Statistics 23, Section 1, Midterm II
Tuesday, November 9, 1999

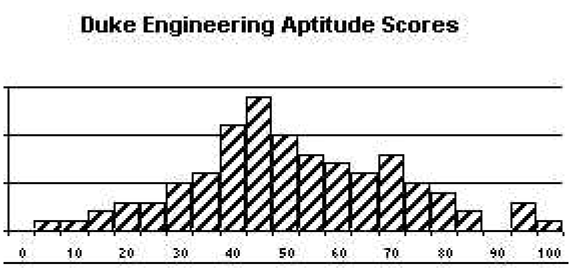
Name: ______________Solution____________________________________

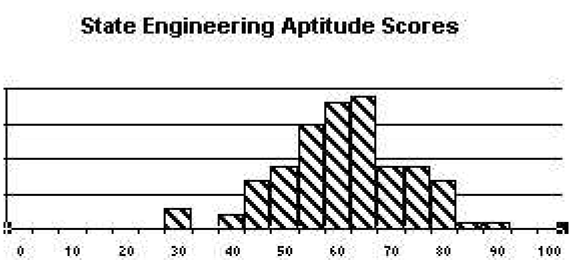
Pledge: I have neither given nor received aid on this examination.

Signature: ____

Instructions: Show all work, but do not do hard arithmetic (an answer like $8.3 + \frac{7}{\sqrt{5.1}}$ is fine).

1. Random samples of 100 Duke Students, and 100 State students were given a test of engineering aptitude. Histograms of the scores are:





a. If, as a project manager, you could hire only one engineer at random from on school or the other, which would you prefer? Why?

[5]

State, since higher on average.

b. If you could carefully select from a large collection of candidates, and it was important to hire the best person possible, which school would you look to? Why?

[5]

Duke, since their top people are the highest.

2. The chance of winning a Lotto game is about 1 in 8 million. Suppose you buy a $2 Lotto ticket, in hopes of winning the $6 million grand prize. Calculate your expected net winnings and interpret the result.

[5]

Net Winnings = -$2 w.p. (8 mil – 1) / 8mil
$5 mil –2 w.p. 1 / 8 mil

Expected Net Winnings = -$2 ((8 mil – 1) / 8mil) + ($5 mil) (1 / 8 mil)
= -$2 + $5 / 8 = -$1.25

3. For $X \sim Bi(16,0.5)$,

a. Will the normal approximation be reasonably accurate? Why or why not?

[5]

EX – 3 sd = 16 (0.5) – 3 sqrt(16(0.5)0.5) = 8 – 6 = 2
EX + 3 sd = 16 (0.5) + 3 sqrt(16(0.5)0.5) = 8 + 6 = 14

Will be reasonably accurate, since 0 < 2 and 14 < 16.

b. Write a formula, that could be used in an Excel formula bar, to calculate $P\{|X-4|>2\}$ using the normal approximation.

[5]

P{|X – 4| > 2} = P{X < 2 or X > 6} = P{X < 1.5} + (1 - P{X < 6.5})

=NORMDIST(1.5,16*0.5,SQRT(16*0.5*0.5),TRUE) +
(1-NORMDIST(6.5,16*0.5,SQRT(16*0.5*0.5),TRUE))

c. Write a formula, that could be used in an Excel formula bar, to find n so that $P\left\{\left|\frac{X}{n}-0.5\right|\le 0.03\right\}=0.90$.

[5]

0.95 = P{Z < 0.03 / sqrt(0.5(0.5)/sqrt(n))} = P{Z < 0.06 sqrt(n)}

so n =(NORMINV(0.95,0,1)/0.06)^2

4. 400 young workers and 300 older workers were asked to rate their job satisfaction on a scale of 0-10. The average scores were 5.8 and 6.4, and the standard deviations were 1.2 and 1.4, respectively.

a. Write a formula that could be put into an Excel formula bar to calculate the probability that the average job satisfaction of the whole population of younger workers is 6 or more.

[5]

Xbar ~ N(5.8, 1.2 / sqrt(400)),

P{Xbar >= 6} = 1 - P{Xbar < 6} =1-NORMDIST(6,5.8,1.2/SQRT(400),TRUE)

b. Write formulas that could be put into an Excel formula bar to calculate the endpoints of a 98% confidence interval for the mean of all younger workers.

[5]

=5.8 - CONFIDENCE(0.02,1.2,400)

=5.8 + CONFIDENCE(0.02,1.2,400)

c. What assumptions about the underlying data are important for the validity of the confidence interval in (c)?

[5]

Data are a random sample, i.e. independent, from same population.

Note: don't need underlying population normal, since n is large.

d. The results of the calculation in (c) is [5.66,5.93]. The results of the parallel calculation for the older workers is [6.21, 6.59]. Does this suggest that the older workers have significantly higher job satisfaction? Why or why not?

[5]

Yes, "likely ranges" don't overlap.

e. Assume that the two confidence intervals in (d) are independent, and write an Excel formula that could be used to calculate the probability that at least one of them will not contain its corresponding population mean.

[5]

Let X = # CI's that don't contain pop. Mean. X ~ Bi(2,0.02)

P{at least one contains} = P{ X >= 1} = 1 – P{X <= 0} =
=1-BINOMDIST(0,2,0.02,TRUE)

5. A random sample of 2000 smoke detectors is checked to find out how often this type of smoke detector is defective. Let X represent the number that are defective. It is believed that 20% of all of this type of smoke detectors are defective.

a. State the name of the probability distribution of X, give the values of the parameters and list the assumptions needed for that distribution.

[5]

Binomial(2000,0.2)

Defectiveness is independent, and always occurs with prob. 0.2

b. Write a formula that could be used in an Excel formula bar to find c so that

$$0.98 = P\left\{\left|\frac{X}{2000} - 0.20\right| < c\right\}.$$

[5]

0.99 = P{Y < c}, where Y ~ N(0,sqrt(0.2*0.8/2000))

so take c =NORMINV(0.95,0,SQRT(0.2*0.8/2000))

c. Write a formula that could be used in an Excel formula bar to calculate, using the normal approximation, the probability that 350 or fewer of the tested smoke detectors are defective.

[5]

X ~ N(2000(0.2), sqrt(2000(0.2)0.8)), P{X <= 350} = P{X <= 350.5}

=NORMDIST(350.5,2000*0.2,SQRT(2000*0.2*0.8),TRUE)

d. The numerical answer to (c) is 0.00282. If only 350 in the sample are defective, is it likely that the defective rate is indeed 20%?

[5]

No, very unlikely. Conclude rate of 20% is wrong.

6. For the 4 data points stored in a Excel Spreadsheet as:

	A	B
7		
8	10.77028718	
9	128.1505783	
10	27.76451918	
11	9.282509842	
12		

a. Write a formula that could be used in an Excel formula bar to calculate the sample median.

[5]

=MEDIAN(A8:A11)

b. Does 19.3 seem to be a reasonable numerical answer to (a)? Why or why not?

[5]

Yes, half way between two central values of 10.7 and 27.7.

c. Suppose 19.3 really is the sample median, and that the sample mean is 44.0. Comment on the expected shape of the probability histogram.

[5]

Mean >> median, so expect strongly right skewed.

d. Write a formula that could be used in an Excel formula bar to calculate the sample standard deviation.

[5]

=STDEV(A8:A11)

e. Does 0.567 seem to be a reasonable numerical answer to (d)? Why or why not?

[5]

No, way too small. Numbers are from 10 to 80 units away from the sample mean of 40.