UNIVERSITY OF TEXAS AT AUSTIN, DEPARTMENT OF MATHEMATICS
M358K - Applied Statistics

PRACTICE PROBEMS FOR IN-TERM III

True/false questions.

Problem 1.1. If a random variable X has a standard normal distribution, then X^2 has a chi-squared distribution with 1 degree of freedom. *True or false?*

Solution: TRUE

Problem 1.2. Let X be a standard normal random variable, and let Y be a $\chi-$squared random variable with one degree of freedom. Assume that X and Y are independent. Then, X/Y is $t-$distributed. *True or false?*

Solution: FALSE
Should be $X/\sqrt{Y}$.

Free-response problems.

Problem 1.3. (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 1.4. (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 1.5. An Airbus a330 can hold 277 passengers. The *Marginaire* airline knows from past experience that only about 90% of the passengers make it to their flights. So, they sell 300 tickets for every Airbus a330 flight.

Assume that passengers travel independently, i.e., the events that individual passengers make it to a flight are independent events. What is the approximate probability that a particular flight gets overbooked? *Note: Don't forget to use the continuity correction.*

Solution: The number of passengers X who do show up for a particular flight is binomial with 300 being the number of trials and 0.90 being the probability of success in every single trial. By the normal approximation to the binomial distribution, we have

$$\mathbb{P}[X > 277] = 1 - \mathbb{P}[X \leq 277.5] \approx 1 - \Phi\left(\frac{277.5 - 300 * 0.9}{\sqrt{300(0.9)(0.1)}}\right) = 1 - \Phi(1.44) = 1 - 0.9251 = 0.0749.$$

Problem 1.6. (10 points) *Source: "Probability and Statistics for Engineers and Scientists" by Walpole, Myers, Myers, and Ye.*
An improvement is a process for manufacturing is being considered. Samples are taken both from parts manufactured using the existing method and the new method. It is found that 75 out of a sample of 1500 items produced using the existing method are defective. It is also found that 80 out of a sample of 2000 items produced using the new method are defective. The two samples are independent.

Find the 90%-confidence interval for the true difference in the proportions of defectives produced using the existing and the new method.

Solution: Let p_1 denote the proportion of defectives resulting from the existing method and let p_2 denote the proportion of defectives resulting from the new method. We are supposed to find the 90%-confidence interval for $p_1 - p_2$.

The sample proportion of defectives for the existing method is $\hat{p}_1 = 75/1500 = 0.05$ and the sample proportion of defectives for the new method is $\hat{p}_2 = 80/2000 = 0.04$. So, the standard error equals

$$\sqrt{\frac{0.05(0.95)}{1500} + \frac{0.04(0.96)}{2000}} = 0.00713.$$

So, with the critical value corresponding to the 90%-confidence being $z^* = 1.645$, we get that the margin of error is

$$1.645(0.00713) = 0.0117.$$

Hence, the confidence interval is

$$0.01 \pm 0.0117 = (-0.0017, 0.0217).$$

Problem 1.7. (15 points)

A casino game involves rolling three dice. The winnings are proportional to the total number of sixes rolled. Suppose a gambler plays the game 150 times, with the following observed counts:

Number of sixes	0	1	2	3
Count	72	51	21	6

Assuming that the die rolls are independent, test the null hypothesis that the dice are all fair. *Note: Keep five decimal places for your expected counts.*

Solution: If the dice were all fair, we would have the following probabilities of the events that a particular number of sixes was rolled:

$$\mathbb{P}[0 \text{ sixes were rolled}] = (5/6)^3,$$
$$\mathbb{P}[1 \text{ six was rolled}] = 3(5/6)^2(1/6),$$
$$\mathbb{P}[2 \text{ sixes were rolled}] = 3(5/6)(1/6)^2,$$
$$\mathbb{P}[3 \text{ sixes were rolled}] = (1/6)^3.$$

So, the expected numbers E_i of times that i sixes in 150 rolls occur (for $i = 0, 1, 2, 3$) are

$$E_0 = 150 \times \frac{125}{6^3} = 86.80556,$$
$$E_3 = 150 \times 3\frac{25}{6^3} = 52.08333,$$
$$E_0 = 150 \times \frac{15}{6^3} = 10.41667,$$
$$E_0 = 150 \times \frac{125}{6^3} = 0.69444.$$

The observed value of the χ^2-statistic is

$$\frac{(86.80556 - 72)^2}{86.80556} + \frac{(52.08333 - 51)^2}{52.08333} + \frac{(10.41667 - 21)^2}{10.41667} + \frac{(0.69446 - 6)^2}{0.6944} = 53.83487.$$

With the number of degrees of freedom is $4 - 1 = 3$, we see that the observed value of the χ^2-statistic exceeds even the critical value 17.73 at the upper-tail probability of 0.0005.

Multiple-choice problems.

Problem 1.8. (5 points) *Source:"Probability and Statistics for Engineers and Scientists" by Walpole, Myers, Myers, and Ye.*
In a simple random sample of 500 households owning televisions in the city of Hamilton, Canada (pop. $536,915$), it is found that 340 subscribe to HBO. Find a 95% confidence interval for the true proportion of households with television which subscribe to HBO.

a. 0.68 ± 0.021
b. 0.68 ± 0.034
c. 0.68 ± 0.041
d. 0.68 ± 0.054
e. None of the above.

