Estadística II
Tema 3. Comparison of two populations

Academic year 2010/11

Chapter 3. Comparison of two populations

Contents

- Comparison of two populations: examples, matched data for experimental reduction of the variability.
- Independent samples:
  - Comparison of the means, equal variances, normal populations.
  - Comparison of the variances in normal populations.
  - Sensitivity of the previous tests.
  - Comparison of the means, large samples.
  - Comparison of proportions, large samples.
- Matched samples, comparison of the means, normal differences.

Chapter 3. Comparison of two populations

Learning objectives

- Know to distinguish when independent or dependent matched samples are being used. Know when is convenient to work with matched samples.
- Know to perform the appropriate hypothesis testing in order to validate or not the specific comparison.
- Know to build the suitable decision rule depending on the test and the case we deal with (assumed hypotheses).
- Know what are the consequences on the conclusions when any assumption is violated.

Chapter 3. Comparison of two populations

Recommended reading

- Meyer, P. "Probabilidad y aplicaciones estadísticas" (1992)
  - Chapter ¿?
- Newbold, P. "Estadística para los negocios y la economía" (1997)
  - Chapter 9 (9.6, 9.7, 9.8)
- Peña, D. "Fundamentos de Estadística" (2001)
  - Chapter 10 (10.5)

Examples

1. A researcher wants to know whether a tax proposal is supported by men and women in the same way.

$$H_0 : p_H = p_M$$
$$H_1 : p_H \neq p_M$$

$p_H =$ men proportion supporting the tax proposal

$p_M =$ women proportion supporting the tax proposal

Effect of social level, education, income level, politic trend:

randomize

Examples

1. A researcher wants to know whether a tax proposal is supported by men and women in the same way.

$$H_0 : p_H = p_M$$
$$H_1 : p_H \neq p_M$$

$$p_H = \text{men proportion supporting the tax proposal}$$
$$p_M = \text{women proportion supporting the tax proposal}$$

Effect of social level, education, income level, politic trend:

randomize

Examples

1. A researcher wants to know whether a tax proposal is supported by men and women in the same way.

$$H_0 : p_H = p_M$$
$$H_1 : p_H \neq p_M$$

$$p_H = \text{men proportion supporting the tax proposal}$$
$$p_M = \text{women proportion supporting the tax proposal}$$

Effect of social level, education, income level, politic trend:

randomize

Examples

1. A researcher wants to know whether a tax proposal is supported by men and women in the same way.

$$H_0 : p_H = p_M$$
$$H_1 : p_H \neq p_M$$

$$p_H = \text{men proportion supporting the tax proposal}$$
$$p_M = \text{women proportion supporting the tax proposal}$$

Effect of social level, education, income level, politic trend:

randomize

Examples

2. For a study it is desired to compare federal and state credit entities in terms of the ratio between the total debts of the entity and its assets.

$$H_0 : \mu_X = \mu_Y$$
$$H_1 : \mu_X \neq \mu_Y$$

$$X = \frac{\text{debts}}{\text{assets}} \text{ for federal entities}$$
$$Y = \frac{\text{debts}}{\text{assets}} \text{ for state entities}$$

Effect of the size and seniority: matched samples

Examples

2. For a study it is desired to compare federal and state credit entities in terms of the ratio between the total debts of the entity and its assets.

$$H_0 : \mu_X = \mu_Y$$
$$H_1 : \mu_X \neq \mu_Y$$

$$X = \frac{\text{debts}}{\text{assets}} \text{ for federal entities}$$
$$Y = \frac{\text{debts}}{\text{assets}} \text{ for state entities}$$

Effect of the size and seniority: matched samples

Examples

2. For a study it is desired to compare federal and state credit entities in terms of the ratio between the total debts of the entity and its assets.

$$H_0 : \mu_X = \mu_Y$$
$$H_1 : \mu_X \neq \mu_Y$$

$$X = \frac{\text{debts}}{\text{assets}} \text{ for federal entities}$$
$$Y = \frac{\text{debts}}{\text{assets}} \text{ for state entities}$$

Effect of the size and seniority: matched samples

Examples

2. For a study it is desired to compare federal and state credit entities in terms of the ratio between the total debts of the entity and its assets.

$$H_0 : \mu_X = \mu_Y$$
$$H_1 : \mu_X \neq \mu_Y$$

$$X = \frac{\text{debts}}{\text{assets}} \text{ for federal entities}$$
$$Y = \frac{\text{debts}}{\text{assets}} \text{ for state entities}$$

Effect of the size and seniority: matched samples

Examples

3. An investor wants to compare the risk associated with two different markets (A and B). Such a risk is measured with the daily fluctuation associated with the prices. In order to do that 21 data are obtained for market A and 16 for market B.

$$H_0 : \sigma_X^2 = \sigma_Y^2$$
$$H_1 : \sigma_X^2 \neq \sigma_Y^2$$

$X =$ daily fluctuation in market A

$Y =$ daily fluctuation in market B

Effect of day: randomize

Effect of the macroeconomic situation: same conditions

Examples

3. An investor wants to compare the risk associated with two different markets (A and B). Such a risk is measured with the daily fluctuation associated with the prices. In order to do that 21 data are obtained for market A and 16 for market B.

$$H_0 : \sigma_X^2 = \sigma_Y^2$$
$$H_1 : \sigma_X^2 \neq \sigma_Y^2$$

$$X = \text{ daily fluctuation in market A}$$
$$Y = \text{ daily fluctuation in market B}$$

Effect of day: randomize
Effect of the macroeconomic situation: same conditions

Examples

3. An investor wants to compare the risk associated with two different markets (A and B). Such a risk is measured with the daily fluctuation associated with the prices. In order to do that 21 data are obtained for market A and 16 for market B.

$$H_0 : \sigma_X^2 = \sigma_Y^2$$
$$H_1 : \sigma_X^2 \neq \sigma_Y^2$$

$$X = \text{ daily fluctuation in market A}$$
$$Y = \text{ daily fluctuation in market B}$$

Effect of day: randomize

Effect of the macroeconomic situation: same conditions

Examples

3. An investor wants to compare the risk associated with two different markets (A and B). Such a risk is measured with the daily fluctuation associated with the prices. In order to do that 21 data are obtained for market A and 16 for market B.

$$H_0 : \sigma_X^2 = \sigma_Y^2$$
$$H_1 : \sigma_X^2 \neq \sigma_Y^2$$

$$X = \text{ daily fluctuation in market A}$$
$$Y = \text{ daily fluctuation in market B}$$

Effect of day: randomize

Effect of the macroeconomic situation: same conditions

Examples

3. An investor wants to compare the risk associated with two different markets (A and B). Such a risk is measured with the daily fluctuation associated with the prices. In order to do that 21 data are obtained for market A and 16 for market B.

$$H_0 : \sigma_X^2 = \sigma_Y^2$$
$$H_1 : \sigma_X^2 \neq \sigma_Y^2$$

$$X = \text{ daily fluctuation in market A}$$
$$Y = \text{ daily fluctuation in market B}$$

Effect of day: randomize
Effect of the macroeconomic situation: same conditions

Examples

3. An investor wants to compare the risk associated with two different markets (A and B). Such a risk is measured with the daily fluctuation associated with the prices. In order to do that 21 data are obtained for market A and 16 for market B.

$$H_0 : \sigma_X^2 = \sigma_Y^2$$
$$H_1 : \sigma_X^2 \neq \sigma_Y^2$$

$$X = \text{ daily fluctuation in market A}$$
$$Y = \text{ daily fluctuation in market B}$$

Effect of day: randomize
Effect of the macroeconomic situation: same conditions

Examples

4. Before launching a very aggressive promotion of a product for stores, the marketing director of a company wants to know whether it is worth (whether the sales of this product are increased in this kind of shops). 50 stores are selected in Madrid to carry out this promotion and the data are collected before the promotion and thereafter.

$$H_0 : \mu_X \geq \mu_Y$$
$$H_1 : \mu_X < \mu_Y$$

$X =$ turnover in stores before the promotion

$Y =$ turnover in stores after the promotion

Effect "launch": matched samples

Effect "site": randomize

Examples

4. Before launching a very aggressive promotion of a product for stores, the marketing director of a company wants to know whether it is worth (whether the sales of this product are increased in this kind of shops). 50 stores are selected in Madrid to carry out this promotion and the data are collected before the promotion and thereafter.

$$H_0 : \mu_X \geq \mu_Y$$
$$H_1 : \mu_X < \mu_Y$$

$X =$ turnover in stores before the promotion

$Y =$ turnover in stores after the promotion

Effect "launch": matched samples

Effect "site": randomize

Examples

4. Before launching a very aggressive promotion of a product for stores, the marketing director of a company wants to know whether it is worth (whether the sales of this product are increased in this kind of shops). 50 stores are selected in Madrid to carry out this promotion and the data are collected before the promotion and thereafter.

$$H_0 : \mu_X \geq \mu_Y$$
$$H_1 : \mu_X < \mu_Y$$

$$X = \text{ turnover in stores before the promotion}$$
$$Y = \text{ turnover in stores after the promotion}$$

Effect "launch": matched samples
Effect "site": randomize

Examples

4. Before launching a very aggressive promotion of a product for stores, the marketing director of a company wants to know whether it is worth (whether the sales of this product are increased in this kind of shops). 50 stores are selected in Madrid to carry out this promotion and the data are collected before the promotion and thereafter.

$$H_0 : \mu_X \geq \mu_Y$$
$$H_1 : \mu_X < \mu_Y$$

$X =$ turnover in stores before the promotion

$Y =$ turnover in stores after the promotion

Effect "launch": matched samples

Effect "site": randomize

Examples

4. Before launching a very aggressive promotion of a product for stores, the marketing director of a company wants to know whether it is worth (whether the sales of this product are increased in this kind of shops). 50 stores are selected in Madrid to carry out this promotion and the data are collected before the promotion and thereafter.

