Print Name: Key

Math 127 – Exam 1 – Summer 2017

Version Wonder Woman

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

Sign Name: Key

The penalty for cheating on this Exam is a grade of 0% for Math 127 Exam 1.

Student Instructions

1. **This test is graded out of 100 points and counts for 20% of your Math 127 grade. Points are in parentheses for each question.**

2. **You can use a calculator, but you cannot use your phone. You can use the calculator on the computers if you wish.**

3. **You will need to use www.statcrunch.com. This is the <u>only</u> permitted webpage.**

4. **You are permitted to use one 8.5" by 11" sheet of notes, front and back. You will submit it with your test.**

   You may <u>not</u> use the pink sheet or copies of the pink sheet.

   You must produce (handwritten or typed up) your own sheet of notes.

   You may <u>not</u> use copies or scans of any instructor-created Math 127 content or answer keys.

5. **Show work or points will be deducted. If you only report an answer and it is wrong, you will receive no credit.**

1. Use the "**Hospital Payments**" dataset for this one.

1a. (3) Variable type – C or I or Q?

I (1/2) "***Provider ID***" I OR C "***Provider Name***" C "***Provider State***"

C "***Provider Zip***" Q "***Total Discharges***" C "***Illness***"

1b. (2) "***The Heart Hospital at Deaconess Gateway LLC***" has one entry. What row? 84980 (2)

1c. (2) Which "***Provider Name***" has the most entries in the entire dataset?

"***Provider Name***": Good Samaritan (1) Number of Entries: 1880 (1) Sum (2)

Both OK

1d. (2) What is "***Maryland's***" most frequent "***Illness***"? G.I. Hem (120) Entries or Renal Failure (6768) Total D.S.

1e. (2) Determine the mean "***Total Discharges***" for "***Diabetes***": 26.39 (2)

1f. (2) Which "***Illness***" has the most "***Total Discharges***"? G.I. Hem (5695) Entries (2) or Kidney/Urinary (273,236) Sum

2. (5) True or False. Write clearly, especially if you are a "**T**" or "**F**" person!

2a. False (1) If every student in Math 127 earns a 100 on this exam, the standard deviation will be 100.

2b. True (1) Lower fences can be negative.

2c. False (1) The gold standard of all possible sampling methods is the convenience sample.

2d. False (1) The value of the 96th percentile must be greater than the value of the 95th percentile.

2e. True (1) "***Credit Card Debt***" for Cecil students is skewed right

3. (4) A Towson student has no "***Student Loan Debt***". Presuming the mean is \$13,506 and the standard deviation is ~~\$1,269.22~~, show the calculation to convert her "***Student Loan Debt***" to a *z*-score.

\$ 12,690.22

(4) $z = \frac{y - \bar{y}}{s} = \frac{0 - 13506}{12690.22} = -1.06$

=

1 zzz Retired

4. Use our "**Calendar Year 2017 Large Survey**" dataset to address the following questions.

4a. (4) Person #98 did not report their "***Commute***". If their z-score was 0, solve for their "***Commute***".

$z = \frac{y - \bar{y}}{s}$ if $z = 0$, Commute must be equal to the mean

His Commute: ~~19.88~~ 19.46 (4)

Bin Width of 1 → Multimodal

4b. (5) Describe the distribution of "***College Credits***". Use bullet points for ease of grading.

Shape: ~~Bimodal~~, skewed Right

Center: Median = 12 credits (1)

Spread: IQR = ~~7~~ 6 credits (1)

Outliers: ~~None~~ (1) 2 high outliers taking 24 credits.

4c. (5) Discuss the independence or dependence of these two categorical variables: "***Global Warming***" vs. "***How Religious***". ~~Make sure to~~ support your claim with appropriate conditional percentages.

(3) Dependent: I grouped by "How Religious"

	Extreme R.	Very R.	Somewhat R.	Not Rel.
% who Think Global Warming Is Real	~~30%~~ 25%	~~48.15%~~ 50.53%	~~59.17%~~ 63.51%	~~78.75%~~ 72.05%

4d. (2) ~~5 OR 4~~ (2) 0 How many respondents are "***Race***" = "***White***", "***Male***", and ***at least 40-years old***?

4e. (2) Give the best measure of center for "***Sleep Hours***"? (Mean) / Median = ~~6.65~~ (2) 6.704

spread

4f. (2) Give the best measure of ~~center~~ for "***Online Time***"? SD / (IQR) = 11 (2)

4g. (4) Using the idea of z-scores, are the following data values unusual? Yes or No.

