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

Name:

TRUE/FALSE

1.1	TRUE	FALSE
1.2	TRUE	FALSE
1.3	TRUE	FALSE
1.4	TRUE	FALSE
1.5	TRUE	FALSE
1.6	TRUE	FALSE
1.7	TRUE	FALSE

MULTIPLE CHOICE

1.11 (5)	a	b	c	d	e
1.12 (5)	a	b	c	d	e
1.13 (5)	a	b	c	d	e
1.14 (5)	a	b	c	d	e
1.15 (5)	a	b	c	d	e
1.16 (5)	a	b	c	d	e

In-term exam II

True/false questions.

Problem 1.1. (2 points) Consider a two-sided hypothesis test for the population mean of a normal population. Then, the power of the test is symmetric with respect to the null mean. *True or false?*

Solution: TRUE

Problem 1.2. (2 points)

Resident statistitian Margie N. Rivera calculated a confidence interval of $[-0.56, 0.88]$. Her assistant boasts: "We should be 95% confident that the **sample average** falls in the provided interval". This is a valid statement. *True or false?*

Solution: FALSE

Problem 1.3. (2 points)

Freddie Threepwood conducts a hypothesis test. He calculates the observed value of the $z-$statistic to be 0.018 (under the null). At the significance level of 0.05, he should reject the null hypothesis.

True or false?

Solution: FALSE

Problem 1.4. (2 points)

Let μ denote the population mean of a normally distributed population model. At a given significance level α, we are testing

$$H_0 : \mu = \mu_0 \quad vs. \quad H_a : \mu < \mu_0.$$

Let μ_a and μ'_a be two values in the alternative such that $\mu_a < \mu'_a$. Then, the power of the test at the alternative μ_a exceeds the power of the test at the alternative μ'_a.

True or false?

Solution: TRUE

Problem 1.5. (2 points)

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 $X/\sqrt{Y}$.

Problem 1.6. (2 points) *Source: Problem 6.21 from the textbook.*

Consider the following two scenarios:

- Take a simple random sample of 100 sophomore students at your college or university.
- Take a simple random sample of 100 students at your college or university.

For each of these samples you record the amount spent on textbooks used for classes during the fall semester. You should suspect that the first sample should have the smaller margin of error.

True or false?

Solution: TRUE

Problem 1.7. (2 points) *Source: Problem 7.63 (a).*
In the usual notation, the researcher says that he wants to test

$$H_0 : \bar{x}_1 = \bar{x}_2 \quad \text{vs.} \quad H_a : \bar{x}_1 \neq \bar{x}_2.$$

His proposal is valid. *True or false?*

Solution: FALSE

Free-response problems.

Problem 1.8. (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 1.9. (10 points) *Source: Problem #7.74 from Moore-McCabe-Craig.*
Researchers were interested in the long-term psychological effects of being on a high-carbohydrate, low-fat (LF) diet versus a high-fat, low-carbohydrate (LC) diet. A total of 106 overweight and obese participants were randomly assigned to one of these two energy-restricted diets. At 52 weeks, 32 LC dieters and 33 LF dieters remained. Mood was assessed using a total mood disturbance score (TMDS), where a lower score is associated with a less negative mood. The sample avergage for the LC group was 47.3 with the sample standard deviation of 28.3. The sample avergage for the LF group was 19.3 with the sample standard deviation of 25.8. At the 0.05 significance level, is there a difference in the mean mood between the two kinds of dieters?

Solution: We are testing

$$H_0 : \mu_{LF} = \mu_{LC} \quad \text{vs.} \quad H_a : \mu_{LC} \neq \mu_{LF}$$

The observed value of the $t-$statistic is

$$t = \frac{\bar{x}_{LC} - \bar{x}_{LF}}{\sqrt{\frac{s_{LF}^2}{n_{LF}} + \frac{s_{LC}^2}{n_{LC}}}} = \frac{47.3 - 19.3}{\sqrt{\frac{28.3^2}{32} + \frac{25.8^2}{33}}} = 4.16481$$

The conservative number of degrees of freedom for the $t-$distribution of our test statistic is $\min(32, 33) - 1 = 31$. However, your $t-$table only contains 30 degrees of freedom, so that's what we'll use. The critical value associates with the upper-tail probability of 0.025 is 2.042. Since the observed value of the $t-$statistic is higher than the critical value, we reject the null hypothesis.

Problem 1.10. (10 points) You believe that the mean pancake consumption at the pancake jamboree is more than 16 per person. So, you decide to test your hypothesis. You model the pancake consumption as normally distributed with an unknown mean and with variance equal to 4. The plan is to collect the information on the number of pancakes consumed from a sample of 64 people. Since you want to have everything ready for the big day, you work out the rejection region right away and you get $(16.4375, \infty)$.

(i) (5 points) What is the significance level used to obtain the above rejection region?

