

FIRST MIDTERM EXAMINATION  
ECON 103, STATISTICS FOR ECONOMISTS

SEPTEMBER 28TH, 2016

**You will have 90 minutes to complete this exam. Graphing calculators, notes, and textbooks are not permitted.**

I pledge that, in taking and preparing for this exam, I have abided by the University of Pennsylvania's Code of Academic Integrity. I am aware that any violations of the code will result in a failing grade for this course.

**Please sign the back of your blue book.**

Question:	1	2	3	4	5	6	7	Total
Points:	10	10	18	10	20	18	14	100
Score:								

**Instructions:** Answer all questions in your blue book. Show your work for full credit but be aware that writing down irrelevant information will not gain you points. Be sure to sign the academic integrity statement in the back of your blue book. Make sure that you have all pages of the exam before starting.

**Warning:** If you continue writing after I call time, even if this is only to fill in your name, twenty-five points will be deducted from your final score. In addition, ten points will be deducted for not signing the back of your blue book.

**CHECKLIST BEFORE CONTINUING:**

- ☐ My name and Penn ID number are on my blue book
- ☐ I have signed the academic integrity statement in the back of the blue book

**1. True or False? If false, briefly explain why.**

- (a) [2 points] Nonsampling error will decrease as the sample size grows.

**Solution:** False. Sampling error will decrease as the sample size grows. Non-sampling error is caused by systematic differences between the sample and the population and as a result, will not decrease as the sample size grows large.

- (b) [2 points] Correlation implies causation.

**Solution:** False. Correlation is simply a measure of whether two variables move in related ways. It does not indicate whether or not there is any kind of causal relationship between the two variables. Subject to confounders.

- (c) [2 points] Statistics uses characteristics of the population to make predictions about the sample.

**Solution:** False. That is what probability does. Statistics uses characteristics of the sample to make predictions about the population from which it was drawn.

- (d) [2 points] Gender is an ordinal variable.

**Solution:** False. Gender is a categorical or nominal variable. There is no natural ordering to gender.

- (e) [2 points] Standard deviation is preferred to variance because it is unitless.

**Solution:** False. Standard deviation is preferred because it takes on the same units as the data. It is not unitless.

**2. Randomized trials**

- (a) [5 points] Describe a double-blind randomized controlled trial.

**Solution:** A double-blind randomized controlled trial consists of creating two groups that are expected to be equivalent across all measures [+2]. Then, treatment is assigned to one group and the control group does not get the treatment [+1]. However, both groups are blind as to whether they are actually receiving the treatment or not (the first “blind” part) [+1].

Then, the experimenters are also blind to which group received the treatment (the second “blind” part) [+1].

- (b) [5 points] Discuss why double-blind randomized controlled trials are superior to observational methods at determining effectiveness or causation.

**Solution:** Observational methods are subject to confounding effects which can cloud the analysis or determination of causation [+2].

RCTs prevent that because by dividing the subjects into two observationally equivalent groups (through randomization), the idea is that any confounding factors would also be even split between the groups, so that would not cause differences between the groups [+2].

Since the confounders have been effectively neutralized, any difference between the groups can be attributed to the treatment [+1].

3. **Summary Statistics:** Lake Wobegone is a “small rural town in central Minnesota where all the women are strong, all the men are good looking, and all the children are above average.” Let’s assume that Lake Wobegone has 100 children living there and children are scored on an integer scale ranging from 1 to 100.

- (a) [7 points] Is it possible that every child in Lake Wobegone is above the *national average*? Why or why not? If not, what is the maximum number of Wobegoner children that can be above the national average? Give an example supporting your solution.

**Solution:** Yes, it is possible that every child in Lake Wobegone is above the national average [+1].

Since there are more children in America than just in Lake Wobegone, all Wobegoner children can be above average without causing a contradiction. In particular, there are millions of children in the United States, so the performance of a small subsample will not meaningfully affect the national average. [+3].

One example would be that we have all 100 Wobegoner children score 100 and 100 other children score 50. The national average would be 75, but all Wobegoner children would be above average [+3].

