LAST NAME (Please Print): KEY

FIRST NAME (Please Print): _____

HONOR PLEDGE (Please Sign): _____

Statistics 111

Midterm 3

- This is a closed book exam.
- You may use your calculator and a single page of notes.
- The room is crowded. Please be careful to look only at your own exam. Try to sit one seat apart; the proctors may ask you to randomize your seating a bit.
- **Report all numerical answers to at least two correct decimal places** or (when appropriate) write them as a fraction.
- All question parts count for 1 point.

The Gamma(α, β) distribution has mean α/β, variance α/β^2, and density function

$$f(x) = \frac{\beta^\alpha}{(\alpha-1)!} x^{\alpha-1} \exp(-\beta x) \text{ for } 0 \le x \text{ and } \alpha > 0, \beta > 0.$$

The Beta(α, β) has mean $\alpha/(\alpha+\beta)$, variance $\alpha\beta/[(\alpha+\beta)^2(\alpha+\beta+1)]$, and density function

$$f(x) = \frac{(\alpha+\beta-1)!}{(\alpha-1)!(\beta-1)!} x^{\alpha-1}(1-x)^{\beta-1} \text{ for } 0 \le x \le 1 \text{ and } \alpha > 0, \beta > 0.$$

$$\int x \exp(ax)\, dx = \left(\frac{x}{a} - \frac{1}{a^2}\right) \exp(ax)$$

The mean and standard deviation of an Exponential r.v. are both $1/\lambda$.

1. You are told that the Hardy-Weinberg equilibrium for blood types is 20% type 0, 30% type A, 40% type B, and the rest are type AB. You want to test whether the Masai are in equilibrium. Among 200 randomly sampled Masai, 35 are type 0, 50 are type A, 90 are type B, and the rest are type AB.

   In words, what is your null hypothesis?

   Masai blood types follow the 2/3/4/1 ratios for O, A, B and AB

   4.79 What is the value of your test statistic?

   The observed values are 35, 50, 90 and 25. The expected are 40, 60, 80 and 20. Sum the (observed - expected)^2/expected to get 4.7917.

   7.81 or 7.82 What is (are) your critical value(s) for a 0.05 level test?

   From the chi-squared table with 4 - 1 = 3 df.

   What conclusion do you reach?

   Do not reject null. There is no evidence that Masai ratios are different from the equilibrium values.

2. A pundit claims that at least 10% more women than men will vote for Hillary Clinton. You want to prove him wrong. You sample 50 men and 80 women at random, and find 30 men and 40 women plan to vote for her.

   In sympbols, what is your null hypothesis? (Subtract men from women.)

   $H_o : p_w - p_m \geq 0.1$

   -2.25 What is the value of your test statistic?

   $ts = \left(\frac{40}{80} - \frac{30}{50} - 0.1\right) / \sqrt{\frac{(0.5)^2}{80} + \frac{.6*0.4}{50}} = -2.2466.$

-1.64 or -1.65 What is (are) your critical value(s) for a 0.05 level test?

From the z-table.

What conclusion do you reach?

Reject the null. Women are less than 10% more likely to support Hillary Clinton.

0.01 What is your significance probability?

This is the area on the z-table below -2.25.

3. 0.59 You want to show that the average IQ at Duke is greater than 115. You know that the standard deviation of IQs at Duke is 16. You sample 100 students. What is the power of a 0.05 level test if the true mean is 118?

$\mathbb{P}[(\bar{X} - 114)/(16/\sqrt{100}) > 1.645] = \mathbb{P}[Z > 1.645 - (118 - 115)/(16/10)] = \mathbb{P}[Z > -0.23] = 0.59$.

176 What sample size would you need to achieve a power of 0.8?

From the z table, $\mathbb{P}[Z > -0.84] = 0.8$, so $-0.84 = 1.645 - (118 - 115)/(16/\sqrt{n})$. Solving gives $n = 175.651$ and we must round up.

4. You want to predict IQ from GPA. From a sample of 15 people, you find that the estimated intercept is 80 and the estimated slope is 10. The mean GPA of the sample is 3.2 and its variance is 2. The correlation is 0.4 and the standard deviation of the residuals is 3.

0.16 What proportion of the variance in IQ is explained by GPA?

The square of the correlation.

119.5 Set a 95% upper confidence bound of the average IQ of people with a GPA of 3.8.

$(80 + 10 * 3.8) + 3 * \sqrt{(1/15) + (3.8 - 3.2)^2/(14 * 2)} * t_{13,0.95}$ where $t_{3,0.95} = 1.771$. You get 119.4983.

122.21 Set a 90% upper prediction bound on the IQ of Aguinaldo, whose GPA is 3.8.

$(80 + 10 * 3.8) + 3 * \sqrt{1 + (1/15) + (3.8 - 3.2)^2/(14 * 2)} * t_{13,0.90}$ where $t_{3,0.90} = 1.350$. You get 122.208.

