MATH 338

EXAM 1

MONDAY, JULY 10, 2017

Your name: _____

Your scores (to be filled in by Dr. Wynne):

Problem 1: ____/8

Problem 2: ____/14

Problem 3: ____/6

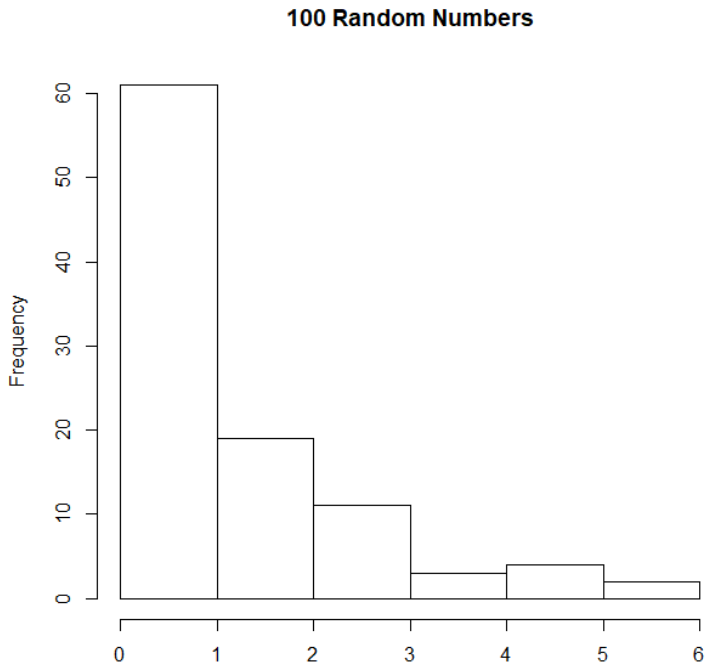
Problem 4: ____/16

Total: ____/44

You have 60 minutes to complete this exam. This exam is closed book and closed notes with the exception of your formula sheet.

For full credit, show all work except for final numerical calculations (which can be done using a scientific calculator).

1. A sample of 100 numbers was generated from a mystery distribution with population mean 1. A histogram of the 100 numbers is shown below. Circle the most correct answer to parts A-H (1 pt each):



A) Did these numbers come from a normal distribution? Probably Probably not

B) The mystery distribution is most likely: Skewed left Symmetric Skewed right

C) The median of the sample is: Less than 1 Between 1 and 2 Greater than 2

D) The third quartile of the sample is: Less than 1 Between 1 and 2 Greater than 2

E) If I created 10,000 random numbers instead of 100, their distribution would be closest to:

Normal Mystery distribution Some other distribution

F) If I created 100 numbers 10,000 times, the distribution of the sample means would be closest to:

Normal Mystery distribution Some other distribution

G) If I created 100 numbers 10,000 times, the sample mean of the sample means would be:

Less than 1 Equal to 1 Greater than 1 Can't determine

H) The area under the density curve describing the population distribution would be:

Less than 1 Equal to 1 Greater than 1 Can't determine

For Part F: Full credit for "Some other distribution" on the assumption that CLT doesn't quite hold with this much skew for n = 100, so the distribution would approach normal but not quite get close enough

2. The table below shows the number of formal votes cast in each of the 11 electoral Divisions in South Australia during the 2016 Australian Federal Election.

Division	Turnout		Division	Turnout
Adelaide	95,544		Makin	94,070
Barker	94,445		Mayo	95,116
Boothby	95,288		Port Adelaide	96,009
Grey	89,876		Sturt	93,032
Hindmarsh	98,032		Wakefield	95,933
Kingston	93,391			

A) [1 pt] Compute the mean number of voters in an electoral division.

Mean = (95544 + 94445 + … + 95933)/11 = 94612.36

Full credit for any reasonable rounding

B) [1 pt] There were a total of 11 electoral Divisions in South Australia in 2016, so the value you calculated in part (A) is a (circle one): statistic parameter

C) [1 pt] Which variable is a label variable (circle one)? Division Turnout Neither

D) [3 pts] Construct a stem-and-leaf plot displaying this data.

1 pt looks like a stem and leaf plot

2 pt all 11 values are represented correctly

D) [8 pts] Compute the five-number summary for this data, and sketch a boxplot. Determine whether outliers exist, and if they do, show them on the boxplot.

0.5 pt: Min = 89,876 0.5 pt: Max = 98,032

1 pt: Q1 = 93,391 1 pt: Median = 95,116 1 pt: Q3 = 95,933

2 pts: IQR = 95,933 – 93,391 = 2542, so outliers are below 93,391 – (1.5)(2542) = 89578 and above 95,933 + (1.5)(2542) = 99,746. Therefore there are no outliers in this data set.