Person #65's "***Work Time***": NO (1) $(0 - 22.83)/16.27 = -1.39$

Person #90's "***Online Time***": YES (1) $(50 - 12.31)/13.31 = 2.83$

Person #148's "***Commute***": NO (1) $(6 - 19.88)/14.67 = -0.93$

Person #264's "***Ideal Children***": YES (1) $(20 - 2.49)/1.91 = 9.17$

ZZZ Retired

5. Use the "**Calendar Year 2017 Library Data**" dataset for this one.

5a. (4) Calculate by hand the cutoffs to be an official low or high outlier for number of "***Pages***". Show your work.

$UF = Q_3 + 1.5(IQR)$ ~~$= 432 + 1.5(208)$~~ ~~$= 744$~~ (2) $= 669$

$402 + 1.5(178)$

$LF = Q_1 - 1.5(IQR)$ ~~$= 224 - 1.5(208)$~~ ~~$= -88$~~ (2) $= -43$

$224 - 1.5(178)$

5b. (2) Now, for "***Length***", how many official low or high outliers do we have?

Official Low Outliers: ~~2~~ 7 (1) Official High Outliers: ~~8~~ 17 (1)

5c. (4) Interpret, with a sentence or two, the 74th percentile for "***Copyright***" year: (4)

74% of the books have Copyright YEAR 2000 or earlier.

26% of books have Copyright 2000 or later.

5d. (4)

~~576~~ (1) Median "***Weight***" for all books 564

~~584~~ (1) Median "***Weight***" for book written by "***Males***" 574

~~1054~~ (1) Median "***Weight***" for books with at least 500 "***Pages***" 1034

~~576~~ (1) Median "***Weight***" for books written in the 1900s (whole century) 572

5e. (2) ~~5~~ 2 (2) 6 How many books had "***Copyright***" = 1977?

5f. (2) Maternal & Child Nursing Care (2) "***Title***" of the heaviest book

5g. (4) Thinking in terms of z-scores, give a range of values for what is not an unusual "***Thickness***". Show work.

Inside $\bar{y} \pm 2(SD)$ not unusual.

1.09 0.450

~~1.12~~ $\pm 2(0.543)$

~~(0.034" to 2.206")~~

0.19 (4) 1.99

a few

6. (6) During the last week, ~~some~~ stolen credit cards have been floating around the Towson Town Center Shopping Mall. Police will take a sample of transactions paid for with credit cards and begin an investigation to track down the criminals.

For each scenario, identify the sampling method. **Pick from**: census, convenience, simple random, systematic, cluster, and stratified.

6a. Cluster (2) There are 180 stores. Police draw three random stores out of a hat because their StatCrunch subscription expired: Brookstone, Forever 21, and Kay Jewelers. The officers then analyze every credit card transaction from those stores during the last week.

6b. Convenience (2) Officer Bardwell has a hunch that criminals shop at the Apple store, so he makes a request to analyze all the transactions from that store. Plus his genius brother works there, so it'll be pretty easy to get all those credit card transactions.

6c. Stratified (2) Though it'll be a ton of work, the whole team decides to take a sample of five random credit card transactions from each and every store.

