

Name: _____

Key

Math 127 Exam 1 Summer 2014

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

Sign Name: _____

Every question is worth 2 points, except those marked * are worth 4 points.

1. The good folks at Tide have a new formulation of detergent and they would like to run a designed experiment to determine if water temperature and / or type of stain impacts the detergent's effectiveness.

Forty identical white cloths (10 each) are stained with either ketchup, mustard, relish, or blackberry jam. Half of the ketchup-stained clothes are washed in hot water, the other half in cold water. This is also done for the mustard, relish, and blackberry jam stained clothes. The treatment assigned to each cloth is done randomly using a computer program.

After washing, the white cloths will be graded twice: once using an instrument called a colorimeter and once eyeballed by an expert grading the cloth on a scale of 1 to 10.

Identify the following:

- a. Experimental units: 40 cloths
- b. Factor(s) and Level(s): Stain: Ketchup, Mustard, Relish, Jam
Temp: Hot, Cold
- c. How many different treatments are there? 8
- d. How many cloths are receiving each treatment? 5
- e. Is there a control group? No
- f. Identify the response variable(s): Colormeter Rating
Expert Rating
- g. Is there replication? Explain. Yes, each treatment given
5 times
- h*. Once we collect our data, explain briefly how you might begin to analyze it for this experiment.
Graph: Side-by-Side Boxplots
Stats: Look at 8 treatment means

2. Invent a dataset with 9 values where the mean is 0 and the median is 1.

1	1	1	1	1	1	1	1	-8
---	---	---	---	---	---	---	---	----

3. Invent a dataset with 8 values where the IQR is larger than the standard deviation.

-100	-100	-100	-100	100	100	100	100
------	------	------	------	-----	-----	-----	-----

Give the values: IQR = 200 Standard Deviation = 106.9045

4. Run a randomization test on StatCrunch to test for the difference between two means.

Open the "In Class Memory Game" dataset from class last week.

We are looking for statistical evidence that people who self-classified as "Alert" perform differently than "Sleepy" people, on average, on the memory game.

- a. What is the skeptical theory for this randomization test?

Both groups do equally as well

- b. What is the alternative theory for this randomization test?

The groups perform differently on average

- c. Give the mean "Level" achieved from both groups:

Alert = 10.118 Sleepy = 8.084

- c. What is the difference in means (take the positive difference): 2.033

- d. Run the Resampling simulation on StatCrunch with at least 10,000 randomizations. You will need to use the Where: "Brain State" = "Alert" and Where: "Brain State" = "Sleepy".

- e. Report the P-Value from your randomization test: 0.0392

- f*. Make a decision and concluding remark for this randomization test.

Reject Skeptical Theory, conclude
the groups have different
mean scores on memory test

ZZZ Retired - Cal. Year 2014 Large Survey ⁵

5. Load up the "Large Survey Math 127" dataset. Show fraction and percentage rounded to two decimal places on all the percentage / proportion questions.

a. What percentage of "Females" are "Carefree"? ~~$\frac{36}{82} = 43.90\%$~~ $\frac{104}{208} = 50\%$

b. What percentage of "Carefree" people are motivated by "Love"? ~~$\frac{20}{65} = 30.8\%$~~ $\frac{56}{176} = 31.82\%$

c. What percentage of students think their community is a "Poor" place to raise children?
 ~~$\frac{5.26}{100} = 5.26\%$~~ ~~$\frac{7}{133} = 5.26\%$~~ $\frac{21}{316} = 6.65\%$

d. How many female respondents are under the age of 40? ~~78~~ 193

e. How many students are official outliers for their "Commute" time? ~~10~~ 27

f. What percentage of students are both Catholic and believe that astrological signs are important?

~~$\frac{5}{125} = 4\%$~~ $\frac{18}{308} = 5.84\%$

g. What is the mean age of the students who smoke? ~~25.5~~ 25.14

h. What is the best measure of center for "Number of Countries"? Give its name and value.

Median = 1

i. What is the best measure of spread for "TV Hours"? Give its name and value.

IQR = 9

j. Describe just the shape of the distribution of "Commute". Unimodal
Skewed Right

k. Describe just the shape of the distribution of "Time Working" each week.

(No Skew) Bimodal (or Multimodal)

l. Determine the median "Days Exercising" for just the students 30 years old or older. 3

m. Argue if pet preference is dependent on or independent of gender. Support with conditional percentages as needed. The variable is named "Cat or Dog".

