STOR 155, Section 1, Final Examination
Thursday, April 30, 2009

Name: _________________Solution_______________________________

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

Signature: ____

Instructions: Do <u>not</u> 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. A company makes 20% of its cars at factory A, and the rest at factory B. Factory A produces 1% lemons, and Factory B produces 2% lemons. A car is chosen at random. What is the probability that:

   a. It came from Factory B?

[5]

+P{B} = 1 – P{not B} = 1 – P{A} = 1 – 0.2 = 0.8

   b. It is a lemon, if it came from Factory B?

[5]

P{L | B} = 0.02

   c. It is a lemon, from Factory B?

[5]

P{L & B} = P{L | B} P{B} = 0.02 * 0.8 = 0.016

   d. It is a lemon?

[5]

P{L} = P{(L & A) or (L & B)} = P{L & A} + P{L & B} – P{(L&A) & (L&B)}
= P{L | A} P{A} + P{L & B} – 0
= (0.01 * 0.2) + (0.02 * 0.8) = 0.002 + 0.016 = 0.018

   e. It came from Factory B, if it is a lemon?

[5]

P{B | L} = P{L & B} / P{L}
= (0.02 * 0.8) / ((0.01 * 0.2) + (0.02 * 0.8)) = 0.016 / 0.018

2. A survey of 2000 student loan borrowers found that 200 had loans totaling more than $40,000.

   a. Give a 99% best guess Confidence Interval for the proportion of all loans totaling more than $40,000.

[5]

X ~ Binom(2000,p), where p = proportion > 40k
phat = X / n = 200 / 2000 = 0.1
margin of error is NORMINV(0.995,0,SQRT(0.1 * 0.9 / 2000))
Left CI = 0.1 - NORMINV(0.995,0,SQRT(0.1 * 0.9 / 2000))
Right CI = 0.1 + NORMINV(0.995,0,SQRT(0.1 * 0.9 / 2000))
(or ± CONFIDENCE(0.01, SQRT(0.1 * 0.9),2000))

   b. Give an Excel expression for the exact p-value for concluding that the proportion of all loans more than $40,000, is at least 5%.

[5]

H_0: $p < 0.05$ H_1: $p \geq 0.05$
p-val = $P\{X \geq 200 \mid p = 0.05\} = 1 - P\{X \leq 199\}$
= 1 – BINOMDIST(199,2000,0.05,true)

   c. Use the Normal approximation to give an alternate answer to (b).

[5]

p-val = $P\{X \geq 200 \mid p = 0.05\} = 1 - P\{X \leq 200\}$
= 1 – NORMDIST(200,2000*0.05,SQRT(2000*0.05*(1 – 0.05)),true)
= 1 – NORMDIST(200,100,SQRT(50),true)

   d. Why is the approximation used in (c) appropriate?

[5]

n * p = 2000 * 0.5 = 1000 > 10
n * (1 – p) = 2000 * (1 – 0.5) = 1000 > 10

   e. What is the 98% conservative margin of error in estimating the proportion of all loans over $40,000?

[5]

NORMINV(0.99,0,SQRT(0.5 * 0.5 / 2000))
= CONFIDENCE(0.02, SQRT(0.5 * 0.5),2000)

3. Scores on tests for a class were:

	A	B	C	D	E
1	1st Exam	153	144	162	127
2	2nd Exam	145	140	143	130
3	Difference	8	4	19	-3

a. Assuming each column represents one student, give a formula for the p-value to show that scores on the 1st exam are significantly higher than those on the 2nd exam.

[5]

$H_0: \mu_1 \leq \mu_2 \quad H_1: \mu_1 > \mu_2$

p-val = P{X_1bar – X_2bar ≥ calculated value | $\mu_1 = \mu_2$}

= TTEST(B1:E1,B2:E2,1,1)

b. Again assuming each column represents one student, give an 80% Confidence Interval for the difference between the mean scores.

[5]

Dbar ± TINV(0.2,n – 1) * sd(Dbar)

AVERAGE(B3:E3) ± TINV(0.2,3) * STDEV(B3:E3) / SQRT(4)

c. Assuming the exam scores come from two different classes, give a formula for the p-value to assess whether exam scores are significantly different between the two exams.

[5]

H0: μ1 = μ2 H1: μ1 ≠ μ2

p-val = P{|X1bar – X2bar |≥ calculated value | μ1 = μ2}

= TTEST(B1:E1,B2:E2,2,3)

d. Write the equation of the least squares regression line, of the 2nd score as a function of the 1st score, in terms of Excel commands

[5]

Y = SLOPE(B2:E2,B1:E1) + INTERCEPT(B2:E2,B1:E1)

e. Write an Excel command to calculate the correlation between exam scores. Will the answer be positive, 0 or negative?

[5]

CORREL(B2:E2,B1:E1), positive

4. For a random variable with distribution:

y	-1	0	1	3
f(y)	0.5	0.2	0.2	0.1

Find:

a. $P\{-1 < Y \leq 2\}$

[5]

= f(0) + f(1) = 0.2 + 0.2 = 0.4

b. $P\{Y = 1 \mid Y > 0\}$

[5]

= P{(Y = 1) & (Y > 0)} / P{Y > 0} = P{Y = 1} / P{Y > 0}
= f(1) / (f(1) + f(3)) = 0.2 / (0.2 + 0.1) = 2 / 3

c. $P\{Y = 1 \mid Y < 0\}$

[5]

0, can't happen at same time

d. The expected value of Y

[5]

EX = sum(X * prob of X) = (-1) * f(-1) + 0 * f(0) + 1 * f(1) + 3 * f(3)
= (-1) * 0.5 + 0 * 0.2 + 1 * 0.2 + 3 * 0.1 = -0.5 + 0.2 + 0.3

e. The standard deviation of Y

[5]

SD(x) = SQRT(sum((X – EX)^2 * prob of X))
SQRT((-1)^2 * 0.5 + 0^2 * 0.2 + 1^2 * 0.2 + 3^2 * 0.1)

5. A TV ad claims that at most 30% of people prefer Brand X. Suppose that 6 out of 10 randomly selected people prefer Brand X.

a. Give an exact p-value to decide whether or not we should dispute the claim.

