Statistical Inference

Hypothesis Testing

BIOSTATISTICS
Biomedical Engineering

Sonia Malefaki

Department of Mechanical Engineering & Aeronautics
University of Patras, Greece.

☎: 2610 997673, ✉: smalefaki@upatras.gr

November 28, 2020

Outline

1. Introduction
2. Hypothesis Testing
   - Introduction to Hypothesis Testing
   - Type I and Type II Errors
   - P-value
   - Hypothesis tests for several parameters of one or two populations
3. Hypothesis testing using SPSS

Introduction

Goal: To draw conclusions about the population parameters from experimental data.

Statistical Inference consists of those methods by which one makes inference or generalization about a population.

Statistical Inference

- Estimation
  - Point estimation
  - Interval estimation
- Hypothesis Testing

Introduction

There is a drug which if treated, early results appear within 10 days. A new drug for the same disease is discovered and the drug company claims that the new drug brings results in less time. So if you know that the average time for the appearance of the first results is 10 days (i.e. $\mu = 10$), you need to check if the new drug acts in less time (i.e. $\mu < 10$).

This is a statistical hypothesis.

A statistical hypothesis is a assertion or conjecture concerning one or more populations.

The truth or falsity of a statistical hypothesis is never known with absolute certainty unless we examine the entire population.

We take a random sample from the population of interest and use the data contained in this sample to provide evidence that either supports or does not support the hypothesis.

▶ Definition

The process of statistical inference by which we arrive at a decision on the selection of one of two opposite hypothesis is called statistical hypothesis testing.

Introduction

The rejection of the hypothesis means that there is a small probability of obtaining the sample information observed when in fact the hypothesis is true.

- $H_0: \ \mu = 10$ **Null hypothesis**
- $H_1: \ \mu < 10$ **Alternative hypothesis**

- The decision to reject or not the null hypothesis H_0, is based on the value of a function of the observations called **test statistic**.
- In order for a function of observations to be a test statistic, it should have a fully specified distribution under the null hypothesis.
- The possible values of the test statistic, are divided into two non overlapping regions, in one region the hull hypothesis is rejected and it is called **critical region** . The last number that we observe in passing into the critical region is called the critical value.

Introduction

- The process of a statistical hypothesis testing

1. The null hypothesis is defined ($H_0 : \theta = \theta_0$)
2. The alternative hypothesis is defined ($H_1 : \theta \neq \theta_0$ or $H_1 : \theta < \theta_0$ or $H_1 : \theta > \theta_0$)
3. The value of the test statistic is computed based on the sample.
4. The critical region is specified
5. Conclusions are extracted.

Type Error I and II

- **Type Error I** (Rejection of the null hypothesis when it is true)

$$a = P(\text{reject } H_0 | H_0 \text{ true})$$

a : level of significance or size of the test.

- **Type Error II** (Nonrejection of the null hypothesis when it is false)

$$\beta = P(\text{nonreject } H_0 | H_0 \text{ false})$$

The probability of a type II error depends on the value of the parameter θ, thus β is a function of θ, i.e. $\beta(\theta)$ with domain the region of values of θ under H_1.

- **Power** of a test is the probability of rejecting H_0 given that a specific alternative is true.

$$\gamma = P(\text{reject } H_0 | H_0 \text{ false}) = 1 - \beta$$

$$\gamma(\theta) = 1 - \beta(\theta)$$

Type Error I and II

- The type I error and type II error are related. A decrease in the probability of one generally results in an increase in the probability of the other.

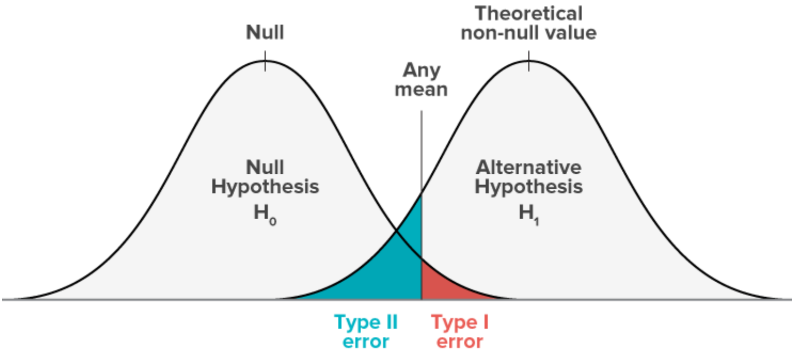
- An increase in the sample size n will reduce a and β simultaneously.