7. We will run a designed experiment as such.

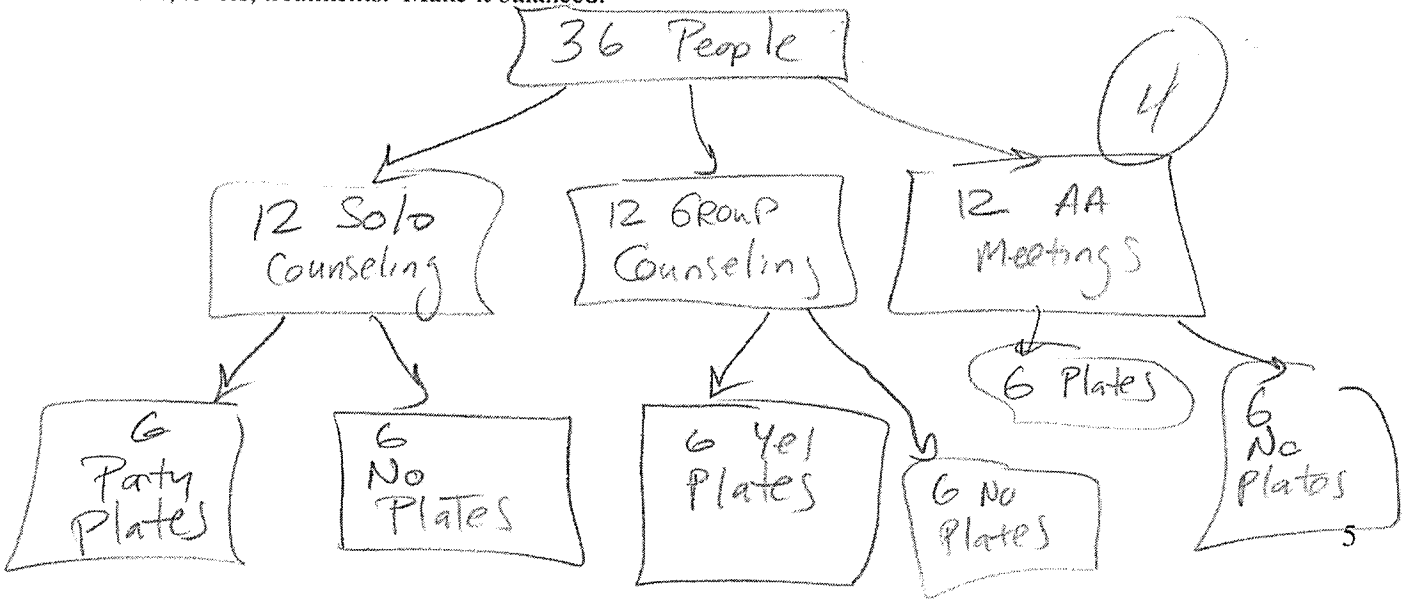
People convicted of drunk driving will be randomly assigned to one of three treatment groups: individual counseling sessions for 6 months, group counseling sessions for 6 months, or AA meetings for 6 months.

Additionally, half of all participants (randomly decided) will be required to put special license plates on their vehicles (nicknamed "Party Plates" in Ohio, it's a real thing, sort of a public shaming / scarlet letter punishment. Google it later…).

7a. (2) If you were in charge, come up with a reasonable response variable to measure the outcome of this experiment.

Reasonable Response Variable: ? Answers Vary: Y/N another DUI or # DAYS Sober OR . . . .

7b. (4) Suppose we had a total of 36 participants in the experiment. Draw out a well-labeled tree diagram to show the factors, levels, treatments. Make it balanced.



ZZZ Retired

8. Back to the "**Calendar Year 2017 Large Survey**" dataset. Give all answers as fractions, then as decimals rounded to four places and then convert the decimals to percentages rounded to the hundredths place.

Example: 42 / 97 = 0.4330 = 43.30%

8a. (2) Percentage of respondents who are "***Very Religious***".

~~55/268~~ ~~96/505~~ = ~~0.2052~~ = ~~20.52%~~ (2)

0.1901 = 19.01%

8b. (2) Percentage of "***Catholics***" who are "***Very Religious***".

~~7/48~~ = ~~0.1458~~ = ~~14.58%~~ (2)

~~10/86~~ 0.1163 = 11.63%

8c. (2) Percentage of respondents who "***Never***" use "***Facebook***" and "***Never***" use "***Instagram***".

~~18/272~~ = ~~0.0662~~ = ~~6.62%~~ (2)

37/507 = 0.0730 = 7.30%

8d. (2) Percentage of respondents who are 67" tall or taller. Use "***Height***" in the "**Large Survey**".

~~102/268~~ = ~~0.3806~~ = ~~38.06%~~ (2)

204/498 = 0.4096 = 40.96%

8e. (2) Percentage of respondents who are "***Catholic***" or "***Protestant***" or "***Christian***".

~~171/269~~ = ~~0.6357~~ = ~~63.57%~~ (2)

307/504 = 0.6091 = 60.91%

9. For this one, use the "**US News National University Rankings**" dataset.

9a. (2) U of IL - Urbana Champaign (2) Which "***IL = Illinois***" school has the largest "***Enrollment***"? (44,942)

9b. (2) $46,678 (2) 80th percentile for "***Tuition in-state***" for just "***Private***" schools

9c. (2) 97.86% (tied) (2) Most common "***Freshmen Retention Rate***" No Unique is OK

9d. (2) Percentage of "***TN = Tennessee***" schools that are "***Public***". (2)

5/7 = 0.7143 = 71.43%

9e. (2) Percentage of schools with at least 30,000 students. (2)

55/267 = 0.2060 = 20.60%