Name: Key

Math 127 Exam 3 Summer 2015

Version G

Oath: *"I will not discuss the exam contents with anyone on Earth until the answer key is posted to BB."*

Sign Name: Key

Permitted Materials: One-sheet of handwritten or typed notes. No copies of published materials.
No pink sheets or photocopies of pink sheets.
The datasets are found on www.statcrunch.com. No other webpages.
Any calculator is permitted. Short bathroom breaks are permitted.
No cell phone calculators.
You must staple your sheet of notes to the exam.

- **Show all work when appropriate. StatCrunch provided numbers are OK always, but on sample sizes and minor algebraic calculations, support your answers.**

- Points are in parentheses for each problem.

- This test is graded out of 100 points and counts for 25% of your Math 127 grade.

- The graded exams are kept on file for at least one year and students are welcome to come see them whenever I am available in my office.

- An answer key will be posted on Blackboard shortly after the testing is completed.

- Exam grades will be posted to Blackboard by Monday, August 3rd (but possibly sooner). I will only be grading in one big swoop and only once all exams are completed and I pick them up Thursday, July 30th. If you take the exam early in the week, it will just sit in the math lab for a few days, FYI.

- Final grade announcements will be posted to Blackboard by Monday, August 3rd (but possibly sooner). Your numerical "Course Grade" on Blackboard is your final grade in Math 127 and you will know your letter grade based on my announcement.

- Letter grades will be posted to MyCecil, but students may see WIP for a few days.

- Good luck on this exam. Good luck in the future. It's been my pleasure to work with you this semester.

1. (2) $t =$ 4.032 Give the t value for $\bar{y} \pm t\left(s/\sqrt{n}\right)$ if $n = 6$ data points, 99% confidence.

$z =$ 1.881 Give the z value for $\hat{p} \pm z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ for 94% confidence.

2. Suppose the mean "***Credit Card Debt***" for **all** Cecil College students is \$1730 with a standard deviation of \$1480 and the shape of the distribution is very skewed right.

2a. (2) What is the minimum required sample size for the Normality of $\bar{y}$ to kick in? $n \geq 30$

2b. (2) Give the mean of $\bar{y}$: $\mu_{\bar{y}} = \mu_y = 1730$

2c. (2) Determine the standard deviation of $\bar{y}$ for sample sizes of $n = 90$. You can round to a whole number.

$$\sigma_{\bar{y}} = \frac{\sigma_y}{\sqrt{n}} = \frac{1480}{\sqrt{90}} \approx 156$$

2d. (2) P(A random sample of 90 students has a mean "***Credit Card Debt***" exceeding \$2000) = 0.0417

2e. (2) Using a common rule of thumb, what would be an unusually low mean "***Credit Card Debt***" from an unbiased group of 90 Cecil College students? Show work.

$\mu_{\bar{y}} - 2\sigma_{\bar{y}} = 1730 - 2(156) =$ \$1418

Other answers OK if reasonably justified.

3. A statistician created a confidence interval for the true proportion of all Marylanders who would vote for Donald Trump for President. The interval was (26.703% to 32.946%) and based on 1425 respondents.

3a. (2) How many in the sample said they'd vote for Trump? $x = 425$

3b. (2) What was the confidence level? 99%

No work is needed to be shown, but here is some space if you need it:

$\hat{p} = \frac{26.703\% + 32.946\%}{2}$

$= 29.8245\%$

$29.8245\% \text{ of } 1425 = 425$

$z = \frac{ME}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} = \frac{0.031215}{\sqrt{\frac{0.298245(1-\ldots)}{1425}}}$

$= 2.576$

so 99%

ZZZ Retired –

4. Open up the "**Calendar Year 2015 Large Survey**" dataset (piano). We would like to test if the mean "***TV Time***" is higher for males than it is for females. Presume the conditions are met.

4a. (3) Give the appropriate summary stats needed to run this test.

