**Statistics 500.002: Applied Statistics**

**Summer 2015 Midterm II (World Campus)**

Name: \_\_\_\_\_\_Daljeet Maken\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Instructions:

1. From the time you first access the test you have **Four hours** to solve and upload the exam.
2. The value of each part of each question is given in parentheses.
3. Work must be shown to earn full credit.
4. Provide justifications for your answers to earn partial credit.
5. Write your answers on the test booklet.
6. You can access all your lesson materials, text, handouts etc. But you may not take help from another person.
7. Calculators may be used and you may use Minitab whenever you want.
8. **The work handed in must be your own.**

***Good Luck!***

1. (**16 pts**) State whether the following statements are true or false:
   1. The main advantage of using a confidence interval rather than a point estimate is that the interval will always cover the population parameter.

FALSE

* 1. If P-value > α, we fail to reject the null hypothesis at level of significance set at α.

TRUE

* 1. Sum of the probabilities of Type I error and Type II error is always 1.

FALSE

* 1. For fixed sample size, confidence intervals with higher confidence level have lower precision.

TRUE

* 1. If null hypothesis is not rejected at level α, then a confidence interval of confidence coefficient 100(1 – α)% will include the null value of the parameter.

FALSE – Correct: True

* 1. If sample size is 30 or more, we can use the t-test for population mean problem even if the sample distribution is slightly different from normal.

TRUE

* 1. When performing a hypothesis test, using the P-value approach will reach the same conclusion as when using the rejection region approach.

TRUE

1. (2 **pts**) A survey of college students reported that 44% of US college students binge drinking. PSU conducted a survey on 2500 students and revealed 1200 admitted they binge drinking. For this case, the notations for population parameter and sample point estimate are:
2. π=0.44 and =0.48
3. µ=0.44 and =0.48
4. π=0.48 and =0.44
5. µ=0.48 and =0.44

**A - π=0.44 and π ̂=0.48**

1. (2 pts) For each high school in the United States, we measure the percentage of seniors that have driver’s license. Suppose it is known that the average percentage is 74%. A random sample of 40 U.S. high schools showed that the mean is 80%. For this problem the notation should be:
2. = 0.74 and = 0.80
3. = 0.80 and = 0.74
4. *π* = 0.80 and = 0.74
5. *π* = 0.74 and = 0.80
6. **= 0.74 and = 0.80**
7. (4 **pts**) A random sample is taken to estimate the population mean weight of fishes in Pacific ocean. Which of the following combinations of sample sizes and confidence level will give you the largest margin of error?
   1. A confidence level of 85% and n = 200.
   2. A confidence level of 90% and n = 100.
   3. A confidence level of 85% and n = 100.
   4. A confidence level of 90% and n = 200.

B - A confidence level of 90% and n = 100.

1. (**4 pts**) An inspector has to decide whether the brake of a car is safe or not. How are you going to set up the hypothesis for the inspector?

Ho : Brake is not safe

Ha : Brake is safe

Explanation: We should set the hypotheses up in a way so that Type I error is more serious. Here we put “Brake is not safe " in  since we consider false rejection of *H*o as a more serious error than failing to reject *H*o.

1. (5 pts) A particular camera is sold at an average price of $190 with a standard deviation of $12. In a sample of 50 randomly selected stores, what is the probability that the sample mean falls within $4 of the population mean?

6. I was asking P(190-4<ybar<190+4). -4pts (Steve Bai)

We have to find the probability that yBar falls within (mu +- 2)

i.e. P(yBar > $188) and P(yBar < $192)

P(z >= (188-190)/(12/sqrt(50))) and P(z <= (192-190)/(12/sqrt(50)))

P(z >= -1.1785) and P(z <= 1.1785)

P(z <= 1.1785) - P(z <= -1.1785)

From Minitab we have P(z <= 1.1785) = 0.880701 and

P(z <= -1.1785) = 0.119299

Therefore the probability asked is 0.880701-0.119299 = 0.761402 or 76.14%

1. (**4 pts**) A sample of n = 300 college students is asked if they believe in extraterrestrial life and 130 of these students say that they do. The data are used to test H0: π = 0.5 versus Ha: π < 0.5, where π is the population proportion who would say they believe in extraterrestrial life. Minitab results are:

Sample X N Sample p 95% CI Z-Value

1 130 300 0.43 (0.38, 0.49) -2.31

What is the correct description of the area that equals the P-value for this problem?

* 1. The area to the left of 0.43 under a standard normal curve
  2. The area between 0.38 and 0.49 under standard Normal curve
  3. The area to the right of −2.31 under the standard normal curve
  4. The area to the left of −2.31 under the standard normal curve

**D**

1. (**10 pts**) The management of Southern Textiles has recently come under fire regarding the supposedly detrimental effects on health caused by its manufacturing process. A social scientist has advanced a theory that the employees who die from natural causes exhibit remarkable consistency in their life-span: The upper and lower limits of their life-spans differ by no more than 550 weeks (about 10.5 years). For a confidence level of 98%, how large a sample should be examined to find the average life-span of these employees within ±30 weeks. (Perform this problem by HAND)

Using the crude method



α = 1-0.98=0.02

Zα/2=Z0.01= 2.326

Margin of error = E = 30

σ≈Range / 4=550 / 4 = 137.5

n = 2.326^2 \* 137.5^2 / 30^2 = 113.653

Sample size should be 114

1. (**6 + 8 = 14 pts**) A community college is considering raising tuition to improve school facilities and they want to know what percentage of students favor the raise. A member of the board took a random sample of 20 students and found that 17 prefer the raise.
   1. Can you use the 1-proportion z-interval to find the confidence interval for the proportion? What will you need to check? If yes, compute the z-interval. If no, suggest an alternative and obtain that confidence interval for the population proportion.
   2. If the setup cost of sampling is very expensive, how many students does the board need to sample so that the proportion can be estimated with 92% confidence with a margin of error of 0.04?