- If the null hypothesis is false, β is a maximum when the true value of a parameter approaches the hypothesized value. The greater the distance between the true value and the hypothesized value, the smaller β will be.

- Preselection of a significant level *a*. *a* is usually small (i.e. 0.05, 0.01, 0.1).

- By selecting also a proper sample size *n*, it is possible to predefine also $\beta(\theta)$.

Type Error I and II



p-value

p-value is defined as the probability of obtaining a result equal to or "more extreme" than what was actually observed, assuming that the null hypothesis is true.

p-value is the lowest level (of significance) at which the observed value of the test statistic is significant.

$$P(Z > |z| | H_0 \text{ true})$$

where z is the value of the test statistic for this random sample.
This probability refers to one-tailed tests.

For the two tailed test the p-value is

$$2 \cdot P(Z > |z| | H_0 \text{ true})$$

Test on a Single mean μ

with known variance

$$H_0: \ \mu = \mu_0 \quad \text{vs} \quad H_1: \mu \neq \mu_0 \quad \text{or} \quad \mu < \mu_0 \quad \text{or} \quad \mu > \mu_0$$

It is known that $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$.
Thus, under the null hypothesis

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$$

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

1. $H_1 : \mu \neq \mu_0$: the critical region is $z < -z_{a/2}$ or $z > z_{a/2}$
2. $H_1 : \mu > \mu_0$: the critical region is $z > z_a$
3. $H_1 : \mu < \mu_0$: the critical region is $z < -z_a$

The power of the test

- $H_0: \ \mu = \mu_0$ vs $H_1: \mu < \mu_0$

$$\gamma(\mu) = \Phi\left(-z_a - \frac{(\mu - \mu_0)\sqrt{n}}{\sigma}\right), \quad \mu < \mu_0$$

- $H_0: \ \mu = \mu_0$ vs $H_1: \mu > \mu_0$

$$\gamma(\mu) = 1 - \Phi\left(z_a - \frac{(\mu - \mu_0)\sqrt{n}}{\sigma}\right), \quad \mu > \mu_0$$

- $H_0: \ \mu = \mu_0$ vs $H_1: \mu \neq \mu_0$

$$\gamma(\mu) = 1 - \Phi\left(z_{a/2} - \frac{(\mu - \mu_0)\sqrt{n}}{\sigma}\right) + \Phi\left(-z_{a/2}\frac{(\mu - \mu_0)\sqrt{n}}{\sigma}\right), \quad \mu \neq \mu_0$$

Exercise

A dealer has ordered a large batch of a product. From the specification, it is known that the weights of the products follow a normal distribution with a mean 2 kg and a standard deviation 50 grams. The dealer suspects that the mean weight of produced pieces is smaller than that in the specifications, and he has decided to accept the consignment received with probability 0.95 when in fact apply the standards by taking a random sample of 16 pieces of it.

1. Define the proper null and alternative hypothesis and also the test statistic and the critical region for the hypothesis test that the dealer wants to do.
2. Compute $\beta(1.98)$.
3. If the true weight is equal to 1.98 kilos with what probability the dealer will reject the batch?
4. From the test of the 16 pieces of this batch, the mean weight is 1.985 kilos, which must be the decision of the dealer?
5. If the probability $\beta(1.98)$ in (2) is too big and we would like to reduce it in order to be equal to 0.10, maintaining the same significance level in the test, how many pieces should be controlled by the dealer from the batch?

Relationship between hypothesis testing and confidence intervals

$$H_0 : \mu = \mu_0 \quad \text{vs} \quad H_1 : \mu \neq \mu_0$$

- When the null hypothesis is not rejected?

$$-z_{a/2} < z < z_{a/2}$$

$$-z_{a/2} < \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} < z_{a/2}$$

$$\bar{x} - \frac{z_{a/2}\sigma}{\sqrt{n}} < \mu_0 < \bar{x} + \frac{z_{a/2}\sigma}{\sqrt{n}}$$

The null hypothesis is accepted in significance level a when the value μ_0 belongs to $100(1-a)\%$ confidence interval on the mean.

- By using confidence interval for two-tailed hypothesis testing, many null hypotheses can be tested simultaneously.