Solution: From the lower bound of the rejection region, we get

$$16.4375 = 16 + z^* \left(\frac{2}{\sqrt{64}}\right) \quad \Rightarrow \quad z^* = 1.75.$$

Hence, the significance level is the upper tail probability associated with the critical value $z^* = 1.75$, i.e., 0.0401.

(ii) (5 points) What is the power of the above test at the alternative mean of 17?
Solution:

$$\mathbb{P}_{\mu=17}\left[\bar{X} > 16.435\right] = \mathbb{P}_{\mu=17}\left[\frac{\bar{X} - 17}{1/4} > \frac{16.435 - 17}{1/4}\right] = 1 - \Phi(-2.25) = \Phi(2.25) = 0.9878.$$

Multiple-choice problems.

Problem 1.11. (5 points) *Source: Ramachandran-Tsokos.*

A dendritic tree is a branched formation that originates from a nerve cell. In order to study brain development, researchers want to examine the brain tissues from adult guinea pigs. At least how many cells must the researchers select (randomly) so as to be 95% confident that the sample mean is within 3.4 cells of the population mean? Assume the **standard deviation** of 10.

(a) 28
(b) 33
(c) 34
(d) 35
(e) None of the above.

Solution: (c)

Problem 1.12. (5 points) The hypotheses $H_0 : \mu = 10$ versus $H_a : \mu \neq 10$ are examined using a sample of size $n = 18$. The one-sample $t-$statistic has the value of $t = -2.05$. Between what two values does the P-value of this test fall?

(a) $0.01 < P-$value < 0.02,
(b) $0.02 < P-$value < 0.025,
(c) $0.025 < P-$value < 0.05,
(d) $0.05 < P-$value < 0.10
(e) None of the above.

Solution: (d)

Problem 1.13. (5 points)

The mean area of the several thousand new apartments is advertized to be at least 1350 square feet. A tenant group thinks that the apartments are smaller than advertised. They hire an engineer to measure a sample of apartments; they should test:

a.: $H_0 : \mu = 1350$ against $H_a : \mu > 1350$.
b.: $H_0 : \mu = 1350$ against $H_a : \mu < 1350$.
c.: $H_0 : \mu < 1350$ against $H_a : \mu = 1350$.
d.: $H_0 : \mu = 1350$ against $H_a : \mu \neq 1350$.
e.: $H_0 : \mu < 1350$ against $H_a : \mu > 1350$.

Solution: b.

Problem 1.14. (5 points)

Lord Clarence Emsworth keeps meticulous records of the feeding patterns of his prized pet hog, *The Empress of Blandings.* He observes that the average weight of slops consumed per diem during the month of November equals 19 lbs with a sample standard deviation of 3lbs. As the incessantly perused tome "The Care of the Pig" states that the daily intake of slops should amount to 20 lbs, Lord Emsworth gets worried.

Luckily, his secretary, the efficient Baxter, knows a bit of statistics and can test whether the Empress is underfed. Which $p-$value will the efficient Baxter report?

a. Between 0 and 0.005.
b. Between 0.005 and 0.01.
c. Between 0.01 and 0.025.

d. Between 0.025 and 0.05.
e. None of the above.

Solution: d.
Actual $p-$value: 0.0395.

Problem 1.15. (5 points)
Bertie and Tuppy are playing a game. Bertie simulates 5 draws from a normal distribution without telling the parameter values to Tuppy. Then, Tuppy calculates the 95% confidence interval for the mean parameter.

Bertie simulated the values which resulted in a sample average of 5.88 and the sample standard deviation of 1.96. Which confidence interval should Tuppy get?

a. 5.88 ± 2.4337
b. 5.88 ± 2.7209
c. 5.88 ± 2.7764
d. 5.88 ± 2.2532
e. None of the above.

Solution: a.

Problem 1.16. (5 points) *Source: "Mathematical Statistics with Applications in R" by Ramachandran and Tsokos.*

A cross is hypothesized to result in a 3 : 1 phenotypic ratio of red-flowered to white-flowered plants. You set up a hypothesis test to test this claim. Suppose your cross actually produces 170 red- and 30 white-flowered plants. What is the $p-$value you obtain?

a. Less than 0.005.
b. Between 0.005 and 0.01.
c. Between 0.01 and 0.025.
d. Between 0.025 and 0.05.
e. None of the above.

Solution: a.
Let p denote the population proportion of red-flowered plants. We are testing

$$H_0 : p = 3/4 \quad \text{vs.} \quad H_a : p \neq 3/4.$$

Under the null, the observed value of the $z-$statistic is

$$z = \frac{\frac{170}{200} - \frac{3}{4}}{\sqrt{\frac{(3/4)(1/4)}{200}}} = 3.266$$

The $p-$value is more than $2(1 - \Phi(3.27)) = 2\Phi(-3.27) = 2(0.0005) = 0.001$.