

Name: Key

Math 127 – Test 1A – Fall 2015

Oath: *"I will not discuss the exam contents with anyone until it is returned to me by my instructor".*

Sign Name: Key

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

Testing Center Staff Instructions

1. One sheet of handwritten or typed notes is OK.

Students may not use the "pink sheet" or any copied or scanned answer keys or Math 127 department documents.

2. Collect the sheet of notes and staple it to the test when submitted.
3. Testing Center issued TI calculator is OK.
4. www.statcrunch.com is required. All other webpages are prohibited.
5. Test must be completed in one sitting, but it is untimed. Very short bathroom breaks are permitted.

Student Instructions

1. This test is graded out of 100 points and counts for 15% of your Math 127 grade.
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 only 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 not use the pink sheet or copies of the pink sheet.

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

You may not 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. The "College Athletic Finances (Through 2014)" dataset contains information for a number of NCAA schools.

"Total Revenue" is the amount of money brought in by the athletic program through ticket sales, merchandising, etc...

"Total Expenses" is the amount of money spent by the athletic program on equipment, coach's salaries, travel, etc...

"Total Subsidy" is the amount of money given to the athletic department by outside government organizations.

1a. (3) Be specific, describe the **Who**: Each school

1b. (3) How many variables in the dataset are:

Quantitative? 1 2 3 4 5 Categorical? 1 2 3 4 5 Identifier? 1 2 3 4 5

No variable can be more than one type. Your answer should add up to 5 because there are 5 variables.

"% Subsidy" is not a variable, it is a mathematical function of "Total Subsidy" and "Total Revenue".

1c. (3) How many schools listed are in the Ohio Valley "Conference"? 11

1d. (3) Which school in the Ohio Valley "Conference" has the highest "Total Revenue"? Murray State
(\$17,217,807)

1e. (3) Create a new variable called "Net Income" which is "Total Revenue" - "Total Expenses". Use Data → Compute Expression.

Determine the mean "Net Income": \$1,500,087

1f. (3) What was the "Net Income" for Ohio State? \$31,295,680

1g. (3) How many schools received a "Total Subsidy" that exceeded the upper fence? 5

1h. (3) How many schools had "Total Revenue" exceeding \$150,000,000? 4

$1.5 \times 10^8 = 150,000,000$

1i. (4) Describe the distribution of the values in column "% Subsidy" with a sentence or two in context. Use only the best summary statistics. ~~Hit all the points discussed in class.~~

Shape: Bimodal (Kupe not comfortable calling it skewed left)

Center: Median = 65.805%, Mean = 53.70%

Spread: IQR = 40.39%, SD = 28.99%

Unusual Features: No official outliers
Two peaks a bit interesting.

~~34 Total~~ ~~ZZZ~~ - Retired

2. Use the "Calendar Year 2015 Library Data" dataset to address the following questions.

2a. (4) Determine the lower fence and upper fence for "Length". Show calculation for credit: How many outliers?

$$LF = Q_1 - 1.5(IQR) = 8.5 - 1.5(1) = 7$$

$$UF = Q_3 + 1.5(IQR) = 9.5 + 1.5(1) = 11$$

Low Outliers: 7

High Outliers: 27

2b. (4) Convert "Pages" for the book "Pizzaro and the Conquest of Peru" to a z-score. Show calculation. It is an unusually long or unusually short book? Comment. PS, it's in row 280.

$$\bar{y} = 338.68 \quad S = 189.45$$

$$Z = \frac{y - \bar{y}}{S} = \frac{153 - 338.68}{189.45} = -0.98$$

Not unusually short. Within ± 2 SD.

2c. (4) One book, "Estate Planning", is missing a value for "Pages". This monster of a book had a z-score of 3.09. Calculate the number of "Pages", showing calculation.

$$Z = \frac{y - \bar{y}}{S}$$

$$y = 338.68 + 3.09(189.45)$$

$$3.09 = \frac{y - 338.68}{189.45}$$

$$y = 924.08 \text{ pages}$$

so 924 pages

2d. (4) Give a range of values for "Pages" that would not be unusual. Use the idea of z-scores.

Inside $\bar{y} \pm 2S$ is not unusual.

$$338.68 \pm 2(189.45)$$

$$338.68 \pm 378.9$$

$$(-40.22, 717.58)$$

So Really (0 pages up to 717 pages)

3. Identify the official sampling methodology for each scenario (simple random, stratified, cluster, systematic, convenience, census, multistage). Answers can be used more than once or not at all.