Test on a Single mean μ

with unknown variance

$$H_0: \ \mu = \mu_0 \quad \text{vs} \quad H_1: \mu \neq \mu_0 \quad \text{or} \quad \mu < \mu_0 \quad \text{or} \quad \mu > \mu_0$$

Under null hypothesis

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1}$$

1. $H_1: \mu \neq \mu_0$ the critical region is $t < -t_{n-1,a/2}$, or $t > t_{n-1,a/2}$
2. $H_1: \mu > \mu_0$ the critical region is $t > t_{n-1,a}$
3. $H_1: \mu < \mu_0$ the critical region is $t < -t_{n-1,a}$

Test on the difference $\mu_1 - \mu_2$ of the means of two independent populations

with known variances

$H_0: \ \mu_1 - \mu_2 = \delta$ vs $H_1: \mu_1 - \mu_2 \neq \delta$ or $\mu_1 - \mu_2 < \delta$ or $\mu_1 - \mu_2 > \delta$

$$Z = \frac{\bar{X} - \bar{Y} - \delta}{\sqrt{\sigma_1^2/n + \sigma_2^2/m}}$$

1. $H_1 : \mu_1 - \mu_2 \neq \delta$ the critical region is $z < -z_{a/2}$ or $z > z_{a/2}$
2. $H_1 : \mu_1 - \mu_2 > \delta$ the critical region is $z > z_a$
3. $H_1 : \mu_1 - \mu_2 < \delta$ the critical region is $z < -z_a$

Test on the difference $\mu_1 - \mu_2$ of the means of two independent populations

with unknown variances but equal

$$T = \frac{\bar{X} - \bar{Y} - \delta}{S_p \sqrt{1/n + 1/m}}$$

where $s_p^2 = \frac{(n-1)s_1^2 + (m-1)s_2^2}{n+m-2}$

1. $H_1 : \mu_1 - \mu_2 \neq \delta$ the critical region is $t < -t_{n+m-2,a/2}$ or $t > t_{n+m-2,a/2}$
2. $H_1 : \mu_1 - \mu_2 > \delta$ the critical region is $t > t_{n+m-2,a}$
3. $H_1 : \mu_1 - \mu_2 < \delta$ the critical region is $t < -t_{n+m-2,a}$

Test on the difference $\mu_1 - \mu_2$ of the means of two independent populations

with unknown variances and not equal

$$T = \frac{\bar{X} - \bar{Y} - \delta}{\sqrt{S_1^2/n + S_2^2/m}}$$

1. $H_1 : \mu_1 - \mu_2 \neq \delta$ the critical region is $t < -t_{v,a/2}$ or $t > t_{v,a/2}$
2. $H_1 : \mu_1 - \mu_2 > \delta$ the critical region is $t > t_{v,a}$
3. $H_1 : \mu_1 - \mu_2 < \delta$ the critical region is $t < -t_{v,a}$

where $v = \frac{\left(\frac{s_1^2}{n} + \frac{s_2^2}{m}\right)^2}{\frac{\left(\frac{s_1^2}{n}\right)^2}{n-1} + \frac{\left(\frac{s_2^2}{m}\right)^2}{m-1}}$

Test on the difference $\mu_1 - \mu_2$ of the means of two dependent populations

$$H_0: \ \mu_1 - \mu_2 = \delta \quad \text{vs} \quad H_1: \mu_1 - \mu_2 \neq \delta \quad \text{or} \quad \mu_1 - \mu_2 < \delta \quad \text{or} \quad \mu_1 - \mu_2 > \delta$$

$d_i = x_i - y_i, \quad i = 1, \ldots, n.$

$$H_0: \ \mu_d = \delta \quad \text{vs} \quad H_1: \mu_d \neq \delta \quad \text{or} \quad \mu_d < \delta \quad \text{or} \quad \mu_d > \delta$$

$$t = \frac{\bar{d} - \delta}{s_d/\sqrt{n}}$$

1. $H_1: \mu_d \neq \delta$ the critical region is $t > |t_{n-1,a/2}|$
2. $H_1: \mu_d > \delta$ the critical region is $t > t_{n-1,a}$
3. $H_1: \mu_d < \delta$ the critical region is $t < -t_{n-1,a}$

Exercises

1. In an experiment for the comparison of the thermal capacity of the carbon extracted from two mines, five measurements were taken from each mine. The measurements in million calories per tonne are given in the following table.

Mine 1	8240	8110	8330	8050	8320
Mine 2	7980	7920	7930	8170	7950

   Is there considerable evidence at 1% significance level that the mean thermal capacity of the carbon is different in two mines?

