

FINAL EXAMINATION
ECON 103, STATISTICS FOR ECONOMISTS

MAY 7TH, 2014

You will have 120 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.

Name: _____

Student ID #: _____

Signature: _____

Question:	1	2	3	4	5	6	Total
Points:	30	20	45	45	20	40	200
Score:							

Instructions: Answer all questions in the space provided, continuing on the back of the page if you run out of space. 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 above and to write your name and student ID number on *each page* in the space provided. Make sure that you have all pages of the exam before starting.

Warning: If you continue writing after we call time, even if this is only to fill in your name, twenty-five points will be deducted from your final score. In addition, two points will be deducted for each page on which you do not write your name and student ID.

1. Let $X_1, X_2, X_3 \stackrel{iid}{\sim} \text{Bernoulli}(1/2)$ and define $Y = X_1 + X_2$ and $Z = X_2 \cdot X_3$.

- (a) (5 points) What is the support of Y ? What is its pmf? Is it a “named” random variable? If so, which one and what are its parameters? Explain briefly.

Solution: Since it is the sum of two independent Bernoulli(1/2) RVs we know from class that $Y \sim \text{Binomial}(n = 2, p = 1/2)$. Its support is $\{0, 1, 2\}$ and its pmf is $p_Y(y) = \binom{2}{y}(1/4)^2$. That is: $p_Y(0) = 1/4$, $p_Y(1) = 1/2$, $p_Y(2) = 1/4$.

- (b) (5 points) Calculate $E[(Y - 1)^2]$.

Solution:

$$\begin{aligned} E[(Y - 1)^2] &= \sum_{\text{all } y} (y - 1)^2 p_Y(y) \\ &= (0 - 1)^2 p_Y(0) + (1 - 1)^2 p_Y(1) + (2 - 1)^2 p_Y(2) \\ &= p_Y(0) + p_Y(2) = 1/4 + 1/4 = 1/2 \end{aligned}$$

- (c) (5 points) What is the support of Z ? What is its pmf? Is it a “named” random variable? If so, which one and what are its parameters? Explain briefly.

Solution: The support of Z is $\{0, 1\}$. Since $Z = 1$ precisely when $X_2 = X_3 = 1$ and X_2, X_3 are independent Bernoulli(1/2), it follows that $p_Z(1) = 1/4$. By the Complement Rule $p_Z(0) = 1 - 1/4 = 3/4$. Thus Z is a Bernoulli(1/4) RV.

- (d) (15 points) Express the joint pmf of Y and Z in tabular form. Place the realizations for Y in the *rows* of the table and the realizations for Z in the *columns*.

Solution: First make a table of all possible realizations for X_1, X_2, X_3 and use it to calculate the corresponding realizations for Y and Z :

X_1	X_2	X_3	Y	Z
0	0	0	0	0
1	0	0	1	0
0	1	0	1	0
0	0	1	0	0
1	1	0	2	0
1	0	1	1	0
0	1	1	1	1
1	1	1	2	1

Since X_1, X_2, X_3 are iid Bernoulli(p), each row of the preceding table is *equally likely* to occur. Thus, we simply need to count up the number of times that each pair of realizations (y, z) occurs and divide by eight to get the joint pmf:

		Z	
		0	1
Y	0	2/8	0
	1	3/8	1/8
	2	1/8	1/8

2. For each statement, indicate whether it is TRUE or FALSE and briefly explain why.

(a) (4 points) For any two RVs X and Y , $E[XY] = Cov(X, Y) - E[X]E[Y]$.

Solution: FALSE. The shortcut rule for covariance is $Cov(X, Y) = E[XY] - E[X]E[Y]$. Rearranging gives $E[XY] = Cov(X, Y) + E[X]E[Y]$. The expression in the problem statement has the wrong sign on $E[X]E[Y]$.

(b) (4 points) If $Z \sim N(0, 1)$ then $P(Z = 0) > P(Z = 5)$.

Solution: FALSE. The probability that a continuous random variable takes on *any particular value* is always zero. Thus $P(Z = 0) = P(Z = 5) = 0$.

(c) (4 points) If I run the command `x <- rnorm(1)` at the R console followed by `replicate(100, x)` the result will be 100 iid $N(0, 1)$ random draws.

Solution: FALSE. This sequence of commands makes a *single* standard normal draw and then prints it out 100 times!

(d) (4 points) I rejected the null hypothesis $H_0: \mu = 5$ against the two-sided alternative $H_1: \mu \neq 5$ with $\alpha = 0.05$. Based on this result, I know that if I were to construct a 95% confidence interval for μ it would not contain zero.

Solution: FALSE. The result of the test tells us that the CI cannot include 5, but tells us nothing about the value zero.

(e) (4 points) The p-value for my test was 0.03. This means that if I had set $\alpha = 0.05$ I would have rejected the null hypothesis.

Solution: TRUE. The p-value is the *minimum* significance level at which we would reject the null. Since $0.05 > 0.03$ we would have rejected at the 5% level.

3. This question is based on a dataframe called **profs** containing the salaries of a random sample of 397 professors at US universities in 2008. Here are the first few rows:

```
> head(profs)
      rank sex salary
1    Prof Male   140
2    Prof Male   173
3 AsstProf Male    80
4    Prof Male   115
5    Prof Male   142
6 AssocProf Male    97
```

Each row contains data for a single professor. The column **rank** is a categorical variable indicating a professor's rank: **Prof** stands for "Full Professor," while **AssocProf** and **AsstProf** stand for "Associate" and "Assistant" Professor, respectively. The column **sex** indicates a professor's sex, **Male** or **Female**, while **salary** gives her salary measured in thousands of US dollars. There are no missing observations.

- (a) (10 points) Write R code to carry out the following operations *separately* for each sex: count the number of professors, calculate the sample mean of **salary**, and calculate the sample variance of **salary**.

Solution: There are many ways to do this. Here's one:

```
male <- subset(profs, sex == "Male")$salary
mean(male)
var(male)
length(male)
female <- subset(profs, sex == "Female")$salary
mean(female)
var(female)
length(female)
```

- (b) (10 points) The results of the preceding part are summarized in the following table

	Female	Male
Sample Size	39	358
Sample Mean	101	115
Sample Variance	676	926

Name: _____

Student ID #: _____

Use this information to construct an approximate 95% confidence interval for the difference of population mean salaries for male and female professors in the US.

Solution: The difference of means is $115 - 101 = 14$ and the standard error is

$$SE = \sqrt{926/358 + 676/39} \approx 4.5$$

so the approximate 95% CI is given by 14 ± 9 or equivalently (9, 23).

- (c) (10 points) Suppose that, rather than constructing a confidence interval, we wanted to carry out a two-sided test of the null hypothesis that the population mean salary is equal across male and female professors. What is the value of your test statistic? What R command would you use to calculate the p-value? Approximately what value would you get if you ran this command?

Solution: The test statistic is the difference of means divided by the standard error: $14/4.5 \approx 3.1$. To calculate the two-sided p-value, we'd use the R command `2 * (1 - pnorm(3.1))`. Since this is the same thing as calculating the area outside of $(-3.1, 3.1)$ under a standard normal distribution, the p-value is very small: less than 0.01. We know this because the probability that a normal RV takes on a value more than 3 std. devs. from its mean is less than 1%.

- (d) (15 points) The following table *breaks down* the results of part (b) by **rank**. It also provides approximate 95% CIs for the difference of means (**Male - Female**):

	Sample Size		Sample Mean		Sample Var.		Approx. 95% CI (Male - Female)
	Female	Male	Female	Male	Female	Male	
AsstProf	11	56	78	81	88	63	3 ± 6
AssocProf	10	54	89	95	327	165	6 ± 12
Prof	18	248	122	127	388	796	5 ± 10

Explain these results. How do they compare to your confidence interval and p-value from parts (b) and (c)? If you find differences, can you propose an explanation? What overall conclusions, if any, can we draw from this dataset?

Solution: When comparing male and female professor salaries across *all* ranks, we found a large and statistically significant difference. In particular, our CI was (\$9000, \$23000) per year. Breaking the comparison down by **rank** we still find that women earn less, on average, in each rank but the differences are not even close to statistically significant. Two things are going on here. First, we see from the table given above that women are disproportionately represented

among the two lower-paying professor ranks: Assistant and Associate. Thus, **rank** was a *confounded* in our initial comparison. However, there's a second effect. Even though the sample mean is lower for women in each **rank**, there are so few women in each of the three categories that we can't get a very precise estimate for the difference of means: the standard errors are large which widens the intervals. It certainly appears that female professors earn less, which could potentially be evidence of discrimination, but we simply don't have enough women in our dataset to get a definitive answer. Ironically, the fact that there are few women in academia, *itself* potential evidence of discrimination, makes it harder to detect if women face salary discrimination!

4. In the basement of the McNeil building I have a storeroom containing thousands of digital scales. I know from past experience that some scales from this manufacturer are defective: they give *biased* readings. All of the scales, both *biased* and *unbiased*, are imperfect in that they make small, independent, normally distributed errors with a standard deviation of two grams. This means that if you weigh the same object repeatedly, you will get a slightly different result each time and the average deviation from the mean will be about two grams. For *unbiased* scales, the errors have zero mean: if you weigh the same object a very large number of times on an *unbiased* scale, the average weight will equal the true weight. For *biased* scales, on the other hand, the errors have mean one gram: if you weigh the same object a very large number of times on a *biased* scale, the average weight will be *one gram higher* than the true weight. For each scale in my storeroom, I carry out the following procedure. First I weigh the same 10 gram mass *repeatedly* and record the results. In total, I make 16 measurements. These constitute a random sample from the following population:

$$X_1, \dots, X_{16} \stackrel{iid}{\sim} N(\mu, \sigma^2 = 4)$$

If the scale I'm testing is *unbiased*, then $\mu = 10$, the true weight. If the scale is *biased* then $\mu = 11$, one gram *higher* than the true weight. I then carry out an *exact* test of the null hypothesis $H_0: \mu = 10$ against the *one-sided* alternative $H_1: \mu > 10$ with $\alpha = 0.025$. If I reject the null, I decide that the scale is defective and throw it away. Otherwise I keep it for use in the candy weighing experiment in Econ 103.

- (a) (10 points) What formula should I use for my test statistic? Approximately what is my critical value? What is my decision rule?

Solution:

$$T_n = \frac{\bar{X} - \mu_0}{SE(\bar{X})} = \frac{\bar{X} - 10}{2/\sqrt{16}} = 2(\bar{X} - 10)$$

The critical value is $\text{qnorm}(0.975) \approx 2$ and my decision rule is to reject the null when $T_n > 2$. Another way of putting this is that we reject the null if $\bar{X} > 11$.

- (b) (5 points) Suppose I'm testing one of the scales and, unbeknownst to me, it happens to be one of the *unbiased* ones. What is the probability that I will reject the null? Explain your answer.

Solution: It is simply 0.025. If the scale is unbiased, then the null is true. By construction, the probability of rejecting a true null is α .

- (c) (15 points) Suppose I'm testing one of the scales and, unbeknownst to me, it happens to be one of the *biased* ones. What is the probability that I will reject the null? Explain your answer.

Solution: If the scale is defective, then $\mu = 11$ so $X_1, \dots, X_{16} \stackrel{iid}{\sim} N(11, \sigma^2 = 4)$ and hence $\bar{X} \sim N(11, \sigma^2 = 1/4)$. Since $T_n = 2(\bar{X} - 10) = 2\bar{X} - 20$, using the fact that linear combinations of normals are normal we find that $T_n \sim N(2, 1)$. We reject the null whenever $T_n > 2$. Since the $N(2, 1)$ RV is symmetric around 2, the probability that this will occur if the scale is in fact biased is exactly $1/2$.

- (d) (5 points) If I wanted to increase the probability that I will successfully detect biased scales, how should I change my testing procedure? Explain.

Solution: This question is equivalent to asking "how can I increase the power of my test?" In general, three things affect the power of a test: sample size, population standard deviation, and α . In this example, the population standard deviation is the measurement error which is outside my control, so I can only adjust α and sample size. If I take *more* than 16 measurements, the power of my test will increase. Similarly, if I increase α , I will use a less stringent critical value and this will also increase the power of my test.

- (e) (5 points) From past experience, I know that roughly 10% of scales produced by this manufacturer are *biased*. If I test a large number of scales, approximately what percentage of them will I end up throwing away? Explain.