- (b) [7 points] Is it possible that every child in Lake Wobegone is above the *average of children in Lake Wobegone*? Why or why not? If not, what is the maximum number of Wobegoner children that can be above the Lake Wobegone average? Give an example supporting your solution.

**Solution:** No, it is not possible that every child in Lake Wobegone can be above the Lake Wobegone average **[+1]**.

If every child was above average (where  $\mu$  is the average), the aggregate score would be at least  $100(\mu + 1)$  while the aggregate score of all children in Lake Wobegone would be  $100\mu$ . Since the populations are the same in this instance, we cannot get different aggregate scores and hence we have a contradiction **[+3]**. The maximum number of children that can be above average would be 99 since we just need one really low score to pull the average down below the average of all other children. We can see this immediately if we have 99 children scoring 100 and one child scoring a 1. the average would be 99, but 99 of the children have scored above average. **[+3]**

- (c) [4 points] Is it possible that every child in Lake Wobegone is above the *median of children in Lake Wobegone*? Why or why not? If not, what is the maximum number of Wobegoner children that can be above the Lake Wobegone median?

**Solution:** The median is the middle observation when observations are ranked in order. For even numbers of observations, it is the average of the middle two observations **[+2]**.

The maximum number of children that could be above the median is 50. Actually, by definition, this will always be the number of children above the median **[+2]**.

4. [10 points] **Regression:** Jay, an economist, is fitting a model of GDP growth and investment growth.

$$y_i = \beta * x_i^2 + e_i$$

where  $y$  is GDP growth and  $x$  is investment growth. Derive the formula for the OLS estimator  $\beta$ . Clearly state the optimization problem and then solve it.

**Solution:** The OLS estimator minimizes the sum of squared deviations

$$\min_{\beta} \sum_{i=1}^n (y_i - \beta x_i^2)^2$$

**[+4, -1 for not specifying what parameter we are minimizing over, -2 for not specifying that we are minimizing over the sum of squared deviations]**

Differentiating with respect to  $\beta$  gives us

$$-2 \sum_{i=1}^n (y_i - \beta x_i^2) x_i^2 = 0$$

**[+3, no credit for differentiating with respect to  $x$  or  $y$ ]**

Solving for  $\beta$  gives us

$$\beta = \frac{\sum_{i=1}^n y_i x_i^2}{\sum_{i=1}^n x_i^4}$$

**[+3, will give partial credit for solving correctly even if differentiation was incorrect]**

5. **Presidential Election Predictions:** According to *The Washington Post*, Professor Allan Lichtman, a political historian at American University, has correctly predicted every presidential election since 1984 based on his “Keys to the White House” model. Since the United States holds presidential elections every 4 years, that means he has correctly predicted 8 presidential elections.

- (a) [2 points] If Professor Lichtman had been randomly guessing the winner of presidential elections, what is the probability that he would have gotten 8 right in a row? Assume there are only two candidates in each election and he flips a fair coin to decide his guess for the winner.

**Solution:** This is simply the probability that he gets a particular sequence of heads and tails. The probability of getting a specific sequence of heads and tails in eight flips is simply  $\frac{1}{2^8} = \frac{1}{256} \approx 0.4\%$  **[+1 point for the right probability and +1 for the right exponent]**

- (b) [8 points] Across the nation, professors try to predict presidential outcomes. Let’s say that 100 professors nationwide are trying to predict presidential elections. What is the chance that at least one professor has called the past eight elections correctly, assuming each adopts the same extensively-researched coin-flipping forecasting strategy? Computing the answer is not required for full credit, but your answer should be an expression that could then be evaluated on a calculator.

**Solution:** The chance that at least one professor has called each election correctly can be found using the Complement Rule. This probability will be the complement to the probability that every professor was wrong. Mathematically, this is

$$P(\text{At Least One Right}) = 1 - P(\text{No One Right})$$

The probability that one professor is wrong is  $1 - P(\text{Right}) = 1 - \frac{1}{256} = \frac{255}{256}$  (from part (a)).

Since each professor is flipping their coins independently, the chance that all

100 are wrong is  $(\frac{255}{256})^{100} \approx 68\%$ .