$$H_0 : \mu_X \geq \mu_Y$$
$$H_1 : \mu_X < \mu_Y$$

$$X = \text{ turnover in stores before the promotion}$$
$$Y = \text{ turnover in stores after the promotion}$$

Effect "launch": matched samples
Effect "site": randomize

Examples

4. Before launching a very aggressive promotion of a product for stores, the marketing director of a company wants to know whether it is worth (whether the sales of this product are increased in this kind of shops). 50 stores are selected in Madrid to carry out this promotion and the data are collected before the promotion and thereafter.

$$H_0 : \mu_X \geq \mu_Y$$
$$H_1 : \mu_X < \mu_Y$$

$$X = \text{ turnover in stores before the promotion}$$
$$Y = \text{ turnover in stores after the promotion}$$

Effect "launch": matched samples
Effect "site": randomize

Examples

5. For an advertising campaign B it is checked whether the turnover increases. A sample of 10 cities with similar behaviour of consumers is considered such that in 5 of then the traditional campaign (campaign A) is followed and in the rest the new campaign, campaign B, is launched.

$$H_0 : \mu_A \geq \mu_B$$
$$H_1 : \mu_A < \mu_B$$

$X =$ turnover with the traditional campaign (A)

$Y =$ turnover with the new campaign (B)

Effect city:

randomize the choice of cities for each of the campaigns

Examples

5. For an advertising campaign B it is checked whether the turnover increases. A sample of 10 cities with similar behaviour of consumers is considered such that in 5 of then the traditional campaign (campaign A) is followed and in the rest the new campaign, campaign B, is launched.

$$H_0 : \mu_A \geq \mu_B$$
$$H_1 : \mu_A < \mu_B$$

$$X = \text{ turnover with the traditional campaign (A)}$$
$$Y = \text{ turnover with the new campaign (B)}$$

Effect city:

randomize the choice of cities for each of the campaigns

Examples

5. For an advertising campaign B it is checked whether the turnover increases. A sample of 10 cities with similar behaviour of consumers is considered such that in 5 of then the traditional campaign (campaign A) is followed and in the rest the new campaign, campaign B, is launched.

$$H_0 : \mu_A \geq \mu_B$$
$$H_1 : \mu_A < \mu_B$$

$$X = \text{ turnover with the traditional campaign (A)}$$
$$Y = \text{ turnover with the new campaign (B)}$$

Effect city:

randomize the choice of cities for each of the campaigns

Examples

5. For an advertising campaign B it is checked whether the turnover increases. A sample of 10 cities with similar behaviour of consumers is considered such that in 5 of then the traditional campaign (campaign A) is followed and in the rest the new campaign, campaign B, is launched.

$$H_0 : \mu_A \geq \mu_B$$
$$H_1 : \mu_A < \mu_B$$

$$X = \text{ turnover with the traditional campaign (A)}$$
$$Y = \text{ turnover with the new campaign (B)}$$

Effect city:

randomize the choice of cities for each of the campaigns

Independent Samples: Comparison of means, equal variances, normal populations

Aim: Given two normal populations with the same variability, such that their mean can be different, it is desired to test the hypothesis of equal means.

$$H_0 : \mu_X = \mu_Y$$
$$H_1 : \mu_X \neq \mu_Y$$

- Let $(X_1, \ldots, X_{n_1})$, $(Y_1, \ldots, Y_{n_2})$ be two s.r.s. of $X \sim N(\mu_X, \sigma^2)$ and $Y \sim N(\mu_Y, \sigma^2)$, respectively, mutually independent.
- Estimator of common variance σ^2:

$$s_P^2 = \frac{(n_1 - 1)s_X^2 + (n_2 - 1)s_Y^2}{n_1 + n_2 - 2}$$

Independent Samples: Comparison of means, equal variances, normal populations

Aim: Given two normal populations with the same variability, such that their mean can be different, it is desired to test the hypothesis of equal means.

$$H_0 : \mu_X = \mu_Y$$
$$H_1 : \mu_X \neq \mu_Y$$

- Let $(X_1, \ldots, X_{n_1})$, $(Y_1, \ldots, Y_{n_2})$ be two s.r.s. of $X \sim N(\mu_X, \sigma^2)$ and $Y \sim N(\mu_Y, \sigma^2)$, respectively, mutually independent.
- Estimator of common variance σ^2:

$$s_P^2 = \frac{(n_1 - 1)s_X^2 + (n_2 - 1)s_Y^2}{n_1 + n_2 - 2}$$

  - It is an unbiased estimator that uses available whole information.
  - It is a weighted estimator of two independent estimators s_X^2 and s_Y^2 with proportional weights with respect to the precision of each estimator.

Independent Samples: Comparison of means, equal variances, normal populations

Aim: Given two normal populations with the same variability, such that their mean can be different, it is desired to test the hypothesis of equal means.

$$H_0 : \mu_X = \mu_Y$$
$$H_1 : \mu_X \neq \mu_Y$$

- Let $(X_1, \ldots, X_{n_1})$, $(Y_1, \ldots, Y_{n_2})$ be two s.r.s. of $X \sim N(\mu_X, \sigma^2)$ and $Y \sim N(\mu_Y, \sigma^2)$, respectively, mutually independent.
- Estimator of common variance σ^2:

$$s_P^2 = \frac{(n_1 - 1)s_X^2 + (n_2 - 1)s_Y^2}{n_1 + n_2 - 2}$$

  - It is an unbiased estimator that uses available whole information.
  - It is a weighted estimator of two independent estimators s_X^2 and s_Y^2 with proportional weights with respect to the precision of each estimator.

Independent Samples: Comparison of means, equal variances, normal populations

Aim: Given two normal populations with the same variability, such that their mean can be different, it is desired to test the hypothesis of equal means.

$$H_0 : \mu_X = \mu_Y$$
$$H_1 : \mu_X \neq \mu_Y$$

- Let $(X_1, \ldots, X_{n_1})$, $(Y_1, \ldots, Y_{n_2})$ be two s.r.s. of $X \sim N(\mu_X, \sigma^2)$ and $Y \sim N(\mu_Y, \sigma^2)$, respectively, mutually independent.
- Estimator of common variance σ^2:

$$s_P^2 = \frac{(n_1 - 1)s_X^2 + (n_2 - 1)s_Y^2}{n_1 + n_2 - 2}$$

  - It is an unbiased estimator that uses available whole information.
  - It is a weighted estimator of two independent estimators s_X^2 and s_Y^2 with proportional weights with respect to the precision of each estimator.

Independent Samples: Comparison of means, equal variances, normal populations

- Basic results:
  - $\frac{(n_1-1)s_X^2}{\sigma^2} \sim \chi^2_{n_1-1}$, $\frac{(n_2-1)s_Y^2}{\sigma^2} \sim \chi^2_{n_2-1}$ mutually independent.
  - If H_0 is true, then $\overline{X} - \overline{Y} \sim N(0, \sigma^2(\frac{1}{n_1} + \frac{1}{n_2}))$
- Test statistic $T(X_1, \ldots, X_{n_1}; Y_1, \ldots, Y_{n_2})$:

$$\frac{\overline{X} - \overline{Y}}{s_P\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{\frac{\overline{X}-\overline{Y}}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}}{\sqrt{\frac{(n_1+n_2-2)s_P^2/\sigma^2}{n_1+n_2-2}}} =$$

$$= \frac{Z}{\sqrt{\chi^2_{n_1+n_2-2}/(n_1+n_2-2)}} \sim_{H_0} t_{n_1+n_2-2}$$

- Critical region

$$R_\alpha = \left\{ (x_1, \ldots, x_{n_1}; y_1, \ldots, y_{n_2}) \,/\, \left| \frac{\overline{X} - \overline{Y}}{s_P\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right| \geq t_{n_1+n_2-2;\, \frac{\alpha}{2}} \right\}$$

- Basic results:
  - $\frac{(n_1-1)s_X^2}{\sigma^2} \sim \chi^2_{n_1-1}$, $\frac{(n_2-1)s_Y^2}{\sigma^2} \sim \chi^2_{n_2-1}$ mutually independent.
  - If H_0 is true, then $\overline{X} - \overline{Y} \sim N(0, \sigma^2(\frac{1}{n_1} + \frac{1}{n_2}))$
- Test statistic $T(X_1, \ldots, X_{n_1}; Y_1, \ldots, Y_{n_2})$:

$$\frac{\overline{X} - \overline{Y}}{s_P\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{\frac{\overline{X}-\overline{Y}}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}}{\sqrt{\frac{(n_1+n_2-2)s_P^2/\sigma^2}{n_1+n_2-2}}} =$$

$$= \frac{Z}{\sqrt{\chi^2_{n_1+n_2-2}/(n_1+n_2-2)}} \sim_{H_0} t_{n_1+n_2-2}$$

- Critical region

$$R_\alpha = \left\{ (x_1, \ldots, x_{n_1}; y_1, \ldots, y_{n_2}) \,/\, \left| \frac{\overline{X} - \overline{Y}}{s_P\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right| \geq t_{n_1+n_2-2;\frac{\alpha}{2}} \right\}$$

