

Name: Key

Math 127 Exam 1A Fall 2014

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

Sign Name: _____

Points are in parentheses for each question.

Students may use:

1. A handheld calculator that is not your cell phone and not the computer calculator.
2. One sheet of hand-written or typed notes on 8.5" by 11" paper. No photos, scans, or copies of any posted materials (like answer keys or completed worksheets).
3. The StatCrunch webpage and datasets.

This exam must be completed in one sitting, but it is untimed.

Good luck!

1:30
mins left
10 left

1:45,
3 left

9-90

11-80

1-70

1-60

1-50

Median 85

Key

ZZZ Retired - Calendar Year 2014 Large Survey²

1. Load up the "~~Large Survey Math 127~~" dataset. Show fraction and percentage rounded to two decimal places on all the categorical variable questions. Round all summary statistics to two decimals if necessary.

- 1a. (2) What percentage of the "**Males**" said that their community is a "**Poor**" place to raise children?

$$\frac{8}{107} = 7.48\%$$

- 1b. (2) What percentage of all respondents are "**Catholic**" and "**Very Religious**"?

$$\frac{7}{310} = 2.26\%$$

- 1c. (2) What percentage of all respondents are motivated by "**Knowledge**" or "**Expression**"?

$$\frac{158}{317} = 49.84\%$$

- 1d. (2) What percentage of those who would "**Google**" a date are "**Female**"?

$$\frac{81}{105} = 77.14\%$$

- 1e. (2) How many students have a "**Commute**" of at least 30 minutes? 53

- 1f. (4) Show calculation. Determine the upper fence for "**Commute**".

$$\begin{aligned} UF &= Q_3 + 1.5(IQR) \\ &= 20 + 1.5(10) = 35 \end{aligned}$$

- 1g. (2) How many students have commutes that are official high outliers? 27

- 1h. (2) What is the mean age of the students who smoke? 25.14

- 1i. (2) What is the best measure of spread for "**Age**" for just the females? Give its name and value.

$$IQR = 7$$

- 1j. (2) What is the best measure of center for "**TV Hours**" for just the males? Give its name and value.

$$Median = 10$$

- 1k. (2) The main shape feature of "**Time Working**" is that the distribution is bimodal.

- 1l. (4) Doctor Climent is in row 291. Convert his age to a z-score. Show calculation.

$$\begin{aligned} Z &= \frac{y - \bar{y}}{s} = \frac{70 - 22.75}{7.77} \\ &= 6.08 \end{aligned}$$

- 1m. (2) On average, how much taller are the males compared to the females? 6.65"

Key

3

2. A piano player has a recurring gig at a lounge every Friday night and he wants to maximize his tips. He has three repertoires he can play – classical pieces, popular cover tunes, and obscure cover tunes. Since he also dabbles in statistics, he will run a designed experiment to determine which repertoire is the best for generating tips. On each of the next 30 Friday nights, he will randomly select one of three repertoires. Also, every other week, he will bring his female lounge singer to accompany him. Disregard the fact that the tips will be split if she is there for the purposes of this experiment.

2a. (2) Experimental units: Each gig

2b. (2) Factor 1 and its levels: Repertoire

Classical Popular Covers Obscure Covers

2c. (2) Factor 2 and its levels: Female Singer

Yes No

2d. (2) List out the 6 treatments: Classical / Yes

Classical / No

Popular Covers / Yes

Popular Covers / No

Obscure Covers / Yes

Obscure Covers / No

2e. (2) Identify the response variable: Tips

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

3a. (2) 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. **Sampling method:** Systematic

3b. (2) 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. **Sampling method:** Stratified

3c. (2) 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. **Sampling method:** Cluster

Answers Var! Key

4

4a. (5) Invent a dataset with 15 values where the third quartile is 5 and the 90th percentile is 10.

5	5	5	5	5	5	5	5	5	5	5	10	10	10
---	---	---	---	---	---	---	---	---	---	---	----	----	----

4b. (5) Invent a dataset with 8 values where the IQR is larger than the standard deviation.

0	0	0	0	1	1	1	1
---	---	---	---	---	---	---	---

Give the values: IQR = 1 Standard Deviation = 0.535

5a. (5) Fire up the "Bachelor's Degree Institutions" dataset. Argue if "Campus Housing" is independent of or dependent on whether or not a school is "Public / Private".

% with housing if Public: 82.3%

% with housing if Private: 63.38%

Dependent: Big increase in % with housing if know it is Public instead of Private.

5b. (5) Same dataset. Describe the distribution of the number of "Undergrads" at these institutions.

Unimodal, Skewed Right

Median = 1334.5

IQR = 3024

Lots of High Outliers (51)
Anything Above 8072

Key

6. Open up the "2010 Hurricanes" dataset. Included are all hurricanes and major tropical storms from the year 2010. Consider this dataset a census, so we have population data, not sample data.

6a. (2) Describe the "Who": Each Hurricane

6b. (5) There are six "Whats" in this dataset. "Dates" is a time / date variable, so we'll call it T. Label the rest as Q, C, or I. We are now taking the stance that I and C answers are completely separate. For clarity, I is no longer a subset of C.