3a. (2) Administrators at Cecil College wanted to take a sample of students to ask, "Are the restrooms on main campus kept adequately clean and stocked with supplies?". They randomly selected 7 classes from the course schedule, visited those classes in person, and sampled every student present with a pencil and paper survey.

Sampling method:

Cluster

3b. (2) Starting in Fall 2015, suppose the Cecil College Fitness Center staff requires every visitor to complete a fitness goals survey before they are allowed to use the facility. Sampling method:

Census

3c. (2) The US Postal Service is considering canceling Saturday delivery, so to collect data, they leave surveys in 100 randomly selected mailboxes for each and every zip code in the entire country. Sampling method:

Stratified

3d. (2) Suppose on the 2nd day of class, instead we did this at the library. We line up at the door of the library, shout in unison, "On your marks! Get set! Go!" and then students run and grab a book off the shelves. Then we record the same variables as we did for our "Calendar Year 2015 Library Data" dataset.

Sampling method:

Convenience

3e. (2) Suppose on the 2nd day of class, instead we did this at the library. Using the library database on MyCecil, which includes a list of every book in the library (about 35,000 volumes), we use a random number generator to determine each student's book. Then we visit the library and find our book. Then we record the same variables as we did for our "Calendar Year 2015 Library Data" dataset.

Sampling method:

Simple Random

4. (3) In words, what is a standard deviation? Give the common language definition. No formulas accepted for credit.

Standard Deviation is "like" the average/
typical distance to the mean -

5. (3) In words, what does "statistically significant" mean? Give the common language definition.

Data that are too unusual to
attribute to chance -

6. (4) In the "Hip Surgery Outcomes" dataset, which "Surgeon" tends to have the longest recovery times and why? Support your answer with statistics.

Hetterberg has highest mean (20.525 days)
and median (21.5 days).
Use measures of central tendency,
not max st!!

ZZZ - Retired -

7. Load up the "Calendar Year 2015 Personality Types" dataset. Show fraction, then decimal with four places rounded correctly, then percentage rounded to two decimal places.

7a. (3) What percentage of the "Females" are "INTJ"?

$$16/224 = 0.0708 = 7.08\%$$

7b. (3) What percentage of respondents are "J" = "Judging"?

$$225/332 = 0.6777 = 67.77\%$$

7c. (3) What percentage of the "ESFJ" are "Male"?

$$5/31 = 0.1613 = 16.13\%$$

7d. (3) What percentage of respondents are "E" = "Extraverted" and "S" = "Sensing"?

$$77/332 = 0.2319 = 23.19\%$$

7e. (3) Make an argument for the independence or dependence of the variables "Thinking / Feeling" versus "Gender". Include supporting conditional percentages for full credit.

Dependent!
66.81% of females are Feeling.
38.68% of males are Feeling.
Huge difference, so gender matters
on this one.

8. (3) Invent a dataset with 9 values with the mean = median = IQR. (All three statistics have the same value.) Put your 9 numbers in ascending order please!

0.5	0.5	0.5	0.5	1	1.5	1.5	1.5	1.5
-----	-----	-----	-----	---	-----	-----	-----	-----

e.g.

ZZZ - Retired

9. Load up the "Calendar Year 2015 Large Survey" dataset. Round all summary statistics to two decimals if necessary.

9a. (2) How many students have a "Commute" of at least 45 minutes? 25

9b. (2) What's the mean "Age" for students who "Smoke"? 26.37

9c. (2) What is the best measure of center for "Number of Tattoos"? Give its name and value.

Median = 0 tattoos

9d. (2) What is the best measure of spread for "Number of Countries"? Give its name and value.

IQR = 2 countries

9e. (2) What is the 90th percentile for "Number of States"? 20 Interpret its value:

90% of our students have been to 20 or fewer States. 10% have been to 20 or more states

9f. (2) How many "Males" use "Pinterest" all the time? 0

9g. (2) Do more "Females" "Rent" or "Own"? Variable is "Living Situation". Own

9h. (2) Which "Gender" has more variation in their "Age"?

Circle:

Male

Female

Why? Justify.

Standard Deviation
IQR

M	F
5.85	8.18
4	6

10. (2) "Maryland Sewer Overflows" dataset. Determine the mean "Duration in Minutes" for just the events from year 2015.

Mean from 2015: 94.55 minutes