Biostatistics [SBE304] (Fall 2019)
Tutorial 6

Test of Hypotheses

Lecturer: Prof. Ayman M. Eldeib TA: Asem Alaa

Tuesday 26th November, 2019

1 Tutorial facts

The problems in this tutorial comprises:

(A) Programming Works:

- **0** programming in-class demos.
- **1** programming homework.
- **0** self-practicing programming works.

(B) Problem Set:

- **13** problems to be solved in-class.
- **3** problems homework.
- **3** self-practicing problems.

Join this GitHub assignment page to create a repository for your submissions: TBA

2 Test of Hypotheses

2.1 Pre-class reading

1. Lecture notes of "Test of Hypotheses" by Prof. Ayman M. Eldeib
2. From Chapter 9 of Montgomery's textbook, read sections 9.1, 9.2, 9.3, & 9.5.

2.2 Chapter overview: Hypothesis Testing

- **Statistical hypothesis(hypothesis)** is a claim or assertion about the value of a single parameter, about the values of several parameters, or about the form of an entire population distribution.
- The **null hypothesis** H_0 is a claim about the value of a population parameter. The **alternate hypothesis** H_1 is a claim opposite to H_0. The **null hypothesis** H_0 is the claim that is initially assumed to be true (the "prior belief" claim). Often called the hypothesis of no change (from current opinion) and will generally be stated as an equality claim, equal to the **null value**. The **alternative hypothesis** or researcher's hypothesis, denoted by H_1 or H_a is the assertion that is contradictory to H_0. The alt hypothesis is often the claim that the researcher would really like to validate.
- **The null hypothesis** will be rejected in favor of the alternative hypothesis only if sample evidence suggests that H_0 is false. If the sample does not strongly contradict H_0, we will continue to believe in the plausibility of the null hypothesis. The two possible conclusions from a hypothesis-testing analysis are then reject H_0 or fail to reject H_0.
- A **test of hypothesis** is a method for using sample data to decide whether to reject H_0. H_0 will be assumed to be true until the sample evidence suggest otherwise. In other words, the null hypothesis will then be rejected if and only if the observed or computed test statistic value falls in the rejection region.
- A **test statistic** is a function of the sample data on which the decision is to be based.
- A **rejection region** is the set of all values of a test statistic for which H_0 is rejected. The basis for choosing a rejection region lies in consideration of the errors that one might be faced with in drawing a conclusion.

- **Type I error**: you reject H_0 when H_0 is true. $P(\text{Type I error}) = P(\text{reject } H_0 | H_0 \text{true}) = \alpha$. The resulting α is called the **significance level** of the test and the corresponding test is called a **level** α **test**. We will use test procedures that give α less than a specified level (0.05 or 0.01).
- **Type II**: fail to reject the null even though it is false.

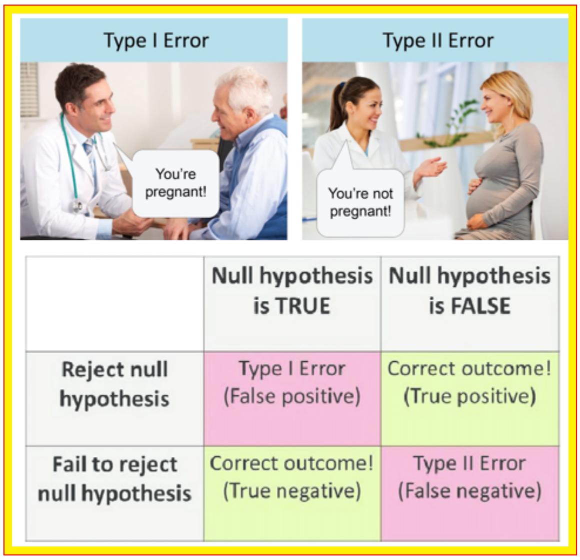

Figure 1: Source: Reddit

$$\alpha = P(\text{type I error}) = P(\text{reject } H_0 \text{ when } H_0 \text{ is true}) \tag{1}$$
$$\beta = P(\text{type II error}) = P(\text{fail to reject } H_0 \text{ when } H_0 \text{ is false}) \tag{2}$$