<u>I</u> Number	<u>I</u> Hurricane	<u>T</u> Dates
<u>Q</u> Max Wind	<u>Q</u> Pressure	<u>C</u> Category

6c. (2) The mean "Max Wind" was 72.14 miles per hour. Is this a Statistic or a Parameter? P

6d. (3) The "Pressure" for Hurricane Earl was 928. Is this unusual or rare in terms of z-scores? Show calculation and make a concluding remark.

-2.08 z-score Concluding Remark: Yes, it is an unusually strong / low pressure storm.

Show work:

$$z = \frac{928 - 979.94}{24.92} = -2.08$$

6e. (10) Pretend that the Weather Channel's instruments were faulty and all "Max Wind" measurements were actually 5 miles per hour stronger than what was recorded in this dataset.

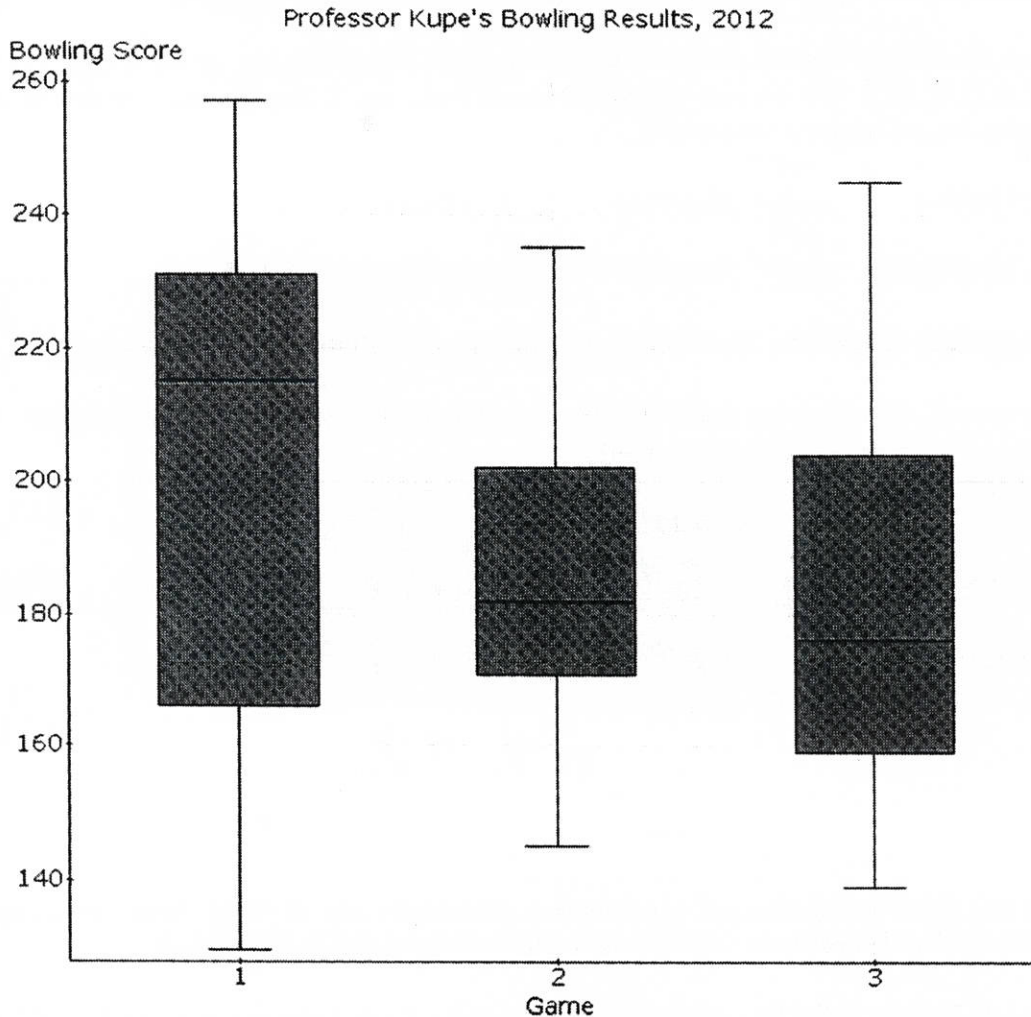
What would happen to the following summary numbers if the correction was made to the dataset?

Circle correct choice for each row:

Mean:	<u>Increase by 5</u>	Decrease by 5	Remain unchanged
IQR:	Increase by 5	Decrease by 5	<u>Remain unchanged</u>
40th Percentile:	<u>Increase by 5</u>	Decrease by 5	Remain unchanged
Maximum:	<u>Increase by 5</u>	Decrease by 5	Remain unchanged
Standard Deviation:	Increase by 5	Decrease by 5	<u>Remain unchanged</u>

Key

7. Circle the best choice for each question below. Use only these boxplots. **This data is different than what is posted in "Kupe Bowling", so loading up that dataset will get you nowhere!** The correct answer is guaranteed to be listed.



You must presume that for each "Game", all the "Scores" are unique and there are no pile-ups of data at any value. These are straightforward questions, not trick questions.

- 7a. (2) Game 1 Median: 166 200 215 231
- 7b. (2) Game 2 IQR: 20 31 90 182
- 7c. (2) Game 3 Q₁: 139 159 176 204
- 7d. (2) Game 1 90th Percentile: 220 231 244 257
- 7e. (2) Game 2 Range: 90 145 235 260
- 7f. (2) Game 3 15th Percentile: 139 150 159 224