Business Analytics

Prof. Phil Jones

Fall, 2016

**Exam #2**

**Instructions:**

1. Please write your names on the cover page.
2. The exam is open book, open note, and open computer.
3. You are expected to do your work with no help from outside

your group.

Note: Problems are worth points as indicated.

Providing short explanations may allow awarding partial credit in some cases.

The exam is due **October 19th**. Please return one exam (hardcopy) at the beginning of class with everyone’s name in your group written on it.

*Good Luck!*

1. Jerry Jacob
2. Danen Sorenson
3. Purna Chandra Kuntla
4. Andrew Paterson
5. (15 points) The hospital in which you are an administrator is concerned about “broken needle” incidents in which a needle used to inject a patient is broken off during the course of the injection. You want to estimate the proportion of injections that result in broken needle incidents.
   1. (3 points) Assuming you knew nothing whatsoever about the proportion of injections that result in broken needle incidents. Conservatively, how large a sample size would you need to take to have the “plus or minus” percentage of error in your estimate be 0.5% or less assuming a 95% confidence interval?

**Error tolerance, E = 0.005**

**Proportion, p = 0.5**

**Confidence Interval = .95**

**Sample size, n = (1.960/E)^2 \* p\*(1-p)**

**= (1.960/0.005)^2 \* 0.5 \* 0.5**

**= 38416 (approximately)**

* 1. (2 points) Suppose you are absolutely certain that the proportion of broken needle incidents in your hospital does not exceed 1%. How large a sample size would you need now to have the “plus or minus” percentage of error in your estimate be 0.5% or less assuming a 95% confidence interval?

**Proportion, p = 0.01**

**Sample size, n = (1.960/E)^2 \* p\*(1-p)**

**= (1.960/0.005)^2 \* 0.01 \* 0.99**

**= 1521.27 (approximately)**

* 1. (5 points) Suppose you take a sample of size 500 and find that exactly 4 of the 500 injections sampled resulted in broken needle incidents. You may safely assume that the remaining 496 did not. What are your estimates for the mean and standard deviation of the sampling distribution? Hint: make sure you know the definition of the standard deviation of the sampling distribution.

**Sample size, n = 500**

**Proportion, p = 4/500 = 0.008**

**Mean = p = 0.008**

**Standard Deviation = SQRT(p \* (1-p)/ n)**

**= SQRT(0.008 \* (0.992)/ 500)**

**= 0.0040 (approximately)**

* 1. (5 points) Using the information from part c above, construct a 95% confidence interval for the percentage of injections that result in broken needle incidents. Based on your confidence interval, what would you say about the claim that the true proportion is greater than 1.5%? Discuss briefly

**If we could reject the null hypothesis; true proportion <= 0.015, then we could confidently say that the proportion is greater than 1.5%**

**Test statistic = (sample mean – hypothesized mean)/ sample standard error**

**= (0.008 – 0.015)/ 0.0040**

**= -1.75**

**For 95% confidence interval = NORMSINV(0.95) = 1.64**

**We Fail to reject the hypothesis since test statistic is lesser than 95% confidence interval. So there isn’t any strong evidence that the true proportion is greater than 1.5%**

**95% Confidence Interval**

**= [0.008 – NORMSINV(0.975)\*0.003984, 0.008 + NORMSINV(0.975)\*0.003984]**

**= [0.000192, 0.0158]**

1. (15 points) You need to determine the average time that a particular accounts payable clerk in your organization takes to process an invoice. You take a sample of 100 invoices and time how long the clerk takes on each invoice. Your sample mean is 34 minutes and your sample standard deviation is 20.

**Sample Size, n = 100**

**Sample mean = 34**

**Sample standard deviation = 20**

* 1. (3 points) What is the sample standard error (eg., your estimate of the standard deviation of the sampling distribution)?

**Sample standard eror, E = sample standard deviation/ SQRT(n)**

**= 20/ 10**

**= 2**

* 1. (5 points) Specify a 95% confidence interval about the mean? That is, determine numerical values for lower and upper fenceposts.

**For 95% confidence interval = NORMSINV(0.975) = 1.96**

**Upper fence post = 34 + (1.96 \* 2) = 37.92**

**Lower fence post = 34 - (1.96 \* 2) = 30.08**

* 1. (2 points) Suppose you want the width of a 98 percent confidence interval to be 2 minutes (error tolerance = 1). How large a sample size would you need?

