Final Examination

Econ 103, Statistics for Economists

May 4th, 2015

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: _____

Signature: _____

Student ID #: _____ Recitation #: _____

Question:	1	2	3	4	5	Total
Points:	30	35	20	35	80	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. Mark each statement as TRUE or FALSE. If FALSE, provide a one sentence explanation.

[3] (a) All else equal, a 99% confidence interval is narrower than a 95% interval.

[3] (b) A p-value > 0.05 implies that we would reject the null hypothesis with $\alpha = 0.05$.

[3] (c) If you reject H_0 with $\alpha = 0.05$, you would also have rejected with $\alpha = 0.1$.

[3] (d) If zero lies outside a 90% confidence interval for μ, this implies that we would reject $H_0 \colon \mu = 0$ with $\alpha = 0.1$ against the two-sided alternative.

[3] (e) If α is the Type I error rate for a hypothesis test, $1 - \alpha$ is the power of that test.

[3] (f) For any *mutually exclusive* events A and B we have $P(A \cup B) = P(A)P(B)$.

[3] (g) For any two events A and B, $P(A|B)/P(B|A) = P(A)/P(B)$.

[3] (h) The pmf $p(x)$ of a discrete random variable X gives $P(X = x)$.

[3] (i) For any continuous random variable X, $P(X \leq 0) = P(X < 0)$.

[3] (j) For any two random variables X and Y, $E[XY] = E[X]E[Y]$.

Name: _____ Student ID #: _____

2. For each part, write your answer in the space provided. No explanation is needed.

[3] (a) What result will I get if I run `pnorm(10, mean = 10, sd = 5)` in R?

[3] (b) Write an R command to calculate the median of a $\chi^2(2)$ random variable.

[3] (c) Approximately what result will I get if I run `qnorm(0.16)` in R?

[3] (d) Given a data table called `grades` with columns `exam1` and `exam2`, write out the full R command to run a regression predicting `exam2` from `exam1`.

[3] (e) Write a single line of R code to display the 4th row of a data table called `studentdata`.

[5] (f) Given a data table called `studentdata` with a column called `exam1`, write a single line of R code to display data for all students who scored above 70 on `exam1`.

[5] (g) Write a single R command to draw three numbers at random from the digits 0–9 *with replacement*.

[5] (h) Write R code to plot the pdf of a standard normal random variable between -3 and 3 using a grid of x-values with a step size of 0.01.

5 (i) Write an R function called `zscores` that takes a vector `x` as its only input and outputs the z-scores of `x`. You may use any R functions that you like in your answer and may assume that there are no missing values.

3. Let $Y_1, \ldots, Y_7 \sim$ iid $N(\mu = -3, \sigma^2 = 9)$.

3 (a) Let $X = 1 + Y_1/3$. What kind of random variable is X? You do not need to explain your answer, but be sure to specify any and all relevant parameters.

3 (b) Let $W = (Y_2 + Y_3 + Y_4 + Y_5 + Y_6 + Y_7)/6$. What kind of random variable is W? You do not need to explain your answer, but be sure to specify any and all relevant parameters.

3 (c) What kind of random variable is $X + W$? You do not need to explain your answer, but be sure to specify any and all relevant parameters.

3 (d) Let $Z = X^2$. What kind of random variable is Z? You do not need to explain your answer, but be sure to specify any and all relevant parameters.

8 (e) Calculate $E[X^2]$. Briefly explain your reasoning.

4. On Monday Rodrigo arrives at the office and informs Yiwen that, over the weekend, he has developed extra-sensory perception (ESP). He claims to be able to predict the outcome of a coin-flip with *better* than 50% accuracy. Yiwen is dubious and proposes and experiment in which Rodrigo will be asked to predict the outcomes of 100 flips of a fair coin. She asks Rodrigo to write down his predictions, in order, and seal them in an envelope. To make sure that the experiment is fair, she enlists Rossa to carry out the 100 coin flips while she and Rodrigo both watch. Rossa then opens the envelope and reads the predictions: Rodrigo has successfully predicted 51 of the 100 coin flips.

15 (a) Rodrigo claims that the results of the experiment prove that he has ESP but Yiwen isn't convinced. She decides to use what we've learned about tests for proportions in Econ 103 to test the null hypothesis that Rodrigo is just guessing the outcomes of the coin flips at random against the two-sided alternative with $\alpha = 0.05$. To ensure that her test is as accurate as possible, Yiwen *fully imposes the null* when specifying her test statistic. What test statistic does Yiwen use? What is its sampling distribution under the null? What is the numeric value of her test statistic? What is the critical value for her test? What is the outcome of the test?