2. In an experiment for the comparison of two teaching methods for mathematics in the first year in the Polytechnic school eight pairs of twins were selected. In each couple applied method A to a child and method B to the other child.The following table shows the grades achieved by children after a fixed period teaching.

Method A	77	74	82	73	87	69	66	80
Method B	72	68	76	68	84	68	61	76

   Is there significant evidence at 10% significance level that the method A is more efficient than the method B?

Exercise 1

	Thermal	Mine
1	8240,00	1,00
2	8110,00	1,00
3	8330,00	1,00
4	8050,00	1,00
5	8320,00	1,00
6	7980,00	2,00
7	7920,00	2,00
8	7930,00	2,00
9	8170,00	2,00
10	7950,00	2,00

$$H_0 : \mu_1 = \mu_2 \quad vs \quad H_1 : \mu_1 \neq \mu_2$$

`Analyze > Compare Means > Independent-Samples T Test...`

Exercise 1

T-Test

Group Statistics

	Mine	N	Mean	Std. Deviation	Std. Error Mean
Thermal	Mine A	5	8210,0000	125,49900	56,12486
	Mine B	5	7990,0000	103,19884	46,15192

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Thermal	Equal variances assumed	,800	,397	3,028	8	,016	220,00000	72,66361	52,43742	387,56258
	Equal variances not assumed			3,028	7,712	,017	220,00000	72,66361	51,34268	388,65732

Levene's Test $(H_0 : \sigma_1 = \sigma_2 \quad vs \quad H_1 : \sigma_1 \neq \sigma_2)$
$p - value = 0.397 > a$, thus, equal variances assumed

p-value (initial test) $=0.016 > 0.01$ thus, we cannot reject the null hypothesis at 1% significance level.

Exercise 1

A	B	diff
77,00	72,00	5,00
74,00	68,00	6,00
82,00	76,00	6,00
73,00	68,00	5,00
87,00	84,00	3,00
69,00	68,00	1,00
66,00	61,00	5,00
80,00	76,00	4,00

$$H_0 : \mu_A = \mu_B \quad vs \quad H_1 : \mu_A \neq \mu_B$$

`Analyze > Compare Means > Paired-Samples T Test...`

Exercise 2

T-Test

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	A	76,0000	8	6,92820	2,44949
	B	71,6250	8	7,00892	2,47803

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	A & B	8	,971	,000

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	A - B	4,37500	1,68502	,59574	2,96629	5,78371	7,344	7	,000

$p - value < 0.001$ thus we reject the null hypothesis

Hypothesis test for the proportion p of a population

$$H_0 : p = p_0 \quad \text{vs} \quad H_1 : p \neq p_0 \quad \text{or} \quad p < p_0 \quad \text{or} \quad p > p_0$$

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

1. $H_1 : p \neq p_0$ the critical region is $z > |z_{a/2}|$
2. $H_1 : p > p_0$ the critical region is $z > z_a$
3. $H_1 : p < p_0$ the critical region is $z < -z_a$

Hypothesis Test for the difference $p_1 - p_2$ of the proportions of two independent populations

$$H_0: \; p_1 - p_2 = \delta \quad \text{vs} \quad H_1: p_1 - p_2 \neq \delta \quad \text{or} \quad p_1 - p_2 < \delta \quad \text{or} \quad p_1 - p_2 > \delta$$

if $\delta \neq 0$

$$z = \frac{\hat{p}_1 - \hat{p}_2 - \delta}{\sqrt{\hat{p}_1(1-\hat{p}_1)/n + \hat{p}_2(1-\hat{p}_2)/m}}$$

if $\delta = 0$

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(1/n + 1/m)}}$$

where $\hat{p} = \frac{x+y}{n+m}$

1. $H_1: p_1 - p_2 \neq \delta$ the critical region is $z > |z_{a/2}|$
2. $H_1: p_1 - p_2 > \delta$ the critical region is $z > z_a$
3. $H_1: p_1 - p_2 < \delta$ the critical region is $z < -z_a$

Exercise

We toss a coin 100 times and we get "heads" 45 times.

1. Can we claim that the coin is not fair at 5% significance level ?
2. Calculate the range of times that you can get "'heads"' in 100, so that the hypothesis that the coin is not fair can not be rejected at 5% significance level.