Male: $n = \cancel{44}\ 101$, $\bar{y} = \cancel{10.386}\ 10.485$, $s = \cancel{6.986}\ 9.147$

Female: $n = \cancel{99}\ 227$, $\bar{y} = \cancel{9.818}\ 9.815$, $s = \cancel{9.941}\ 9.992$

Diff: $\cancel{0.568}\ 0.67$

4b. (2) Hypotheses: $H_0: \mu_M = \mu_F$ vs. $H_A: \mu_M > \mu_F$

4c. (4) Give the test statistic and the P-value. Technology is OK, please uncheck "Pool Variances".

Test Statistic: $t = \cancel{0.391}\ 0.595$ P-value: $\cancel{0.3481}\ \cancel{0.35}\ 0.2762$

4d. (2) Make a decision: Fail to Reject H_0

4e. (2) Write a conclusion in context.

No evidence to say that males at Cecil watch more TV, on average, than females at Cecil.

4f. (2) Interpret the test statistic with a sentence in context.

Our difference of $\cancel{0.568}\ 0.67$ was $\cancel{0.391}\ 0.595$ standard errors above the hypothesized difference of 0.

4g. (2) Interpret the standard error with a sentence in context.

With repeated samples, we'd expect the difference in means to vary by about $\cancel{1.452}\ 1.124$ Hours.

4h. (2) Interpret the P-value with a sentence in context.

If males and females watched the same amount of TV on average, we'd get a difference of $\cancel{0.568}\ 0.67$ hours, or one even bigger, $\cancel{34.81\%}\ 27.62\%$ of the time.

Not on Exam 3 Any Longer

5. A linear regression equation was fit to the data in "**Roller Coasters**" on StatCrunch. "***Speed***" is in miles per hour. "***Drop***" is in feet.

Simple linear regression results:
Dependent Variable: Speed
Independent Variable: Drop
Speed = 37.529503 + 0.18599121 Drop
Sample size: 61
R (correlation coefficient) = 0.92132219
R-sq = 0.84883459
Estimate of error standard deviation: 4.4083553

5a. (2) Predict "***Speed***" for a 195 foot drop: 73.798 mph

5a. (2) Interpret the slope with a sentence in context: For each extra one foot of drop, we expect speed to increase by 0.186 mph.

5b. (2) Interpret R^2 with a sentence in context: 84.88% of the variation in Speed is explained by Drop, and 15.12% is explained by other variables.

5c. (2) Interpret s_e with a sentence in context: On average, our predicted speeds are off by ~4.41 mph when using Drop as the x-variable.

5d. (2) Interpret the *y*-intercept with a sentence in context: X=0' drop is probably extrapolation and would make for a very boring rollercoaster! No realistic meaning.

6. (3) The producers of ***Trainwreck***, starring Amy Schumer, rated R, want to estimate the mean age of people who come out to the theatre to see the movie. If they'd like to estimate the mean age to within 2 years with 99% confidence, how many people will they need to survey? You'll need a reasonable estimate for the standard deviation, so use the $\frac{\text{Range}}{6}$ estimate as mentioned in class.

ME = 2

$z = 2.576$

$n = \left[\frac{2.576(11.7)}{2}\right]^2 = 227.09$

So $n = 228$

Say Range = 88 − 18 = 70

$SD \approx \frac{70}{6} \approx 11.7$

σ	8	8.5	9	9.5	10	10.5	11.	11.5
n	107	120	135	150	166	183	201	220

7. (3) Research shows 1/3 of all new marriages started online. Does that pattern hold in Cecil County? To start, we will need to collect a sample of marriages, but how many? Using 95% confidence and a margin of error of 5%, estimate the required sample size.

$$n = \frac{z^2 \hat{p}(1-\hat{p})}{(ME)^2} = \frac{1.96^2 (1/3)(2/3)}{(0.05)^2} \approx 341.5$$

so $n = 342$

8. At Firebirds, 5% of all people order the salmon. Expecting $n = 600$ people over the weekend, answer the following questions.

8a. (3) Management will cook up the sampling distribution model for the proportion of people ordering salmon. Determine the mean and standard deviation of that model.