5 (b) Rodrigo objects to Yiwen's procedure claiming that she should have tested against the *one-sided* alternative. Re-do the preceding part using the one-sided alternative and briefly explain which, if any, of the following items will change: the test statistic, the sampling distribution of the test statistic under the null, the numeric value of the test statistic, the critical value, and the outcome of the test.

15 (c) Let p denote Rodrigo's accuracy in predicting coin flips and assume, for the sake of argument, that he really does have ESP so that $p > 1/2$. (For example, if he correctly predicts the outcome of a coin flip 60% of the time, then $p = 0.6$.) Calculate the approximate power of a one-sided test with $\alpha = 0.16$ based on Yiwen's experiment as a function of p. Note that your answer should be an R command *that depends on* p, not a specific numeric value.

5. This question concerns a data table called `earnings` containing data on the height in inches (`height`), sex (`female` = 1 denotes female), and annual earnings in US dollars (`earn`) of a random sample of 1379 individuals. Here are the first few rows:

```
    earn height female
1  50000     74      0
2  60000     66      1
3  30000     64      1
4  50000     63      1
5  51000     63      1
6   9000     64      1
```

and here are some summary statistics:

	earn	height	female
Mean	20015	67	0.62
Median	16400	66	1
S.D.	19764	4	0.48

To answer this question you will need to consult the regression results that appear on the last page of this exam. (You may want to tear out the page of regression results to avoid having to flip back-and-forth.)

5 (a) Is there evidence of skewness in `height` or `earnings`? Explain briefly.

5 (b) About how many of the individuals in this dataset are female? Explain briefly.

10 (c) Who's taller on average in this dataset: males or females? About how much taller? Does the difference in sample means in this dataset provide compelling evidence of a difference of mean heights in the population from which these individuals were sampled? Support your answer by constructing an approximate 95% confidence interval for the difference of population mean heights (female minus male) and interpreting the results.

15 (d) Who earns more on average in this dataset: males or females? About how much more? Does the difference in sample means in this dataset provide compelling evidence of a difference of mean earnings in the population from which these individuals were sampled? Support your answer by testing the null hypothesis that males and females earn the same amount against the two-sided alternative with $\alpha = 0.05$ and interpreting the results. Be sure to specify the value of the test statistic, the decision rule, critical value, and outcome of the test. Approximately what is the associated p-value?

5 (e) What is the sample correlation between `height` and `earn`? Whare the units of this summary statistic?

5 (f) What is the sample covariance between `height` and `earn`? What are the units of this summary statistic?

5 (g) What is the value of the estimated intercept for the regression that uses `height` to predict `earn`? What are its units? Briefly explain the meaning of this estimate.

5 (h) What is the value of the estimated slope for the regression that uses `height` to predict `earn`? What are its units? Briefly explain the meaning of this estimate.

5 (i) Construct an approximate 95% confidence interval for the regression slope from the previous part and interpret it. In particular, do you find evidence that there is a positive relationship between height and income in the population?

5 (j) Which more accurately predicts `earnings`: `height` or `female`? Explain briefly.

15 (k) Kevin argues that the results discussed above provide evidence that women are discriminated against in the labor market. He says "the only reason that there is a positive relationship between height and earnings is that women are systematically and unfairly paid less than men and women also happen to be shorter, on average." Amanda does not believe that women are discriminated against and argues that the situation is exactly the *reverse* of what Kevin claims. She says: "taller people earn more than shorter people since being tall is valuable in many careers. Since women are shorter than men, this fully explains why they earn less, on average." Use what you have learned in Econ 103 along with the *full* set of regression results on the final page of this exam to discuss Kevin and Amanda's interpretation of the regressions results and whether you agree with either, both or neither of them. You will be graded on the clarity of your answer and the extent to which it incorporates the tools and concepts we studied in the course. Write your answer using bullet points with *no more than five bullets*: all other things equal, a concise answer will be treated more favorably.

```
lm(formula = earn ~ female)
            coef.est  coef.se
(Intercept)  28926.92    811.86
female      -14307.02   1028.65
---
n = 1379, k = 2
residual sd = 18513.23, R-Squared = 0.12
```

```
lm(formula = height ~ female)
            coef.est coef.se
(Intercept) 70.05     0.12
female      -5.54     0.15
---
n = 1379, k = 2
residual sd = 2.70, R-Squared = 0.50
```

```
lm(formula = earn ~ height)
            coef.est  coef.se
(Intercept) -84078.32   8901.10
height        1563.14    133.45
---
n = 1379, k = 2
residual sd = 18853.92, R-Squared = 0.09
```

```
lm(formula = earn ~ height + female)
            coef.est  coef.se
(Intercept)  -9636.63  12953.75
height         550.54    184.57
female      -11254.57   1448.89
---
n = 1379, k = 3
residual sd = 18460.37, R-Squared = 0.13
```