Hypothesis test for the variance σ^2 of a population

$$H_0 : \sigma^2 = \sigma_0^2 \quad \text{vs} \quad H_1 : \sigma^2 \neq \sigma_0^2 \quad \text{or} \quad \sigma^2 < \sigma_0^2 \quad \text{or} \quad \sigma^2 > \sigma_0^2$$

$$X^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

$(n-1)s^2 = \sum (x_i - \bar{X})^2 = \sum x_i^2 - n\bar{X}^2$

1. $H_1 : \sigma^2 \neq \sigma_0^2$ the critical region is

$$X^2 < X^2_{n-1,1-a/2}, \text{ or } X^2 > X^2_{n-1,a/2}$$

2. $H_1 : \sigma^2 > \sigma_0^2$ the critical region is $X^2 > X^2_{n-1,a}$
3. $H_1 : \sigma^2 < \sigma_0^2$ the critical region is $X^2 < X^2_{n-1,1-a}$

Hypothesis test for the ratio σ_2^2/σ_1^2 of the variances of two independent populations

$$H_0 : \sigma_2^2/\sigma_1^2 = 1 \quad \text{vs} \quad H_1 : \sigma_2^2/\sigma_1^2 \neq 1 \quad \text{or} \quad \sigma_2^2/\sigma_1^2 < 1 \quad \text{or} \quad \sigma_2^2/\sigma_1^2 > 1$$

$$F = \frac{s_1^2}{s_2^2}$$

1. $H_1 : \sigma_2^2/\sigma_1^2 \neq 1$ the critical region is

$$F < F_{n-1,m-1,1-a/2} \text{ or } F > F_{n-1,m-1,a/2}$$

2. $H_1 : \sigma_2^2/\sigma_1^2 > 1$ the critical region is $F < F_{n-1,m-1,1-a}$
3. $H_1 : \sigma_2^2/\sigma_1^2 < 1$ the critical region is $F > F_{n-1,m-1,a}$

$$H_0 : \sigma_2^2/\sigma_1^2 = \lambda_0 \quad \text{vs} \quad H_1 : \sigma_2^2/\sigma_1^2 \neq \lambda_0 \quad \text{or} \quad \sigma_2^2/\sigma_1^2 < \lambda_0 \quad \text{or} \quad \sigma_2^2/\sigma_1^2 > \lambda_0$$

$$F = \frac{s_1^2}{s_2^2} \cdot \lambda_0$$

Exercises

1. From a new method for determining the metal melting arose the following measurements for manganese:

   $$1267,\ 1262,\ 1267,\ 1263,\ 1258,\ 1263,\ 1268.$$

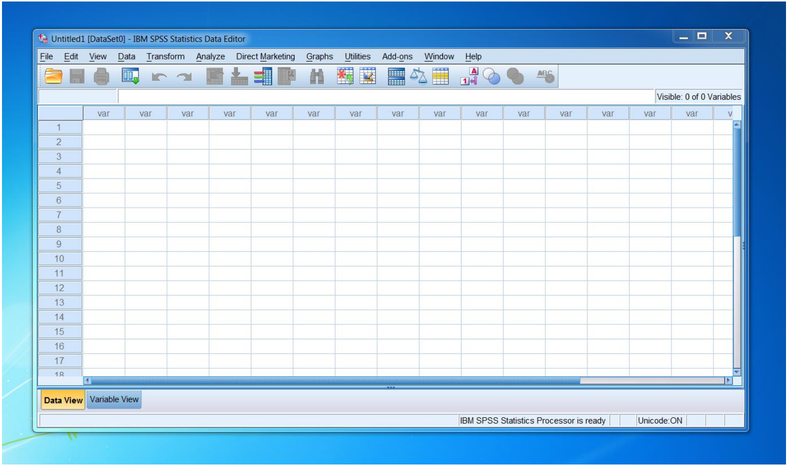
   Test if the new method is safe at 5% significance level, since the melting point of manganese is known and is equal to $1260^o C$.

2. The following data are the fracture loads (in tn/cm^2) of two different types of composites used for aircraft construction.

Type I	1.2	0.3	0.8	0.5	0.4	0.9	1.0
Type II	1.4	1.5	1.1	1.0	0.8	1.7	0.9

   Can we consider that the two types of composites have the same mean fracture load at 5% significance level?

SPSS for Windows



Reading

- *Biostatistics*, A Foundation for Analysis in the Health Sciences, W.W. Daniel and C.L. Cross

  http://informatika.uvlf.sk/subory/prezentacie%20zas/book%201.pdf
- *Probability & Statistics for Engineers and Scientists* R.E.Walpole, R.H. Myers, S.L.Myers, K.Ye

  http://folk.ntnu.no/jenswerg/40CEFd01.pdf