STOR 155, Section 1, Midterm II
Tuesday, April 7, 2009

Name: _____Solution_____

Pledge: I have neither given nor received aid on this examination.

Signature: _____

Instructions: Do not do any actual numerical calculations. Answers in a form that you would type into an Excel field, such as "=28*SQRT(82)^2", with a *working* answer, are expected.

1. Measurements of C-Reactive Protein in a set of children are record in cells A1-A25 of an Excel spreadsheet. The 5 number summary for this data set is: 0, 0.1, 0.6, 3.1, 27.

[25]

a. Write the Excel command that was used to compute the 3rd Quartile.

[3]

=QUARTILE(A1:A25,3)

b. Does this distribution appear to be skewed, and if so, which way?

[4]

Yes, right skewed

c. What is the median of this distribution?

[3]

0.6

d. Is the mean expected to be smaller or larger than the median?

[3]

Larger, since right skewed

e. Why is the Inter Quartile Range equal to 3?

[4]

IQR = Q3 – Q1 = 3.1 – 0.1 = 3

f. How do you know there is at least one outlier in this data set?

[4]

Max data = 27 >> Q3 + 1.5 * IQR = 3.1 + 1.5 * 3

g. Suppose the normal Q-Q plot of the log10 of the data roughly follows a straight line. What fraction of the log10 transformed data are within 3 standard deviations of their mean (i.e. mean of the log10 transformed data)?

[4]

Since data approximately normal, use 68-95-99.7 rule: 99.7%

2. Sheila may suffer from high blood glucose levels. This level is classified as high when it is above 150. Sheila's levels vary according to the normal distribution, with mean 160 and s.d. 16.

[25]

a. If a single measurement is made, what is the probability that Sheila is diagnosed as having a high blood glucose level?

[5]

Blood glucose measurement = $X \sim N(160,16)$
$P\{X > 150\} = 1 - P\{X \leq 150\}$ = 1 - NORMDIST(150,160,16,true)

b. What should the number L_1 be, so that there is only probability 0.1 of Sheila's blood glucose level rising above L_1?

[5]

$0.1 = P\{X > L_1\}$
$0.9 = P\{X \leq L_1\}$
So L_1 = NORMINV(0.9,160,16)

c. If measurements are made on four separate days, and the mean result is compared with the criterion, what is the probability that Sheila is diagnosed with high blood glucose?

[5]

Average of 4 blood glucose measurements = $Xbar \sim N(160,16/sqrt(4)) = N(160,8)$
$P\{Xbar > 150\} = 1 - P\{Xbar \leq 150\}$ = 1 - NORMDIST(150,160,8,true)

d. What should the number L_4 be, so that there is only probability 0.1 of the average of four of Sheila's blood glucose levels rising above L_4?

[5]

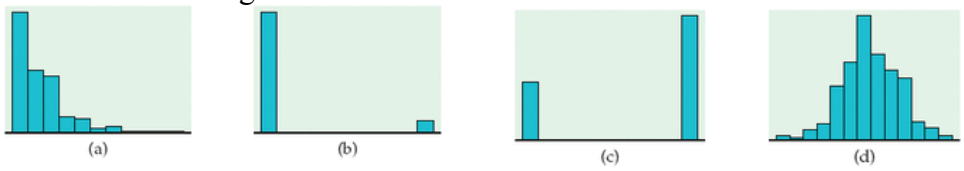
$0.1 = P\{X > L_4\}$
$0.9 = P\{X \leq L_4\}$
So L_4 = NORMINV(0.9,160,8)

a. Use the 68-95-99.7 rule to calculate what percent of single observations are expected to be at 176 or above.

[5]

176 = 160 + 16 = mean + 1 sd. But 68% is between mean +- 1 sd.
So 32% is outside, and 16% is bigger, i.e. > mean + 1 sd

3. A survey of a large college class asked four questions and the numerical responses were summarized in these histograms:



i. Which histogram most likely summarizes results to the question: Are you male or female ? (male = 0, female = 1)

[3] (c)

ii. Which histogram most likely summarizes results to the question: Are you right or left handed? (right = 0, left = 1)

[3] (b)

iii. Which histogram most likely summarizes results to the question: How many minutes do you study on a typical weeknight?

[3] (a)

iv. Which histogram most likely summarizes results to the question: What is your height in inches?

[3] (d)

v. Which histogram most closely follows a normal distribution?

[2] (d)

vi. Which histogram is most likely to have an outlier, using the rule given in class?

[2] (b)

vii. Which histogram best fits the description “left skewed”?

[2] (c)

viii. Which histogram best fits the description “right skewed”?

[2] (a)

ix. Which histogram is most likely to give a linear Normal Quantile plot?

[2] (d)

x. Which histogram is most likely to have an Inter Quartile Range of 0?

[3] (b)

4. A college offers admission to 1000 students. Past experience indicates that about 60% of those offered will accept. Assume that students make their decisions independently.

[25]

a. What is the distribution (please give the name and also the values of the parameters) of the number who accept.

[4]

Binomial(1000,0.6)

b. What is the expected number that will accept?

[3]

EX = n * p = 1000 * 0.6 = 600

c. What is the standard deviation of the number that will accept?

[3]

SD(X) = SQRT(n * p * (1 – p)) = SQRT(1000 * 0.6 * (1 – 0.6))

d. What is the exact probability that at least 650 students will accept?

[3]

$P\{X \geq 650\} = 1 - P\{X \leq 649\}$ = 1 – BINOMDIST(649,1000,0.6,true)

e. Why can the probability in (d) be approximated using the normal distribution?

[4]

np = 1000 * 0.6 > 10
n(1 – p) = 1000 * (1 – 0.6) > 10

f. Calculate the probability in (d), using a normal approximation.

[4]

$P\{X \geq 650\} = 1 - P\{X \leq 649\}$ =
= 1 - NORMDIST(649,1000 * 0.6,sqrt(1000 * 0.6 * 0.4),true)
= 1 - NORMDIST(649,600,sqrt(240),true)

g. What is the standard error of the proportion of students who accept?

[4]

= SQRT(p * (1 – p) / n) = SQRT(0.6 * 0.4 / 1000)