 g1
```

DO FILE

```
* Code for ECO375, homework 1

************************************

* Start the log
capture log close
capture noisily log using "code_hw1_2023.log", replace


************************************
```

```
clear all
set more off


************************************
* PROBLEM 4

use eco375hw1data2023, clear

* 4(a)
sum montreal vancouver toronto

* 4(b)
reg rain umbrella

* 4(c)
sum cold
gen aboveavgcold = 0
replace aboveavgcold = 1 if cold > 65.15
reg pizza aboveavgcold

* 4(d)
gen logpoutine = ln(poutine)
gen logmontreal = ln(montreal)
reg poutine logmontreal
reg logpoutine montreal
reg logpoutine logmontreal


************************************
* PROBLEM 5

* 5(a)
clear all
set seed 1234
cap program drop regra1
program regra1, rclass
drop _all
set obs 400
gen x = rnormal(-1,1)
gen u = rnormal(0,1)
gen y = 2 + 3* x + u
reg y x
return scalar b0=_b[_cons]
return scalar b1=_b[x]
end
regra1
```

```
scatter y x
graph export "graphics/scatter.eps", replace

* 5(b)
cap program drop regra2
program regra2, rclass
drop _all
set obs 200
gen x = rnormal(-1,1)
gen u = rnormal(0,1)
gen y = 2 + 3* x + u
reg y x
return scalar b0=_b[_cons]
return scalar b1=_b[x]
end
simulate "regra2" b0=r(b0) b1=r(b1), reps(200)
sum b0 b1
histogram b0
graph export "graphics/hist0.eps", replace
histogram b1
graph export "graphics/hist1.eps", replace

* 5(c)
cap program drop regra3
program regra3, rclass
drop _all
set obs 400
gen x = rnormal(-1,1)
gen u = rnormal(0,1)
gen y = 2 + 3* x + u
reg x y
return scalar g0=_b[_cons]
return scalar g1=_b[y]
end
simulate "regra3" g0=r(g0) g1=r(g1), reps(200)
sum g0 g1


************************************
* End the log

capture log close
```

LOG FILE

```
.
.
```

```
. ************************************
.
. clear all

. set more off

.
.
. ************************************
. * PROBLEM 4
.
. use eco375hw1data2023, clear

.
. * 4(a)
. sum montreal vancouver toronto

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
    montreal |         13    13.53846    26.13623          3        100
   vancouver |         13    12.46154    26.96793          2        100
     toronto |         13    19.53846    24.46976          7        100

.
. * 4(b)
. reg rain umbrella

      Source |       SS           df       MS      Number of obs   =        13
-------------+----------------------------------   F(1, 11)        =      8.28
       Model |  1263.46567         1  1263.46567   Prob > F        =    0.0150
    Residual |  1678.84202        11  152.622002   R-squared       =    0.4294
-------------+----------------------------------   Adj R-squared   =    0.3775
       Total |  2942.30769        12  245.192308   Root MSE        =    12.354

------------------------------------------------------------------------------
        rain | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
    umbrella |    .627965   .2182541     2.88   0.015     .1475911    1.108339
       _cons |   34.01729    16.0835     2.12   0.058    -1.382269    69.41684
------------------------------------------------------------------------------

.
. * 4(c)
. sum cold

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
        cold |         13    65.15385    17.09213         24        100
```

```
. gen aboveavgcold = 0

. replace aboveavgcold = 1 if cold > 65.15
(8 real changes made)

. reg pizza aboveavgcold

      Source |       SS           df       MS      Number of obs   =        13
-------------+----------------------------------   F(1, 11)        =      0.24
       Model |  96.4923077         1  96.4923077   Prob > F        =    0.6344
    Residual |      4437.2        11  403.381818   R-squared       =    0.0213
-------------+----------------------------------   Adj R-squared   =   -0.0677
       Total |  4533.69231        12  377.807692   Root MSE        =    20.084

------------------------------------------------------------------------------
       pizza | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
aboveavgcold |        5.6   11.44985     0.49   0.634    -19.60095    30.80095
       _cons |       72.4   8.982002     8.06   0.000     52.63075    92.16925
------------------------------------------------------------------------------


.
. * 4(d)
. gen logpoutine = ln(poutine)

. gen logmontreal = ln(montreal)

. reg poutine logmontreal

      Source |       SS           df       MS      Number of obs   =        13
-------------+----------------------------------   F(1, 11)        =     31.11
       Model |  4389.14202         1  4389.14202   Prob > F        =    0.0002
    Residual |  1551.78106        11  141.071006   R-squared       =    0.7388
-------------+----------------------------------   Adj R-squared   =    0.7151
       Total |  5940.92308        12  495.076923   Root MSE        =    11.877

