Name: Key **Math 127 – Test 1A – Summer 2015**

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

Sign Name: _____

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 <u>not</u> 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. **<u>www.statcrunch.com</u> 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. You can use a calculator, but you cannot use your phone. You can use the calculator on the computers if you wish.

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

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

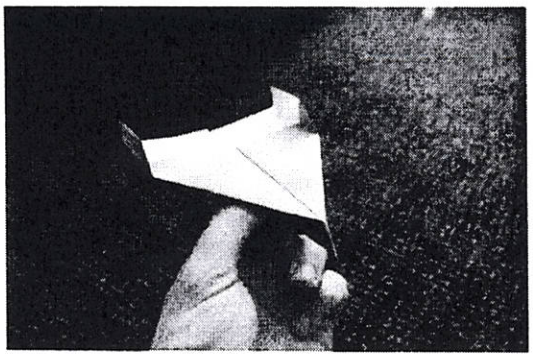
   You may **<u>not</u>** use the pink sheet or copies of the pink sheet.

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

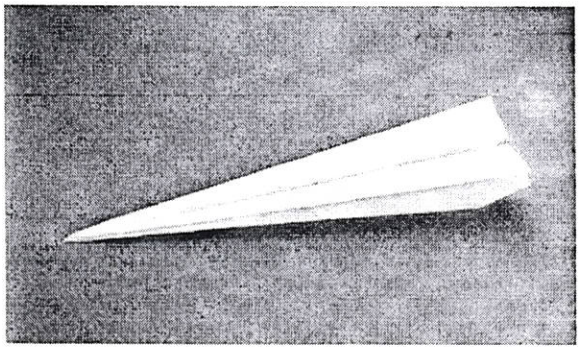
   You may **<u>not</u>** use copies or scans of any instructor-created Math 127 content or answer keys.

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

1. The "**Paper Planes**" dataset contains the results of a randomized experiment. The variable "***Distance***" is measured in feet from the throw point. The student running this experiment wanted to see if "***Design***" mattered and if adding "***Flaps***" mattered for how far a plane could be expected to fly.



This is a glider with flaps. It could also be without flaps.



This is a dart without flaps. Flaps could be added.

Identify the following:

1a. (2) First factor with levels: Design ~ Glider, Dart

1b. (2) Second factor with levels: Flaps ~ Yes, No

1c. (3) Response variable: Distance (ft)

1d. (2) How many different treatments are there? Circle: 1 2 3 (4) 64

1e. (4) Give the treatment means:

Dart with flaps: 21.13 ✓ Dart without flaps: 21.98

Glider with flaps: 13.45 Glider without flaps: 10.00

OK but wrong: Dart Flaps: 21.55
Glider Flaps: 11.72

2. Identify the official sampling methodology for each scenario (simple random, stratified, cluster, systematic, convenience, census, multistage).

2a. (3) At StatCrunch U, the 46,000 students were numbered, went home and put on their numbered T-shirts, came back to campus, and then using a random number generator, $n = 32$ students were selected.
Sampling method: Simple Random

2b. (3) The registrar at Cecil College keeps grade records for every single student that has ever attended this institution. **Sampling method:** Census

2c. (3) Professor Kupe has his financial records for every purchase he has made in the last 8 years on www.mint.com. He takes every 13th purchase starting at the beginning of the list to determine his median purchase amount.
Sampling method: Systematic

2d. (3) Suppose on the 2nd day of class, instead we did this at the library. With 20 aisles and 8 bookcases per aisle, there are 160 total bookcases in our library. We draw a random number between 1 and 160, go find that bookcase, and take every book found on that bookcase as our sample. Then we record the same variables as we did for our "**Calendar Year 2015 Library Data**" dataset.
Sampling method: Cluster

3. (3) Circle the only correct expression.

$P_{50} < \text{Median} < Q_2$

$P_{50} \leq \text{Median} \leq Q_2$

$P_{50} = \text{Median} = Q_2$

4. (2) Which of the following statistics **could** take on negative values? Circle all that are correct.

median	mean	standard deviation	IQR	minimum
range	Q_1	z-score	Q_3	maximum

5. (3) In words, what is a z-score? Give the common language definition. No formulas accepted for credit.

It is the number of standard deviations a data value lies from the mean

6a. (10) Describe the distribution of "***Recovery (in days)***" in the "**Hip Surgery Outcomes**" dataset. Use the values of the best summary statistics in your write up. Write in sentences in the context of the problem.