$\mu_{\hat{p}} = p = 0.05$

$$\sigma_{\hat{p}} = \sqrt{\frac{0.05(0.95)}{600}} \approx 0.0089$$

8b. (3) A news story breaks on Wednesday that salmon is no longer good for you. In fact, it is downright unhealthy. Using your model, determine the **<u>number of salmon orders</u>** over the weekend that would convince you statistically that fewer people are ordering salmon.

Easiest: $p - 2\sigma_{\hat{p}} = 0.05 - 2(0.0089)$

$= 0.0322$

$600 \times 0.0322 = 19.32$ so fewer than 19 or 20 orders

ZZZ Retired –

9. Open up the "**Calendar Year 2015 Personality Types**" dataset. We would like to test if less than half of all Cecil College "***Females***" are "***Extraverted***" = "***E***". Conditions are met.

9a. (2) Hypotheses: H_0: $p = 0.5$ H_A: $p < 0.5$

9b. (2) Give the value of the sample proportion, fraction and percentage: ~~$\hat{p} = \frac{44}{101} = 0.4356$~~ $\hat{p} = \frac{108}{226} = 0.4779$

9c. (4) Test Statistic: ~~$z = -1.294$~~ -0.665

P-Value: ~~0.0979~~ 0.253

9d. (2) Decision: Reject H_0 (Fail to Reject H_0)

9e. (2) Write a conclusion in context: No strong evidence to say that less than 1/2 of all CC females are extraverted.

ZZZ Retired –

10. Open up the "**Calendar Year 2015 Personality Types**" dataset. We would like to test if a higher proportion of "***Females***" at Cecil College are "***Feeling***" = "***F***" when compared to the males.

10a. (2) Hypotheses: H_0: $p_F = p_M$ H_A: $p_F > p_M$ 151/226 = 0.6681

10b. (2) Give the value of the sample proportion, fraction and percentage: $\hat{p}_F = \frac{69}{101} \approx 0.6337$

$\hat{p}_M = \frac{25}{55} \approx 0.4545$ 41/106 = 0.3868

Diff: ~~0.1792~~ 0.2813

10c. (4) Test Statistic: ~~$z = 4.159$~~ 4.84

P-Value: ~~0.0154~~ < 0.0001

10d. (2) Decision: (Reject H_0) Fail to Reject H_0

10e. (2) Write a conclusion in context: There is evidence that a higher proportion of females are "Feeling" when compared to the males

10f. (2) If we made a mistake, what kind, and what would it imply, in the context of the problem?

Type I. In reality, no difference between males & females on this personality trait.

11. Use the "**Calendar Year 2015 Large Survey**" dataset (piano) (last time, promise). Are students spending on average less than 10 hours per week "***Online Time***"?

11a. (2) Hypotheses: H_0: $\mu = 10$ Hours H_A: $\mu < 10$ Hours

11b. (2) Summarized Data:

$n = \cancel{144}\ 330$, $\bar{y} = \cancel{8.89}\ 10.073$, $s = \cancel{10.098}\ 11.449$

11c. (3) What three conditions must be met to proceed with the hypothesis test? Explain if each is met. The variable is quantitative, so no need to go over that one.

Condition 1: Unbiased Sample? Yes, most likely.

Condition 2: $n = \cancel{144}\ 330 < 10\%$ of all CC? Yes ✓ It's OK

Condition 3: $n = \cancel{144}\ 330$ exceeds 30 ✓ (Need it, skewed right) 0.115

11d. (2) Test Statistic: $t = \cancel{-1.32}$ P-value: $\cancel{0.0944}\ 0.5459$

11e. (2) Decision using a 5% significance level: Fail to Reject H_0

11f. (2) Concluding remark using a 5% significance level: No evidence that the mean "Online Time" is under 10 hours at CC.

11g. (2) Interpret a 95% confidence interval with a sentence in the context of the problem:

We are 95% confident that on average, students spend between $\cancel{7.23}\ 8.83$ and $\cancel{10.55}\ 11.31$ Hours online each week.

12. (2) Explain what is meant by "statistical significance".

Data too unusual to attribute to chance.

(We would reject the null, but that's not what stat. sig. means.)

Kupe — Out!!