------------------------------------------------------------------------------
     poutine | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
 logmontreal |    20.5043   3.675987     5.58   0.000     12.41351    28.59509
       _cons |   -10.1444   7.902576    -1.28   0.226    -27.53785    7.249055
------------------------------------------------------------------------------

. reg logpoutine montreal

      Source |       SS           df       MS      Number of obs   =        13
-------------+----------------------------------   F(1, 11)        =     26.71
```

```
       Model |  2.04324557         1  2.04324557   Prob > F        =    0.0003
    Residual |  .841432366        11  .076493851   R-squared       =    0.7083
-------------+----------------------------------   Adj R-squared   =    0.6818
       Total |  2.88467794        12  .240389828   Root MSE        =    .27658

------------------------------------------------------------------------------
  logpoutine | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
    montreal |    .015788   .0030548     5.17   0.000     .0090645    .0225115
       _cons |   3.040751   .0871467    34.89   0.000     2.848943     3.23256
------------------------------------------------------------------------------

. reg logpoutine logmontreal

      Source |       SS           df       MS      Number of obs   =        13
-------------+----------------------------------   F(1, 11)        =     17.56
       Model |  1.77362389         1  1.77362389   Prob > F        =    0.0015
    Residual |  1.11105405        11  .101004914   R-squared       =    0.6148
-------------+----------------------------------   Adj R-squared   =    0.5798
       Total |  2.88467794        12  .240389828   Root MSE        =    .31781

------------------------------------------------------------------------------
  logpoutine | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
 logmontreal |    .412179   .0983618     4.19   0.002     .1956862    .6286718
       _cons |   2.449057   .2114565    11.58   0.000     1.983644    2.914469
------------------------------------------------------------------------------

.
.
. ************************************
. * PROBLEM 5
.
. * 5(a)
. clear all

. set seed 1234

. cap program drop regra1

. program regra1, rclass
  1.         drop _all
  2.         set obs 400
  3.         gen x = rnormal(-1,1)
  4.         gen u = rnormal(0,1)
  5.         gen y = 2 + 3* x + u
  6.         reg y x
  7.         return scalar b0=_b[_cons]
```

```
  8.         return scalar b1=_b[x]
  9. end

. regra1
Number of observations (_N) was 0, now 400.

      Source |       SS           df       MS      Number of obs   =       400
-------------+----------------------------------   F(1, 398)       =   4073.75
       Model |  4102.83131         1  4102.83131   Prob > F        =    0.0000
    Residual |  400.840944       398  1.00713805   R-squared       =    0.9110
-------------+----------------------------------   Adj R-squared   =    0.9108
       Total |  4503.67226       399  11.2873991   Root MSE        =    1.0036

------------------------------------------------------------------------------
           y | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
           x |   3.068917   .0480826    63.83   0.000     2.974389    3.163444
       _cons |   2.126651   .0700762    30.35   0.000     1.988885    2.264416
------------------------------------------------------------------------------

. scatter y x

. graph export "graphics/scatter.eps", replace
file graphics/scatter.eps saved as EPS format

.
. * 5(b)
. cap program drop regra2

. program regra2, rclass
  1.         drop _all
  2.         set obs 200
  3.         gen x = rnormal(-1,1)
  4.         gen u = rnormal(0,1)
  5.         gen y = 2 + 3* x + u
  6.         reg y x
  7.         return scalar b0=_b[_cons]
  8.         return scalar b1=_b[x]
  9. end

. simulate "regra2" b0=r(b0) b1=r(b1), reps(200)

Command:      regra2
Statistics:   b0         = r(b0)
              b1         = r(b1)

. sum b0 b1
```

```
    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
          b0 |        200    1.983599    .1009261   1.680729   2.248271
          b1 |        200    2.999378    .0676116   2.837306   3.211396

. histogram b0
(bin=14, start=1.6807289, width=.04053872)

. graph export "graphics/hist0.eps", replace
file graphics/hist0.eps saved as EPS format

. histogram b1
(bin=14, start=2.8373065, width=.02672069)

. graph export "graphics/hist1.eps", replace
file graphics/hist1.eps saved as EPS format

.
. * 5(c)
. cap program drop regra3

. program regra3, rclass
  1.         drop _all
  2.         set obs 400
  3.         gen x = rnormal(-1,1)
  4.         gen u = rnormal(0,1)
  5.         gen y = 2 + 3* x + u
  6.         reg x y
  7.         return scalar g0=_b[_cons]
  8.         return scalar g1=_b[y]
  9. end

. simulate "regra3" g0=r(g0) g1=r(g1), reps(200)

Command:      regra3
Statistics:   g0          = r(g0)
              g1          = r(g1)

. sum g0 g1

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
          g0 |        200   -.6989677    .0168848  -.7387857  -.6438376
          g1 |        200    .2999472    .0051233   .2885587   .3156957

.
.
. ************************************
```

```
. * End the log
.
. capture log close
```