Giving every single statistic that StatCrunch can produce will result in a penalty.

Also, determine the official number of outliers as governed by the fences. Tell where outliers start. Include all values that are outliers in your write up.

Finally, let's agree that for all intents and purposes, "***Recovery (in days)***" is pretty darn symmetric.

(1) Unimodal and Symmetric

(1) Mean = 19.44 days (1)

(1) Standard Deviation = 5.84 days (1)

(1) High outliers start at 34 days -
we have two: both 35.9 days (1) (1)

Low outliers start at 5.2 days (1) -
we have five: 2.6, 2.7, 3.8, 4.4, 5.1 (2)

Space to calculate fences by hand:

$UF = Q_3 + 1.5(IQR) = 23.2 + 1.5(7.2) = 34$

$LF = Q_1 - 1.5(IQR) = 16 - 1.5(7.2) = 5.2$

6b. (4) Looks like there was a tie for the longest recovery time. Convert that value to a z-score. Show work.

$$z = \frac{y - \bar{y}}{s} = \frac{35.9 - 19.44}{5.84} = 2.818$$

6c. (4) A patient's recovery time was lost. If his z-score was 0, what was his recovery time? 19.44 days

ZZZ - Retired

7. Load up the "**Calendar Year 2015 Large Survey**" dataset. Show fraction, then decimal, then percentage rounded to two decimal places on all the categorical variable questions.

7a. (3) What percentage of the ***"Females"*** drink "***Very Often***"?

~~5/98 = 0.0510 = 5.1%~~ 11/229 = 4.8%

7b. (3) What percentage of all respondents are "***Extremely Religious***"?

~~10/145 = 0.0690 = 6.9%~~ 15/333 = 4.5%

7c. (3) What percentage of our "***Married***" students think "***Marriage is Obsolete***"?

~~3/16 = 0.1875 = 18.75%~~

5/37 = 0.1351 = 13.51%

7d. (3) One student didn't answer "***Facebook***". How often does she drink "***Alcohol***"? Sometimes

7e. (3) Show calculation. Professor Kupe is in row 1. Using *z*-scores, is the "***Number of States***" he has visited unusually high?

$z = \frac{y - \bar{y}}{s} = \frac{39 - 10.27}{7.13} = 4.029$ ~~$\frac{39 - 11.12}{7.84} = 3.56$~~

Oh yeah! Extremely high # (for this dataset!)

7f. (3) Interpret with a sentence, the 90th percentile for "***Number of Tattoos***":

$P_{90} = 4$

90% of our students have 4 or fewer tattoos.
10% have 4 or more tattoos.

7g. (3) What is the mean "***Height***" of the females? ~~64.34"~~ 64.43"

7h. (3) What is the best measure of center for "***TV Time***"? Give its name and value.

Median = 8 hours

7i. (3) What is the best measure of spread for "***TV Time***"? Give its name and value.

IQR = 8 hours

7j. (3) Are a majority of our respondents in the 20 to 29 age bracket? Circle: Yes (No)

45.81%

(~~48.97%~~ is not more than half.)

Sum must be $11(6.5) = 71.5$

8a. (3) Invent a dataset with 11 values with a mean of **exactly** 6.5 and a Q_3 of **exactly** 8.

-8.5	8	8	8	8	8	8	8	8	8	8

e.g. answers not unique ↕

8b. (3) Invent a dataset with a standard deviation of exactly 1, as computed by StatCrunch:

$-1, -1, 0, 1, 1$ works

Recall $s = \sqrt{\frac{\Sigma(y-\bar{y})^2}{n-1}}$

9. (4) Fire up the "**NYPD January 2012**" dataset. Included are the police interrogation records from all 69,073 police interactions that month. Argue if getting frisked is independent of or dependent on gender. Support with the proper conditional and / or marginal percentages.

Totally dependent!
30.5% of females were frisked.
57.85% of males were frisked.
Gender totally matters!

10. In the "**Honshu Japan Earthquake**" dataset, we have the 446 earthquakes that occurred in the region immediately before and after the magnitude 8.9 quake on March 11, 2011.

10a.(3) Give the "***Who***": Each Earthquake

10b.(3) The variable "***NST***" is the number of earthquake stations that reported earthquake activity for each particular earthquake. Is the variable categorical or quantitative?

Categorical Quantitative

10c.(3) Give the mean "***Magnitude***" for just the earthquakes that occurred on March 14th. 4.982