**For 98% confidence interval = NORMSINV(0.99) = 2.33**

**Sample size, n = (2.33 \* standard dev. / Error tolerance) ^ 2**

**= (2.33 \* 20 / 1) ^ 2**

**= 2164.8 (approximately)**

* 1. (5 points) Suppose your company has established a standard that, on average, each AP clerk should be able to process invoices in 30 minutes or less. Does the information obtained with the sample of 100 invoices allow you to conclude at the 95% level that this particular clerk does not meet standard? Explain.

**If we could reject the null hypothesis: Invoice processing time <=30, then we could say that the processing time is 30 minutes or more.**

**Test statistic = 34 – 30/ 2 = 2**

**For 95% confidence interval = NORMSINV(0.95) = 1.64**

**Since test statistic is greater than 1.64, we Reject the null hypothesis; which means the clerk does not meet the standard.**

**As we increase the value of n, the Test statistic will keep increasing. Whether the sample size is 100 or any higher value, we don’t get any significant evidence to Fail to reject the above hypothesis.**

3. (10 points) Suppose you want to compute the upper and lower bounds of a 92 percent confidence interval. To do so, you need to compute the z-value (how many standard deviations away from the mean the upper and lower confidence limits lie). The correct Excel command to do so is (choose the single best answer):

1. Normsinv(.84)
2. Normsinv(.88)
3. Normsinv(.92)
4. **Normsinv(.96)**
5. Normsinv(.98)
6. None of the percentages are correct, but you should use the normsinv function.
7. One of the percentages is correct, but you should use the normsdist function instead of the normsinv function.
8. (15 points) Sam, Sadie, and Sonorra compete in a 3-person relay race against Jackie, Jo, and Jasmine. In the relay race, each person runs two “legs” of the relay for a total of 6 “legs”, and the winning team is the one that finishes first. Based on past history, the times of each contestant on her leg of the relay is a normally distributed random variable. The averages and standard deviations for the various contestants are shown in the following table:

|  |  |  |
| --- | --- | --- |
| Contestant | Average Time | Standard Deviation |
| Sam  Sadie  Sonorra  Jackie  Jo  Jasmine | 9.5  9  8  10.5  8.5  9 | 1.75  1  1.5  2.5  2  2.25 |

What is the probability that Sam, Sadie, and Sonorra beat Jackie, Jo, and Jasmine?

**Team1=Sam + Sadie + Sonorra**

**Team2= Jackie + Jo + Jasmine**

***Mean* = *a μx + b μy + c μz*and**

***Var = a2 σx2 + b2 σy2+ c2 σz2***

**a=b=c=2**

**Team1 mean=2\*9.5+2\*9+2\*8=53**

**Team1 Std. dev = =SQRT((2\*1.75)^2+(2\*1)^2+(2\*1.5)^2) = 5.024938**

**Team2 mean=2\*10.5+2\*8.5+2\*9=56**

**Team2 Std. dev = =SQRT((2\*2.5)^2+(2\*2)^2+(2\*2.25)^2) = 7.826238**

**W=Team1-Team2; if W<0 Team1 wins**

**W mean = 53-56=-3**

**W Std. Dev = SQRT((Team1 Std dev)^2 + (Team2 Std dev)^2)**

**=SQRT(5.024938^2 + 7.826238^2)**

**W Std. Dev   = 9.300538**

**We seek P (W<=0) = NORMDIST(0, -3, 9.300538, 1)**

**= 0.626487**

**There is 62.65% chance for Team1 win.**

1. (15 points) Suppose it is known with certainty that the height of men (in inches) in a particular population is normally distributed with a mean of 71 and a standard deviation of 7.

**Mean of the population = 71**

**Standard Deviation = 7**

1. (3 points) In a random sample of 1,000,000 men, how many would you expect to have heights less than 58 inches?

**P(X<=58) = NORMDIST(58,71,7,1) = 31645.42**

**People with height less than 58 inches = 31645.42**

1. (3 points) In a random sample of 1,000,000 men, how many would you expect to have heights exceeding 84 inches?

