

## MIDTERM 2, ANSWERS.

1. The nicotine content in cigarettes of a certain brand is Normally distributed with mean (in milligrams)  $\mu$  and standard deviation  $\sigma = 0.2$ . The brand advertises that the mean nicotine content of their cigarettes is 1.5, but measurements on a random sample of 400 cigarettes of this brand gave a mean of  $\bar{x} = 1.51$ . Is this a strong evidence that the mean nicotine content is actually higher than advertised?

### **Solution:**

1.  $H_0: \mu = 1.5$

$H_a: \mu > 1.5$

2.  $z = (1.51 - 1.5) / (0.2 / \sqrt{400}) = 1$

3. P-value = 0.1589

4. P-value is large, therefore we cannot reject  $H_0$ , i.e. there is **no evidence that the mean nicotine content is higher than advertised.**

2. A researcher wishes to compare the average amount of time spent doing sports by students in a suburban school with that in a large city school. She selects a random sample of 19 suburban students and finds that their mean time spent doing sports is 7 hours per week, with a standard deviation of 3 hours. She also selects a random sample of 22 students from a large city school and finds that they spend on average 4 hours per week doing sports, with a standard deviation of 2 hours. Assume that t procedures are safe to use.

a) Was the data in this study collected with one-sample design, matched pairs design or the two-sample design?

**two-sample design**

b) Do these data support the claim that the average time devoted to sports is different for the suburban and city students?

### **Solution:**

|   |  |          |  |                                  |
|---|--|----------|--|----------------------------------|
| 1 |  | suburban |  | $n_1 = 19, \bar{x}_1 = 7, s = 3$ |
| 2 |  | city     |  | $n_2 = 22, \bar{x}_2 = 4, s = 2$ |

1.  $H_0: \mu_1 = \mu_2$

$H_a: \mu_1 \neq \mu_2$

2.  $t = 3.71, df = 18$

3.  $0.0005 * 2 < P\text{-value} < 0.001 * 2$ , i.e.  $.001 < P\text{-value} < 0.002$

4. P-value is small, i.e. we reject  $H_0$ , in other words, according to the data, **the average time devoted to sports is different for suburb and city students**

c) Use a 99% confidence interval to describe in words the difference between the average times spent doing sports by students in the suburbs and in the city.

99% confidence interval:  $(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{s_1^2/n_1 + s_2^2/n_2}$

$t^* = 2.878$  (df=18, C=99%), conf. interval:  $3 \pm 2.33$ , or **0.67 to 5.33**

**This means that we are 99% confident that the true difference in average times of doing sports lies somewhere between 0.67 and 5.33 hours per week.**

d) Give a 99% confidence interval for the average time spent doing sports by city students.

99% confidence interval:  $\bar{x}_2 \pm t^* s_2 / \sqrt{n_2}$

$t^*=2.831$  (df=21, C=99%)

conf. interval:  $4 \pm 1.21$ , or **[2.79 to 5.21] hours per week**

3. A food company developing a new breakfast drink is interested in whether preference for the new product is related to a person's gender. Of the 162 male volunteers who tried the drink, 65 liked it. Of the 135 female volunteers, 74 liked it.

a) Give a 95% confidence interval for the difference in proportions of men and women who like the drink.

|   |       |                                  |                      |
|---|-------|----------------------------------|----------------------|
| 1 | women | $n_1 = 135$ , success count = 74 | $p_1 = 0.5481$       |
| 2 | men   | $n_2 = 162$ , success count = 65 | $\hat{p}_2 = 0.4012$ |

95% confidence interval using  $z^*=1.96$ :

**0.0340 to 0.2598** or there are between 3.4% and 25.98% more individuals who like the drink among women than there are among men.

b) Is the difference significant at  $\alpha = 0.05$  level?

To do this, use the test of significance: 1.  $H_0: p_1 = p_2$

$H_a: p_1 \neq p_2$

2.  $p = 0.468$

$z = 2.53$

3. P-value:  $2 * 0.0057 = 0.0114$

4. **0.0114 < 0.05, i.e  $H_0$  should be rejected in favor of  $H_a$ . Yes, the difference is significant at 0.05 level.**

4. A news program asks its viewers to call in and express their opinion about proposed budget cuts for city libraries. They receive 1230 calls with 589 people supporting the budget cuts.

a) A statistician viewer knows that these data should not be used to give a confidence interval for the proportion of people supporting the cuts. What about this study makes him uneasy?

**The fact that the sample is not a simple random sample:** the viewers who gave their answers were not selected by a randomized procedure, but rather chose to call in (they may not be representative of the overall population, because they may feel stronger about the issue, or because they were the people viewing the particular news program, etc.)

b) If you decide to study the question yourself, how big a sample would you need to give a 95% confidence interval with 0.02 margin of error? Assume that you can use  $p^*=0.5$ .

$n = (z^*/m)^2 p^*(1-p^*)$ , where  $z^*=1.96$ ,  $m=.02$ ,  $p^*=0.5$

We need a sample with **n=2401**