Hence, the chance that at least one professor was right is  $1 - (\frac{255}{256})^{100} \approx 32\%$   
**[+2 for each step, -2 for forgetting Complement Rule]**

- (c) [5 points] There are likely more than 100 professors predicting presidential elections (not counting private forecasting companies and individuals). If there are  $n$  professors trying to predict elections, what is the expression for how we would calculate the chance that at least one is right at predicting the past eight elections?

**Solution:** Using the same explanation as above, based on the Complement Rule, we know that

$$P(\text{At Least One Right}) = 1 - P(\text{No One Right})$$

The probability that one professor is wrong is simply  $\frac{255}{256}$  (from part (a)). Since each professor is flipping their coins independently, the chance that all 100 are wrong is  $(\frac{255}{256})^n$ . Hence, the chance that at least one professor was right is  $1 - (\frac{255}{256})^n$  **[+1 for each of the steps, -2 for forgetting Complement Rule]**

- (d) [5 points] In reality, people do not use coin flips to make their predictions (or at least they do not admit to it). We have  $n$  people across the country making predictions for the election. Assuming each model has probability  $p$  of predicting each election correctly and each model is independent of the others, what is the probability that at least one model has predicted the past 8 elections correctly?

**Solution:** Using similar logic as outlined above, the Complement Rule is essential to solving this problem. Generalizing the solutions to the previous parts, we can express this probability as  $1 - (1 - p^8)^n$ .  $p^8$  is the chance that a person has been right for the past 8 elections.  $1 - p^8$  is then the probability that a person has been wrong at least once over the past 8 elections.  $(1 - p^8)^n$  is the probability that all  $n$  people have been wrong at least once in their predictions. Finally  $1 - (1 - p^8)^n$  is the probability that at least one person has been right over the past 8 elections. **[+1 for each step, -2 for forgetting Complement Rule]**

6. **R Programming and Statistical Analysis:** Sandy, an economist, is researching the gender-wage gap.

- (a) [3 points] Before starting her analysis, she wants to make sure R is working properly. She types the following into R. What does she expect R to return?

```
x <- 2
y <- 1
z <- x * y
z + 2 == 4
```

**Solution:** TRUE

- (b) [3 points] She downloads the data and stores it in a data table called `wageData`. She then types `head(wageData)` and the following is displayed:

	Gender	Wage	Age
1:	F	20	25
2:	M	18	26
3:	F	10	27
4:	F	NA	28
5:	M	15	29
6:	F	18	30

She then types `wageData[, mean(Wage, na.rm = FALSE), by = Gender]`. What is she trying to compute?

**Solution:** She is calculating the average wage by gender in her data set.

- (c) [3 points] What will R return and why? How should she modify her code to fix this?

**Solution:** R will return NA because the NA's were included in the mean calculation. **[+2 for knowing it would give NA]**

In order to get the mean without an error, she will have to ignore the NA's and she should change her code to read `wageData[Gender == "F", mean(Wage, na.rm = TRUE)]` **[+1 for changing to TRUE]**

- (d) [2 points] She then runs `lm(data = wageData, Wage ~ Gender)`. What does the `lm()` command do? If you do not remember what this command does, I will also accept the R code that would pull up the documentation.

**Solution:** This command performs a linear regression for the model  $Wage_i = a + b \cdot Gender_i$ . The code to pull up the documentation is either `help(lm)` or `?lm`.

- (e) [5 points] Sandy wants to estimate the following regression  $Wage_i = a + b \cdot Age_i$ . In this regression how are  $b$  and the correlation related? Do not simply state the

formula. Start with one of the equations for  $b$  and demonstrate how to rearrange it to include the correlation.