- **P-value** (or observed significance level) is the probability, calculated assuming that H_0 is true, of obtaining a value of the test statistic at least as contradictory to H_0 as the value calculated from the available sample. It is also the smallest significance level at which one can reject H_0. In other words, suppose we have observed a realization $x_{obs} = (x1, \dots, xn)$ of our random sample $X_1, \cdots, X_n \sim f(x, \theta)$. We wish to investigate the compatibility of the null hypothesis, with the observed data. We do so by comparing the probability distribution of the test statistic $T(X1, \cdots, Xn)$ with its observed value $t_{obs} = T(x_{obs})$, assuming H_0 to be true. As a measure of compatibility, we calculate

  $$p(xobs) = \text{p-value} = P(T(X1, \dots, X_n) \geq t_{obs} | H_0)$$

  In general, report the p-value. When it is less than 5% or 1 %, the result is statistically significant.

2.2.1 General Procedure for Hypothesis Tests

1. **Parameter of interest:** From the problem context, identify the parameter of interest.
2. **Null hypothesis,** H_0**:** State the null hypothesis, H_0.
3. **Alternative hypothesis,** H_1**:** Specify an appropriate alternative hypothesis, H_1.
4. **Test statistic:** Determine an appropriate test statistic.
5. **Reject** H_0 **if:** State the rejection criteria for the null hypothesis.
6. **Computations:** Compute any necessary sample quantities, substitute these into the equation for the test statistic, and compute that value.
7. **Draw conclusions:** Decide whether or not H_0 should be rejected and report that in the problem context.

Generally, once the experimenter (or decision maker) has decided on the question of interest and has determined the design of the experiment (that is, how the data are to be collected, how the measurements are to be made, and how many observations are required), only three steps are really required:

1. Specify the test statistic to be used (such as Z_0).
2. Specify the location of the critical region (two-tailed, upper-tailed, or lower-tailed).
3. Specify the criteria for rejection (typically, the value of α or the P-value at which rejection should occur).

Remarks

- In the best of all possible worlds, test procedures for which neither type of error is possible could be developed. However, this ideal can be achieved only by basing a decision on an examination of the entire population. The difficulty with using a procedure based on sample data is that because of sampling variability, an unrepresentative sample may result, e.g., a value of $\bar{X}$ that is far from μ or a value of $\hat{p}$ that differs considerably from p.

- Suppose an experiment and a sample size are fixed and a test statistic is chosen. Then decreasing the size of the rejection region to obtain a smaller value of α results in a larger value of β for any particular parameter value consistent with H_a. In other words, once the test statistic and n are fixed, there is no rejection region that will simultaneously make both α and all β's small. A region must be chosen to effect a compromise between α and β.

2.2.2 Tests about a Population Mean

Case 1: A Normal Population with a Known σ Rejection Regions:

$H_0 : \mu = \mu_0$	H_1		P Value
Reject H_0:	$\mu \neq \mu_0$	if $z_0 = \left\lvert \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \right\rvert > z_{\frac{\alpha}{2}}$	$2(1 - \Phi(z_0))$
Reject H_0:	$\mu < \mu_0$	if $z_0 = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} < -z_\alpha$	$\Phi(z_0)$
Reject H_0:	$\mu > \mu_0$	if $z_0 = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} > z_\alpha$	$1 - \Phi(z_0)$

Table 1: Rejection regions when known σ

Probability of a Type II Error for a Two-Sided Test on the Mean, Variance Known

$$\beta = \Phi(z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma}) - \Phi(-z_{\alpha/2} - \frac{\delta\sqrt{n}}{\sigma})$$

where $\delta = \mu - \mu_0$

Sample Size for a Two-Sided Test on the Mean, Variance Known

$$n \approx \frac{(z_{\alpha/2} + z_\beta)^2\sigma^2}{\delta^2}$$

where $\delta = \mu - \mu_0$

Sample Size for a One-Sided Test on the Mean, Variance Known

$$n \approx \frac{(z_\alpha + z_\beta)^2\sigma^2}{\delta^2}$$

where $\delta = \mu - \mu_0$

Case 2: A Normal Population with an Unknown σ Rejection Regions:

$H_0 : \mu = \mu_0$	H_1		P Value
Reject H_0:	$\mu \neq \mu_0$	if $t_0 = \left\lvert \frac{\bar{x}-\mu_0}{s/\sqrt{n}} \right\rvert > t_{\alpha/2,n-1}$	$2(1 - F_{\text{student}}(t_0, n-1))$
Reject H_0:	$\mu < \mu_0$	if $t_0 = \frac{\bar{x}-\mu_0}{s/\sqrt{n}} < -t_{\alpha,n-1}$	$F_{\text{student}}(t_0, n-1)$
Reject H_0:	$\mu > \mu_0$	if $t_0 = \frac{\bar{x}-\mu_0}{s/\sqrt{n}} > t_{\alpha,n-1}$	$1 - F_{\text{student}}(t_0, n-1)$

Table 2: Rejection regions when unknown σ

2.3 Problem Set

Hypothesis Testing

1. Problem

A hypothesis will be used to test that a population mean equals 10 against the alternative that the population mean is greater than 10 with unknown variance. What is the critical value for the test statistic T_0 for the following significance levels?

a. $\alpha = 0.01$ and n = 20

b. $\alpha = 0.05$ and n = 12

c. $\alpha = 0.10$ and n = 15

1. Solution

2. Problem

State whether each of the following situations is a correctly stated hypothesis testing problem and why

a. $H_0 : \mu = 25, H_1 : \mu \neq 25$

b. $H_0 : \sigma > 10, H_1 : \sigma = 10$

c. $H_0 : \bar{x} = 50, H_1 : \bar{x} \neq 50$

d. $H_0 : p = 0.1, H_1 : p = 0.5$

e. $H_0 : s = 30, H_1 : s > 30$

2. Solution

a. Yes, because the hypothesis is stated in terms of the parameter of interest, inequality is in the alternative hypothesis, and the value in the null and alternative hypotheses matches.

b. No, because the inequality is in the null hypothesis.

c. No, because the hypothesis is stated in terms of the statistic rather than the parameter.

d. No, the values in the null and alternative hypotheses do not match and both of the hypotheses are equality statements.

e. No, because the hypothesis is stated in terms of the statistic rather than the parameter.

3. Problem

A textile fiber manufacturer is investigating a new drapery yarn, which the company claims has a mean thread elongation of 12 kilograms with a standard deviation of 0.5 kilograms. The company wishes to test the hypothesis $H_0 : \mu = 12$ against $H_1 : \mu < 12$, using a random sample of four specimens

a. What is the type I error probability if the critical region is defined as $\bar{x} < 11.5$ kilograms?

b. Find β for the case in which the true mean elongation is 11.25 kilograms.

c. Find β for the case in which the true mean is 11.5 kilograms.

3. Solution

a.

$$\begin{aligned}
\alpha &= P(\text{reject } H_0 \text{ when } H_0 \text{ is true }) \\
&= P(\ \bar{x} \leq 11.5 \text{ when } \mu = 12) \\
&= P(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \leq \frac{11.5 - 12}{0.5/\sqrt{4}}) \\
&= P(Z \leq -2) \\
&= 0.02275
\end{aligned}$$

b.

$$\begin{aligned}\beta &= P(\text{accept } H_0 \text{ when } \mu = 11.25\ )\\ &= P(\bar{x} \geq 11.5 | \mu = 11.25)\\ &= P(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \geq \frac{11.5-11.25}{0.5/\sqrt{4}})\\ &= P(Z \geq 1)\\ &= 1 - P(Z \leq -1) = 0.15866\end{aligned}$$

c.

$$\begin{aligned}\beta &= P(\text{accept } H_0 \text{ when } \mu = 11.5\ )\\ &= P(\bar{x} \geq 11.5 | \mu = 11.5)\\ &= P(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \geq \frac{11.5-11.5}{0.5/\sqrt{4}})\\ &= P(Z \geq 0)\\ &= 0.5\end{aligned}$$

4. Problem

Repeat Exercise 3 using a sample size of $n = 16$ and the same critical region.

