University of Texas at Austin, Department of Mathematics
M358K - Applied Statistics

Practice Problems for the final exam.

Free-response problems.

Problem 4.1. (15 points) *Source: Ramachandran-Tsokos.*
A study of two kinds of machine failures shows that 58 failures of the first kind took on the average 79.7 minutes to repair with a sample standard deviation of 18.4 minutes, whereas 71 failures of the second kind took on average 87.3 minutes to repair with a sample standard deviation of 19.5 minutes. Find a 99% confidence interval for the difference between the true average amounts of time it takes to repair failures of the two kinds of machines.

Solution: The standard error of the difference in sample means is

$$\sqrt{\frac{18.4^2}{58}+\frac{19.5^2}{71}} = 3.3456.$$

So, the 99%-confidence interval for $\mu_1 - \mu_2$ is

$$(79.7 - 87.3) \pm 2.575(3.3456) = (-16.2149, 1.01486).$$

Problem 4.2. (8 points) To write an article about Denver for a tourist magazine you would like to estimate the average nightly cost for a hotel room in the Denver area. You are willing to assume that the nighly cost of a room is normally distributed.

You open up the yellow pages and take a random sample of hotels. The sample of 16 hotels gives an average nightly cost of \$55.98 and a **sample** standard deviation of \$12. Estimate the mean nightly cost and include a 95% confidence interval.

Solution: The number of degrees of freedom of the $t-$distribution is $16 - 1 = 15$. The critical value associated with this distribution and the confidence level of 95% is 2.131. Since the sample standard deviation equals \$12, the standard error equals $12/\sqrt{16} = 3$. So, the confidence interval we are looking for is

$$\mu = 55.98 \pm 2.131 \times 3 = 55.98 \pm 6.393.$$

Problem 4.3. (10 points) It is claimed that the bags of chocolate chips available in Costco contain at least 4 pounds (64 ounces).

A random sample of 50 bag measurements resulted in the sample average of 62 and the sample standard deviation of 8. Please, test the hypothesis that the bags contain at least 64 ounces of delicious chocolate chips at the significance level of 5%.

Solution:
Let μ denote the population mean of the weight of bags of chocolate chips. The hypotheses are:

$$H_0 : \mu = 64 \quad vs. \quad H_a : \mu < 64.$$

Since the population standard deviation is not given, we should use the $t-$test. However, with the sample size of 50 we can be comfortable enough using the $z-$test. The observed value of the test-statistic is

$$z = \frac{62-64}{8/\sqrt{50}} = -\frac{5\sqrt{2}}{4} = -1.768.$$

On the other hand, the critical value associated with a left-sided hypothesis test with a 5% confidence level is -1.645. Since the observed value of the test statistic falls below the critical value, we **reject the null hypothesis**.

Problem 4.4. *Source: Ramachandran-Tsokos.*
In a salary equity study of faculty at a certain university, sample salaries of 50 male assistant professors and 50 female assistant professors yielded the following basic statistics:

- for male assistant professors, the sample mean was \$46,400 with a sample standard deviation of \$360;
- for female assistant professors, the sample mean was \$46,000 with a sample standard deviation of \$220.

Assume independent samples. Test the hypothesis that the mean salary of male assistant professors exceeds the mean salary of female assistant professors at the significance level of 0.05.

Solution: Let μ_f be that mean salary of female assistant professors and let μ_m be the mean salary of male assistant professors. We are testing

$$H_0 : \mu_m = \mu_f \quad \text{vs.} \quad H_a : \mu_m > \mu_f.$$

It is appropriate to use the $t-$test since the sample sizes exceed 30. In obvious notation, the observed value of the test statistic is

$$t = \frac{\bar{x}_m - \bar{x}_f}{\sqrt{\frac{s_m^2}{n_m} + \frac{s_f^2}{n_f}}} = \frac{46400 - 46000}{\sqrt{\frac{(360)^2}{50} + \frac{(220)^2}{50}}} = 6.704015.$$

The number of degrees of freedom is $\min(n_m, n_f) - 1 = 49$. The upper tail probability is $1 - pt(6.704015, df = 49) = 0.00000000953624$. The result is evidently statistically significant.

