Pledge:

2/22/2005 MATH 271 Name: Solution
Dr. Lunsford Test 1 100 Points Possible

I. Multiple Choice. For each statement/question please circle the best choice for an answer. (3 points each – 15 total)

1. *Using the same sample data*, a 99% confidence interval for a population mean will be ________ than a 95% confidence interval.

$\bar{x}$ still center

$z_{.01/2} > z_{.05/2}$

(a) longer (but generally centered at the same point)
(b) shorter (but generally centered at the same point)
(c) about the same
(d) longer (but generally centered at a different point)
(e) shorter (but generally centered at a different point)

2. If one is using the same confidence level to compute a confidence interval for a population mean (assume σ known), then a confidence interval obtained from a random sample of size $n = 35$ will be _____ than a confidence interval for a random sample of size $n = 350$.

same

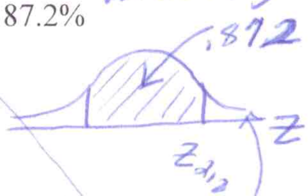
same

$\bar{x} \pm \underbrace{z_{\alpha/2} \frac{\sigma}{\sqrt{n}}}_{\text{becomes smaller as } n \text{ increases...}}$

will be a different value for a different sample in general! ($\bar{x}$ is a random variable!)

(a) longer (but generally centered at the same point)
(b) shorter (but generally centered at the same point)
(c) about the same
(d) longer (but generally centered at a different point)
(e) shorter (but generally centered at a different point)

3. If you want to compute a z confidence interval for a population mean with an 87.2% level of confidence, then $z_{\alpha/2} =$ invNorm(.936)



$.064 = \frac{.128}{2}$

(a) 1.1264 (b) 1.1359 (c) 1.5220 (d) 1.6543

4. A z test statistic for a two-tailed (or sided) significance test for a population mean is $z = -1.4$. Which value below is the p-value for the test:

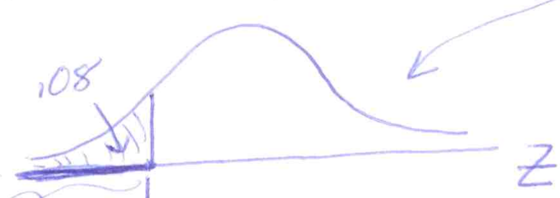
(a) 0.08076 (b) 0.83849 (c) 0.16151 (d) 0.04038 (e) None of these

2(normalcdf(-1E99, -1.4))

5. If you want to use a z-test to test $H_0 : \mu = \mu_0$ against $H_a : \mu < \mu_0$ then you will reject the null hypothesis in favor of the alternative at the $\alpha = 0.08$ level if

(a) $z < -1.405$ (b) $z < -1.751$ or $z > 1.751$ (c) $z > 1.405$

(d) $z < 0.08\%$ (e) None of these



critical region

$-1.405 =$ invNorm(.08)

Problem I. Below are the final grades for 21 randomly chosen students from some of Dr. L.'s recent MATH 171 classes. Note that the data is given in ascending order for your convenience. (48 points total)

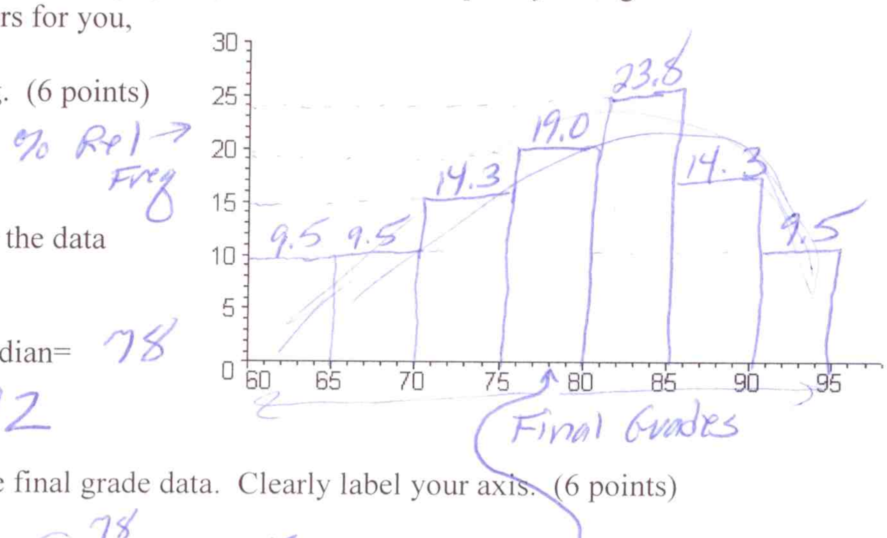
62, 63, 67, 69, 70, 72, 73, 75, 78, 78, 78, 80, 81, 83, 84, 84, 86, 86, 87, 90, 92