**P(X>=84) = 1 - NORMDIST(84,71,7,1) = 31645.42**

**People with height greater than 84 inches = 31645.42**

1. (3 points) In a random sample of 1,000,000 men, how many would you expect to have heights between 61 and 81 inches?

**P(X<=81) - P(X<=61) = NORMDIST(81,71,7,1) - NORMDIST(61,71,7,1)**

**= 0.846873**

**People with height between 61 and 81 inches = 846,873**

1. (3 points) Imagine you took a random sample of 10 men from this population and computed their average height. What is the probability that the average height of the 10 men is between 63 and 79 inches?

**Standard deviation of the sample distribution = 7/SQRT(10) = 2.21**

**P(X<=79) - P(X<=63) = NORMDIST(79,71,2.21,1) - NORMDIST(63,71,2.21,1) = 99.97%**

1. (3 points) Same as part d except that now your sample size is 100.

**Standard deviation of the sample distribution = 7/SQRT(100) = 0.7**

**P(X<=79) - P(X<=63) = NORMDIST(79,71,0.7,1) - NORMDIST(63,71,0.7,1) = 100%**

1. (5 points) Suppose you have a standard normal distribution. You wish to compute the probability of being more than 4.5 standard deviations to the right of the mean. The correct excel function to use is: (choose the best answer)
2. Normdist
3. **Normsdist**
4. Normsinv
5. Norminv
6. Binomdist
7. (5points) Suppose A and B are percentages and A < B. Then (choose the best answer):
8. **normsinv(A) < normsinv(B)**
9. normsinv(A) >normsinv(B)
10. You need to know what A and B are to tell
11. (20 points) Your hospital is quite concerned over the amount of time between taking blood samples and when the lab results are available. The accompanying excel spreadsheet presents a set of data that have been collected. Each row in the table represents a sample of 4 observations from a single day. There are 50 rows (days) in the set of samples. Thus, there are a total of 200 observations in all.
    1. (5 points) What are the mean and standard deviation of the 200 observations? Based on these two calculations, estimate the standard deviation of the sampling distribution if your sample is of size 4. Is it necessarily the case that your estimate will equal the actual standard deviation of the sampling distribution?

**Mean = 122.195**

**Standard Deviation = 30.53**

**Std Dev of sampling distribution = 30.53/SQRT(4) = 15.265**

**Is it not necessarily the case that your estimate will equal the actual standard deviation of the sampling distribution**

* 1. (5 points) Compute a 95% confidence interval for the average time it takes to get results back from the lab. Hint: here use as your sample all 200 observations and a sample size of 200.

**Mean = 122.195**

**Standard Deviation of sampling distribution = 30.53/ SQRT(200) = 2.16**

**95% Confidence = NORMSINV(0.975) = 1.96**

**Lower fence post = 122 – (2.16 \* 1.96) = 117.77**

**Upper fence post = 122 + (2.16 \* 1.96) = 126.23**

* 1. (5 points) Compute a 95% confidence interval for the sample mean if your sample size is 4.

**Mean = 122.195**

**Standard Deviation of sampling distribution = 30.53/ SQRT(4) = 15.265**

**95% Confidence = NORMSINV(0.975) = 1.96**

**Lower fence post = 122 – (15.265 \* 1.96) = 92.08**

**Upper fence post = 122 + (15.265 \* 1.96) = 151.92**

* 1. (5 points) Suppose you wish to continue taking a sample of size four each day and computing the sample mean. A frequently used tool in process management is the control chart which is set up by computing a 3-sigma confidence interval (by using 3 instead of normsinv(.975) in your calculation for the confidence interval) for the sample average of a sample of size four. The chart is set up by drawing horizontal lines representing the LCL and UCL of your 3-sigma confidence interval and then plotting the sample mean of each sample on the chart. If a sample mean plots within the two horizontal lines, the process is said to remain in control, and if the sample mean plots outside the horizontal lines, the process said to be out of control. Compute the LCL and UCL for a 3-sigma control chart based upon the data given. Suppose the very next day your sample consists of the following observations: 200, 255, 107, 95. What would you conclude?

**Confidence Interval: 122.195+3(Std. Dev.)**

**Upper: 122.195+3\*15.27 = 167.99**

**Lower: 122.195 – 3\*15.27=76.385**

**The mean of the sample observations, 200, 255, 107, 95 is 164.25, lies between LCL and UCL. So the process is in control.**