2:57
3:12

ST 305: Exam 1

Spring 2019

By handing in this completed exam, I state that I have neither given nor received assistance from another person during the exam period. I have used no resources other than the exam itself and the basic mathematical functions of a calculator (ie, no notes, electronic communication, notes stored in calculator memory, etc.) Using your calculator for values from probability distributions like the normal or t is OK; however, if you are doing a calculation from a normal distribution show your work all the way to the point of calculating z-scores. I have not copied from another person's paper. I understand that the penalty if I am found guilty of any such cheating will include failure of the course and a report to the NCSU Office of Student Conduct. **I understand that I must show all work/calculations, even if they seem trivial, to get credit for my answers.**

Name: KEY

ID#: _____

$\bar{x} = \frac{1}{n}\sum x_i$ $s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$ $Z = \frac{X-\mu}{\sigma}$		

7:57

Definitions. (5 points each) **Clearly define each of the following terms.**

1. Variable:

   a characteristic of an individual

2. Distribution:

   the possible values of a variable, along with the frequencies of those values

3. Resistant measure:

   a measure unaffected by skewness or outliers

Multiple Choice. (5 points each) **Select the <u>one</u> best answer.**

4. The 5-number summary does NOT include

A
a. the standard deviation
b. the median
c. the upper (third) quartile

5. When displaying a distribution, you are most likely to use a stemplot if

B
a. the data set contains only discrete, quantitative values
b. the data set is small
c. the data set contains one or more outliers

6. The standard normal distribution

B
a. has mean equal to 1
b. has standard deviation equal to 1
c. can have any possible value for the mean and standard deviation

7. When a distribution is roughly symmetric, the mean will usually be

A
a. approximately equal to the median
b. less than the median
c. greater than the median

8. Which of the following properties of distributions is typically described with a number?

A
a. center
b. shape
c. symmetry

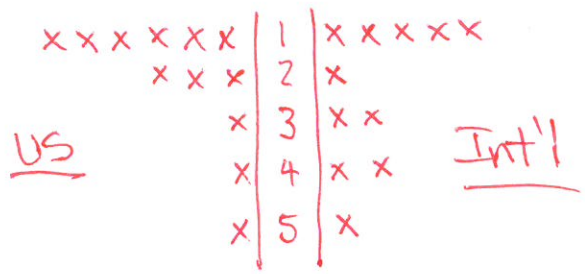
For the remaining problems, SHOW YOUR WORK. Numerical answers with no supporting work or explanation will receive zero credit, even if the calculations are trivial.

9. The following table shows the number of years current US and International graduate students have been in the Bioinformatics PhD program at NC State:

International: 2, 1, 3, 4, 4, 5, 1, 1, 1, 3, 1
US: 3, 1, 1, 2, 2, 1, 1, 1, 2, 5, 4, 1

a) Display the data graphically in a way that allows easy comparisons of the distributions for US and International students. (10 points)

One possibility is this:

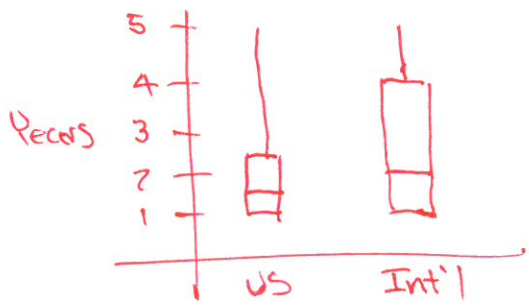


You might also make variations of histograms (side-by-side) or of stemplots.

b) Compute the 5-number summaries for the two groups (Int'l and US), and display them with boxplots on a single graph. (10 points)

US: 1 1 1 1 1 1 2 2 2 3 4 5
min, $Q_1 = 1$, $M = 1.5$, $Q_3 = 2.5$, max

Int'l: 1 1 1 1 1 2 3 3 4 4 5
min, Q_1, M, Q_3, max



c) What were the **variables** in the data set above? (5 points)

Nationality (US / Int'l)
# of years in program

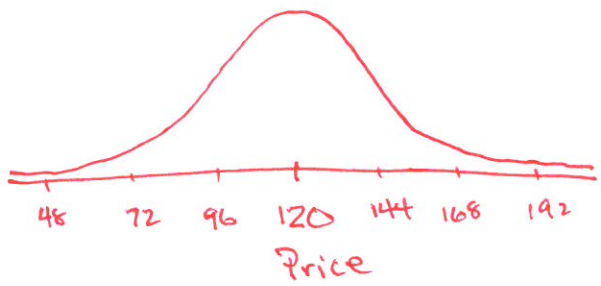
d) What were the **individuals** in the data set? (5 points)

Grad student in the NCSU Bioinformatics PhD program

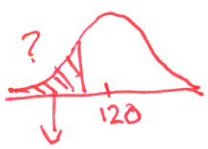
10. The selling prices for vintage fountain pens on eBay auctions follows a distribution that is approximately normal with mean \$120 and standard deviation \$24.

a. **Carefully and accurately** draw and label the density curve describing the selling prices of these pens. (5 points)

$X \sim N(120, 24)$



b. What percentage of these pens sell for \$100 or less? (5 points)

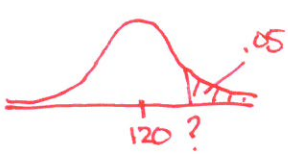


.2033

$$z = \frac{x-\mu}{\sigma} = \frac{100-120}{24} = -.833$$

$$\Rightarrow \boxed{20.33\%}$$

c. I want to focus on the most expensive pens. If I'm only interested in the top 5% of prices, above what price should I be looking? (5 points)



The area right of 1.645 is $\approx .05$

$\uparrow$
z

$$z = \frac{x-\mu}{\sigma} \Rightarrow 1.645 = \frac{x-120}{24}$$

$$\boxed{x = \$159.48}$$

10. What are the four characteristics we always want to mention when describing the distribution of a quantitative variable? (5 points)

1- shape

2- center

3- spread

4- outliers (if present)

11. The 5-number summary for the waiting times (in minutes) of 238 patients visiting a local Urgent Care center is: 0, 8, 11, 18, 82. The mean and standard deviation of the times are 16 and 10.

a. Is this distribution most likely symmetric, skewed left, or skewed right? Justify your answer briefly. (5 points)

Skewed right. We see evidence of this in the form of the boxplot, and the mean (16) is considerably larger than the median.

b. Provide appropriate numerical measures of the center and spread of the distribution. Justify your choices. (5 points)

5 # summary : 0 8 11 18 82

min Q_1 M Q_3 max

or : center : M = 11

spread : IQR = 18 − 8 = 10