Solution: By the Law of Total Probability,

$$\begin{aligned} P(\text{Reject}) &= P(\text{Reject}|\text{Biased})P(\text{Biased}) + P(\text{Reject}|\text{Unbiased})P(\text{Unbiased}) \\ &= 1/2 \times 1/10 + 25/1000 \times 9/10 \\ &= 725/10000 = 7.25\% \end{aligned}$$

- (f) (5 points) Continuing from the previous part, among the scales that I end up throwing away, approximately what percentage of them will *actually* be biased? Explain.

Solution: By Bayes' Rule,

$$\begin{aligned} P(\text{Biased}|\text{Reject}) &= \frac{P(\text{Reject}|\text{Biased})P(\text{Biased})}{P(\text{Reject})} \\ &= \frac{1/20}{725/10000} = 100/145 \approx 69\% \end{aligned}$$

5. (20 points) Suppose that $X_1, \dots, X_n \stackrel{iid}{\sim} \text{Bernoulli}(p)$, and that n is sufficiently large that the sampling distribution of \hat{p} is approximately normal by the CLT. Write an R function called `prop.test` that returns the p-value for the *refined test* of $H_0: p = p_0$ against $H_1: p \neq p_0$. Your function should take two arguments. The first, `x`, is the sample data: a vector of zeros and ones corresponding to the realizations x_1, \dots, x_n of X_1, \dots, X_n . The second is `p.0`, a value between zero and one that indicates *which* null hypothesis the user wishes to test. For example if `p.0 = 0.5` then we are testing $H_0: p = 0.5$. You may assume that there are no missing values in `x`.

Solution:

```
prop.test <- function(x, p.0){
  p.hat <- mean(x)
  SE.0 <- sqrt(p.0 * (1 - p.0) / length(x))
  test.stat <- abs((p.hat - p.0) / SE.0)
  p.value <- 2 * (1 - pnorm(test.stat))
  return(p.value)
}
```

6. Garth is concerned that students registered for Monday recitation sections of Econ 103

may have an unfair advantage on quizzes because they have additional days to study compared to students in Friday sections. To decide whether he should adjust the course curve to compensate, he fits a number of linear regression models, the results of which appear in Table 1. Each regression is calculated using a dataframe called `gradebook`:

	quiz.avg	midterm.avg	Monday
1	71	85.8	1
2	80	74.6	1
.	.	.	.
.	.	.	.
.	.	.	.
82	58	66.3	0
83	73	96.1	1

(To preserve privacy the preceding values are fake but the regression results given below are based on the real course gradebook.) Each row of `gradebook` corresponds to a student in Econ 103. The column `quiz.avg` contains that student's overall quiz average in percentage points, while `midterm.avg` contains the average of her scores on the two midterms, also in percentage points. Finally, `Monday` is a dummy variable that takes on the value one for students in Monday recitation sections, zero otherwise.

- (a) (5 points) Which set of regression results allows us to determine the sample mean of `quiz.avg` for Monday and Friday students separately? What is the sample mean of each group? What is the difference of means?

Solution: Based on the results of Regression 2, the sample mean of `quiz.avg` is approximately 63.8% for students registered for Friday recitations and 68.6% for students registered for Monday recitations. On average, Monday students do a little under 5 percentage points better.

- (b) (5 points) Continuing from the previous part, construct a 95% confidence interval for the difference of means between Monday and Friday students. Is the difference statistically significant at the 5% level? Explain briefly.

Solution: From the results for Regression 3, the difference of sample means is approximately 4.9 with a standard error of about 2.7. Thus, an approximate 95% confidence interval for the difference of population means is 4.9 ± 5.4 , in other words $(-0.5, 10.3)$. This confidence interval includes zero, but just barely: although the observed difference is not quite statistically significant at the 5% level, it is very nearly so.

- (c) (5 points) What is the sample correlation between `quiz.avg` and `midterm.avg`? Approximately how accurately does a simple linear regression model with the predictor `midterm.avg` *alone* predict `quiz.avg`?

Solution: The R-squared for Regression 3 is 0.21. Taking the square root, we find that the sample correlation is approximately 0.46. The residual standard deviation is just over 10, so this model predicts `quiz.avg` to an accuracy of approximately 10 percentage points.