Independent. Pie charts

Very Close.

Cannot Find a 10% across groups by gender.

6. In the "**Pizza**" dataset, we have 29 brands of grocery store pizza, along with a few pertinent variables. Use the summary statistics for just the "**Cheese**" pizzas to answer the following questions. Do not add or subtract any data values to the dataset when calculating your summary statistics.

- a. A new **cheese** pizza might be introduced to the market. It has 252 calories per serving. Is that an unusually low- or high-calorie pizza? Show calculations as needed.

$$\bar{y} = 339.24 \quad S = 28.17$$

$$z = \frac{252 - 339.24}{28.17} = -3.097$$

Yes, very rare.

- b. Another **cheese** pizza has a z-score of -0.94. Determine its "**Calorie**" count. Show work.

$$-0.94 = \frac{y - 339.24}{28.17}$$

$$y = 312.76 \text{ calories}$$

- c*. Determine the values of the lower and upper fences for the "**Calories**" of the "**Cheese**" pizzas. Show work. Are there any official outliers? If so, list them out by calorie count.

$$UF = 364 + 1.5(39) = 422.50$$

$$LF = 325 - 1.5(39) = 266.50$$

No Outliers

- d. What percentage of "**Cheese**" pizzas cost more than one dollar per serving? Show fraction and percentage rounded to two decimal places.

$$\frac{5}{17} = 29.41\%$$

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

- a. Online dating habits among male college students were analyzed with a sample of size 200 from the male population of size 7000. The sample was chosen using every 35th name on the list of all 7000 students.

Systematic

- b. A company is being audited and the auditors will pull from the pool of 5500 accounts receivable records. The total pool can be broken down into groupings by dollar amount: Under \$1000, from \$1000 to \$5000, and over \$5000. The auditors will randomly select 100 records (using a computer program) from each grouping.

Stratified

- c. At Cecil College, the Registrar has access to every student who has ever graduated. They can be identified using their Student ID numbers. She will select 135 student files at random using a computer program and perform a degree audit to ensure our students with degrees have actually met the degree requirements.

Simple random

- d. The county supervisors are curious if the rural residents of Cecil County would support the installation of a sewer system. The supervisors divide up the rural parts of the county into 195 zones, and then randomly select (using a computer program) six zones. Once the zones are selected, every household in those zones will be visited and surveyed.

Cluster

- 8*. In the "Darts" dataset, a student ran an experiment to test her accuracy of throwing a dart. The response variable is "Accuracy", measured inches to the bulls eye, and smaller numbers are better. Her dominant hand is her right hand, so that one should be better on average.

She stood at three different "Distances" and used both of her "Hands" for six throws each.

Determine the mean "Distance" for each of the six treatments. Which treatment is best?

$$\bar{Y}_{R, \text{CLOSE}} = 2.13$$

$$\bar{Y}_{L, \text{CLOSE}} = 3.45$$

$$\bar{Y}_{R, \text{MEDIUM}} = 3.58$$

$$\bar{Y}_{L, \text{MEDIUM}} = 4.95$$

$$\bar{Y}_{R, \text{FAR}} = 3.9$$

$$\bar{Y}_{L, \text{FAR}} = 7.52$$

9. The general manager at Wal-Mart has records for every employee at the Elkton location. Included in the files are social security number, hourly pay rate, number of tax exemptions, whether or not the employee is full- or part-time, and department.

Describe the population:

All Elkton WM employees

Describe the "Who":

Each employee

List the "What(s)" and label as C, Q, or I

SSN (I)
Pay Rate (Q) # Tax Ex (Q)
Full / Part (C) Dept. (C)

10. When sampling a population, there are three main ideas that were talked about in the notes and the videos. Name two of them.

Randomize, Sample Size Matters, Examine Part
of the whole

- 11*. Using the "NYPD January 2012" dataset, analyze the 69,073 interactions police had that month. Is there a dependency between "Race" and whether or not someone was "Frisked"? Support with conditional proportions.

% White Frisked: ~~21043~~ 2577 / 5997 = 42.97%
 % Black Frisked: 21043 / 36548 = 57.58%

Big difference, Race matters.
Dependent.

- 12*. Using the "Kiplinger's" dataset, give a range of values for "Admission Rate" that would not be unusual.

$$\bar{y} \pm 2S$$

$$57.95 \pm 2(15.71)$$

$$26.37\% \text{ to } 89.53\%$$