NAME Key

NOVEMBER 17, 2014

MATH 245 - PROBABILITY AND STATISTICS

LONG EXAM III

Instructions: *Please include all relevant work to get full credit. Define your random variables properly and write your solutions using proper notations. Round off your final answers to four decimal places. Finally, encircle your final answers.*

1. Write **True** if the statement is **always** true, otherwise, write **False**. [12]

T 1. Increasing the level of confidence, while keeping everything else the same, will result in wider confidence intervals.

F 2. Increasing the sample size, while keeping everything else the same, will result in wider confidence intervals.

T 3. Decreasing the level of significance (α) in hypothesis testing, while keeping everything else the same, will also decrease the power of the test.

T 4. Sending an innocent person to jail is an example of a type I error.

T 5. When the null hypothesis is rejected, then it is impossible to have committed a type II error.

F 6. Bigger $p-$values implies more evidence against the null hypothesis.

2. Decide which inferential method is most appropriate to apply for each of the following practical research questions. Write the letter of the most appropriate statistical procedure next to each experiment or study question. Each procedure may be used more than once or not at all. [10]

A) One-sample $t-$test for a mean.
B) One-sample $z-$test for a proportion.
C) One-sample χ^2-test for a variance.
D) Two-sample independent $t-$test for means.
E) Paired differences $t-$test.
F) Two-sample $z-$test for proportions.
G) C.I. for a proportion.
H) C.I. for a mean.
I) C.I. for difference of means using independent samples.
J) C.I. for the mean of paired differences.

C 1. As part of the quality control procedure, a manufacturing company requires that the standard deviation of the breaking strength of a randomly selected items from the production line is no higher than a specific value.

E 2. To test the effectiveness of a new drug, the glucose level of all 25 diabetic patients were recorded before and 30 minutes after the new drug was taken.

D 3. Test the hypothesis that high school principals in big cities have the same mean annual salary as those in smaller towns.

H 4. What is the mean annual salary of a high school principal in Wisconsin?

G 5. What percentage of high school principals in the U.S. are males?

3. A certain airline reports that the length of delay (X) of their flights is uniformly distributed from 0 to 40 minutes. Recall that if $X \sim Unif[c, d]$, then $E(X) = \frac{c+d}{2}$ and $SD(X) = \frac{d-c}{\sqrt{12}}$.

a. What is the sampling distribution of $\bar{X}$, the average delay of 50 randomly selected flights from this airline? Specify the mean and standard deviation of this sampling distribution. [5]

$E(X) = \frac{0+40}{2} = 20$ and $SD(X) = \frac{40-0}{\sqrt{12}} \approx 11.55$

$\bar{X} \approx \text{Normal}\left(\mu_{\bar{X}} = 20, \ \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{11.55}{\sqrt{50}} \approx 1.63\right)$

b. What result/theorem can you use to justify your answer in part (a). How do you know that this result/theorem applies to this situation? [2]

Central Limit Theorem because $n > 30$.

c. What is the probability that the average delay time of the 50 randomly selected flights is more than 24 minutes? [5]

$$P(\bar{x} > 24) = P\left(\frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} > \frac{24-20}{1.63}\right)$$

$$= P(Z > 2.45)$$

$$= 1 - .9929$$

$$= .0071$$

4. A biologist wants to estimate the mean weight of adult squirrels in Wisconsin. If the population standard deviation of weights of adult squirrels in Wisconsin is known to be $\sigma = 0.2$ pounds, how many adult squirrels should she sample to get a 95% confidence interval for the mean weight of adult squirrels in Wisconsin with a margin of error of not more than 0.05 pounds? [4]

$$n = \left(\frac{z \cdot \sigma}{M}\right)^2 = \left(\frac{1.96 * 0.2}{.05}\right)^2 \approx 61.46$$

Round up $n = 62$

5. In the previous problem, suppose the biologist do not know the value of σ and decided to sample only 25 adult squirrels from Wisconsin. The sample mean weight of these 25 squirrels is 1.36 pounds and the sample standard deviation is 0.3 pounds. Assume that the weight of adult squirrels in Wisconsin is approximately normal.

