Lecture 8: Hypothesis testing part II

Statistical Methods for Data Science

Yinan Yu

Department of Computer Science and Engineering

November 29, 2021

Today

1. Test statistics and hypothesis tests
   - z-test
   - One-sample t-test
   - Two-sample t-test
   - Paired t-test

Learning outcome

- Be able to explain the following hypothesis tests
  - One-sample and two-sample z-test
  - One-sample and two-sample t-test
  - Paired t-test

  For each of these tests, be able to describe the typical set up for the experiment, the general purpose of the test, data produced by the experiment, random variables, parameter of interest, null hypothesis, alternative hypothesis, test statistic, null distribution, the computation of *p*-value
- ... more to come (to be updated)

Today

1. Test statistics and hypothesis tests
   - z-test
   - One-sample t-test
   - Two-sample t-test
   - Paired t-test

Disclaimer

- Recall that in this course, we only consider H_0 **with an equal sign in them**, i.e. the **null distribution is fully specified**; the description of H_0 is based on this assumption

Disclaimer

- Recall that in this course, we only consider H_0 **with an equal sign in them**, i.e. the **null distribution is fully specified**; the description of H_0 is based on this assumption
- For **symmetric null distributions**, e.g. **standard Gaussian distribution**, **student's t distribution**, **binomial distribution with** $p = 0.5$, etc, we only illustrate examples with the two-tailed alternative hypothesis H_A in this lecture without loss of generality; the one-tailed version can be easily derived

Disclaimer

- Recall that in this course, we only consider H_0 **with an equal sign in them**, i.e. the **null distribution is fully specified**; the description of H_0 is based on this assumption
- For **symmetric null distributions**, e.g. **standard Gaussian distribution**, **student's t distribution**, **binomial distribution with $p = 0.5$**, etc, we only illustrate examples with the two-tailed alternative hypothesis H_A in this lecture without loss of generality; the one-tailed version can be easily derived
- For the **exact binomial test with $p \neq 0.5$**, the null distribution is not symmetric; in this case, the computation of the two-tailed p-value is not uniquely defined; in this lecture, we will not go into details for these cases; we will only look at the one-tailed tests for asymmetric binomial null distributions

Disclaimer

- Recall that in this course, we only consider H_0 **with an equal sign in them**, i.e. the **null distribution is fully specified**; the description of H_0 is based on this assumption
- For **symmetric null distributions**, e.g. **standard Gaussian distribution**, **student's t distribution**, **binomial distribution with** $p = 0.5$, etc, we only illustrate examples with the two-tailed alternative hypothesis H_A in this lecture without loss of generality; the one-tailed version can be easily derived
- For the **exact binomial test with** $p \neq 0.5$, the null distribution is not symmetric; in this case, the computation of the two-tailed p-value is not uniquely defined; in this lecture, we will not go into details for these cases; we will only look at the one-tailed tests for asymmetric binomial null distributions
- For each hypothesis test, the purpose of the Python code snippet is to provide a better understanding of the calculation; in practice, there are alternative libraries and built-in functions for these tests that might result in a more compact implementation

Disclaimer (cont.)

For each of the hypothesis tests we introduce, we present the following components:

Disclaimer (cont.)

For each of the hypothesis tests we introduce, we present the following components:

- **Typical set up for the experiment**
  - **Test subjects**, e.g. the number of samples, the number of groups, etc
  - Description of the **experiment** and the **result**
  - Description of the **data type** produced in the result

Disclaimer (cont.)

For each of the hypothesis tests we introduce, we present the following components:

- **Typical set up for the experiment**
  - **Test subjects**, e.g. the number of samples, the number of groups, etc
  - Description of the **experiment** and the **result**
  - Description of the **data type** produced in the result
- **Purpose**: the general purpose of the test

Disclaimer (cont.)

For each of the hypothesis tests we introduce, we present the following components:

- **Typical set up for the experment**
  - **Test subjects**, e.g. the number of samples, the number of groups, etc
  - Description of the **experiment** and the **result**
  - Description of the **data type** produced in the result
- **Purpose**: the general purpose of the test
- **Data**: symbolic description of the data produced by the experiment
- **Random variable** and **assumption** corresponding to the data

Disclaimer (cont.)

For each of the hypothesis tests we introduce, we present the following components:

- **Typical set up for the experiment**
  - **Test subjects**, e.g. the number of samples, the number of groups, etc
  - Description of the **experiment** and the **result**
  - Description of the **data type** produced in the result
- **Purpose**: the general purpose of the test
- **Data**: symbolic description of the data produced by the experiment
- **Random variable** and **assumption** corresponding to the data
- **Parameter of interest** and the **estimates**

Disclaimer (cont.)

For each of the hypothesis tests we introduce, we present the following components:

- **Typical set up for the experiment**
  - **Test subjects**, e.g. the number of samples, the number of groups, etc
  - Description of the **experiment** and the **result**
  - Description of the **data type** produced in the result
- **Purpose**: the general purpose of the test
- **Data**: symbolic description of the data produced by the experiment
- **Random variable** and **assumption** corresponding to the data
- **Parameter of interest** and the **estimates**
- **Hypotheses** H_0 and H_A

Disclaimer (cont.)

For each of the hypothesis tests we introduce, we present the following components:

- **Typical set up for the experiment**
  - **Test subjects**, e.g. the number of samples, the number of groups, etc
  - Description of the **experiment** and the **result**
  - Description of the **data type** produced in the result
- **Purpose**: the general purpose of the test
- **Data**: symbolic description of the data produced by the experiment
- **Random variable** and **assumption** corresponding to the data
- **Parameter of interest** and the **estimates**
- **Hypotheses** H_0 and H_A
- **Test statistic**
- **Null distribution**
  - PDF/PMF: description of the PDF/PMF
  - Python: code snippet of the PDF/PMF

Disclaimer (cont.)