(a) Complete the following frequency table. Please keep all class widths the same length. (10 points)

Class Limits for X (Final Grades)	**Frequency**	**Relative Frequency (in % - round to one decimal place)**
$60 \le X < 65$	2	9.5
$65 \le X < 70$	2	9.5
$70 \le X < 75$	3	14.3
$75 \le X < 80$	4	19.0
$80 \le X < 85$	5	23.8
$85 \le X < 90$	3	14.3
$90 \le X < 95$	2	9.5

$\rightarrow \frac{2}{21} \cdot 100$

(b) Use the frequency table in Part (a) to graph a percent relative frequency histogram. Although I have provided numbers for you, please label the axes with what variable/quantity you are plotting. (6 points)

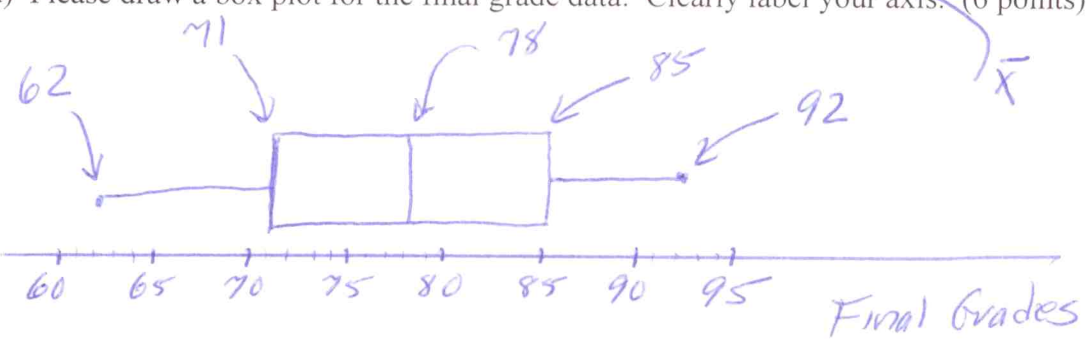


(c) Find the following values for the data above (7 points):

Min= 62 Q_1= 71 Median= 78

Q_3= 85 Max= 92

(d) Please draw a box plot for the final grade data. Clearly label your axis. (6 points)



(e) Given that $\bar{x} = 78.00$ for the final grade data, discuss the shape of the data. You should reference your descriptive statistics and graphs above. (4 points)

Since $\bar{x} = 78 =$ med and the box plot is relatively symmetric (median basically in center of box - whiskers about the same length) and the histogram shows only one peak - I would say this data is basically symmetric & unimodal. (Some students said a slight left skew... based on shape of histogram...)

(f) Given that $s = 8.59$ for the final grade data (and recall $\bar{x} = 78.00$ from above), find a 95% confidence interval for the average final grade in all of Dr. L.'s recent MATH 171 classes. Clearly indicate what type of confidence interval you are finding and why. Clearly indicate your answer. (4 points)

We are finding a t-interval b/c σ is unknown (note that from part (e) it also seems "reasonable" to assume the data is coming from a normal population), 95% conf. interval is $(74.09, 81.91)$