4. Solution

a. $\alpha \approx 0$

b. $\beta = 0.02275$

c. $\beta = 0.5$

5. Problem

In Exercise 3, calculate the p-value if the observed statistic is:

a. $\bar{x} = 11.25$

b. $\bar{x} = 11.0$

c. $\bar{x} = 11.75$

5. Solution

a. p-value = 0.00135

b. p-value = 0.000033

c. p-value = 0.158655

6. Problem

In Exercise 3, calculate the probability of a type II error if the true mean elongation is 11.5 kilograms and

a. $\alpha = 0.05$ and $n = 4$

b. $\alpha = 0.05$ and $n = 16$

c. Compare the values of β calculated in the previous parts. What conclusion can you draw?

6. Solution

a. 0.3594

b. 0.0082 (corrected)

c. .

7. Problem

A manufacturer is interested in the output voltage of a power supply used in a PC. Output voltage is assumed to be normally distributed with standard deviation 0.25 volt, and the manufacturer wishes to test $H_0 : \mu = 5$ volts against $H_1 : \mu \neq 5$ volts, using n = 8 units.

a. The acceptance region is $4.85 \leq \bar{x} \leq 5.15$. Find the value of α.

b. Find the power of the test for detecting a true mean output voltage of 5.1 volts.

7. Solution

a. $\alpha = 0.0910$

b. $1 - \beta = 0.29$

8. Problem

The proportion of adults living in Tempe, Arizona, who are college graduates is estimated to be $p = 0.4$. To test this hypothesis, a random sample of 15 Tempe adults is selected. If the number of college graduates is between 4 and 8, the hypothesis will be accepted; otherwise, you will conclude that $p \neq 0.4$

a. Find the type I error probability for this procedure, assuming that p = 0.4.

b. Find the probability of committing a type II error if the true proportion is really p = 0.2

8. Solution

a. 0.2937

b. 0.2572

9. Problem

The proportion of residents in Phoenix favoring the building of toll roads to complete the freeway system is believed to be p = 0.3. If a random sample of 10 residents shows that one or fewer favor this proposal, we will conclude that p < 0.3.

a. Find the probability of type I error if the true proportion is p = 0.3.

b. Find the probability of committing a type II error with this procedure if p = 0.2.

c. What is the power of this procedure if the true proportion is p = 0.2?

9. Solution

a. 0.08379

b. 0.78524

c. 0.21476

Hypothesis testing on population mean, variance σ known

10. Problem

State the null and alternative hypothesis in each case.

a. A hypothesis test will be used to potentially provide evidence that the population mean is more than 10.

b. A hypothesis test will be used to potentially provide evidence that the population mean is not equal to 7.

c. A hypothesis test will be used to potentially provide evidence that the population mean is less than 5

10. Solution

11. Problem

A hypothesis will be used to test that a population mean equals 5 against the alternative that the population mean is less than 5 with known variance σ. What is the critical value for the test statistic Z_0 for the following significance levels?

a. 0.01

b. 0.05

c. 0.10

11. Solution

a. -2.33

b. -1.64

c. -1.29

12. Problem

A manufacturer produces crankshafts for an automobile engine. The crankshafts wear after 100,000 miles (0.0001 inch) is of interest because it is likely to have an impact on warranty claims. A random sample of n = 15 shafts is tested and $\bar{x} = 2.78$. It is known that $\sigma = 0.9$ and that wear is normally distributed.

a. Test $H_0 : \mu = 3$ versus $H_1 : \neq 3$ using $\alpha = 0.05$.

b. What is the power of this test if $\mu = 3.25$?

c. What sample size would be required to detect a true mean of 3.75 if we wanted the power to be at least 0.9?

12. Solution

a. fail to reject H_0

b. $\beta = 0.80939$, so the power of the test equals $1 - \beta = 0.19$

c. 16

13. Problem

The life in hours of a battery is known to be approximately normally distributed with standard deviation σ = 1.25 hours. A random sample of 10 batteries has a mean life of $x = 40.5$ hours

a. Is there evidence to support the claim that battery life exceeds 40 hours? Use $\alpha = 0.5$.

b. What is the P-value for the test in part (a)?

c. What is the β-error for the test in part (a) if the true mean life is 42 hours?

d. What sample size would be required to ensure that β does not exceed 0.10 if the true mean life is 44 hours?

e. Explain how you could answer the question in part (a) by calculating an appropriate confidence bound on battery life

13. Solution

a. fail to reject H_0

b. P-value = 0.1038

c. β = 0.000325

d. n = 1

e. Hint: calculate 95% one-sided confidence interval on the sample mean.