Solution: c.
The observed proportion of HBO subscribers is $\hat{p} = 340/500 = 0.68$. So, the standard error equals

$$\sqrt{\frac{0.68(0.32)}{500}} = 0.0209.$$

With the critical value associated with the 95%-confidence level equal to 1.96, we get the margin of error equal to

$$1.96(0.0209) = 0.041.$$

Hence, our confidence interval is 0.68 ± 0.041.

Problem 1.9. (5 points) *Source:"Probability and Statistics for Engineers and Scientists" by Walpole, Myers, Myers, and Ye.*

A commonly prescribed drug for relieving nervous tension is declared to be effective in 60% of patients. Experimental results with a **new** drug administered to a SRS of 100 patients show that 70 received relief. To answer the question whether the new drug is truly superior, you calculate the $p-$value. What do you get?

a. 0.0146
b. 0.0207.
c. 0.0292
d. 0.0414
e. None of the above.

Solution: b.
Let p denote the true proportion of people who receive relief when administered the new drug. We are testing

$$H_0 : p = 0.06 \quad \text{vs.} \quad H_a : p > 0.6.$$

The sample proportion of successes is $\hat{p} = 0.7$. The observed value of the $z-$statistic, under the null, equals

$$z = \frac{0.7 - 0.6}{\sqrt{\frac{0.6(0.4)}{100}}} = 2.04.$$

Consulting the tables, we get the $p-$value of

$$1 - \Phi(2.04) = 1 - 0.9793 = 0.0207.$$

Problem 1.10. (5 points) In 1956 Middletown, Lynd and Lynd conducted a sociological study in which questionnaires were administered to 784 white high school students. They were asked *"which* 2 *of the given* 10 *attributes where most desirable in their fathers."*

Among other things, and along with the students' genders, it was tallied how many of them mentioned *"being a college graduate"* as one of the 2 chosen desirable qualities. The following two-way table contains the resulting counts:

	Male	Female	Total
Mentioned	86	55	141
Not mentioned	283	360	643
Total	369	415	784

The question we can try to answer using the above data is whether males and females value this particular attribute differently. What is the conclusion of your hypothesis test of independence? *Note: When you calculate, keep four places after the decimal point for expected counts.*

a. The $p-$value is less than 0.001.
b. The $p-$value is between 0.001 and 0.005.
c. The $p-$value is between 0.005 and 0.01.
d. The $p-$value is between 0.01 and 0.02.
e. None of the above.

Solution: a.
We are testing

$$\begin{aligned} H_0 &: \quad \text{Gender and the given attribute are independent.} \\ &vs. \\ H_a &: \quad \text{Gender and the given attribute are not independent.} \end{aligned}$$

The expected counts (under the null hypothesis are)

$$\begin{aligned} E_{11} &= \frac{(141)(369)}{784} = 66.3635, \\ E_{12} &= \frac{(141)(415)}{784} = 74.6365, \\ E_{11} &= \frac{(643)(369)}{784} = 302.6365, \\ E_{11} &= \frac{(643)(415)}{784} = 340.3635. \end{aligned}$$

The observed value of the χ^2-statistic is

$$q^2 = \frac{(86 - 66.3635)^2}{66.3635} + \frac{(55 - 74.6365)^2}{74.6365} + \frac{(283 - 302.6365)^2}{302.6365} + \frac{(360 - 340.3635)^2}{340.3635} = 13.3836.$$

Now, we consult the χ^2-tables from your textbook for one degree of freedom. The $p-$value is less than 0.001.

Problem 1.11. 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 1.12. 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.

Problem 1.13. *Source: "Probability and Statistical Inference" by Hogg, Tanis, and Zimmerman.*
Let p_m and p_f be the population proportions of male and female sparrows who return to their hatching site. You want to test whether the two proportions are different. The observed number of males who returned is 124 out of 894, while the observed number of females who returned is 70 out of 700. What is your decision for this hypothesis test?

(a) Reject at the 0.01 significance level.

(b) Fail to reject at the 0.01 significance level; reject at the 0.025 significance level.

(c) Fail to reject at the 0.025 significance level; reject at the 0.05 significance level.

(d) Fail to reject at the 0.05 significance level; reject at the 0.10 significance level.

(e) None of the above.

Solution: (b)
We are testing

$$H_0 : p_m = p_f \quad vs. \quad H_a : p_m \neq p_f.$$

The observed proportions are

$$\hat{p}_m = \frac{124}{894} = 0.1387 \quad \text{and} \quad p_f = \frac{70}{700} = 0.10.$$

The pooled proportion estimate is

$$\hat{p} = \frac{124 + 70}{894 + 700} = 0.1217.$$

The observed value of the $z-$statistic is

$$z = \frac{\hat{p}_m - \hat{p}_f}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.1387 - 0.1}{\sqrt{0.1217(1-0.1217)\left(\frac{1}{894} + \frac{1}{700}\right)}} = 2.3454.$$

Since this is a two-tailed test, we have that the $p-$value equals $2\Phi(-2.3454)$. This value is between $2\Phi(-2.34)$ and $2\Phi(-2.35)$. Using the standard normal tables, we conclude that the $p-$value is between $2(0.0094)$ and $2(0.0096)$, i.e., between 0.0188 and 0.0192.