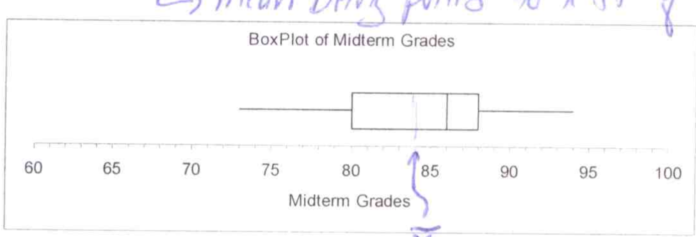
(g) Write a probability statement and an English sentence using the information in your confidence interval in part (b). Clearly indicate what any variables you use represent. (4 points)

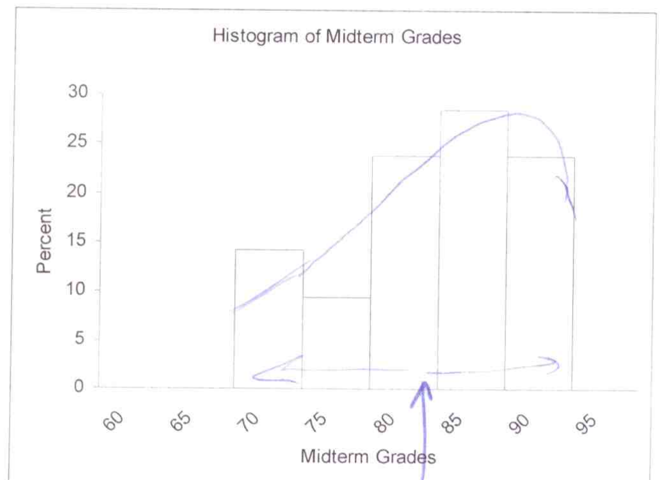
$P(74.09 < \mu < 81.91) = .95$, We are 95% confident that the average final grade in all of Dr. L's recent MATH 171 classes is between 74.09 and 81.91.

(h) Below is a histogram and box plot of the Midterm Grades for the same random sample of students in part (a). The sample mean of the Midterm Grade data is 84.10. Based on the histogram and boxplot below, circle the description that best describes the shape of the Midterm Grade data: (3 points)

Left Skewed   Right Skewed   Uniform   Symmetric Unimodal   Bimodal

↳ mean being pulled to left of median



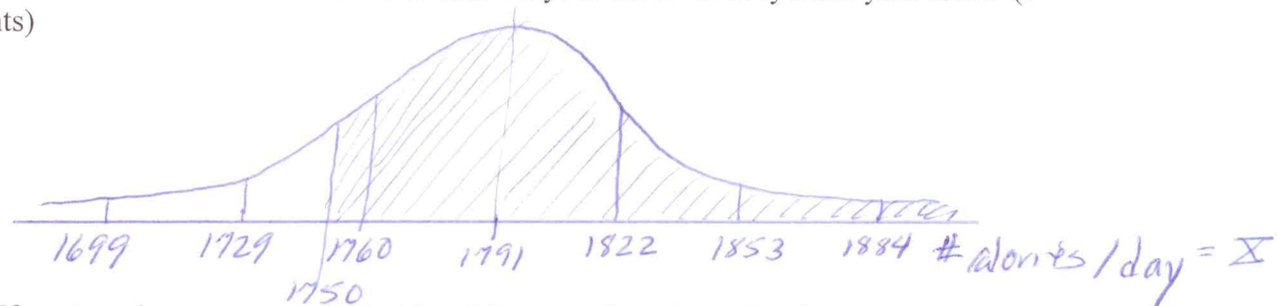


(i) Based on the histograms, boxplots, and descriptive statistics for the Midterm Grade and Final Grade data sets, which data set do you think has the larger standard deviation? Why? (4 points)

The Final Grade data appears to have a larger standard deviation. This is because that data set has more data further from the mean than the Midterm Grade data set.