2 pts: sketch an appropriate box-and-whiskers plot

3. Read the following excerpt from a published study's Methods section:

"The experiment manipulated the extent to which people (N = 689,003) were exposed to emotional expressions in their News Feed. This tested whether exposure to emotions led people to change their own posting behaviors…Two parallel experiments were conducted for positive and negative emotion: One in which exposure to friends' positive emotional content in their News Feed was reduced, and one in which exposure to negative emotional content in their News Feed was reduced. In these conditions, when a person loaded their News Feed…each emotional post had between a 10% and 90% chance (based on their User ID) of being omitted from their News Feed for that specific viewing.…Posts were determined to be positive or negative if they contained at least one positive or negative word, as defined by Linguistic Inquiry and Word Count software (LIWC2007) word counting system…such that no text was seen by the researchers. As such, it was consistent with Facebook's Data Use Policy, to which all users agree prior to creating an account on Facebook, constituting informed consent for this research. Both experiments had a control condition, in which a similar proportion of posts in their News Feed were omitted entirely at random...Participants were randomly selected based on their User ID, resulting in a total of ~155,000 participants per condition who posted at least one status update during the experimental period."

The subjects were not aware that News Feeds had been manipulated until this study was published.

A) [1.5 pts] What were the cases in this study, and how many were there?

The cases were 689,003 Facebook users who posted at least one status update

B) [1.5 pts] From the study methods below, circle all that were used in this study:

Blinding Randomization Matching

0.5 per answer (circling first two, leaving last uncircled)

C) [1 pt] What was the factor in this experiment?

Removing emotional vs. random posts (full credit for anything dealing with the type of post being removed)

D. [2 pt] Argue for or against the statement, "The researchers should have told people that they were manipulating their News Feeds before they started collecting data."

For: This experiment is a sketchy use of the Data Use policy; subjects could still be blind if they knew they were in the study, but not which group they were in.

Against: Subjects could behave differently if they knew their News Feeds were being manipulated, or there could be a lack of ecological validity; generally, in A/B tests users don't know that they're being manipulated (or what the manipulation is), so this study follows the "rules" of A/B testing.

4. It is 1936 and someone is flooding the antiquities market with golden idols from Peru. After stealing Forrestal's work, you estimate that the weight of the idols is approximately normally distributed with mean 2.6 pounds and standard deviation 0.13 pounds.

A) [4 pts] The golden idol in the Temple of the Warriors weighs 3 pounds. Assuming that your estimations are correct, what percentage of golden idols weigh more than this one?

1 pt work within the z-score framework

1.5 pt obtain z = x – mu/sigma = (3 – 2.6)/0.13 = 3.08

1.5 pt use the table to find that the cumulative proportion is about 0.9987, so about 1 – 0.9987 = 0.0013 or 0.1% of golden idols weigh more than this one

B) [3 pts] Rene Belloq steals the idol from you. Since he is French, he uses kilograms instead of pounds (1 kg = 2.2 pounds). Express the distribution of idol weights in a way that Belloq can understand.

1 pt the distribution is still normal

1 pt the new mean is 2.6*(1 kg/2.2 lbs) = 1.18 kg

1 pt the new sd is 0.13*(1 kg/2.2 lbs) = 0.06 kg

C) [2 pts] Will Belloq get a different answer to part (A)? Why or why not?

1 pt No

1 pt because the z-score for the idol he stole is the same whether the weight is expressed in pounds or kilograms.

D) [7 pts] You find a set of 20 idols in another temple. The total weight of the idols is 50.85 pounds. What is the probability that the mean weight of a random sample of 20 idols is smaller than the mean weight of the idols in this temple?

1 pt compute the sample mean = 50.85/20 = 2.5425

3 pt the distribution of the sample means is N(2.6, 0.13/sqrt(20)) = N(2.6, 0.029)

1 pt z = (2.5425 – 2.6)/(0.029) = -1.98

1 pt cumulative proportion is about 0.0228

1 pt write conclusion: The probability that the mean weight of a random sample of 20 idols is smaller than the mean weight of the sample we observed is 0.0228

Extra Space. The tables below show a number of critical values *z* for the standard normal variable $Z \sim N(0, 1)$ and the corresponding cumulative proportions, corresponding to P(Z ≤ z).

z-score	Cumulative Proportion
-3.00	0.0013
-2.50	0.0062
-2.00	0.0228
-1.65	0.0495
-1.28	0.1003
-1.00	0.1587
-0.67	0.2514

z-score	Cumulative Proportion
0.67	0.7486
1.00	0.8413
1.28	0.8997
1.65	0.9505
2.00	0.9772
2.50	0.9938
3.00	0.9987

The rest of this space to be used for extra work: