People's Democratic Republic of Algeria
University Med Khider of Biskra
Faculty of SNVSTU

Hypothesis Testing

Dr. Ben Gherbal Hanane

Lecture: Biostatistics — Level: M1
Email: hanane.bengherbal@univ-biskra.dz

Let H_0 be a hypothesis concerning a population. Based on the results of samples drawn from this population, we are led to accept or reject the hypothesis H_0. The decision rules are called statistical tests. The principle of hypothesis testing is to set a working hypothesis and to predict the consequences of this hypothesis for the population or the sample. We compare these predictions with the observations and conclude by accepting or rejecting the working hypothesis according to objective decision rules. We distinguish two types of hypotheses, only one of which is true.

H_0 denotes the so-called null hypothesis and H_1 the so-called alternative hypothesis. H_1 is the negation of H_0; it is equivalent to saying "H_0 is false". The decision to reject H_0 means that H_1 holds.

Risk of a test: The random nature of the sample (observations) can distort the final decision (to reject or not reject the hypothesis H_0). Indeed, when we reject H_0 while H_0 is true, or accept H_1 while it is false, then we have a probability α (generally fixed at $\alpha = 0.05$, 0.01 or 0.1) of making an error; α is called the type I error:

$$\alpha = P(\text{reject } H_0 \,|\, H_0 \text{ true}).$$

Another situation in which we can make a decision error is when we accept H_0 while H_0 is false. In this case, we have a probability β of making an error; this is called the type II error:

$$\beta = P(\text{accept } H_0 \,|\, H_1 \text{ true}).$$

Therefore,

$$1-\alpha = P(\text{accept } H_0 \,|\, H_0 \text{ true}) \quad \text{and} \quad 1-\beta = P(\text{reject } H_0 \,|\, H_1 \text{ true}).$$

α: significance level of the test; $1-\alpha$: its confidence level; and $1-\beta$: the power of the test.

Types of tests:

$$\text{Two-sided test: } H_0 : \theta = \theta_0 \text{ against } H_1 : \theta \neq \theta_0.$$

$$\text{One-sided test: } \begin{cases} H_0 : \theta = \theta_0 \text{ against } H_1 : \theta < \theta_0 \text{ (left one-sided test)}, \\ H_0 : \theta = \theta_0 \text{ against } H_1 : \theta > \theta_0 \text{ (right one-sided test)}. \end{cases}$$

1. Goodness-of-Fit Tests (Comparing an observed mean with a theoretical mean)

Goodness-of-fit tests are intended to check whether a sample can be considered as drawn from a given population or representative of that population, with respect to a parameter such as the mean, variance, or observed frequency. In this part, we address the mean parameter.

Let X be a random variable observed on a population and following a normal distribution, and let a sample be drawn from this population. We test:

$$\textbf{Hypotheses:} \quad H_0 : \mu = \mu_0 \text{ against } H_1 : \mu \neq \mu_0 \quad (\text{other cases: } \mu < \mu_0 \text{ or } \mu > \mu_0).$$

To test this hypothesis, there are two statistics: the population variance σ^2 is known (Z test), or the variance is unknown and must be estimated (T test). The table below reports a summary of hypothesis tests for a single-population mean in the different situations (two-sided and one-sided (right and left)):

	σ **known,** $X \sim \mathcal{N}(\mu, \sigma)$	σ **unknown,** $n < 30$	σ **unknown,** $n > 30$
$H_0 : \mu = \mu_0$	$z = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \to \mathcal{N}(0,1)$	$t = \dfrac{\overline{X} - \mu_0}{S/\sqrt{n}}$ (Student)	$z = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \to \mathcal{N}(0,1)$
$H_1 : \mu < \mu_0$	Reject H_0 if $z < -z_\alpha$	Reject H_0 if $t < -t_{\alpha,n-1}$; check -2α	Reject H_0 if $z < -z_\alpha$
$H_1 : \mu > \mu_0$	Reject H_0 if $z > z_\alpha$	Reject H_0 if $t > t_{\alpha,n-1}$; check 2α	Reject H_0 if $z > z_\alpha$
$H_1 : \mu \neq \mu_0$	Reject H_0 if $z \notin (-z_{\alpha/2}, z_{\alpha/2})$; check $1 - \alpha/2$	Reject H_0 if $t \notin (-t_{\alpha/2,n-1}, t_{\alpha/2,n-1})$; check α	Reject H_0 if $z \notin (-z_{\alpha/2}, z_{\alpha/2})$

Table 1: Transposed rejection criteria for hypothesis tests of a population mean.

Example 1. *Blood glucose in a population follows a normal distribution with mean $\mu_0 = 1$ g/l and standard deviation $\sigma_0 = 0.1$ g/l. We record glucose levels for 9 patients and obtain $\overline{X} = 1.12$ g/l. Is this sample representative of the population?*

Solution 1. *Since the population standard deviation σ_0 is known, we can use a Z test to determine whether the sample is representative of the population. The hypotheses are:*

$$H_0 : \mu = 1 \quad \textit{against} \quad H_1 : \mu \neq 1.$$

The Z test statistic is computed as follows:

$$z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{1.12 - 1}{0.1/\sqrt{9}} = 3.6.$$

If α is not specified, we choose $\alpha = 0.05$, hence $z_{\alpha/2} = 1.96$ (since we look up $1 - \alpha/2 = 0.975$). As $z > z_{\alpha/2}$, i.e., $z \notin (-z_{\alpha/2}, z_{\alpha/2})$, we reject H_0 at the error risk $\alpha = 0.05$. Therefore, this sample is not representative of the reference population with mean $\mu = 1$ g/l.

Example 2. *To study a batch of pills, we randomly sample 10 pills and weigh them. The observed weights (in grams) are:*

0,81	0,84	0,83	0,80	0,85	0,86	0,85	0,83	0,84	0,80

Is the observed mean compatible with 0,83 g at error risk 0.02?

Solution 2. *To determine whether the observed mean is compatible with 0,83 g ($\alpha = 0.02$), we carry out a T test. Steps and calculations:*

Hypotheses: $H_0 : \mu = 0{,}83$ g *against* $H_1 : \mu \neq 0{,}83$ g.

Sample size: $n = 10$.

Sample mean: $\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i = \dfrac{1}{10}\sum_{i=1}^{10} X_i = 0.831.$

Estimator of the standard deviation:

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2} = 0.2131.$$

Test statistic: $t = \dfrac{\overline{X} - \mu_0}{S/\sqrt{n}} = 0.1483.$

Degrees of freedom (df): $df = n - 1 = 10 - 1 = 9.$

Critical value: *from Student's t-table,* $t_{(\alpha, n-1)} = t_{(0.02, 9)} = 2.8214.$

As $t \in (-t_{(\alpha,n-1)}, t_{(\alpha,n-1)})$*, the hypothesis* H_0 *is accepted. The observed mean is compatible with* 0,83 *g at* $\alpha = 0.02$*.*

2. Homogeneity Test

Homogeneity tests are intended to compare two populations using an equivalent number of samples. In this case, the theoretical law of the parameter under study (for example (μ, σ^2)) is unknown at the population level. In this part of the course, we test equality between two variances and two means.

2.1. Comparing two variances

Let X be a random variable observed on two populations following normal laws $\mathcal{N}(\mu_1, \sigma_1^2)$ and $\mathcal{N}(\mu_2, \sigma_2^2)$, and let two independent samples be drawn from these two populations. We assume that the two samples come from two populations whose variances are equal:

$$\textbf{Hypotheses:} \quad H_0 : \sigma_1^2 = \sigma_2^2 \text{ against } H_1 : \sigma_1^2 \neq \sigma_2^2.$$

The statistic

$$F = \frac{S_1^2}{S_2^2} = \frac{\frac{n_1}{n_1-1} S_{e1}^2}{\frac{n_2}{n_2-1} S_{e2}^2}$$

follows a Fisher–Snedecor law with $(n_1 - 1, n_2 - 1)$ degrees of freedom, where n_1 is the size of sample 1 and n_2 the size of sample 2.

Remarks:

- The larger of the two estimated variances must be in the numerator and the smaller in the denominator ($S_1^2 > S_2^2$ by convention).
- To apply this test, the random variable X must follow $\mathcal{N}(\mu, \sigma^2)$, and the two samples must be independent.

Application and decision: The value of the statistic F is compared with the value F_α read from the Fisher–Snedecor table for a fixed two-sided error risk $\alpha/2$ and $(n_1 - 1, n_2 - 1)$ degrees of freedom.

- If $F \geq F_\alpha$, the hypothesis H_0 is rejected at risk α: the two samples come from populations with statistically different variances.
- If $F < F_\alpha$, the hypothesis H_0 is accepted: the two samples come from populations with the same variance.

Example 3 (Calcium and blood pressure). *Data collected in a study on calcium supplements and their effects on blood pressure. A placebo group and a calcium group began the study with a blood pressure measurement. Results:*

	Size	*Std. dev.*
placebo	$n_1 = 13$	$S_1 = 9.46$
calcium	$n_2 = 15$	$S_2 = 8.469$

Test the claim that the two samples come from populations with the same standard deviation ($\alpha = 0.05$)*.*

Solution 3. *The samples come from normal populations:*

$$\textbf{\textit{Hypotheses:}} \quad H_0 : \sigma_1^2 = \sigma_2^2 \text{ against } H_1 : \sigma_1^2 \neq \sigma_2^2.$$

We have

$$F = \frac{S_1^2}{S_2^2} = \frac{9.46^2}{8.469^2} = 1.248.$$

This is a two-sided test with area 0.025 *(*$0.05/2$*). Using the Fisher table with* 0.025, $df_1 = 12$, $df_2 = 14$, $F(12, 14) = 3.0502$.

Conclusion: $F = 1.248 < 3.0502$, *hence we accept* H_0.

Remark 1. *For "*$H_0 : \sigma_1^2 = \sigma_2^2$ *against* $H_1 : \sigma_1^2 > \sigma_2^2$*", use the value* F_α *from the Fisher–Snedecor table for error risk* α *and* (n_1-1, n_2-1) *degrees of freedom. For "*$H_0 : \sigma_1^2 = \sigma_2^2$ *against* $H_1 : \sigma_1^2 < \sigma_2^2$*", use* F_α *with* $(n_2 - 1, n_1 - 1)$ *degrees of freedom, with* $F = \dfrac{S_2^2}{S_1^2}$.

2.2. Comparing two means

In this section, we are interested in the homogeneity of two populations with respect to the mean. We focus on the case where the two samples are large and may come from any distribution, and on the case where the two samples are Gaussian and of small size.

Hypothesis / Test criterion	$n_1, n_2 > 30$ **(Large samples)**	$n_1, n_2 < 30$ **(Small samples)**
$H_0 : \mu_1 = \mu_2$ **(Test statistic)**	$u = \dfrac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$, $\quad u = \dfrac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim \mathcal{N}(0,1)$	Under H_0 and $\sigma_1 = \sigma_2 = \sigma$: $t = \dfrac{\overline{X}_1 - \overline{X}_2}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$
$H_1 : \mu_1 < \mu_2$	Reject H_0 if $u \leq -u_\alpha$; check $1 - \alpha$	Reject H_0 if $t \leq -t_\alpha$
$H_1 : \mu_1 > \mu_2$	Reject H_0 if $u \geq u_\alpha$; check $1 - \alpha$	Reject H_0 if $t \geq t_\alpha$
$H_1 : \mu_1 \neq \mu_2$ **(Two-sided test)**	Reject H_0 if $u \notin (-u_{\alpha/2}, u_{\alpha/2})$; check $1 - \alpha/2$	Reject H_0 if $t \notin (-t_{\alpha/2}, t_{\alpha/2})$

Table 2: Rejection criteria for hypothesis tests comparing two population means.

S is the estimator of σ:

$$S = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}.$$

Example 4. *The lifetime of electronic components from factory A follows a normal distribution* $\mathcal{N}(\mu_1, \sigma_1 = 30)$ *and those from factory B follow* $\mathcal{N}(\mu_2, \sigma_2 = 32)$. *The lifetime of a sample of size* $n_1 = 150$ *from factory A has mean* $\overline{X}_1 = 1852$ *hours, and that of a sample of size* $n_2 = 90$ *from factory B has mean* $\overline{X}_2 = 1840$. *At the* 5% *significance level, can we conclude that there is a significant difference between the lifetimes of components from factories A and B?*

Solution 4. *We want to test whether the two populations have equal means.*

Hypotheses:

$$H_0 : \mu_1 = \mu_2 \quad vs \quad H_1 : \mu_1 \neq \mu_2, \quad \alpha = 0.05.$$

$\overline{X}_1 = 1852,\ n_1 = 150,\ \sigma_1 = 30;\ \ \overline{X}_2 = 1840,\ n_2 = 90,\ \sigma_2 = 32.$

$$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{30^2}{150} + \frac{32^2}{90}} \approx 4.1687.$$

Test statistic:

$$z = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{1852 - 1840}{4.1687} \approx 2.879.$$

Critical value: $z_{0.025} = 1.96$. *Since* $|z| = 2.879 > 1.96$, *we reject* H_0.

Conclusion: *The difference between the lifetimes of components from factories A and B is statistically significant at the* 5% *level.*

Example 5. *The weight of a medicine packaged in boxes is distributed as* $\mathcal{N}(\mu, \sigma)$. *Two samples of respective sizes* $n_1 = 12$ *and* $n_2 = 18$ *have means* $\overline{X}_1 = 22.235$ *and* $\overline{X}_2 = 21.988$, *and standard deviations* $S_1 = 0.18$ *and* $S_2 = 0.23$, *respectively. Can we conclude that these two samples come from the same population at the* 5% *significance level?*

Solution 5. *We want to test whether the two samples come from populations with the same mean.*

Hypotheses:

$$H_0 : \mu_1 = \mu_2 \quad vs \quad H_1 : \mu_1 \neq \mu_2, \quad \alpha = 0.05.$$

$\overline{X}_1 = 22.235,\ n_1 = 12,\ S_1 = 0.18;\ \ \overline{X}_2 = 21.988,\ n_2 = 18,\ S_2 = 0.23.$

$$S = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{11(0.18)^2 + 17(0.23)^2}{28}} \approx 0.212.$$

$$S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 0.212\sqrt{\frac{1}{12} + \frac{1}{18}} \approx 0.079.$$

Test statistic:

$$t = \frac{\overline{X}_1 - \overline{X}_2}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{22.235 - 21.988}{0.079} \approx 3.13, \quad df = n_1 + n_2 - 2 = 28.$$

Critical value: $t_{0.025,28} \approx 2.048$. *Since* $|t| = 3.13 > 2.048$, *we reject* H_0.

Conclusion: *The two samples do not come from the same population at the* 5% *level.*

Additional Examples

Example 6 (Blood pressure treatment). *A new drug is intended to reduce systolic blood pressure. In a clinical trial, the mean pressure before treatment is known to be* 150 *mmHg. A sample of* 20 *patients treated with the new drug shows an average of* $\overline{X} = 145$ *mmHg and a sample standard deviation* $S = 10$ *mmHg. At the* 5% *significance level, can we conclude that the drug effectively lowers blood pressure?*

Example 7 (Effect of fertilizer on plant height). *Two groups of tomato plants are cultivated under identical conditions except for the fertilizer used. Fertilizer A:* $n_1 = 16$ *plants,* $\overline{X}_1 = 42.3$ *cm,* $S_1 = 3.2$ *cm. Fertilizer B:* $n_2 = 14$ *plants,* $\overline{X}_2 = 39.8$ *cm,* $S_2 = 2.8$ *cm. At* $\alpha = 0.05$, *test whether the fertilizers lead to significantly different mean plant heights.*

Example 8 (Bacterial growth rate). *The normal growth rate of a bacterial strain in culture is* $\mu_0 = 0.54$ *(optical density/hour). After adding a new nutrient, ten cultures are tested and give an average rate of* $\overline{X} = 0.57$ *with* $S = 0.03$*. At* $\alpha = 0.01$*, can we conclude that the nutrient increases the mean growth rate?*

Example 9 (Protein concentration in serum). *A diagnostic test claims that the mean serum protein concentration in healthy adults is* 7.1 *g/dL. In a sample of* 30 *patients, the measured mean is* 6.95 *g/dL with* $\sigma = 0.4$ *g/dL. At the* 1% *level, test whether this sample can be considered representative of the healthy population.*

Example 10 (Comparison of enzyme activity). *Two enzyme extraction methods are tested for their catalytic activity (measured in µmol/min). Method 1:* $n_1 = 12$, $\overline{X}_1 = 82.6$, $S_1 = 6.4$*; Method 2:* $n_2 = 10$, $\overline{X}_2 = 78.2$, $S_2 = 5.9$*. At* $\alpha = 0.05$*, can we conclude that the two methods yield the same mean activity?*

Example 11 (Variance of seed germination times). *Two seed varieties are tested for uniformity in germination time (days). Variety A:* $n_1 = 15$, $S_1 = 1.9$*; Variety B:* $n_2 = 12$, $S_2 = 1.1$*. At* $\alpha = 0.05$*, test whether the variability in germination time differs between the two varieties.*

Example 12 (Paired data – before/after training). *Ten biology students measure their reaction time before and after a cognitive training program. Before: 0.51, 0.49, 0.54, 0.50, 0.53, 0.47, 0.52, 0.49, 0.50, 0.51 s. After: 0.47, 0.44, 0.50, 0.47, 0.49, 0.43, 0.48, 0.45, 0.46, 0.47 s. At* $\alpha = 0.05$*, test whether the training significantly improves mean reaction time.*

Example 13 (Effect of temperature on enzyme activity). *An enzyme was tested at two different temperatures (25°C and 37°C) to measure its reaction rate (µmol/min). At 25°C:* $n_1 = 8$, $\overline{X}_1 = 52.4$, $S_1 = 4.1$*. At 37°C:* $n_2 = 8$, $\overline{X}_2 = 59.3$, $S_2 = 3.8$*. At* $\alpha = 0.05$*, test whether temperature has a significant effect on the enzyme's mean activity.*

Example 14 (Cholesterol levels before and after treatment). *A group of* 12 *patients follows a new diet intended to lower cholesterol. Cholesterol (mg/dL) before treatment: 240, 235, 260, 245, 238, 250, 255, 248, 243, 230, 240, 247. After six weeks: 225, 220, 240, 230, 225, 235, 232, 228, 224, 218, 222, 229. At the* 5% *significance level, test whether the diet significantly reduces mean cholesterol.*

Example 15 (Genetic diversity in two populations). *Two populations of a plant species were analyzed for genetic diversity (Shannon index). Population 1:* $n_1 = 20$, $\overline{X}_1 = 1.85$, $S_1 = 0.18$*. Population 2:* $n_2 = 18$, $\overline{X}_2 = 1.71$, $S_2 = 0.15$*. At* $\alpha = 0.01$*, can we conclude that the two populations have different levels of genetic diversity?*