**Solution:** If you start with the following equation for  $b$ ,

$$b = \frac{\sum_{i=1}^n (x - \bar{x})(y - \bar{y})}{\sum_{i=1}^n (x - \bar{x})^2}$$

Then, you can divide the top and the bottom by  $\frac{1}{n-1}$  to get

$$b = \frac{s_{xy}}{s_x^2}$$

[+3]

Then, using the definition of correlation  $r = \frac{s_{xy}}{s_x s_y}$ , you can multiply the above expression by  $\frac{s_y}{s_y}$  to get

$$b = \frac{s_{xy}s_y}{s_y s_x^2} = r \frac{s_y}{s_x}$$

[+2]

(f) [2 points] She runs the following commands in R

```
> cov(wageData$Age, wageData$Wage, use = "complete.obs")
[1] -1.6
> cor(wageData$Age, wageData$Wage, use = "complete.obs")
[1] -0.1979083
> var(wageData$Age, na.rm = TRUE)
[1] 3.5
> var(wageData$Wage, na.rm = TRUE)
[1] 15.2
> mean(wageData$Age, na.rm = TRUE)
[1] 27.5
> mean(wageData$Wage, na.rm = TRUE)
[1] 16.2
```

If she were to estimate the following regression  $Wage_i = a + b \cdot Age_i$ , what would her coefficients be? You do not have to calculate the exact numbers, but you should have an expression that can be computed using a calculator.

**Solution:** We know that  $b = \frac{s_{xy}}{s_x}$ . Since we were given the covariance and the variance, we find that  $b = \frac{-1.6}{\sqrt{3.5}}$ . We also know that  $a = \bar{y} - b\bar{x}$ . Hence, we find that  $a = 16.2 - \frac{-1.6}{\sqrt{3.5}} * 27.5$  [+1 for  $a$ , +1 for  $b$ ]



## 7. Probability:

- (a) [2 points] State the mathematical definition of conditional probability

**Solution:** Conditional probability is simply  $P(A|B) = \frac{P(A \cap B)}{P(B)}$

- (b) [2 points] State Bayes' rule (mathematically).

**Solution:** Bayes' rule is  $P(B|A) = \frac{P(A|B) * P(B)}{P(A)}$

- (c) [10 points] In "Silver Blaze," Sherlock Holmes is investigating the mysterious disappearance of a race horse and the murder of its trainer. Near the stables, there is a watchdog which will bark whenever a stranger approaches. However, at night, the dog is only 80% accurate. That is, if a stranger approaches, 80% of the time it will bark and 20% of the time it will stay silent. If a familiar person approaches, the dog will stay silent 80% of the time and bark 20% of the time. Let's assume that 90% of the suspects are familiar to the dog and 10% are strangers.

One of the most famous Sherlock Holmes exchanges is below:

**Gregory (Scotland Yard Detective):** Is there any other point to which you wish to draw my attention?

**Holmes:** To the curious incident of the dog in the night-time.

**Gregory:** The dog did nothing in the night-time.

**Holmes:** That was the curious incident.

If a stranger had committed this crime, was this actually a curious incident? Compute the corresponding probability that a stranger committed the crime using the information that the dog did not bark.

**Solution:** We are asked to compute  $P(\text{stranger}|\text{silent})$ . In order to compute this probability, we must use Bayes' rule, which gives us the following relationship:

$$P(\text{stranger}|\text{silent}) = \frac{P(\text{silent}|\text{stranger}) * P(\text{stranger})}{P(\text{silent})}$$

**[+2 for correct specification of Bayes' rule]**

From the information given in the question, we know that

$$P(\text{silent}|\text{stranger}) = 0.2$$

$$P(\text{stranger}) = 0.1$$

**[+2 for proper use of given information]**

The last piece of information we need to get is the  $P(\text{silent})$ . In order to get this, we must use the Law of Total Probability, which gives us the following relationship

$$\begin{aligned} P(\text{silent}) &= P(\text{silent}|\text{stranger}) * P(\text{stranger}) + P(\text{silent}|\text{familiar}) * P(\text{familiar}) \\ &= 0.2 * 0.1 + 0.8 * 0.9 = 0.74 \end{aligned}$$

**[+2 for specifying use of the Law of Total Probability, +2 for properly using it]**

Combining all of these together, we get

$$P(\text{stranger}|\text{silent}) = \frac{P(\text{silent}|\text{stranger}) * P(\text{stranger})}{P(\text{silent})} = \frac{0.2 * 0.1}{0.74} \approx 2.7\%$$

So yes, if a stranger had committed the crime, this was quite a curious incident.

**[+2 for properly computing using Bayes' rule]**