Independent Samples: Comparison of means, equal variances, normal populations

- Basic results:
  - $\frac{(n_1-1)s_X^2}{\sigma^2} \sim \chi^2_{n_1-1}$, $\frac{(n_2-1)s_Y^2}{\sigma^2} \sim \chi^2_{n_2-1}$ mutually independent.
  - If H_0 is true, then $\overline{X} - \overline{Y} \sim N(0, \sigma^2(\frac{1}{n_1} + \frac{1}{n_2}))$
- Test statistic $T(X_1, \ldots, X_{n_1}; Y_1, \ldots, Y_{n_2})$:

$$\frac{\overline{X} - \overline{Y}}{s_P\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{\frac{\overline{X}-\overline{Y}}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}}{\sqrt{\frac{(n_1+n_2-2)s_P^2/\sigma^2}{n_1+n_2-2}}} =$$

$$= \frac{Z}{\sqrt{\chi^2_{n_1+n_2-2}/(n_1+n_2-2)}} \sim_{H_0} t_{n_1+n_2-2}$$

- Critical region

$$R_\alpha = \left\{ (x_1, \ldots, x_{n_1}; y_1, \ldots, y_{n_2}) \,/\, \left| \frac{\overline{X} - \overline{Y}}{s_P\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \right| \geq t_{n_1+n_2-2; \frac{\alpha}{2}} \right\}$$

Independent Samples: Comparison of means, equal variances, normal populations

- What if we want to perform one-sided testing?

$$\begin{array}{l} H_0 : \mu_X \leq \mu_Y \\ H_1 : \mu_X > \mu_Y \end{array} \quad R_\alpha = \left\{ (x_1, \ldots, x_{n_1}; y_1, \ldots, y_{n_2}) \,/\, \frac{\overline{X} - \overline{Y}}{s_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} > t_{n_1+n_2-2;\alpha} \right\}$$

$$\begin{array}{l} H_0 : \mu_X \geq \mu_Y \\ H_1 : \mu_X < \mu_Y \end{array} \quad R_\alpha = \left\{ (x_1, \ldots, x_{n_1}; y_1, \ldots, y_{n_2}) \,/\, \frac{\overline{X} - \overline{Y}}{s_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} < -t_{n_1+n_2-2;\alpha} \right\}$$

- What if we want to perform one-sided testing?

$$\begin{array}{l} H_0 : \mu_X \leq \mu_Y \\ H_1 : \mu_X > \mu_Y \end{array} \quad R_\alpha = \left\{ (x_1, \ldots, x_{n_1}; y_1, \ldots, y_{n_2}) \,/\, \frac{\overline{X} - \overline{Y}}{s_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} > t_{n_1+n_2-2;\alpha} \right\}$$

$$\begin{array}{l} H_0 : \mu_X \geq \mu_Y \\ H_1 : \mu_X < \mu_Y \end{array} \quad R_\alpha = \left\{ (x_1, \ldots, x_{n_1}; y_1, \ldots, y_{n_2}) \,/\, \frac{\overline{X} - \overline{Y}}{s_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} < -t_{n_1+n_2-2;\alpha} \right\}$$

Independent Samples: Comparison of means, equal variances, normal populations

- What if we want to perform one-sided testing?

$$H_0 : \mu_X \leq \mu_Y \qquad H_1 : \mu_X > \mu_Y \qquad R_\alpha = \left\{ (x_1, \ldots, x_{n_1}; y_1, \ldots, y_{n_2}) \,/\, \frac{\overline{X} - \overline{Y}}{s_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} > t_{n_1+n_2-2;\alpha} \right\}$$

$$H_0 : \mu_X \geq \mu_Y \qquad H_1 : \mu_X < \mu_Y \qquad R_\alpha = \left\{ (x_1, \ldots, x_{n_1}; y_1, \ldots, y_{n_2}) \,/\, \frac{\overline{X} - \overline{Y}}{s_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} < -t_{n_1+n_2-2;\alpha} \right\}$$

- What if we want to perform one-sided testing?

$$H_0 : \mu_X \leq \mu_Y \quad H_1 : \mu_X > \mu_Y \qquad R_\alpha = \left\{ (x_1, \ldots, x_{n_1}; y_1, \ldots, y_{n_2}) \,/\, \frac{\overline{X} - \overline{Y}}{s_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} > t_{n_1+n_2-2;\alpha} \right\}$$

$$H_0 : \mu_X \geq \mu_Y \quad H_1 : \mu_X < \mu_Y \qquad R_\alpha = \left\{ (x_1, \ldots, x_{n_1}; y_1, \ldots, y_{n_2}) \,/\, \frac{\overline{X} - \overline{Y}}{s_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} < -t_{n_1+n_2-2;\alpha} \right\}$$

- What if we want to test with a general difference $d_0 \geq 0$?

$H_0 : \mu_X - \mu_Y = d_0$ $H_1 : \mu_X - \mu_Y \neq d_0$	$H_0 : \mu_X - \mu_Y \leq d_0$ $H_1 : \mu_X - \mu_Y > d_0$	$H_0 : \mu_X - \mu_Y \geq d_0$ $H_1 : \mu_X - \mu_Y < d_0$
$T(X_1, \ldots, X_{n_1}; Y_1, \ldots, Y_{n_2}) = \frac{\overline{X} - \overline{Y} - d_0}{s_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim_{H_0} t_{n_1+n_2-2}$		

Independent Samples: Comparison of means, equal variances, normal populations

- What if we want to perform one-sided testing?

$$\begin{array}{l} H_0 : \mu_X \leq \mu_Y \\ H_1 : \mu_X > \mu_Y \end{array} \quad R_\alpha = \left\{ (x_1, \ldots, x_{n_1}; y_1, \ldots, y_{n_2}) \,/\, \frac{\overline{X} - \overline{Y}}{s_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} > t_{n_1+n_2-2;\alpha} \right\}$$

$$\begin{array}{l} H_0 : \mu_X \geq \mu_Y \\ H_1 : \mu_X < \mu_Y \end{array} \quad R_\alpha = \left\{ (x_1, \ldots, x_{n_1}; y_1, \ldots, y_{n_2}) \,/\, \frac{\overline{X} - \overline{Y}}{s_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} < -t_{n_1+n_2-2;\alpha} \right\}$$

- What if we want to test with a general difference $d_0 \geq 0$?

$H_0 : \mu_X - \mu_Y = d_0$ $H_1 : \mu_X - \mu_Y \neq d_0$	$H_0 : \mu_X - \mu_Y \leq d_0$ $H_1 : \mu_X - \mu_Y > d_0$	$H_0 : \mu_X - \mu_Y \geq d_0$ $H_1 : \mu_X - \mu_Y < d_0$

$$T(X_1, \ldots, X_{n_1}; Y_1, \ldots, Y_{n_2}) = \frac{\overline{X} - \overline{Y} - d_0}{s_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim_{H_0} t_{n_1+n_2-2}$$

Independent Samples: Comparison of means, equal variances, normal populations

- What if we want to perform one-sided testing?

$$\begin{array}{l} H_0 : \mu_X \leq \mu_Y \\ H_1 : \mu_X > \mu_Y \end{array} \quad R_\alpha = \left\{ (x_1, \ldots, x_{n_1}; y_1, \ldots, y_{n_2}) \,/\, \frac{\overline{X} - \overline{Y}}{s_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} > t_{n_1+n_2-2;\alpha} \right\}$$

$$\begin{array}{l} H_0 : \mu_X \geq \mu_Y \\ H_1 : \mu_X < \mu_Y \end{array} \quad R_\alpha = \left\{ (x_1, \ldots, x_{n_1}; y_1, \ldots, y_{n_2}) \,/\, \frac{\overline{X} - \overline{Y}}{s_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} < -t_{n_1+n_2-2;\alpha} \right\}$$

- What if we want to test with a general difference $d_0 \geq 0$?

<table>
<tr>
<td>$H_0 : \mu_X - \mu_Y = d_0$
$H_1 : \mu_X - \mu_Y \neq d_0$</td>
<td>$H_0 : \mu_X - \mu_Y \leq d_0$
$H_1 : \mu_X - \mu_Y > d_0$</td>
<td>$H_0 : \mu_X - \mu_Y \geq d_0$
$H_1 : \mu_X - \mu_Y < d_0$</td>
</tr>
<tr>
<td colspan="3">$T(X_1, \ldots, X_{n_1}; Y_1, \ldots, Y_{n_2}) = \dfrac{\overline{X} - \overline{Y} - d_0}{s_P \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim_{H_0} t_{n_1+n_2-2}$</td>
</tr>
</table>

Example 5

- Suppose that $X \sim N(\mu_A, \sigma^2)$, $Y \sim N(\mu_B, \sigma^2)$.
- For the two s.r.s. the following values of turnover are obtained:

CAMPAIGN A	16	14	42	38	23
CAMPAIGN B	61	33	37	63	65

- Test statistic: $T = \frac{\overline{X} - \overline{Y}}{s_P \sqrt{\frac{2}{5}}}$.