a. Construct and interpret a 95% confidence interval for the mean weight of squirrels in Wisconsin. What is the margin of error? *Use 4 decimal places.* [6]

$$\bar{x} \pm 2.064 \frac{s}{\sqrt{n}} = 1.36 \pm (2.064)\left(\frac{.3}{\sqrt{25}}\right) \quad \Rightarrow M.E. = 0.1238$$

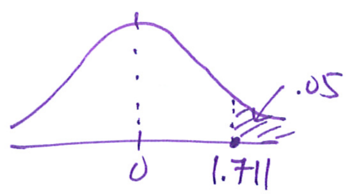
$$= 1.36 \pm 0.1238$$

$$= [1.2362, 1.4838]$$

We are 95% confident that the mean weight of adult squirrels in Wisconsin is between 1.24 and 1.48 pounds.

b. Using $\alpha = 0.05$, conduct a complete test of hypothesis to test $H_0: \mu = 1.3$ versus $H_1: \mu > 1.3$. Include all 6 steps. Do you reject the null hypothesis? Write a practical conclusion in the context of this problem. [8]

1. $H_0: \mu = 1.3$ vs $H_1: \mu > 1.3$
2. $\alpha = 0.05$
3. $T = \frac{\bar{x} - \mu}{s/\sqrt{n}} \sim t_{df = n-1 = 24}$



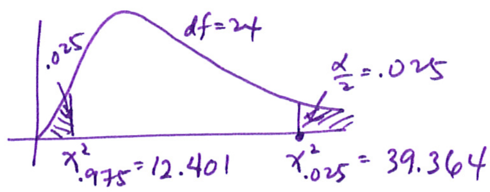
4. Reject Ho if $T_{obs} > 1.711$.
5. $T_{obs} = \frac{1.36 - 1.3}{.3/\sqrt{25}} \approx 1$
6. Sine $T_{obs} = 1 \not> 1.711$, we do not reject Ho. Therefore, we did not find enough evidence to say that the mean weight of squirrels in WI is > 1.3 lbs

c. Using $\alpha = 0.05$, conduct a complete test of hypothesis to test $H_0 : \sigma = .2$ versus $H_1 : \sigma \neq .2$. Include all 6 steps. Do you reject the null hypothesis? Write a practical conclusion in the context of this problem. [8]

1. $H_0 : \sigma = .2$ vs $H_1 : \sigma \neq .2$

2. $\alpha = 0.05$

3. $\chi^2 = \dfrac{(n-1)s^2}{\sigma^2} \sim \chi^2_{df = n-1 = 24}$



4. Reject H_0 if $\chi^2_{obs} < 12.401$ or > 39.364

5. $\chi^2_{obs} = \dfrac{(25-1)(.3^2)}{(.2)^2} \approx 54$

6. Since $\chi^2_{obs} = 54 > 39.36$, we reject H_0. Therefore, we found sufficient evidence to conclude that the s.d. of squirrel weights in Wisconsin is higher than .2.

d. If the biologist decides to sample 1 additional squirrel, construct a 95% prediction interval for its weight. *Use 4 decimal places.* [6]

$$\bar{x} \pm t_{\frac{\alpha}{2}}\, s\sqrt{1+\tfrac{1}{n}} = 1.36 \pm (2.064)(.3)\sqrt{1+\tfrac{1}{25}}$$

$$= 1.36 \pm .6315$$

$$= [.7285,\ 1.9915]$$

6. The Greystone Department Store age study provided the following data on the ages of customers from independent random samples taken at two store locations.

