

# Second Course in Statistics: lecture 13

## Statistical inference

- Review: confidence interval estimation
- Hypothesis testing: more examples
- P-value
- Two types of errors
- Power of the test

# Statistical inference

## Review: Hypothesis testing procedure

1. State  $H_0$  and  $H_a$
2. Choose proper test statistic and specify its sampling distribution under the  $H_0$
3. Find out the rejection region according to level of significance  $\alpha$  and  $H_a$
4. Compare observed sample with rejection region and draw a conclusion.

# Statistical inference

## Hypothesis testing

### Manuscript example continued

An archaeologist discovers a short manuscript containing 30 words and 198 letters in total. There are two written versions of the language. In the early form of the language the mean word length is 6.2 letters with standard deviation 2.5; in the late form certain words were given prefixes, raising the mean length to 7.6 letters but leaving the standard deviation unaltered.

Assume that this manuscript was taken from early form of the language. Assess this claim by hypothesis testing at 5% level of significance.

# Statistical inference

## Hypothesis testing

### Solution

#### 1. Specify Hypotheses:

- $H_0: \mu = 6.2$
- $H_a: \mu > 6.2$  or  $\mu = 7.6$

#### 2. Test statistic :

- $\bar{X} \sim N\left(6.2, \frac{2.5^2}{\sqrt{30}}\right)$  or
- $Z = \frac{\bar{X} - 6.2}{2.5/\sqrt{30}} \sim N(0, 1)$
- Observed  $\bar{x} = 198/30 = 6.6$  or
- the observed  $z_{\text{ob}} = \frac{6.6 - 6.2}{2.5/\sqrt{30}} = 0.8764$

# Statistical inference

## Hypothesis testing

### Solution continued

#### 3 Rejection region:

- Under the sampling distribution of  $\bar{X}$ , rejection region is
$$\bar{X} > \mu + z_{0.05} \frac{\sigma}{\sqrt{30}} = 6.2 + 1.645 \times \frac{2.5}{\sqrt{30}} = 6.951$$
- Under the sampling distribution of  $Z$ , rejection region is just
$$Z > z_{0.05} = 1.645$$

#### 4 Comparison and conclusion :

- Since the observed  $\bar{x} = 6.6 < 6.951$ , we cannot reject  $H_0$  at **5% level of significance** and we conclude that we haven't found critical evidence to prove that this manuscript is from late language form.
- The same conclusion is drawn when comparing the value of  $z$ .

# Statistical inference

## Hypothesis testing

### Example 2

The average (mean) live weights of a farmer's steers prior to slaughter was 380 pounds in past years. This year his 50 steers were fed on a new diet. Suppose that we consider these 50 steers fed on the new diet as a random sample taken from a population of all possible steers that may be fed the diet now or in the future. The mean weight of steers fed on new diet is 390 pounds. Suppose the standard deviation of last year is 35.2 pounds. Use  $\alpha = 0.01$  to test the research hypothesis that the mean live weight for steers on the new diet is greater than 380. Make a decision by comparing the test statistics with the critical value.

# Statistical inference

## Hypothesis testing

### Solution

#### 1. Specify Hypotheses:

- $H_0: \mu = 380$
- $H_a: \mu > 380$

#### 2. Test statistic :

- $Z = \frac{\bar{X} - 380}{35.2/\sqrt{50}} \sim N(0, 1)$
- $z_{\text{ob}} = \frac{390 - 380}{35.2/\sqrt{50}} = 2.0088$

#### 3. Rejection region:

- If  $Z > z_{0.01} = 2.33$ , we can reject  $H_0$ , otherwise we cannot reject  $H_0$

# Statistical inference

## Hypothesis testing

### Solution continued

#### 4 Comparison and conclusion:

- Since  $z_{\text{ob}} < 2.33$ , we cannot reject  $H_0$  at 1% significance level and conclude that **statistically** new diet has not increased the steers' weight.

#### Caution

- We might be tempted to accept the null hypothesis  $\mu = 380$ . The problem with this statement is that we do not know the **type II error**, **the probability of incorrectly accepting the null hypothesis**.
- In the situation of unknown type II error, we recommend stating that there is insufficient evidence to support rejecting the null hypothesis.
- If type II error can be calculated and is reasonably small, then  $H_0$  is accepted.