$\overline{x} = 26{,}6$ $\qquad$ $\overline{y} = 51{,}8$

$$s_X^2 = \frac{\sum_{i=1}^5 x_i^2 - 5\overline{x}^2}{4} = 162{,}8 \qquad s_Y^2 = \frac{\sum_{i=1}^5 y_i^2 - 5\overline{y}^2}{4} = 239{,}2$$

$$s_P^2 = \frac{4s_X^2 + 4s_Y^2}{8} = 201$$

$$t = \frac{26{,}6 - 51{,}8}{\sqrt{(201 \cdot 2)/5}} = -2{,}81$$

Example 5

- Suppose that $X \sim N(\mu_A, \sigma^2)$, $Y \sim N(\mu_B, \sigma^2)$.
- For the two s.r.s. the following values of turnover are obtained:

CAMPAIGN A	16	14	42	38	23
CAMPAIGN B	61	33	37	63	65

- Test statistic: $T = \frac{\overline{X} - \overline{Y}}{s_P \sqrt{\frac{2}{5}}}$.

$$\overline{x} = 26{,}6 \qquad \overline{y} = 51{,}8$$

$$s_X^2 = \frac{\sum_{i=1}^5 x_i^2 - 5\overline{x}^2}{4} = 162{,}8 \qquad s_Y^2 = \frac{\sum_{i=1}^5 y_i^2 - 5\overline{y}^2}{4} = 239{,}2$$

$$s_P^2 = \frac{4s_X^2 + 4s_Y^2}{8} = 201$$

$$t = \frac{26{,}6 - 51{,}8}{\sqrt{(201 \cdot 2)/5}} = -2{,}81$$

Example 5

- Suppose that $X \sim N(\mu_A, \sigma^2)$, $Y \sim N(\mu_B, \sigma^2)$.
- For the two s.r.s. the following values of turnover are obtained:

CAMPAIGN A	16	14	42	38	23
CAMPAIGN B	61	33	37	63	65

- Test statistic: $T = \frac{\overline{X} - \overline{Y}}{s_P\sqrt{\frac{2}{5}}}$.

$$\overline{x} = 26{,}6 \qquad \overline{y} = 51{,}8$$

$$s_X^2 = \frac{\sum_{i=1}^{5} x_i^2 - 5\overline{x}^2}{4} = 162{,}8 \qquad s_Y^2 = \frac{\sum_{i=1}^{5} y_i^2 - 5\overline{y}^2}{4} = 239{,}2$$

$$s_P^2 = \frac{4s_X^2 + 4s_Y^2}{8} = 201$$

$$t = \frac{26{,}6 - 51{,}8}{\sqrt{(201 \cdot 2)/5}} = -2{,}81$$

Example 5 (cont.)

- At α significance level, we reject $H_0 : \mu_A \geq \mu_B$ if $t = \frac{\overline{x}-\overline{y}}{s_P\sqrt{\frac{2}{5}}} = -2{,}81 < -t_{8;\alpha}$

$$t_{8;0,01} = 2{,}896 \quad t_{8;0,05} = 1{,}860 \quad t_{8;0,1} = 1{,}397$$

H_0 is rejected at $\alpha = 0{,}1; 0{,}05$ significance levels, and it is not rejected for $\alpha = 0{,}01$.

- The p-value of the hypothesis testing is:

$$p = Pr\{t_8 \leq -2{,}81\} = Pr\{t_8 \geq 2{,}81\} \in (0{,}01; 0{,}025)$$

Example 5 (cont.)

- At α significance level, we reject $H_0 : \mu_A \geq \mu_B$ if $t = \frac{\overline{x}-\overline{y}}{s_P\sqrt{\frac{2}{5}}} = -2{,}81 < -t_{8;\alpha}$

$$t_{8;0{,}01} = 2{,}896 \quad t_{8;0{,}05} = 1{,}860 \quad t_{8;0{,}1} = 1{,}397$$

  H_0 is rejected at $\alpha = 0{,}1; 0{,}05$ significance levels, and it is not rejected for $\alpha = 0{,}01$.
- The p-value of the hypothesis testing is:

$$p = Pr\{t_8 \leq -2{,}81\} = Pr\{t_8 \geq 2{,}81\} \in (0{,}01; 0{,}025)$$

Example 5 (cont.)

- At α significance level, we reject $H_0 : \mu_A \geq \mu_B$ if $t = \frac{\overline{x} - \overline{y}}{s_P \sqrt{\frac{2}{5}}} = -2{,}81 < -t_{8;\alpha}$

$$t_{8;0{,}01} = 2{,}896 \quad t_{8;0{,}05} = 1{,}860 \quad t_{8;0{,}1} = 1{,}397$$

  H_0 is rejected at $\alpha = 0{,}1; 0{,}05$ significance levels, and it is not rejected for $\alpha = 0{,}01$.

- The p-value of the hypothesis testing is:

$$p = Pr\{t_8 \leq -2{,}81\} = Pr\{t_8 \geq 2{,}81\} \in (0{,}01; 0{,}025)$$

Independent Samples: Comparison of variances, normal populations

Aim: Given 2 normal populations, it is desired to test the hypothesis of equal variances.

$$H_0 : \sigma_X^2 = \sigma_Y^2$$
$$H_1 : \sigma_X^2 \neq \sigma_Y^2$$

- Let $(X_1, \ldots, X_{n_1})$, $(Y_1, \ldots, Y_{n_2})$ be two s.r.s. of $X \sim N(\mu_X, \sigma_X^2)$ e $Y \sim N(\mu_Y, \sigma_Y^2)$, respectively, mutually independent.
- Basic result: $\frac{(n_1-1)s_X^2}{\sigma_X^2} \sim \chi^2_{n_1-1}$, $\frac{(n_2-1)s_Y^2}{\sigma_Y^2} \sim \chi^2_{n_2-1}$ indep.

$$\frac{s_X^2/\sigma_X^2}{s_Y^2/\sigma_Y^2} \sim F_{(n_1-1, n_2-1)}$$

- Test statistic: If H_0 is true:

$$T(X_1, \ldots, X_{n_1}; Y_1, \ldots, Y_{n_2}) = \frac{s_X^2}{s_Y^2} \sim_{H_0} F_{(n_1-1, n_2-1)}$$

Independent Samples: Comparison of variances, normal populations

Aim: Given 2 normal populations, it is desired to test the hypothesis of equal variances.

$$H_0 : \sigma_X^2 = \sigma_Y^2$$
$$H_1 : \sigma_X^2 \neq \sigma_Y^2$$

- Let $(X_1, \ldots, X_{n_1})$, $(Y_1, \ldots, Y_{n_2})$ be two s.r.s. of $X \sim N(\mu_X, \sigma_X^2)$ e $Y \sim N(\mu_Y, \sigma_Y^2)$, respectively, mutually independent.
- Basic result: $\frac{(n_1-1)s_X^2}{\sigma_X^2} \sim \chi^2_{n_1-1}$, $\frac{(n_2-1)s_Y^2}{\sigma_Y^2} \sim \chi^2_{n_2-1}$ indep.

$$\frac{s_X^2/\sigma_X^2}{s_Y^2/\sigma_Y^2} \sim F_{(n_1-1, n_2-1)}$$

- Test statistic: If H_0 is true:

$$T(X_1, \ldots, X_{n_1}; Y_1, \ldots, Y_{n_2}) = \frac{s_X^2}{s_Y^2} \sim_{H_0} F_{(n_1-1, n_2-1)}$$

Independent Samples: Comparison of variances, normal populations

Aim: Given 2 normal populations, it is desired to test the hypothesis of equal variances.

$$H_0 : \sigma_X^2 = \sigma_Y^2$$
$$H_1 : \sigma_X^2 \neq \sigma_Y^2$$

- Let $(X_1, \ldots, X_{n_1})$, $(Y_1, \ldots, Y_{n_2})$ be two s.r.s. of $X \sim N(\mu_X, \sigma_X^2)$ e $Y \sim N(\mu_Y, \sigma_Y^2)$, respectively, mutually independent.
- Basic result: $\frac{(n_1-1)s_X^2}{\sigma_X^2} \sim \chi^2_{n_1-1}$, $\frac{(n_2-1)s_Y^2}{\sigma_Y^2} \sim \chi^2_{n_2-1}$ indep.

$$\frac{s_X^2/\sigma_X^2}{s_Y^2/\sigma_Y^2} \sim F_{(n_1-1, n_2-1)}$$

- Test statistic: If H_0 is true:

$$T(X_1, \ldots, X_{n_1}; Y_1, \ldots, Y_{n_2}) = \frac{s_X^2}{s_Y^2} \sim_{H_0} F_{(n_1-1, n_2-1)}$$

Independent Samples: Comparison of variances, normal populations

- Critical Region

$$R_\alpha = \left\{ (x_1, \ldots, x_{n_1}; y_1, \ldots, y_{n_2}) \,/\, \frac{s_X^2}{s_Y^2} \leq F_{(n_1-1, n_2-1); 1-\frac{\alpha}{2}} \text{ or } \frac{s_X^2}{s_Y^2} \geq F_{(n_1-1, n_2-1); \frac{\alpha}{2}} \right\}$$

- One-sided tests:

$$H_1 : \sigma_X^2 > \sigma_Y^2 \Rightarrow R_\alpha = \{ \frac{s_X^2}{s_Y^2} \geq F_{(n_1-1, n_2-1); \alpha} \}$$

$$H_1 : \sigma_X^2 < \sigma_Y^2 \Rightarrow R_\alpha = \{ \frac{s_X^2}{s_Y^2} \leq F_{(n_1-1, n_2-1); 1-\alpha} \}$$

Independent Samples: Comparison of variances, normal populations

- Critical Region

$$R_\alpha = \left\{ (x_1, \ldots, x_{n_1}; y_1, \ldots, y_{n_2}) \;/\; \frac{s_X^2}{s_Y^2} \leq F_{(n_1-1, n_2-1); 1-\frac{\alpha}{2}} \text{ or } \frac{s_X^2}{s_Y^2} \geq F_{(n_1-1, n_2-1); \frac{\alpha}{2}} \right\}$$

- One-sided tests:

$$H_1 : \sigma_X^2 > \sigma_Y^2 \Rightarrow \; R_\alpha = \{ \frac{s_X^2}{s_Y^2} \geq F_{(n_1-1, n_2-1); \alpha} \}$$

$$H_1 : \sigma_X^2 < \sigma_Y^2 \Rightarrow \; R_\alpha = \{ \frac{s_X^2}{s_Y^2} \leq F_{(n_1-1, n_2-1); 1-\alpha} \}$$

Independent Samples: Comparison of variances, normal populations

Example 3

In order to compare the risk of markets A and B 21 data are obtained for market A and 16 for market B. It is obtained:

Mercado A	Mercado B
$\overline{x}_A = 0{,}3$	$\overline{x}_B = 0{,}4$
$s_A = 0{,}25$	$s_B = 0{,}45$

- Test statistic: $T = \frac{s_A^2}{s_B^2} \sim_{H_0} F_{(20,15)}$
- It is obtained $t = \left(\frac{0{,}25}{0{,}45}\right)^2 = 0{,}309$
- Critical region:

$$R_\alpha = \{t \leq F_{(20,15);1-\frac{\alpha}{2}} \text{ or } t \geq F_{(20,15);\frac{\alpha}{2}}\}$$

We only have one-tailed tables at 5 % and 1 %, What do we do?

Independent Samples: Comparison of variances, normal populations

Example 3

In order to compare the risk of markets A and B 21 data are obtained for market A and 16 for market B. It is obtained:

Mercado A	Mercado B
$\overline{x}_A = 0{,}3$	$\overline{x}_B = 0{,}4$
$s_A = 0{,}25$	$s_B = 0{,}45$

- Test statistic: $T = \frac{s_A^2}{s_B^2} \sim_{H_0} F_{(20,15)}$
- It is obtained $t = \left(\frac{0{,}25}{0{,}45}\right)^2 = 0{,}309$
- Critical region:

$$R_\alpha = \{t \leq F_{(20,15);1-\frac{\alpha}{2}} \text{ or } t \geq F_{(20,15);\frac{\alpha}{2}}\}$$

We only have one-tailed tables at 5 % and 1 %, What do we do?

Independent Samples: Comparison of variances, normal populations

Example 3

In order to compare the risk of markets A and B 21 data are obtained for market A and 16 for market B. It is obtained:

Mercado A	Mercado B
$\overline{x}_A = 0{,}3$	$\overline{x}_B = 0{,}4$
$s_A = 0{,}25$	$s_B = 0{,}45$

- Test statistic: $T = \frac{s_A^2}{s_B^2} \sim_{H_0} F_{(20,15)}$
- It is obtained $t = \left(\frac{0{,}25}{0{,}45}\right)^2 = 0{,}309$
- Critical region:

$$R_\alpha = \{t \leq F_{(20,15);1-\frac{\alpha}{2}} \text{ or } t \geq F_{(20,15);\frac{\alpha}{2}}\}$$

We only have one-tailed tables at 5 % and 1 %, What do we do?

Independent Samples: Comparison of variances, normal populations

Example 3

In order to compare the risk of markets A and B 21 data are obtained for market A and 16 for market B. It is obtained:

Mercado A	Mercado B
$\overline{x}_A = 0{,}3$	$\overline{x}_B = 0{,}4$
$s_A = 0{,}25$	$s_B = 0{,}45$

- Test statistic: $T = \frac{s_A^2}{s_B^2} \sim_{H_0} F_{(20,15)}$
- It is obtained $t = \left(\frac{0{,}25}{0{,}45}\right)^2 = 0{,}309$
- Critical region:

$$R_\alpha = \{t \leq F_{(20,15);1-\frac{\alpha}{2}} \text{ or } t \geq F_{(20,15);\frac{\alpha}{2}}\}$$

We only have one-tailed tables at 5 % and 1 %, What do we do?

Independent Samples: Comparison of variances, normal populations

Example 3 (cont.)

- If we have a computer: statistics package, or Excel, for obtaining the critical values, or for calculating the p-value:

$$p = \text{mín}\Big(2Pr\{T \leq 0{,}309 \mid H_0\}, 2Pr\{T \geq 0{,}309 \mid H_0\}\Big) =$$
$$= 2F_{(20,15)}(0{,}309) = 2 \cdot 0{,}0077677 = 0{,}01553$$

  What are the significance levels such that it is not rejected H_0?

- What if we have no computer? Perform one-sided test with $H_1 : \sigma_1^2 > \sigma_2^2$ by considering always the estimation with the greatest value in the numerator. In such a case, $s_B > s_A \Rightarrow$

$$H_0 : \sigma_B^2 \leq \sigma_A^2$$
$$H_1 : \sigma_B^2 > \sigma_A^2$$

  Now $t = \frac{1}{0{,}309} = 3{,}236$, and we can use the tables to find out $F_{(15,20);0{,}05} = 2{,}20$, $F_{(15,20);0{,}01} = 3{,}09$ What is the conclusion?

Independent Samples: Comparison of variances, normal populations

Example 3 (cont.)

- If we have a computer: statistics package, or Excel, for obtaining the critical values, or for calculating the p-value:

$$p = \text{mín}\Big(2Pr\{T \leq 0{,}309 \mid H_0\}, 2Pr\{T \geq 0{,}309 \mid H_0\}\Big) =$$
$$= 2F_{(20,15)}(0{,}309) = 2 \cdot 0{,}0077677 = 0{,}01553$$

What are the significance levels such that it is not rejected H_0?

- What if we have no computer? Perform one-sided test with $H_1 : \sigma_1^2 > \sigma_2^2$ by considering always the estimation with the greatest value in the numerator. In such a case, $s_B > s_A \Rightarrow$

$$H_0 : \sigma_B^2 \leq \sigma_A^2$$
$$H_1 : \sigma_B^2 > \sigma_A^2$$

Now $t = \frac{1}{0{,}309} = 3{,}236$, and we can use the tables to find out $F_{(15,20);0{,}05} = 2{,}20$, $F_{(15,20);0{,}01} = 3{,}09$ What is the conclusion?

Independent Samples: Comparison of variances, normal populations

Example 3 (cont.)

- If we have a computer: statistics package, or Excel, for obtaining the critical values, or for calculating the p-value:

$$p = \text{mín}\Big(2Pr\{T \leq 0{,}309 \mid H_0\}, 2Pr\{T \geq 0{,}309 \mid H_0\}\Big) =$$
$$= 2F_{(20,15)}(0{,}309) = 2 \cdot 0{,}0077677 = 0{,}01553$$

  What are the significance levels such that it is not rejected H_0?

- What if we have no computer? Perform one-sided test with $H_1 : \sigma_1^2 > \sigma_2^2$ by considering always the estimation with the greatest value in the numerator. In such a case, $s_B > s_A \Rightarrow$

$$H_0 : \sigma_B^2 \leq \sigma_A^2$$
$$H_1 : \sigma_B^2 > \sigma_A^2$$

  Now $t = \frac{1}{0{,}309} = 3{,}236$, and we can use the tables to find out $F_{(15,20);0{,}05} = 2{,}20$, $F_{(15,20);0{,}01} = 3{,}09$ What is the conclusion?

Independent Samples: Comparison of variances, normal populations

Example 3 (cont.)

- If we have a computer: statistics package, or Excel, for obtaining the critical values, or for calculating the p-value:

$$p = \text{mín}\Big(2Pr\{T \leq 0{,}309 \mid H_0\}, 2Pr\{T \geq 0{,}309 \mid H_0\}\Big) =$$
$$= 2F_{(20,15)}(0{,}309) = 2 \cdot 0{,}0077677 = 0{,}01553$$

  What are the significance levels such that it is not rejected H_0?

- What if we have no computer? Perform one-sided test with $H_1 : \sigma_1^2 > \sigma_2^2$ by considering always the estimation with the greatest value in the numerator. In such a case, $s_B > s_A \Rightarrow$

$$H_0 : \sigma_B^2 \leq \sigma_A^2$$
$$H_1 : \sigma_B^2 > \sigma_A^2$$

  Now $t = \frac{1}{0{,}309} = 3{,}236$, and we can use the tables to find out $F_{(15,20);0{,}05} = 2{,}20$, $F_{(15,20);0{,}01} = 3{,}09$ What is the conclusion?

Independent samples: Sensitivity of hypothesis testing

Aim: What are the consequences on the conclusions obtained when the working hypotheses are not held?

- Lack of Normality
  - Comparison of means: from the CLT we know that the means have always an approximated normal distribution. BE CAREFUL!!! outliers.
  - Comparison of variances: high sensitivity.
- Heteroscedasticity
- Lack of random sample: Very sensitive
  Randomization Principle: It prevents the systematic bias when assigning the sampling units. Useful for avoiding detection of differences associated with another factors.

Independent samples: Sensitivity of hypothesis testing

Aim: What are the consequences on the conclusions obtained when the working hypotheses are not held?

- Lack of Normality
  - Comparison of means: from the CLT we know that the means have always an approximated normal distribution. BE CAREFUL!!! outliers.
  - Comparison of variances: high sensitivity.
- Heteroscedasticity
  - Type I error (α): low sensitivity with similar sample sizes. High sensitivity for very different sample sizes (greater than double). Type II error (β): high sensitivity (the probability of not detecting differences increases)
- Lack of random sample: Very sensitive
  Randomization Principle: It prevents the systematic bias when assigning the sampling units. Useful for avoiding detection of differences associated with another factors.

Independent samples: Sensitivity of hypothesis testing

Aim: What are the consequences on the conclusions obtained when the working hypotheses are not held?

- Lack of Normality
  - Comparison of means: from the CLT we know that the means have always an approximated normal distribution. BE CAREFUL!!! outliers.
  - Comparison of variances: high sensitivity.
- Heteroscedasticity
  - Type I error (α): low sensitivity with similar sample sizes. High sensitivity for very different sample sizes (greater than double)
  - Type II error (β): high sensitivity (the probability of not detecting differences increases)
- Lack of random sample: Very sensitive
  Randomization Principle: It prevents the systematic bias when assigning the sampling units. Useful for avoiding detection of differences associated with another factors.

Independent samples: Sensitivity of hypothesis testing

Aim: What are the consequences on the conclusions obtained when the working hypotheses are not held?

- Lack of Normality
  - Comparison of means: from the CLT we know that the means have always an approximated normal distribution. BE CAREFUL!!! outliers.
  - Comparison of variances: high sensitivity.
- Heteroscedasticity
  - Type I error (α): low sensitivity with similar sample sizes. High sensitivity for very different sample sizes (greater than double)
  - Type II error (β): high sensitivity (the probability of not detecting differences increases)
- Lack of random sample: Very sensitive

  Randomization Principle: It prevents the systematic bias when assigning the sampling units. Useful for avoiding detection of differences associated with another factors.

Independent samples: Sensitivity of hypothesis testing

Aim: What are the consequences on the conclusions obtained when the working hypotheses are not held?

- Lack of Normality
  - Comparison of variances: high sensitivity
- Heteroscedasticity
  - Type I error (α): low sensitivity with similar sample sizes. High sensitivity for very different sample sizes (greater than double)
  - Type II error (β): high sensitivity (the probability of not detecting differences increases)
- Lack of random sample: Very sensitive

  Randomization Principle: It prevents the systematic bias when assigning the sampling units. Useful for avoiding detection of differences associated with another factors.

Independent samples: Sensitivity of hypothesis testing

Aim: What are the consequences on the conclusions obtained when the working hypotheses are not held?

- Lack of Normality
- Heteroscedasticity
  - Type I error (α): low sensitivity with similar sample sizes. High sensitivity for very different sample sizes (greater than double)
  - Type II error (β): high sensitivity (the probability of not detecting differences increases)
- Lack of random sample: Very sensitive
  Randomization Principle: It prevents the systematic bias when assigning the sampling units. Useful for avoiding detection of differences associated with another factors.

Independent samples: Sensitivity of hypothesis testing

Aim: What are the consequences on the conclusions obtained when the working hypotheses are not held?

- Lack of Normality
- Heteroscedasticity
  - Type II error (β): high sensitivity (the probability of not detecting differences increases)
- Lack of random sample: Very sensitive
  Randomization Principle: It prevents the systematic bias when assigning the sampling units. Useful for avoiding detection of differences associated with another factors.

Independent Samples: Comparison of means, large samples

Aim: Given 2 populations, we desire to test the hypothesis of equal means

$$H_0 : \mu_X = \mu_Y$$
$$H_1 : \mu_X \neq \mu_Y$$

- Let $(X_1, \ldots, X_{n_1})$, $(Y_1, \ldots, Y_{n_2})$ be two s.r.s. of X and Y, respectively, mutually independent, with n_1 y n_2 large enough.
- Basic result: Approximate method (CLT)

$$T(X_1, \ldots, X_{n_1}; Y_1, \ldots, Y_{n_2}) = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{s_X^2}{n_1} + \frac{s_Y^2}{n_2}}} \sim_{H_0} N(0,1)$$

Independent Samples: Comparison of means, large samples

Aim: Given 2 populations, we desire to test the hypothesis of equal means

$$H_0 : \mu_X = \mu_Y$$
$$H_1 : \mu_X \neq \mu_Y$$

- Let $(X_1, \ldots, X_{n_1})$, $(Y_1, \ldots, Y_{n_2})$ be two s.r.s. of X and Y, respectively, mutually independent, with n_1 y n_2 large enough.
- Basic result: Approximate method (CLT)

$$T(X_1, \ldots, X_{n_1}; Y_1, \ldots, Y_{n_2}) = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{s_X^2}{n_1} + \frac{s_Y^2}{n_2}}} \sim_{H_0} N(0, 1)$$

Independent Samples: Comparison of means, large samples

Aim: Given 2 populations, we desire to test the hypothesis of equal means

$$H_0 : \mu_X = \mu_Y$$
$$H_1 : \mu_X \neq \mu_Y$$

- Let $(X_1, \ldots, X_{n_1})$, $(Y_1, \ldots, Y_{n_2})$ be two s.r.s. of X and Y, respectively, mutually independent, with n_1 y n_2 large enough.
- Basic result: Approximate method (CLT)

$$T(X_1, \ldots, X_{n_1}; Y_1, \ldots, Y_{n_2}) = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{s_X^2}{n_1} + \frac{s_Y^2}{n_2}}} \sim_{H_0} N(0,1)$$

Independent Samples: Comparison of means, large samples

- In general, for $d_0 \geq 0$:

$$T(X_1, \ldots, X_{n_1}; Y_1, \ldots, Y_{n_2}) = \frac{\overline{X} - \overline{Y} - d_0}{\sqrt{\frac{s_X^2}{n_1} + \frac{s_Y^2}{n_2}}} \sim_{H_0} N(0,1)$$

$H_1 : \mu_X - \mu_Y \neq d_0$	$H_1 : \mu_X - \mu_Y > d_0$	$H_1 : \mu_X - \mu_Y < d_0$
$R_\alpha = \left\{ \lvert T \rvert \geq z_{\frac{\alpha}{2}} \right\}$	$R_\alpha = \{ T \geq z_\alpha \}$	$R_\alpha = \{ T \leq -z_\alpha \}$

Independent samples: Comparison of proportions, large samples

Aim: Given 2 populations, it is desired to test the hypothesis that the proportion of elements with a specific attribute is the same in both populations.

$$H_0 : p_X = p_Y$$
$$H_1 : p_X \neq p_Y$$

- Let $(X_1, \dots, X_{n_1})$, $(Y_1, \dots, Y_{n_2})$ two s.r.s. of both populations that are mutually independent, with r_X and r_Y being the number of observations with such an attribute in each sample.

$$\text{Sampling proportions: } \hat{p}_X = \frac{r_X}{n_1}, \quad \hat{p}_Y = \frac{r_Y}{n_2}$$

Independent samples: Comparison of proportions, large samples

Aim: Given 2 populations, it is desired to test the hypothesis that the proportion of elements with a specific attribute is the same in both populations.

$$H_0 : p_X = p_Y$$
$$H_1 : p_X \neq p_Y$$

- Let $(X_1, \ldots, X_{n_1})$, $(Y_1, \ldots, Y_{n_2})$ two s.r.s. of both populations that are mutually independent, with r_X and r_Y being the number of observations with such an attribute in each sample.

Sampling proportions: $\hat{p}_X = \dfrac{r_X}{n_1}, \quad \hat{p}_Y = \dfrac{r_Y}{n_2}$

Independent samples: Comparison of proportions, large samples

If H_0 is true:

- The best estimator of common proportion p_0 is:

$$\hat{p}_0 = \frac{r_X + r_Y}{n_1 + n_2}$$

- $\hat{p}_X - \hat{p}_Y$ r.v. with $E(\hat{p}_X - \hat{p}_Y) = 0$ and $V(\hat{p}_X - \hat{p}_Y) = V(\hat{p}_X) + V(\hat{p}_Y)$, that is estimated with:

$$\hat{V}(\hat{p}_X - \hat{p}_Y) = \frac{\hat{p}_0(1-\hat{p}_0)}{n_1} + \frac{\hat{p}_0(1-\hat{p}_0)}{n_2}$$

- If n_1 and n_2 are large enough $\Rightarrow$ CLT

$$\frac{\hat{p}_X - \hat{p}_Y}{\sqrt{\hat{p}_0(1-\hat{p}_0)}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim_{H_0} N(0,1)$$

Independent samples: Comparison of proportions, large samples

If H_0 is true:

- The best estimator of common proportion p_0 is:

$$\hat{p}_0 = \frac{r_X + r_Y}{n_1 + n_2}$$

- $\hat{p}_X - \hat{p}_Y$ r.v. with $E(\hat{p}_X - \hat{p}_Y) = 0$ and $V(\hat{p}_X - \hat{p}_Y) = V(\hat{p}_X) + V(\hat{p}_Y)$, that is estimated with:

$$\hat{V}(\hat{p}_X - \hat{p}_Y) = \frac{\hat{p}_0(1 - \hat{p}_0)}{n_1} + \frac{\hat{p}_0(1 - \hat{p}_0)}{n_2}$$

- If n_1 and n_2 are large enough $\Rightarrow$ CLT

$$\frac{\hat{p}_X - \hat{p}_Y}{\sqrt{\hat{p}_0(1 - \hat{p}_0)}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim_{H_0} N(0, 1)$$

Independent samples: Comparison of proportions, large samples

If H_0 is true:

- The best estimator of common proportion p_0 is:

$$\hat{p}_0 = \frac{r_X + r_Y}{n_1 + n_2}$$

- $\hat{p}_X - \hat{p}_Y$ r.v. with $E(\hat{p}_X - \hat{p}_Y) = 0$ and $V(\hat{p}_X - \hat{p}_Y) = V(\hat{p}_X) + V(\hat{p}_Y)$, that is estimated with:

$$\hat{V}(\hat{p}_X - \hat{p}_Y) = \frac{\hat{p}_0(1-\hat{p}_0)}{n_1} + \frac{\hat{p}_0(1-\hat{p}_0)}{n_2}$$

- If n_1 and n_2 are large enough $\Rightarrow$ CLT

$$\frac{\hat{p}_X - \hat{p}_Y}{\sqrt{\hat{p}_0(1-\hat{p}_0)}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim_{H_0} N(0,1)$$

Independent samples: Comparison of proportions, large samples

In general:

$T(X_1,\ldots,X_{n_1};Y_1,\ldots,Y_{n_2})=\dfrac{\hat{p}_X-\hat{p}_Y}{\sqrt{\hat{p}_0(1-\hat{p}_0)}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$		
$H_1: p_X \neq p_Y$	$H_1: p_X > p_Y$	$H_1: p_X < p_Y$
$R_\alpha=\left\{\lvert T\rvert \geq z_{\frac{\alpha}{2}}\right\}$	$R_\alpha=\{T\geq z_\alpha\}$	$R_\alpha=\{T\leq -z_\alpha\}$