Problem II. According to the publication *Dietary Reference Intakes for Energy, Carbohydrate, Fiber, Fat, Fatty Acids, Cholesterol, Protein, and Amino Acids (Macronutrients)* published by the Food and Nutrition Board of the Institute of Medicine, 2002, the mean number of calories consumed by women in the United States who are 19 to 30 years of age is 1791 calories per day. The standard deviation is 31 calories. Assume that the number of calories consumed per day is normally distributed. Please use this information to answer the following questions. Please show all work including any functions you use (with input) on your TI calculator. (37 points total)

(a) Draw a graph of the distribution of the number of calories per day consumed by women in the United States. You should show the mean amount and values at at ± 1, ± 2, and ± 3 standard deviations from the mean on your axes. Clearly label your axes. (5 points)

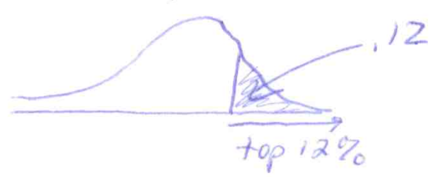


(b) If an American woman who is 19 to 30 years of age is randomly chosen, what is the probability that the number of calories she consumes per day is at least 1750? Please represent this probability on your graph in part (a) above. (4 points)

$P(X \geq 1750) = \boxed{.9070}$

↳ normalcdf(1750, 1E99, 1791, 31)

(c) What is the smallest amount of calories per day an American women in this age range would consume to be considered among the top 12% of all women (in terms of amount of calories consumed per day)? (5 points)



invNorm(.88, 1791, 31)

= $\boxed{1827.42 \text{ calories / day}}$

(d) If 100 American women in this age range are randomly chosen, what is the probability that the group's average amount of calories consumed per day will between 1785 and 1800 calories? (5 points)

$P(1785 \leq \bar{X} \leq 1800) = \boxed{.9717}$

$\bar{X}$ is $N(\mu_{\bar{X}} = 1791, \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{31}{10} = 3.1)$

normalcdf(1785, 1800, 1791, 3.1)

$\mu_0 = 1791$ (average for general)

$\sigma_{general} = 31 = \sigma_{athletes}$

Problem II, continued.

(e) A researcher believes that American women athletes who are 19 to 30 years of age consume significantly more calories per day, on average, than American women in general from that same age group. She will assume the data given in part (a) is accurate for the general population of American women who are age 19 to 30 years and that the population of American women athletes will have the same variance in their calorie consumption as the general population. To test her hypothesis, the researcher *randomly selects 200 American women athletes* who are 19 to 30 years of age and finds that the average number of calories consumed per day for the group is 1797. Formulate the researcher's claim in terms of a hypothesis test. Clearly indicate what any variables you use represent and which hypothesis is the claim. (5 points)

$\bar{x}$

H_0: $\mu = 1791 \leftarrow \mu_0$ $\qquad$ H_a: $\mu > 1791$ (Claim)

Where μ is the average # of calories consumed per day by women athletes.

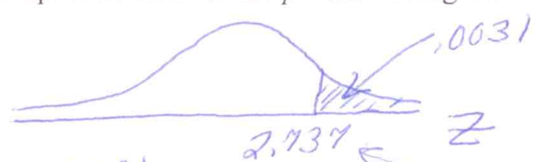
(f) What test statistic will you use to conduct the above hypothesis test? Clearly indicate the value of the test statistic and the values of all variables needed to compute the test statistic. Also indicate any assumption(s) you need to make in order to use the test statistic you have chosen. (4 points)

Since n is large ($n=200$) and $\sigma_{athletes}$ is assumed to be known, we will use a z-score (ie a z-test)

$$z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{1797-1791}{31/\sqrt{200}} = 2.737$$

(c) Find the *p*-value for the data. State the meaning of the *p*-value in English in the context of this hypothesis test and draw a pictorial representation of the *p*-value using the appropriate test statistic curve. (5 points)

p-value $= .0031$



If we assume that the average # of calories consumed per day by athletes is 1791, then the probability of getting a result "as extreme as our data in the direction of the alternative hypothesis ($\mu > 1791$)" is .0031 (i.e. only a .31% chance of getting a result this extreme if H_0 true!)

(d) If you conduct this test at the $\alpha = 0.05$ level of significance, then what is your conclusion (both in terms of the hypotheses and the claim) and why? (4 points)

Since $p < \alpha$ we reject the null hypothesis in favor of the alternative hypothesis. Thus the data supports the researcher's claim that women athletes consume more calories per day, on average, than the general population of women from the same age group (19-30).