# Statistical inference

## Hypothesis testing: p-value

### P-value

Previously conclusion is made by comparing the observed sample value with the critical value specified by the rejection region. Another popular methods is to compare p-value with level of significance. This method turns out to be a practical convention in the research since the result can be easily produced by many statistical software tools.

### Definition

**P-value** is the probability, computed assuming that  $H_0$  is true, that the test statistic would take a value as extreme or more extreme than that actually observed.

# Statistical inference

## Hypothesis testing: p-value

### Making conclusion by p-value

- The key to calculating the p-value is the sampling distribution of the test statistic.
- Comparing the p value with level of significance is exactly the same as comparing the observed sample value with rejection region.
- If p-value is smaller than significance level  $\alpha$ , we can reject  $H_0$ , otherwise we cannot.
- The smaller the p-value, the stronger the evidence against  $H_0$  provided by the data.

# Statistical inference

## Hypothesis testing: p-value

### P-value in example 2

- Calculate  $P(\bar{X} > 390)$  under the null hypothesis or equivalently  $P(Z > 2.0088)$

- 

$$\begin{aligned} \text{p-value} &= P(\bar{X} > 390) \\ &= P\left(\frac{\bar{X} - 380}{35.2/\sqrt{50}} > \frac{390 - 380}{35.2/\sqrt{50}}\right) \\ &= P(Z > 2.0088) = 0.0222 \end{aligned}$$

- Conclusion: Since p-value is larger than 1% level of significance, we cannot reject  $H_0$  at 1% level of significance.

# Statistical inference

Hypothesis testing: p-value

## Example 3

Do middle-aged male executives have different average blood pressure than the general population? The national center for Health Statistics reports that the mean systolic blood pressure for males 35 to 44 years of age is 128 and the standard deviation in this population is 15. The medical director of a company looks at the medical records of 72 company executives in this age group and finds that the mean systolic blood pressure in this sample is  $\bar{x} = 126.07$ . Is this the evidence that executive blood pressure differ from the national average?

Make a decision by calculating p value.

# Statistical inference

Hypothesis testing: p-value

## Solution

- $H_0 : \mu = 128$

- $H_a : \mu \neq 128$

- Test statistic:  $Z = \frac{\bar{X} - 128}{15/\sqrt{72}}$

Observed  $z_{\text{ob}} = \frac{126.07 - 128}{15/\sqrt{72}} = -1.092$

- p-value =  $2P(Z < -1.09) = 2 \times 0.1379 = 0.2758$
- Conclusion: Since p-value is larger than 10%, we cannot reject  $H_0$  at 10% level of significance and conclude that executive blood pressure does not differ from national level.

# Statistical inference

## Hypothesis testing: two types of errors

### Type I Error (rejection error)

A Type I error is committed if we reject the null hypothesis when it is true. The probability of a Type I error is denoted by symbol  $\alpha$ .

### Type II Error (acceptance error)

A Type II error is committed if we accept the null hypothesis when it is false and the research hypothesis is true. The probability of a Type II error is denoted by symbol  $\beta$ .

# Statistical inference

Hypothesis testing: two types of errors

## Two-way decision process

Decision	Null Hypothesis	
	True	False
Reject $H_0$	Type I error $(\alpha)$	Correct $(1 - \beta)$
Accept $H_0$	Correct $(1 - \alpha)$	Type II error $(\beta)$

# Statistical inference

## Hypothesis testing: two types of errors

### Remarks

- Type I error can be translated as  $P(X \in \text{rejection region} \mid H_0)$
- Type II error can be translated as  $P(X \in \text{acceptance region} \mid H_a)$
- It is impossible to determine the acceptance and rejection region so as to minimize both  $\alpha$  and  $\beta$  simultaneously. For a fixed sample size  $n$  and fixed value under  $H_a$ , as we change the rejection region to increase  $\alpha$ , then  $\beta$  decreases, and vice versa.
- Rejection error is of much concern so that type I error is specified by the experimenter.  $\alpha$  is usually chosen to be 0.01, 0.05, 0.10.



# Statistical inference

Hypothesis testing: calculation of type II error

## Steps

1. State  $H_0$  and  $H_a$ .  $H_a$  is usually a particular alternative we want to detect.
2. Choose proper significance level  $\alpha$ .
3. Find acceptance region of  $\bar{X}$  under  $H_0$
4. Calculation of type II error is straight forward by definition, i.e. the probability of  $\bar{X}$  in the acceptance region found in step 3 when  $H_a$  is true.