5. An unknown proportion p of people are descended from Caligula. Your prior belief about that proportion is uniform. You sample 3 people and DNA testing reveals that all of them are related to him.

Beta(4,1) What is your new distribution for p?

Since the uniform prior on [0, 1] is Beta(1,1), the Beta-Binomial conjugate model applies.

0.84 If loss is proportional to absolute error, what is your guess about p ?

The median. Since the beta density is $f(p) = 4p^3$ for $0 \leq p \leq 1$, then the cumulative distribution function is $F(p) = p^4$. Solve $0.5 = m^4$ to get $m = 0.84$.

0.8 If loss is proportional to squared error, what is your guess about p?

The mean, or (4/(4+1).

0.76 What is the chance that $p > 0.7$?

$\mathbb{P}[p > 0.7] = \int_{0.7}^{1} 4p^3 \, dp = 0.76$.

6. 0.28 A sample of 20 children in a class of 35 third-graders finds that 8 want to become super villains. Set a 95% lower confidence bound on the proportion of students who yearn for comic book criminality.

$L = (8/20) - \sqrt{(0.4 * 0.6)/20} * \sqrt{(35-20)/(35-1)} * 1.645 = 0.2803.$

7. You want to show that male singers make more money than female singers. You observe what is paid to each in famous duets from their last gig (in thousands):

	male	female	difference
Michael & Janet	20	18	2
Donny & Marie	10	9 *	1
Captain & Tennille	5	6	-1
Dr. Dre & Queen Pen	15	13	2
Tony & Madonna	12	12	0

1.63 What is your test statistic?

The variance of the differences is 1.2, so $ts = (0.8 - 0)/\sqrt{1.2/5} = 1.633$, where 0.8 is the average difference.

2.13 What is (are) your critical value(s) if $\alpha = 0.05$?

$t_{4,0.95}$

What was the advantage of studying duets?

Pairing controlled for the commercial success of the singer.

8. You want to decide whether there is a relationship between major and dating activity.

	no date this week	one date	more than one
statistics	10	5	0
biology	0	5	10
mathematics	10	10	10

In words, what is your alternative hypothesis?

There is some relationship between major and dating.

20 What is your test statistic?

Independence test. The table of expected values is:

	no date this week	one date	more than one
statistics	5	5	5
biology	5	5	5
mathematics	10	10	10

and so the sum of (observed - expected)^2/expected is 20.

9.49 What is (are) your critical value(s) for a 0.05 level test?

$\chi^2_{4,0.95} = 9.488$.

9. You want to set a confidence interval on the interquartile range (IQR) of the amount of sleep students get. A random sample of 80 students has an interquartile range of 6. Twenty bootstrap samples of those 80 students finds the following IQRs:

7, 4, 6, 9, 8, 8.5, 2, 2.5, 11, 7, 9.5, 4.5, 9, 10, 3, 1.5, 9, 8, 8.5, 4

1.75 Set a 95% one-sided lower CI with the percentile bootstrap.

The two lowest values are 1.5 and 2, so their midpoint is 1.75

Find U in a 90% two-sided CI with the pivot bootstrap. U = 10.25

The formula is 2 * pe - L where pe is 6 and L, from before, is 1.75

10. 550 You want to predict the number of tickets sold from the cost of the ticket (C), the popularity of the performer (P) and the size of the venue (S). For a sample of 10 concerts, STATA finds that $\hat{\beta}_0 = 800$, $\hat{\beta}_C = -10$, $\hat{\beta}_P = 20$, $\hat{\beta}_S = 0.1$, and their standard errors are 25, 4, 15, and 6, respectively. What is the predicted number of tickets for a Joan Baez performance at Page Hall, when the ticket costs \$40, Ms. Baez has popularity 2.5, and the Page Hall can seat 1000?

    550 = 800 - 10 * 40 + 20 * 2.5 + 0.1 * 1000.

    1.33 You want to test whether Popularity should be in the model. What is your test statistic?

    $(pe - null\ value)/stderr = (20 - 0)/15 = 1.33.$

    ± 2.45 What is (are) your critical value(s) when the Type I error is 0.05?

    Two-sided test, cv is $t_{6,0.025}$ and $t_{6,0.975}$

    yes Should you remove Popularity from the model?

11. $L = \beta_0 + \frac{\beta_1}{D^3} + \epsilon$ The lifespan L of a soap bubble is linearly related to the inverse of its volume. For a soap bubble with diameter D, write the nonlinear regression model, including a noise term. Besides D and the ϵ, use only β_0 and β_1.

12. List all and only the true statements. A, D, E

    A. Any two of n, Type I error, and Type II error determine the third.

    B. As points cluster more tightly around a line, correlation increases.

    C. It is an ecological correlation if the X values are measured with error.

    D. Sir Francis Galton invented birth-and-death processes.

    E. René Descartes was an artillery officer.