I, II, IV

Mine Çağlar

May 7, 2010

Math 202: Statistics for Social Sciences

Fall 2010 EXAM 2

Calculator OK, 90 min.

Instructions: There are five parts to this exam I-V. Please inspect the exam and make sure you have all 5 pages of questions. Do all your work on these pages. If you use the back of a page, make sure to indicate that.

Remember: ***You must show your work to get proper credit. In hypothesis testing questions, show all steps of the test and state your conclusion in plain English.***

Academic Honesty Code: Koç University Academic Honesty Code stipulates that "copying from others or providing answers or information, written or oral, to others is cheating." By taking this exam, you are assuming full responsibility for observing the Academic Honesty Code.

NAME: KEY

Part I:	/20
Part II:	/20
Part III:	/20
Part IV:	/20
Part V:	/20
Total:	/100

Formulas:

Confidence interval (CI) : $\overline{X} \mp z\, SE$ or $\overline{X} \mp t\, SE$ or $\hat{p} \mp z\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$

$(\hat{p}_1 - \hat{p}_2) \mp z\, SE_{2\text{-}sample}$ or $(\overline{X}_1 - \overline{X}_2) \mp z\, SE_{2\text{-}sample}$ or $(\overline{X}_1 - \overline{X}_2) \mp t\, SE_{2\text{-}sample}$

Standard Error (SE) : $SE_{2\text{-}sample} = \sqrt{(SE_1)^2 + (SE_2)^2}$, $SE = \sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$ for percentages,

and $SE = \dfrac{SD}{\sqrt{n}}$ for averages.

Standard Deviation (SD) :

sum all (entry-average)^2 values,
then divide by n-1, and then take the square root
Here, entry: each observation in the data set, n: sample size

Part I. **(20 points)** Last year the average cost of making a movie was $55 million. This year, a random sample of 15 recent action movies had an average production cost of $62 million with a standard deviation of $30 million. Assume that the cost of making a movie has a normal distribution.

1. (2 points) What is the variable under study?

Production cost ~~of~~ a movie

2. (12 points) Can it be concluded that it costs more than average (of all movies) to produce an action movie?

$H_0: \mu = 55$

$H_a: \mu > 55$

$SE = \frac{30}{\sqrt{15}} \cong 7.75$

$t = \frac{62-55}{7.75} = 0.90$ $\quad d.f. = n-1 = 14$

$\Rightarrow$ P-value $> 10\%$

Large P-value $\Rightarrow$ Do not reject H_0.

So, it costs the same amount to produce an action movie as others, on the average.

3. (6 points) Construct a 95% confidence interval for the mean production cost of an action movie.

$\bar{x} \mp t\, SE$ $\quad t_{14, 2.5\%} = 2.14$

$62 \mp 2.14\,(7.75)$

$\Rightarrow [45.4, 78.6]$ million dollars.

Part II. (20 points) The following table shows the results of a randomized controlled study on nicotinic acid, which is a drug for coronary (heart) disease. "Number" denotes the sample size, and "Deaths" denotes the rate of deaths in each group.

Nicotinic Acid		Placebo	
Number	Deaths	Number	Deaths
58	11%	93	14%

1. (16 points) Does Nicotinic Acid have a significant effect in reducing the percentage of deaths due to coronary heart disease?

$H_0: p_1 = p_2$

$H_a: p_1 < p_2$

$$SE_1 = \sqrt{\frac{(0.11)(0.89)}{58}} \cong 0.041 = 4.1\%$$

$$SE_2 = \sqrt{\frac{(0.14)(0.86)}{93}} \cong 0.036 = 3.6\%$$

$$SE_{diff.} = \sqrt{(4.1)^2 + (3.6)^2}\,\% \cong 5.5\%$$

$$z = \frac{11\% - 14\%}{5.5\%} = \frac{-3\%}{5.5\%} \cong -0.55$$

$$\text{Table} \Rightarrow \text{P-value} = \frac{100 - 41.77}{2}\% = 29.115\%$$

$\Rightarrow$ Do not reject H_0 as P-value $> 5\%$.

No, Nicotinic acid does not have a significant effect in reducing the percentage of deaths due to coronary heart disease.