9a.. you should use minitab to compute exact CI. -3pts (Steve Bai)

a) Assumptions needed to check before one can use the one-sample z-interval for π

nπ^≥5 and n(1−π^)≥5

In our case π^ = 17/20 = 0.85

n π^ = 20 \* 0.85 = 17 > 5

n(1−π^) = 20 \* (1-0.85) = 3 < 5

Since our test fails we have to adjust our computation:

π^adj = (n+3/8)/(n+3/4) = (20+3/8)/(20+3/4) = 0.982

The 95% confidence interval will be ((α/2)1/n, 1) = ((.05/2)1/20, 1) = (0.832, 1)

b) Since the setup cost is very expensive, we will use the conservative method:



E=0.04 α = 0.08, Z(α/2) = Z(0.04) = 1.7507 (from minitab)

piHat = 0.5

n = (1.7507)^2 \* 0.5 \* 0.5 / 0.04^2 = 478.8985141

Therefore sample size = 479

1. (**12 + 2 = 14 pts**) Sports car owners complain that the state vehicle inspection station judges their car differently from family-styled cars. Previous records indicate that 30% of all passenger cars fail the inspection on the first time through. In a random sample of 150 sports cars, 60 failed the inspection on the first time through.
   1. Is there sufficient evidence to indicate that the percentage of first failures for sports cars is higher than passenger cars? Report the P-value of the test. (setup hypotheses, check any conditions in order to perform the test, and compute the p-value)
   2. If Type I error for the test is set at 1% what would be your conclusion?

a)  (or )



This is a one-sided test and(assumed since not provided). So we reject Ho if p-value ≤ 0.05



Test Statistic: Z\* = ( - πo) /, where  = √[ πo (1 - πo) / n] =

= (0.4-0.3) / sqrt(0.3\*0.7/150) = 2.6726

p-value = P(Z > observed z)=P(Z>2.6726)= 1-0.9962 = 0.0038 < α = 0.05, so we can reject Ho.

So there is sufficient evidence to indicate that the percentage of first failures for sports cars is higher than passenger cars at 

b) At an  we will reject Ho if p-value ≤ 0.01. Since we got p-value 0.0038 < α = 0.01, so we can reject Ho.

So there is sufficient evidence to indicate that the percentage of first failures for sports cars is higher than passenger cars at 

1. (**6 pts**) A soap company promoted a new and expensive antibacterial soap. A consumer testing agency questions whether the new soap is a substantial improvement over ordinary soap. A procedure to exam how much bacteria killed by soap is to place a soap solution onto a petri dish and add E coli bacteria. Then count the number bacteria after 24 hours. From previous studies, the mean bacteria count is 33 for ordinary soap. The agency ran the tests on the new soap in n=35 petri dishes and they got a mean bacterial count of 31.2 with a standard deviation of 8.4. This result failed to reject the H0: µ≥33, and concluded that there is no sufficient evidence to show the new soap is more effective.

Suppose that agency thinks that company will take legal action (against agency’s conclusion) if the new soap has a population mean bacterial count that is much less than 33, say 28. Thus, the agency wants to know the Type II error in its test if the population mean is 28 or smaller, i.e determine β(28). Perform this with MINITAB!

H0: µ≥33, Ha: µ<33

n = 35, α = 0.05, SD = 8.4

Difference to detect = 33-28 =5

**Power and Sample Size**

1-Sample t Test

Testing mean = null (versus < null)

Calculating power for mean = null + difference

α = 0.05 Assumed standard deviation = 8.4

Sample

Difference Size Power

5 35 0.0000002

The power is very low and we cannot accept the null hypothesis since the possible Type II error is β = 1-0.0000002 = 0.9999998

The possible Type II error is too high.

11. have you wonder why the power is so small? diff should be 28-33=-5, not +5. -3pts (Steve Bai)

1. (**12 + 5 = 17 pts**) An auto mechanic would like to estimate the average time it takes to replace a carburetor. In a random sample of 18 cars, it took an average of 50 minutes. The standard deviation is 7 minutes.
   1. Find a 90 % confidence interval for the population mean.
   2. What assumption do you need to check so that you can perform (a)? Indicate how are you going to check it if you have the complete data?

a) Data given as follows:

n = 18, df = 17, yBar = 50, s = 7  
tα/2 = t0.05 = 1.7396 (from minitab)

So the interval is: 50 ± 1.7396 \* 7/sqrt(18) which is 50 ± 2.8702

Or, it can be represented as (47.1298, 52.8702)

b) Conditions to check When One Can Use t-Procedure for Population Mean

Conditions to use the t-interval are:

* When sample size is less than 15, use t-interval procedure only when population is very close to normal.
* When sample size is between 15 and 30, it can be used if the variable is not far from normal.
* When sample size is large, we can always use t-interval if there are no extreme outliers that cannot be removed.

In order to check the normality of the data, we will use the normal probability plot. If the points fall within the confidence limits, then the data may come from a normal distribution. If we cannot use the t-interval then we may look for a more robust procedure such as the one-sample Wilcoxon procedure.

***Congratulations! You have completed Midterm II !!!***