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. (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 1.4. (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 1.5. (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 1.6. (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 1.7. (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

H_0 : Gender and the given attribute are independent.

vs.

H_a : Gender and the given attribute are not independent.

The expected counts (under the null hypothesis are)

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

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 1.8. *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.

Problem 1.9. A car manufacturer claims that the mean time until the car battery needs to be replaced is five years. From past experience, the lifetime of a car battery is modeled as normal with a **known** standard deviation of one year. An environmental institute wants to test the car manufacturer's claim. They collect the data from 49 cars and find the sample average of 4.7 years. What is their decision going to be at the 2% significance level?

(a) Reject the null hypothesis.
(b) Fail to reject the null hypothesis.
(c) Accept the null hypothesis.
(d) Reject the alternative hypothesis.
(e) None of the above.

Solution: (a)
The environmental institute is testing

$$H_0 : \mu = 5 \quad vs. \quad H_a : \mu < 5.$$

Under the null hypothesis, the z-score corresponding to the given sample average is

$$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{4.7 - 5}{\frac{1}{\sqrt{49}}} = -2.1.$$

The p-value is, according to the standard normal tables,

$$\mathbb{P}[Z < -2.1] = 0.0179.$$

where $Z \sim N(0, 1)$. So, the null hypothesis is rejected at the 2% significance level.

Problem 1.10. In a hand sanitizer production facility, a machine is operated whose job is to fill the hand-sanitizer bottles with exactly 8 oz of the precious liquid. You are wondering whether the machine is correctly calibrated. From past experience, you know that you can model the amount in every bottle as normal with a known standard deviation of 1/4 oz. You are going to sample 100 bottles to test whether the machine is properly calibrated. If you choose that you are going to use a 1% significance level, what is the associated rejection region (in real units)?

(a) $[0, 7.9356] \cup [8.0644, \infty)$
(b) $(7.9356, 8.0644)$
(c) $[0, 7.951]$
(d) Not enough information is given.
(e) None of the above.

Solution: (a)
Let the unknown mean amount of hand sanitizer per bottle be denoted by μ. Then the distribution of the amount of hand sanitizer in a randomly chosen bottle can be written as

$$X \sim Normal(mean = \mu, sd = 1/4).$$

We need to test

$$H_0 : \mu = \mu_0 = 8 \quad vs. \quad H_a : \mu \neq \mu_0 = 8.$$

The rejection region for this two-sided test will be of the form (in our usual notation)

$$RR = \left(-\infty, \mu_0 + z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)\right] \cup \left[\mu_0 + z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right), \infty\right)$$

with $z_{\alpha/2} = \Phi^{-1}(0.005) = -2.576$. We have that

$$\mu_0 + z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) = 8 - 2.576\left(\frac{1/4}{\sqrt{100}}\right) = 7.9356,$$
$$\mu_0 - z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) = 8 + 2.576\left(\frac{1/4}{\sqrt{100}}\right) = 8.0644.$$

Of course, the amount in any bottle is at least 0. So, our rejection region is

$$RR = [0, 7.9356] \cup [8.0644, \infty).$$

We can say that the ∞ above is actually a placeholder for the capacity of a bottle.

Problem 1.11. (5 points) Suppose that we have a random sample of size 25 from a normal population with an unknown mean μ and a standard deviation of 4. Bertram Wooster was in charge of testing the hypotheses

$$H_0 : \mu = 10 \quad vs. \quad H_a : \mu > 10.$$

Before he went on vacation, he obtained the rejection region $[11.2, \infty)$. What is the power of the above test at the alternative mean $\mu_a = 11$?

(a) 0.4013

(b) 0.4503

(c) 0.5120

(d) 0.6368

(e) None of the above.

Solution: (a)

Under the given alternative mean, the distribution of the sample mean is

$$\bar{X} \sim Normal\left(mean = 11, sd = \frac{4}{\sqrt{25}} = \frac{4}{5}\right).$$

The power of the test at the alternative mean $\mu_a = 11$ is the probability that $\bar{X}$ falls into the given rejection region $[11.2, \infty)$. We have

$$\mathbb{P}[\bar{X} \geq 11.2] = \mathbb{P}\left[\frac{\bar{X} - 11}{0.8} \geq \frac{11.2 - 11}{0.8}\right] = \mathbb{P}\left[Z \geq 0.25\right]$$

where $Z \sim N(0, 1)$. Our answer is

$$\mathbb{P}[\bar{X} \geq 11.2] = \mathbb{P}\left[Z \geq 0.25\right] = 1 - \Phi(0.25) = 1 - 0.5987 = 0.4013.$$