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 $X_1, X_2, X_3 \overset{iid}{\sim} \text{Bernoulli}(1/2)$ 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.

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

   (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.

   (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*.

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, $E[XY] = Cov(X, Y) - E[X]E[Y]$.

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

   (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.

   (d) (4 points) I rejected the null hypothesis $H_0\colon \mu = 5$ against the two-sided alternative $H_1\colon \mu \neq 5$ 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.

   (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.

3. This question is based on a data table 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`.

(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

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.

(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?

(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
	`Female`	`Male`	`Female`	`Male`	`Female`	`Male`	`(Male - Female)`
`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 to 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?

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 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:

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

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 $H_0\colon \mu = 10$ against the *one-sided* alternative $H_1\colon \mu > 10$ 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?

(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.

Name: ____________________ Student ID #: ____________________

(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.

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

(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.

(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.

5. (20 points) Suppose that $X_1, \ldots, X_n \overset{iid}{\sim} \text{Bernoulli}(p)$, and that n is sufficiently large that the sampling distribution of $\widehat{p}$ is approximately normal by the CLT. Write an R function called `prop.test` that returns the p-value for the *refined test* of $H_0\colon p = p_0$ against $H_1\colon p \neq p_0$. Your function should take two arguments. The first, `x`, is the sample data: a vector of zeros and ones corresponding to the realizations $x_1, \ldots, x_n$ of $X_1, \ldots, X_n$. 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 $H_0\colon p = 0.5$. You may assume that there are no missing values in `x`.

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 data table 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 students separately? What is the sample mean of each group? What is the difference of means?

(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.

(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`?

(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.

Name: ______________________ Student ID #: ______________________

(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.

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
```