2. (4 points) Complete the following sentence about the interpretation of the *P*-value:

The *P*-value (of the test performed in Question 1) corresponds to the probability of observing __3__% difference, when actually the true difference in the two population percentages is __0__.

Part III. (20 points) On the basis of a random sample of size 50, a 95% confidence interval for population mean μ is found to be [3.2, 14.8].

1. (5 points) Does the sample mean lie in the interval [3.2, 14.8]? If yes, find it. If not, explain.

Yes, because the formula is $\bar{X} \mp z\,SE$.
So, $\bar{X}$ is in the middle of this interval.

$$\bar{X} = \frac{3.2+14.8}{2} = \frac{18}{2} = 9$$

2. (5 points) Does the population mean lie in the interval [3.2, 14.8]? If yes, find it. If not, explain.

We do not know for sure; it may or may not.
We know that it lies in this interval with 95% probability.

3. (5 points) If a second random sample of the same size 50 is taken, would the new 95% symmetric confidence interval for the population mean be different? Why or why not?

It would be different because the new sample mean and standard deviation will be different in the formula

$$\bar{X} \mp z \frac{SD}{\sqrt{n}}$$

even if z and n remain the same.

4. (5 points) In repeated sampling, suppose 20 such confidence intervals are constructed from different samples. On the average, how many of these intervals do you expect to cover the population mean?

They are all 95% CI's.
So each one covers μ with 95% chance.
We expect $20 \times 95\% = 19$ of them to cover μ.

Part IV. (20 points) Public health officials believe that 18% of all high school students smoke at least one pack of cigarettes a day. A random sample of 300 high school students showed that 75 students smoked at least one pack of cigarettes a day.

1. (12 points) Does the evidence indicate that the true percentage is higher than what the public health officials believe?

$H_0: p = 18\%$ $\qquad \hat{p} = \frac{75}{300} = 25\%$

$H_a: p > 18\%$

$SE = \sqrt{\frac{(0.18)(0.72)}{300}} \simeq 2.1\%$

$z = \frac{25\% - 18\%}{2.1\%} = \frac{7\%}{2.1\%} = 3.33$

$\Rightarrow$ P-value $= \frac{100 - 99.919}{2}\% = 0.0405\% < 5\%$

$\Rightarrow$ Reject Ho. The true percentage of students who smoke at least one pack of cigarettes a day is significantly higher than 18%.

2. (6 points) Construct a 99% confidence interval for the percentage of all high school students, who smoke at least one pack of cigarettes a day.

$25\% \mp z \cdot SE$ $\qquad z = 2.60$ from z-table.

$SE = \sqrt{\frac{(0.25)(0.75)}{300}} = 0.025 = 2.5\%$

$\Rightarrow 25\% \mp (2.60)\, 2.5\% \Rightarrow [18.5\%, 31.5\%]$

3. (2 points) What is the confidence interval in Question 2 constructed for?

population percentage

Fill in the two blanks above by using ONLY TWO of the following words:
mean, population, sample, percentage.

Part V. (20 points) The Association Romance Writers conducted a research to compare the price of romance novels against all others (in dollars). The members of the association suspect that the romance novels are cheaper than any other novel.

The following is a MINITAB output for this study where "All" refers to a random sample containing mixed type of novels and "Romance" refers to a random sample of romance novels, independent from the other sample.

```
Variable      N     Mean    SE Mean
All          31     18.6      1.1
Romance      35     15.0      1.4
```

1. (15 points) Are the members of the association right in their suspicion? Test at $\alpha = 1\%$.

$$SE_{diff.} = \sqrt{(1.1)^2 + (1.4)^2} = 1.8$$

$H_0: \mu_1 = \mu_2$

$H_a: \mu_1 > \mu_2$

$$z = \frac{18.6 - 15.0}{1.8} = 2$$

$$\Rightarrow \text{P-value} = \frac{100\% - 95.45\%}{2} \cong 2.28\%$$

$\Rightarrow$ P-value $> 1\%$ $\Rightarrow$ Do not reject Ho.

There is no significant difference between the average prices of romance novels and other novels. The members are not right in their suspicion at 1% level of significance.

2. (5 points) Sketch an approximate histogram for the difference of the mean prices of all novels and romance novels. Show 3 points on the *x*-axis.