Problem 4.5. (10 points) It is rumored that *Olivander's Wands* produces wands which affect the scores on the *Nastily Exhausting Wizarding Test (NEWT)* in unexpected ways. A scientific study (pioneered by certain muggle-borns) was inspired by this "observational study".

An experienced team is now trying to estimate the effect of the length of the wand used on the *NEWT* score. The two variables of interest are the flexibility of the wand x having values on the *Peverell* scale from 0 to 10, and the results on the *NEWTs* denoted by y on the usual scale from 0 to 100.

i. (2 points)
What is the explanatory variable and what is the response in the above situation?
Solution: Explanatory: wand flexibility; response: *NEWT* score.

ii. (5 points)
These are the values of the summary statistics for last year's results (in our usual notation):

$$\bar{x} = 7.7, \bar{y} = 77,$$
$$s_x = 1.4, s_y = 7, r = 0.7 \quad \text{(correlation between } x \text{ and } y\text{)}$$

You know that the equation of the least-squares regression line for predicting y form x is of the form

$$\hat{y} = bx + a$$

with $b = r\frac{s_y}{s_x}$. Find the slope b and intercept a consistent with the above data.
Solution:
From the provided formula for the slope, we get

$$b = r\frac{s_y}{s_x} = 0.7 \times \frac{7}{1.4} = 3.5.$$

We also know that the linear regression line must pass through the point $(\bar{x}, \bar{y})$. Hence,

$$a = \bar{y} - b\bar{x} = 77 - 3.5 \times 7.7 = 50.05.$$

So, $\hat{y} = 50.05 + 3.5x$ is the equation of the regression line.

iii. (3 points) Use the equation you obtained above to predict the result Harry Potter would have had on the *NEWTs* knowing that the flexibility of his wand is 8.
Solution: For $x = 8$, we have

$$\hat{y} = 50.05 + 3.5 \times 8 = 78.05.$$

Multiple-choice problems.

Problem 4.6. An experiment was designed to test whether people's reaction times to an orange light are different from their reaction times to a blue light. Upon being signaled with a light, the subjects would hit a button and the reaction time (in seconds) would be recorded. The reaction time (in seconds) of 16 subjects was recorded. The average reaction time for the blue light was 0.2025 seconds, and the average reaction time for the orange light was 0.1380 seconds. The sample standard deviation of the differences between reaction times was 0.0565. What is the 80%-confidence interval for the difference in mean reaction times? Assume the normal model for the reaction times.

(a) 0.0645 ± 0.0189

(b) 0.0645 ± 0.0122

(c) 0.2025 ± 0.0189

(d) 0.1380 ± 0.0122

(e) None of the above.

Solution: (a)
The point estimate for the mean difference in reaction times is $\bar{d} = 0.2025 - 0.1380 = 0.0645$. Since the sample size is 16, the number of degrees of freedom in our $t-$distribution is $16 - 1 = 15$. The critical value for the 80% confidence level, according to our tables, equals $t^* = 1.341$. So, the confidence interval is

$$0.0645 \pm (1.341)\left(\frac{0.0565}{\sqrt{16}}\right) = 0.0645 \pm 0.0189.$$

Problem 4.7. Twenty-five fortunate middle schoolers were put on an intense rope-jumping regimen in the hope of improving their times in the 40-yard dash. Assume that the distribution of the differences in the run times is normal. Let μ_d be the mean difference between the "pre-run" (before the regimen) and "post-run" (after the regimen). We want to test

$$H_0 : \mu_d = 0 \quad vs. \quad H_a : \mu_d > 0.$$

The observed average difference in run times was $\bar{x}_d = 0.0854$ while the sample standard deviation was $s_d = 0.2432$. What can you say about the $p-$value for this hypothesis test?

(a) It's below 0.01.

(b) It's between 0.01 and 0.02.

(c) It's between 0.02 and 0.025.

(d) It's between 0.025 and 0.05.

(e) None of the above.

Solution: (d)
The observed value of the $t-$statistic, under the null hypothesis, is

$$t = \frac{0.0854 - 0}{\frac{0.2432}{\sqrt{25}}} = 1.7558.$$

The number of degrees of freedom of the $t-$distribution of our test statistic is $25 - 1 = 24$. Consulting our $t-$tables, we conclude that the $p-$value is between 0.025 and 0.05. *Note: If one uses* **R***, one gets* 0.04593959.