- (d) (10 points) Garth fears that confounders could be responsible for the difference in average quiz scores between Monday and Friday recitation sections. For example, if Differential Equations meets on Friday from 10-11:30am, then students taking this course will be *forced* to register for a Monday recitation section of Econ 103. Since taking difficult math courses is likely correlated with greater mathematical ability and hence grades in Econ 103, this could bias the comparison. Accordingly, Garth decides to *control* for differences in students' ability using `midterm.avg`, fitting a regression in which `quiz.avg` is linearly related to `midterm.avg` but the intercept is allowed to *differ* for Monday and Friday students. The slope is *not* allowed to vary across groups. Which of the four regression models given below contains the results? Explain the fitted relationships given by the model and briefly discuss the results in light of Garth's concerns.

Solution: The results appear in Regression 4. For students registered for a Friday recitation, the fitted linear relationship is approximately

$$\text{quiz.avg} = 39.75 + 0.35 \times \text{midterm.avg}$$

whereas for students registered for a Monday recitation it is

$$\text{quiz.avg} = 43.42 + 0.35 \times \text{midterm.avg}$$

Thus, for two students with the *same* value for `midterm.avg` we would predict that the student in the Monday recitation will have a `quiz.avg` that is approximately 3.7 points *higher*. We see that controlling for ability via `midterm.avg` results in a *smaller* predicted difference between Monday and Friday recitation sections: 3.7 versus 4.9 points. Constructing a 95% confidence interval for the difference of intercepts, we have 3.67 ± 4.8 or equivalently $(-1.13, 8.47)$. There is still some tenuous evidence suggesting that Monday recitations perform better, but it is much weaker after controlling for `midterm.avg`.

- (e) (15 points) Before making a final decision about whether to adjust the recitation

quiz scores, Garth decides to try one more thing. Rather than comparing Monday recitations to Friday recitations, he calculates the sample mean of `quiz.avg` and the associated standard error for *each* recitation section:

FridayA	MondayA	MondayB	FridayB	MondayC
60.25	66.33	67.73	68.50	71.17
(2.4)	(2.6)	(2.3)	(4.3)	(2.2)

The standard errors appear in parentheses. (*The letter designations are arbitrary: for privacy reasons I don't want to reveal precisely which recitation section is which.*) Based on this new information, the full set of regression results, and your answers to the preceding parts, do you think Garth should adjust `quiz.avg` upwards for students registered for Friday recitation sections? If so, by how much? Use what you've learned in Econ 103 to make your case.

Solution: It doesn't matter whether students come down *pro* or *con*: the point is to make sure that they say something reasonable, indicating that they understand the statistical results presented in the question. At a minimum however, they should note that the difference of means between Monday and Friday recitations is being driven almost entirely by **FridayA**. Indeed, **FridayB** had the second highest average overall. Ideally, they should note that the differences in means between recitation sections are small compared to the standard errors, although I haven't given enough information to construct CIs for the differences.

Table 1: Regression Results

Regression 1:

```
lm(formula = quiz.avg ~ Monday + midterm.avg + Monday:midterm.avg,  
    data = gradebook)
```

	coef.est	coef.se
(Intercept)	39.19	10.39
Monday	4.46	12.48
midterm.avg	0.36	0.15
Monday:midterm.avg	-0.01	0.18

n = 83, k = 4

residual sd = 10.15, R-Squared = 0.23

Regression 2:

```
lm(formula = quiz.avg ~ Monday, data = gradebook)
```

	coef.est	coef.se
(Intercept)	63.77	2.19
Monday	4.86	2.65

n = 83, k = 2

residual sd = 11.19, R-Squared = 0.04

Regression 3:

```
lm(formula = quiz.avg ~ midterm.avg, data = gradebook)
```

	coef.est	coef.se
(Intercept)	41.31	5.72
midterm.avg	0.37	0.08

n = 83, k = 2

residual sd = 10.17, R-Squared = 0.21

Regression 4:

```
lm(formula = quiz.avg ~ Monday + midterm.avg, data = gradebook)
```

	coef.est	coef.se
(Intercept)	39.75	5.77
Monday	3.67	2.40
midterm.avg	0.35	0.08

n = 83, k = 3

residual sd = 10.09, R-Squared = 0.23

Name: _____

Student ID #: _____