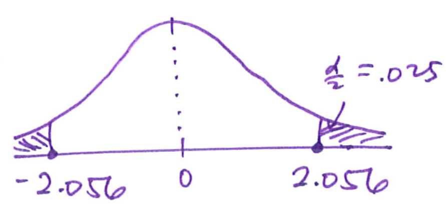
Location	n	$\bar{x}$ (in years)	s
Inner-City	16	40	9
Suburban	12	35	10

a. Let μ_1 be the mean age of customers at the Inner-city location and μ_2 at the suburban location. Formulate the hypotheses that could be used to detect a difference between the population mean ages of customers at the two stores. [2]

$H_0 : \mu_1 = \mu_2$ vs $H_1 : \mu_1 \neq \mu_2$

b. If we can assume that the population standard deviations are equal ($\sigma_1 = \sigma_2$). Specify the appropriate test statistic and its distribution. Using $\alpha = 0.05$, state your rejection rule. [6]

$$T = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{df = n_1 + n_2 - 2 = 26}$$



Reject H_0 if $|T_{obs}| > 2.056$

c. Calculate the value of the test statistic. Do you reject the null hypothesis. Write practical conclusion in the context of the problem? [8]

$$S_p^2 = \frac{(16-1)(9^2)+(12-1)(10^2)}{16+12-2} = \frac{2315}{26} \approx 89.04 \Rightarrow S_p = 9.44$$

$$T_{obs} = \frac{(40-35)-(0)}{9.44\sqrt{\frac{1}{16}+\frac{1}{12}}} = \frac{5}{3.605} \approx 1.387$$

Since $T_{obs} = 1.387 \not> 2.056$, we do not reject H_0.

Therefore, we did not find enough evidence that the mean age of customer differ between the 2 locations.

d. Construct and interpret a 95% confidence interval for $\mu_1 - \mu_2$. [6]

$$(\bar{X}_1 - \bar{X}_2) \pm t_{\frac{\alpha}{2}} \cdot S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} = (40-35) \pm 2.056(9.44)\sqrt{\frac{1}{16}+\frac{1}{12}}$$
$$= 5 \pm 7.41$$
$$= [-2.41, 12.41]$$

We are 95% confident that $\mu_1 - \mu_2$ is between -2.41 and 12.41

e. Explain how the confidence interval of part (d) supports your conclusion in part (c). [2]

Since 0 is in the C.I. for $\mu_1 - \mu_2$, then we can't reject $H_0: \mu_1 - \mu_2 = 0$.

7. Let $X_1, X_2, \ldots, X_n$ be a random sample taken from a population with mean μ and variance σ^2. Prove that $S^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)^2\right]$ is an unbiased estimator of σ^2. [12]

Proof: First, note that $V(Y) = E(Y^2) - \mu_Y^2 \Rightarrow E(Y^2) = \sigma_Y^2 + \mu_Y^2$.

$$\Rightarrow E(S^2) = \frac{1}{n-1}\left[E\left(\sum x_i^2\right) - \frac{1}{n}E\left(\sum x_i\right)^2\right]$$
$$= \frac{1}{n-1}\left[\sum E(x_i^2) - \frac{1}{n}\left[V\left(\sum x_i\right) + \left(E\left(\sum x_i\right)\right)^2\right]\right]$$
$$= \frac{1}{n-1}\left[\sum\left(\underbrace{V(x_i)}_{\sigma^2} + \underbrace{\mu_{x_i}^2}_{\mu^2}\right) - \frac{1}{n}\left[\sum \underbrace{V(x_i)}_{\sigma^2} + \left(\sum \underbrace{E(x_i)}_{\mu}\right)^2\right]\right]$$
$$= \frac{1}{n-1}\left[(n\sigma^2 + n\mu^2) - \frac{1}{n}\left(n\sigma^2 + (n\mu)^2\right)\right]$$
$$= \frac{1}{n-1}\left[n\sigma^2 + \cancel{n\mu^2} - \sigma^2 - \cancel{n\mu^2}\right] = \frac{1}{n-1}\left[n\sigma^2 - \sigma^2\right]$$
$$= \frac{\sigma^2\cancel{(n-1)}}{\cancel{n-1}} = \sigma^2. \blacksquare$$