[5]

H_0: $p \leq 0.3$ H_1: $p > 0.3$, let X = # prefer ~ Binom(10,p)
p-val = $P\{X \geq 6 \mid p = 0.3\} = 1 - P\{X \leq 5\}$
= 1 – BINOMDIST(5,10,0.3,true)

b. If the p-value in part a turns out to be 0.03, give a "yes-no" conclusion.

[5]

Have strong evidence claim is wrong

c. If the p-value in part a turns out to be 0.03, give a gray level conclusion.

[5]

Moderately strong evidence. Significant, but not overwhelming.

d. Give a 98% conservative confidence interval for the proportion of people that prefer Brand X.

[5]

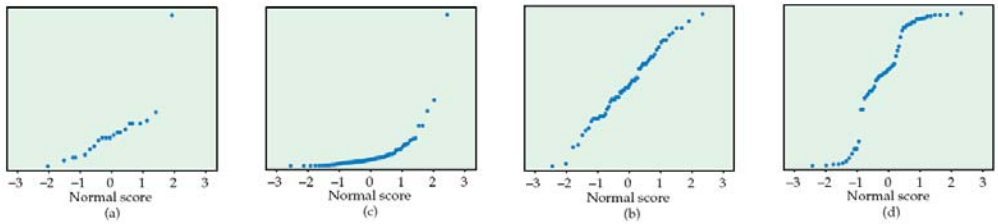
0.6 ± CONFIDENCE(0.02,SQRT(0.5 * (1-0.5)),10)

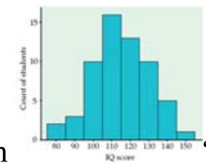
e. How large a sample (in the best guess sense) is needed so that with probability 90%, the estimated proportion of people that prefer Brand X is within 0.01 of the actual number?

[5]

(NORMINV(0.95,0,1) / m)^2 * p * (1 – p)
= (NORMINV(0.95,0,1) / 0.01)^2 * 0.6 * (1 – 0.6)
OR: = (NORMINV(0.95,0,1) / 0.01)^2 * 0.3 * (1 – 0.3)

6. A set of 4 Normal Quantile plots (in scrambled order, <u>watch the labels</u>) are:

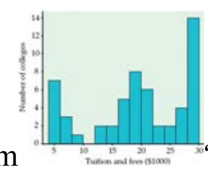


i. Which most likely is from a data set of IQ scores with histogram  ?
[3]

b

ii. Which most likely is from a data set of 21 car mileages of 2-seater cars?
[3]

a

iii. Which most likely is from a data set of tuition charges with histogram  ?
[3]

d

iv. Which most likely is from a data set of 80 phone call lengths with histogram  ?
[4]

c

v. Which most likely is from a Normally distributed data set?
[4]

b

vi. Which most likely is from a Normal distribution, with a single outlier?
[4]

a

vii. Which most likely is from a distribution with multiple clusters?
[4]

d

7. Length of horse pregnancies vary according to a roughly Normal distribution, with mean 340, and standard deviation 5.

   a. Use the 68-95-99.7 rule to indicate the range which contains 95 % of the data.

[5]

= mean ± 2 sd = 340 ± 2 * 5 = 340 ± 10 = 330 – 360

   b. Use the 68-95-99.7 rule to indicate which % of pregnancies last at least 335 days.

[5]

≥ mean – 1 sd
But 67% are within mean ± 1 sd
So 32% are outside mean ± 1 sd
So 16% are less than mean – 1 sd
So 84% are greater than mean – 1 sd

   c. Give an Excel command to answer part (a).

[5]

For X ~ Norm(340,5)
0.95 = P{-C < X - 340 < C}
So 0.975 = P{X - 340 < C}
So C = NORMINV(0.975,0,5), and range is: 340 – C to 340 + C
OR: Lower End: NORMINV(0.025,340,5)
Upper End: NORMINV(0.975,340,5)

   d. Give an Excel command to answer part (b).

[5]

P{X ≥ 335} = 1 – P{X < 335}
= NORMDIST(335,340,5,true)

   e. How large a sample should be used to be 98% sure of estimating the true mean within 0.1?

[5]

n = (NORMINV(0.99,0,1) * σ / m)^2)^2
= (NORMINV(0.99,0,1) * 5 / 0.1)^2

8. Gas mileages for a vehicle, after a random sample of fill-ups are:

	A	B	C	D	E	F
1	41.5	50.7	36.8	44.2	45	37.4

a. Find the sample mean and standard deviation.

[5]

AVERAGE(A1:F1)
STDEV(A1:F1)

b. Find the 60% margin of error in estimation of the population mean.

[5]

= TINV(0.4,5) * STDEV(A1:F1) / SQRT(6)

c. Give a 60% Confidence Interval for the population mean.

[5]

= AVERAGE(A1:F1) ± TINV(0.4,5) * STDEV(A1:F1) / SQRT(6)

d. Find the p-value to test whether the population mean is less than 40.

[5]

H: $\mu \geq 40$ H: $\mu < 40$
p-val = P{Xbar < AVERAGE | $\mu = 40$}
= TDIST(ABS(AVERAGE(A1:F1) – 40) / (STDEV(A1:F1) / SQRT(6)),5,1)

e. Briefly state (5 words or less) the needed assumptions in parts (c) and (d).

[5]

Individuals normally distributed