Independent samples: Comparison of proportions, large samples

Example 1

- Suppose that $X \sim Ber(p_H)$, $Y \sim Ber(p_M)$. It is desired to test:

$$H_0 : p_H = p_M$$
$$H_1 : p_H \neq p_M$$

- A s.r.s. of 800 men revealed that 320 of them supported the proposition, and also 150 women from a s.r.s. of 500 women.
- Test statistic: $T = \frac{\hat{p}_H - \hat{p}_M}{\sqrt{\hat{p}_0(1-\hat{p}_0)}\sqrt{\frac{1}{800}+\frac{1}{500}}}$.

$$\hat{p}_H = \frac{320}{800} = 0{,}4, \quad \hat{p}_M = \frac{150}{500} = 0{,}3$$

$$\hat{p}_0 = \frac{320+150}{800+500} = 0{,}3615$$

Independent samples: Comparison of proportions, large samples

Example 1

- Suppose that $X \sim Ber(p_H)$, $Y \sim Ber(p_M)$. It is desired to test:

$$H_0 : p_H = p_M$$
$$H_1 : p_H \neq p_M$$

- A s.r.s. of 800 men revealed that 320 of them supported the proposition, and also 150 women from a s.r.s. of 500 women.
- Test statistic: $T = \frac{\hat{p}_H - \hat{p}_M}{\sqrt{\hat{p}_0(1-\hat{p}_0)}\sqrt{\frac{1}{800}+\frac{1}{500}}}$.

$$\hat{p}_H = \frac{320}{800} = 0{,}4, \quad \hat{p}_M = \frac{150}{500} = 0{,}3$$

$$\hat{p}_0 = \frac{320+150}{800+500} = 0{,}3615$$

Independent samples: Comparison of proportions, large samples

Example 1

- Suppose that $X \sim Ber(p_H)$, $Y \sim Ber(p_M)$. It is desired to test:

$$H_0 : p_H = p_M$$
$$H_1 : p_H \neq p_M$$

- A s.r.s. of 800 men revealed that 320 of them supported the proposition, and also 150 women from a s.r.s. of 500 women.
- Test statistic: $T = \frac{\hat{p}_H - \hat{p}_M}{\sqrt{\hat{p}_0(1-\hat{p}_0)}\sqrt{\frac{1}{800}+\frac{1}{500}}}$.

$$\hat{p}_H = \frac{320}{800} = 0{,}4, \quad \hat{p}_M = \frac{150}{500} = 0{,}3$$

$$\hat{p}_0 = \frac{320+150}{800+500} = 0{,}3615$$

Independent samples: Comparison of proportions, large samples

Example 1 (cont.)

- $$t = \frac{0{,}4 - 0{,}3}{\sqrt{0{,}3615(1 - 0{,}3615)}\sqrt{\frac{1}{800} + \frac{1}{500}}} = \frac{0{,}1}{0,02738} = 3{,}65$$
- $z_{0{,}005} = 2{,}57 \Rightarrow$ we reject H_0 at $\alpha = 0{,}01$ level.
- What do we do for $\alpha = 0{,}05; 0{,}1$?
- What can you say about the p-value of the test?
- If we build a 95 % CI for $p_H - p_M$, does 0 belong to the CI?

Independent samples: Comparison of proportions, large samples

Example 1 (cont.)

- $$t = \frac{0{,}4 - 0{,}3}{\sqrt{0{,}3615(1-0{,}3615)}\sqrt{\frac{1}{800}+\frac{1}{500}}} = \frac{0{,}1}{0,02738} = 3{,}65$$
- $z_{0{,}005} = 2{,}57 \Rightarrow$ we reject H_0 at $\alpha = 0{,}01$ level.
- What do we do for $\alpha = 0{,}05; 0{,}1$?
- What can you say about the p-value of the test?
- If we build a 95 % CI for $p_H - p_M$, does 0 belong to the CI?

Independent samples: Comparison of proportions, large samples

Example 1 (cont.)

- $$t = \frac{0{,}4 - 0{,}3}{\sqrt{0{,}3615(1 - 0{,}3615)}\sqrt{\frac{1}{800} + \frac{1}{500}}} = \frac{0{,}1}{0,02738} = 3{,}65$$
- $z_{0{,}005} = 2{,}57 \Rightarrow$ we reject H_0 at $\alpha = 0{,}01$ level.
- What do we do for $\alpha = 0{,}05; 0{,}1$?
- What can you say about the p-value of the test?
- If we build a 95 % CI for $p_H - p_M$, does 0 belong to the CI?

Independent samples: Comparison of proportions, large samples

Example 1 (cont.)

- $$t = \frac{0{,}4 - 0{,}3}{\sqrt{0{,}3615(1 - 0{,}3615)}\sqrt{\frac{1}{800} + \frac{1}{500}}} = \frac{0{,}1}{0,02738} = 3{,}65$$
- $z_{0{,}005} = 2{,}57 \Rightarrow$ we reject H_0 at $\alpha = 0{,}01$ level.
- What do we do for $\alpha = 0{,}05; 0{,}1$?
- What can you say about the p-value of the test?
- If we build a 95 % CI for $p_H - p_M$, does 0 belong to the CI?

Independent samples: Comparison of proportions, large samples

Example 1 (cont.)

- $$t = \frac{0{,}4 - 0{,}3}{\sqrt{0{,}3615(1 - 0{,}3615)}\sqrt{\frac{1}{800} + \frac{1}{500}}} = \frac{0{,}1}{0,02738} = 3{,}65$$
- $z_{0{,}005} = 2{,}57 \Rightarrow$ we reject H_0 at $\alpha = 0{,}01$ level.
- What do we do for $\alpha = 0{,}05; 0{,}1$?
- What can you say about the p-value of the test?
- If we build a 95 % CI for $p_H - p_M$, does 0 belong to the CI?

Matched samples, comparison of means, normal differences

Example 4

Before launching a very aggressive promotion of a product for stores, the marketing director of a company wants to know whether it is worth (whether the sales of this product are increased in this kind of shops). 50 stores are selected in Madrid to carry out this promotion and the data are collected before the promotion and thereafter.

Matched data

They come from a measurement of the same variable in the same individual just before and after applying a treatment.

Matched samples, comparison of means, normal differences

Example 4

Before launching a very aggressive promotion of a product for stores, the marketing director of a company wants to know whether it is worth (whether the sales of this product are increased in this kind of shops). 50 stores are selected in Madrid to carry out this promotion and the data are collected before the promotion and thereafter.

Matched data

They come from a measurement of the same variable in the same individual just before and after applying a treatment.

Matched samples, comparison of means, normal differences

Aim

Deal with a couple of measures taken in very similar conditions in order to make comparison of two experimental units that are a priori as equal as possible.

Why?

- Reduce population variability: to detect differences
- Control the effect of another factors: to avoid blaming the differences on other factors (another way?)

Matched samples, comparison of means, normal differences

Aim

Deal with a couple of measures taken in very similar conditions in order to make comparison of two experimental units that are a priori as equal as possible.

Why?

- Reduce population variability: to detect differences
- Control the effect of another factors: to avoid blaming the differences on other factors (another way?)

Matched samples, comparison of means, normal differences

Aim

Deal with a couple of measures taken in very similar conditions in order to make comparison of two experimental units that are a priori as equal as possible.

Why?

- Reduce population variability: to detect differences
- Control the effect of another factors: to avoid blaming the differences on other factors (another way?)

Matched samples, comparison of means, normal differences

Example 2

For a study it is desired to compare federal and state credit entities in terms of the ratio between the total debts of the entity and its assets.

Aim

We want to control the effect of another factors: size and seniority.
Deal with a couple of measures taken in very similar conditions in order to make comparison of two experimental units that are a priori as equal as possible.

Matched dependent samples

145 pairs of credit entities were chosen. Each pair contained one state unit and one federal unit. The matching was performed such that the 2 members were as similar as posible in size and seniority

Matched samples, comparison of means, normal differences

Example 2

For a study it is desired to compare federal and state credit entities in terms of the ratio between the total debts of the entity and its assets.

Aim

We want to control the effect of another factors: size and seniority. Deal with a couple of measures taken in very similar conditions in order to make comparison of two experimental units that are a priori as equal as possible.

Matched dependent samples

145 pairs of credit entities were chosen. Each pair contained one state unit and one federal unit. The matching was performed such that the 2 members were as similar as posible in size and seniority

Matched samples, comparison of means, normal differences

Example 2

For a study it is desired to compare federal and state credit entities in terms of the ratio between the total debts of the entity and its assets.

Aim

We want to control the effect of another factors: size and seniority.
Deal with a couple of measures taken in very similar conditions in order to make comparison of two experimental units that are a priori as equal as possible.

Matched dependent samples

145 pairs of credit entities were chosen. Each pair contained one state unit and one federal unit. The matching was performed such that the 2 members were as similar as posible in size and seniority

Matched samples, comparison of means, normal differences

Any other option?

Add the information about the size and seniority in the analysis

Analysis of Variance

Furthermore, it permits to extend the test of equality of means in normal populations to $k > 2$ populations with equal variances.

Matched samples, comparison of means, normal differences

Any other option?

Add the information about the size and seniority in the analysis

Analysis of Variance

Furthermore, it permits to extend the test of equality of means in normal populations to $k > 2$ populations with equal variances.