14. Problem

Supercavitation is a propulsion technology for undersea vehicles that can greatly increase their speed. It occurs above approximately 50 meters per second when pressure drops sufficiently to allow the water to dissociate into water vapor, forming a gas bubble behind the vehicle. When the gas bubble completely encloses the vehicle, **supercavitation** is said to occur. Eight tests were conducted on a scale model of an undersea vehicle in a towing basin with the average observed speed $\bar{x} = 102.2$ meters per second. Assume that speed is normally distributed with known standard deviation $\sigma = 4$ meters per second.

a. Test the hypothesis $H_0 : \mu = 100$ versus $H_1 : \mu < 100$ using $\alpha = 0.05$.

b. What is the P-value for the test in part (a)?

c. Compute the power of the test if the true mean speed is as low as 95 meters per second.

d. What sample size would be required to detect a true mean speed as low as 95 meters per second if you wanted the power of the test to be at least 0.85?

e. Explain how the question in part (a) could be answered by constructing a one-sided confidence bound on the mean speed.

14. Solution

a. .

b. 0.94

c. 0.02938

d. 5

e. Hint: calculate 95% one-sided confidence interval on the sample mean.

15. Problem

Medical researchers have developed a new artificial heart constructed primarily of titanium and plastic. The heart will last and operate almost indefinitely once it is implanted in the patient's body, but the battery pack needs to be recharged about every 4 hours. A random sample of 50 battery packs is selected and subjected to a life test. The average life of these batteries is 4.05 hours. Assume that battery life is normally distributed with standard deviation $\sigma = 0.2$ hours.

a. Is there evidence to support the claim that mean battery life exceeds 4 hours? Use $\alpha = 0.05$.

b. What is the P-value for the test in part (a)?

c. Compute the power of the test if the true mean battery life is 4.5 hours.

d. What sample size would be required to detect a true mean battery life of 4.5 hours if you wanted the power of the test to be at least 0.9?

e. Explain how the question in part (a) could be answered by constructing a one-sided confidence bound on the mean life.

15. Solution

a. .

b. 0.04

c. 0

d. 2

e. Hint: calculate 95% one-sided confidence interval on the sample mean.

16. Problem

The bacterial strain Acinetobacter has been tested for its adhesion properties. A sample of five measurements gave readings of 2.69, 5.76, 2.67, 1.62 and 4.12 dyne-cm^2. Assume that the standard deviation is known to be 0.66 dyne-cm^2 and that the scientists are interested in high adhesion (at least 2.5 dyne-cm^2).

a. Should the alternative hypothesis be one-sided or two-sided?

b. Test the hypothesis that the mean adhesion is 2.5 dyne-cm^2.

c. What is the P-value of the test statistic?

16. Solution

a. .

b. evidence rejects H_0

c. 0.002

17. Problem

Suppose that the population mean of systolic blood pressure in the US is 115. We hypothesize mean systolic blood pressure is lower than 115 among people who consume a small amount (e.g., around 3.5 ounces) of dark chocolate every day. Assume that systolic blood pressure, X, in this population has a $N(\mu, \sigma^2)$ distribution. To evaluate our hypothesis, we randomly selected 100 people, who include a small amount of dark chocolate in their daily diet, and measured their blood pressure. If the sample mean is $\bar{x} = 111$ and the sample variance is s = 32, can we reject the null hypothesis at 0.1 confidence level?

17. Solution

18. Problem

18. Solution

19. Problem

19. Solution

3 R Programming Language

3.1 R Programming Problems

1. Problem

Using the `Pima.tr` dataset from `MASS` library (for description of this data run `?Pima.tr` in the R-console; make sure to load the library first by `library(MASS)` and `data = Pima.tr` to load the data in a new variable):

a. Test the hypothesis that the population mean of diastolic blood pressure for Pima Indian women is 70, such that $H_0 : \mu = 70$ and $H_1 : \mu \neq 70$.

b. What is the P-value of the test statistic?

c. Verify your previous answers using `t.test` function (check documentation `?t.test`)