# Statistical inference

Hypothesis testing: calculation of type II error

## Example: mean live weight of steers

Suppose the actual mean live weight of steers on the new diet is 387.  $\sigma = 35.2$   $n = 50$   $\alpha = 0.01$

## Solution

1.  $H_0 : \mu = 380$   
 $H_a : \mu = 387$
2. Acceptance region under  $H_0$  :

$$\bar{X} < \mu_0 + z_{0.01} \frac{\sigma_X}{\sqrt{n}} = 380 + 2.33 \frac{35.2}{\sqrt{50}} = 391.5988$$

3. Type II error:

$$\beta = P(\bar{X} < 391.5988 | \mu_a = 387)$$

# Statistical inference

Hypothesis testing: calculation of type II error

## Solution continued

3 Type II error:

$$\begin{aligned}\beta &= P(\bar{X} < 391.5988 \mid \mu_a = 387) \\ &= P\left(\frac{\bar{X} - \mu_a}{\sigma_X/\sqrt{50}} < \frac{391.5988 - \mu_a}{\sigma_X/\sqrt{50}}\right) \\ &= P\left(Z < \frac{391.5988 - 387}{35.2/\sqrt{50}}\right) \\ &= P(Z < 0.9238) \\ &= 0.8218\end{aligned}$$

# Statistical inference

Hypothesis testing: calculation of type II error

## Example continued: mean live weight of steers

Suppose the actual mean live weight of steers on the new diet is 395.  $\sigma = 35.2$   $n = 50$   $\alpha = 0.01$

### Solution

1.  $H_0 : \mu = 380$   
 $H_a : \mu = 395$
2. Acceptance region under  $H_0$  :

$$\bar{X} < 391.5988$$

3. Type II error:

$$\beta = P(\bar{X} < 391.5988 \mid \mu_a = 395)$$

# Statistical inference

Hypothesis testing: calculation of type II error

## Solution continued

### 3 Type II error:

$$\begin{aligned}\beta &= P(\bar{X} < 391.5988 \mid \mu_a = 395) \\ &= P\left(\frac{\bar{X} - \mu_a}{\sigma_X/\sqrt{50}} < \frac{391.5988 - \mu_a}{\sigma_X/\sqrt{50}}\right) \\ &= P\left(Z < \frac{391.5988 - 395}{35.2/\sqrt{50}}\right) \\ &= P(Z < -0.6832) \\ &= 1 - 0.7517 \\ &= 0.2483\end{aligned}$$

# Statistical inference

Hypothesis testing: calculation of type II error

## Observations

- Type II error depends on how close the actual mean is to  $\mu_0 = 380$ . The further away the actual mean from  $\mu_0$ , the smaller the  $\beta$ .
- The whole process of determining type II error is a test of "what-if" type of process.
- Graphic presentation is very helpful to understand the situation.

# Statistical inference

Hypothesis testing: calculation of type II error

## Conclusion of mean steer weight example

- Previously observed  $\bar{X} = 390$  does not fall in rejection region, we conclude that there was insufficient evidence to support rejecting  $H_0$  because  $\beta$  is unknown.
- If  $\mu_a = 395$  and we are willing to risk at  $\beta = 0.25$  (falsely accepting  $H_0$ ) we can accept the null hypothesis  $\mu_0 = 380$
- If  $\mu_a = 387$  and we are still willing to risk at  $\beta = 0.25$  we cannot accept the null hypothesis  $\mu_0 = 380$

# Statistical inference

## Hypothesis testing: power of the test

### Definition

The probability that a test at  $\alpha$  level rejects  $H_0$  when  $H_a$  (for a particular value of the parameter) is true is called the power of the test to detect this particular alternative.

It can be translated as  $P(X \in \text{rejection region} \mid H_a)$

Power of the test = 1 – type II error =  $1 - \beta$

### Example of mean steer weight

If  $\mu_a = 395$ ,

$$\text{power of the test} = 1 - \beta = 1 - 0.2483 = 0.7517$$



# Statistical inference

## Hypothesis testing: power of the test

### Remarks

- The larger the power of the test, the better decision can be made.
- For a fixed sample size and fixed value of parameter under  $H_a$ , when  $\alpha$  is determined by the experimenter, power of the test of preset, so is type II error.