Matched samples, comparison of means, normal differences

Aim: Given 2 populations it is desired to test the hypothesis of equal means.

$$H_0 : \mu_X = \mu_Y$$
$$H_1 : \mu_X \neq \mu_Y$$

- Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ a s.r.s. from a normal bivariate distribution with parameters μ_X, μ_Y, σ_X^2, σ_Y^2 and ρ. The univariate s.r.s. $D_i = X_i - Y_i$, $i = 1, \ldots, n$ with normal distribution is enough.
- If H_0 is true, then $\overline{D}$ is normal with $E(\overline{D}) = 0$ and $V(\overline{D}) = \frac{\sigma_X^2 + \sigma_Y^2 - 2\sigma_X\sigma_Y\rho}{n}$.
- Test statistic

$$T(D_1, \ldots, D_n) = \frac{\overline{D}}{s_D/\sqrt{n}} \sim_{H_0} t_{n-1}$$

where $s_D^2 = \hat{V}(D)$ is the sample quasivariance of the differences:

$$s_D^2 = \frac{\sum_{i=1}^{n}(D_i - \overline{D})^2}{n-1} = \frac{\sum_{i=1}^{n} D_i^2 - n\overline{D}^2}{n-1}$$

Matched samples, comparison of means, normal differences

Aim: Given 2 populations it is desired to test the hypothesis of equal means.

$$H_0 : \mu_X = \mu_Y$$
$$H_1 : \mu_X \neq \mu_Y$$

- Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ a s.r.s. from a normal bivariate distribution with parameters μ_X, μ_Y, σ_X^2, σ_Y^2 and ρ.
  The univariate s.r.s. $D_i = X_i - Y_i$, $i = 1, \ldots, n$ with normal distribution is enough.
- If H_0 is true, then $\overline{D}$ is normal with $E(\overline{D}) = 0$ and $V(\overline{D}) = \frac{\sigma_X^2 + \sigma_Y^2 - 2\sigma_X\sigma_Y\rho}{n}$.
- Test statistic

$$T(D_1, \ldots, D_n) = \frac{\overline{D}}{s_D/\sqrt{n}} \sim_{H_0} t_{n-1}$$

where $s_D^2 = \widehat{V}(D)$ is the sample quasivariance of the differences:

$$s_D^2 = \frac{\sum_{i=1}^n (D_i - \overline{D})^2}{n-1} = \frac{\sum_{i=1}^n D_i^2 - n\overline{D}^2}{n-1}$$

Matched samples, comparison of means, normal differences

Aim: Given 2 populations it is desired to test the hypothesis of equal means.

$$H_0 : \mu_X = \mu_Y$$
$$H_1 : \mu_X \neq \mu_Y$$

- Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ a s.r.s. from a normal bivariate distribution with parameters μ_X, μ_Y,σ_X^2, σ_Y^2 and ρ.
  The univariate s.r.s. $D_i = X_i - Y_i$, $i = 1, \ldots, n$ with normal distribution is enough.
- If H_0 is true, then $\overline{D}$ is normal with $E(\overline{D}) = 0$ and $V(\overline{D}) = \frac{\sigma_X^2 + \sigma_Y^2 - 2\sigma_X\sigma_Y\rho}{n}$.
- Test statistic

$$T(D_1, \ldots, D_n) = \frac{\overline{D}}{s_D/\sqrt{n}} \sim_{H_0} t_{n-1}$$

where $s_D^2 = \hat{V}(D)$ is the sample quasivariance of the differences:

$$s_D^2 = \frac{\sum_{i=1}^n (D_i - \overline{D})^2}{n-1} = \frac{\sum_{i=1}^n D_i^2 - n\overline{D}^2}{n-1}$$

Matched samples, comparison of means, normal differences

Aim: Given 2 populations it is desired to test the hypothesis of equal means.

$$H_0 : \mu_X = \mu_Y$$
$$H_1 : \mu_X \neq \mu_Y$$

- Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ a s.r.s. from a normal bivariate distribution with parameters μ_X, μ_Y,σ_X^2, σ_Y^2 and ρ.
  The univariate s.r.s. $D_i = X_i - Y_i$, $i = 1, \ldots, n$ with normal distribution is enough.
- If H_0 is true, then $\overline{D}$ is normal with $E(\overline{D}) = 0$ and $V(\overline{D}) = \frac{\sigma_X^2 + \sigma_Y^2 - 2\sigma_X\sigma_Y\rho}{n}$.
- Test statistic

$$T(D_1, \ldots, D_n) = \frac{\overline{D}}{s_D/\sqrt{n}} \sim_{H_0} t_{n-1}$$

where $s_D^2 = \hat{V}(D)$ is the sample quasivariance of the differences:

$$s_D^2 = \frac{\sum_{i=1}^n (D_i - \overline{D})^2}{n-1} = \frac{\sum_{i=1}^n D_i^2 - n\overline{D}^2}{n-1}$$

Matched samples, comparison of means, normal differences

Aim: Given 2 populations it is desired to test the hypothesis of equal means.

$$H_0 : \mu_X = \mu_Y$$
$$H_1 : \mu_X \neq \mu_Y$$

- Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ a s.r.s. from a normal bivariate distribution with parameters μ_X, μ_Y,σ_X^2, σ_Y^2 and ρ.
  The univariate s.r.s. $D_i = X_i - Y_i$, $i = 1, \ldots, n$ with normal distribution is enough.
- If H_0 is true, then $\overline{D}$ is normal with $E(\overline{D}) = 0$ and $V(\overline{D}) = \frac{\sigma_X^2 + \sigma_Y^2 - 2\sigma_X\sigma_Y\rho}{n}$.
- Test statistic

$$T(D_1, \ldots, D_n) = \frac{\overline{D}}{s_D/\sqrt{n}} \sim_{H_0} t_{n-1}$$

where $s_D^2 = \hat{V}(D)$ is the sample quasivariance of the differences:

$$s_D^2 = \frac{\sum_{i=1}^n (D_i - \overline{D})^2}{n-1} = \frac{\sum_{i=1}^n D_i^2 - n\overline{D}^2}{n-1}$$

Matched samples, comparison of means, normal differences

In general:

$T(D_1, \ldots, D_n) = \dfrac{\overline{D} - d_0}{s_D/\sqrt{n}}$		
$H_1 : \mu_X - \mu_Y \neq d_0$	$H_1 : \mu_X - \mu_Y > d_0$	$H_1 : \mu_X - \mu_Y < d_0$
$R_\alpha = \left\{ \lvert T \rvert \geq t_{n-1; \frac{\alpha}{2}} \right\}$	$R_\alpha = \{ T \geq t_{n-1;\alpha} \}$	$R_\alpha = \{ T \leq -t_{n-1;\alpha} \}$

Matched samples, comparison of means, normal differences

Example 2

- For the aforementioned sample:
  145 pairs of credit entities were chosen. Each pair contained a state unit and a federal unit. The matching was performed such that the 2 members were as similar as possible in size and seniority
  The mean of the differences (federal minus state) was 0,0518, with a standard deviation equal to 0,3055.
- Test statistic: $t = \frac{0{,}0518}{0{,}3055/\sqrt{145}} = 2{,}0417$
- $n-1$ is very high, we can work with the critical values of the normal distribution and approximate the p-value of the test by:

$$p - value = 2P\{Z \geq 2{,}04\} = 2 \cdot 0,0207 = 0,0414$$

Matched samples, comparison of means, normal differences

Example 2

- For the aforementioned sample:
  145 pairs of credit entities were chosen. Each pair contained a state unit and a federal unit. The matching was performed such that the 2 members were as similar as possible in size and seniority
  The mean of the differences (federal minus state) was 0,0518, with a standard deviation equal to 0,3055.
- Test statistic: $t = \frac{0{,}0518}{0{,}3055/\sqrt{145}} = 2{,}0417$
- $n-1$ is very high, we can work with the critical values of the normal distribution and approximate the p-value of the test by:

$$p - value = 2P\{Z \geq 2{,}04\} = 2 \cdot 0,0207 = 0,0414$$

Matched samples, comparison of means, normal differences

Example 2

- For the aforementioned sample:
  145 pairs of credit entities were chosen. Each pair contained a state unit and a federal unit. The matching was performed such that the 2 members were as similar as possible in size and seniority
  The mean of the differences (federal minus state) was 0,0518, with a standard deviation equal to 0,3055.
- Test statistic: $t = \frac{0{,}0518}{0{,}3055/\sqrt{145}} = 2{,}0417$
- $n-1$ is very high, we can work with the critical values of the normal distribution and approximate the p-value of the test by:

$$p - value = 2P\{Z \geq 2{,}04\} = 2 \cdot 0,0207 = 0,0414$$

Comparison of two populations

Summary for two independent s.r.s., two-sided tests

Difference of	Hypothesis	Statistic	Critical region
Means	Normal data Equal variances	$\dfrac{\overline{X}-\overline{Y}}{s_P\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim_{H_0} t_{n_1+n_2-2}$	$\{\lvert T\rvert \geq t_{n_1+n_2-2;\frac{\alpha}{2}}\}$
	Not normal data Large samples	$\dfrac{\overline{X}-\overline{Y}}{\sqrt{\frac{s_X^2}{n_1}+\frac{s_Y^2}{n_2}}} \sim_{H_0} N(0,1)$	$\{\lvert T\rvert \geq z_{\frac{\alpha}{2}}\}$
Proportions	Large samples	$\dfrac{\hat{p}_X-\hat{p}_Y}{\sqrt{\hat{p}_0(1-\hat{p}_0)}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim_{H_0} N(0,1)$	$\{\lvert T\rvert \geq z_{\frac{\alpha}{2}}\}$
Variances	Normal data	$\dfrac{s_X^2}{s_Y^2} \sim_{H_0} F_{(n_1-1,n_2-1)}$	$\{T \leq F_{(n_1-1,n_2-1);1-\frac{\alpha}{2}}$ or $T \geq F_{(n_1-1,n_2-1);\frac{\alpha}{2}}\}$

$$s_P^2 = \frac{(n_1-1)s_X^2+(n_2-1)s_Y^2}{n_1+n_2-2}$$