For each of the hypothesis tests we introduce, we present the following components:

- **Typical set up for the experiment**
  - **Test subjects**, e.g. the number of samples, the number of groups, etc
  - Description of the **experiment** and the **result**
  - Description of the **data type** produced in the result
- **Purpose**: the general purpose of the test
- **Data**: symbolic description of the data produced by the experiment
- **Random variable** and **assumption** corresponding to the data
- **Parameter of interest** and the **estimates**
- **Hypotheses** H_0 and H_A
- **Test statistic**
- **Null distribution**
  - PDF/PMF: description of the PDF/PMF
  - Python: code snippet of the PDF/PMF
- p-**value**
  - Definition: an expression of p-value in terms of a probability
  - Python: code snippet to illustrate the computation of the p-value (see page 5)

z-test

One-sample z-test

- **Typical set up for the experiment**:

One-sample z-test

- **Typical set up for the experiment**:
  - **One sample** of independent test subjects, e.g. a sample of patients, a sample of customers, etc

One-sample z-test

- **Typical set up for the experiment**:
  - **One sample** of independent test subjects, e.g. a sample of patients, a sample of customers, etc
  - Run the same experiment on each subject and collect the outcomes, e.g. give a new drug to a sample of patients and measure the effect on each individual patient; test a new web design on a sample of customers and record the time they spend on the web page, etc

One-sample z-test

- **Typical set up for the experiment**:
  - **One sample** of independent test subjects, e.g. a sample of patients, a sample of customers, etc
  - Run the same experiment on each subject and collect the outcomes, e.g. give a new drug to a sample of patients and measure the effect on each individual patient; test a new web design on a sample of customers and record the time they spend on the web page, etc
  - The result contains one i.i.d. sample with **continuous numerical values**

One-sample z-test

- **Typical set up for the experiment**:
  - **One sample** of independent test subjects, e.g. a sample of patients, a sample of customers, etc
  - Run the same experiment on each subject and collect the outcomes, e.g. give a new drug to a sample of patients and measure the effect on each individual patient; test a new web design on a sample of customers and record the time they spend on the web page, etc
  - The result contains one i.i.d. sample with **continuous numerical values**
- **Purpose**: to test if the mean of the result differs from a predefined constant

One-sample z-test

- **Typical set up for the experiment**:
  - **One sample** of independent test subjects, e.g. a sample of patients, a sample of customers, etc
  - Run the same experiment on each subject and collect the outcomes, e.g. give a new drug to a sample of patients and measure the effect on each individual patient; test a new web design on a sample of customers and record the time they spend on the web page, etc
  - The result contains one i.i.d. sample with **continuous numerical values**
- **Purpose**: to test if the mean of the result differs from a predefined constant
- **Data**: $x_1, \cdots, x_N$, e.g. blood pressure after taking a new drug

One-sample z-test

- **Typical set up for the experiment**:
  - **One sample** of independent test subjects, e.g. a sample of patients, a sample of customers, etc
  - Run the same experiment on each subject and collect the outcomes, e.g. give a new drug to a sample of patients and measure the effect on each individual patient; test a new web design on a sample of customers and record the time they spend on the web page, etc
  - The result contains one i.i.d. sample with **continuous numerical values**
- **Purpose**: to test if the mean of the result differs from a predefined constant
- **Data**: $x_1, \cdots, x_N$, e.g. blood pressure after taking a new drug
- **Random variable** and **assumption**: $X_1, \cdots, X_N$
  - X_i i.i.d.
  - X_i Gaussian or large N (CLT)
  - X_i standard deviation σ known

One-sample z-test

- **Typical set up for the experiment**:
  - **One sample** of independent test subjects, e.g. a sample of patients, a sample of customers, etc
  - Run the same experiment on each subject and collect the outcomes, e.g. give a new drug to a sample of patients and measure the effect on each individual patient; test a new web design on a sample of customers and record the time they spend on the web page, etc
  - The result contains one i.i.d. sample with **continuous numerical values**
- **Purpose**: to test if the mean of the result differs from a predefined constant
- **Data**: $x_1, \cdots, x_N$, e.g. blood pressure after taking a new drug
- **Random variable** and **assumption**: $X_1, \cdots, X_N$
  - X_i i.i.d.
  - X_i Gaussian or large N (CLT)
  - X_i standard deviation σ known
- **Parameter of interest**: μ

One-sample z-test

- **Typical set up for the experiment**:
  - **One sample** of independent test subjects, e.g. a sample of patients, a sample of customers, etc
  - Run the same experiment on each subject and collect the outcomes, e.g. give a new drug to a sample of patients and measure the effect on each individual patient; test a new web design on a sample of customers and record the time they spend on the web page, etc
  - The result contains one i.i.d. sample with **continuous numerical values**
- **Purpose**: to test if the mean of the result differs from a predefined constant
- **Data**: $x_1, \cdots, x_N$, e.g. blood pressure after taking a new drug
- **Random variable** and **assumption**: $X_1, \cdots, X_N$
  - X_i i.i.d.
  - X_i Gaussian or large N (CLT)
  - X_i standard deviation σ known
- **Parameter of interest**: μ
- **Parameter estimate**: $\bar{x}$, $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/N)$

One-sample z-test

- **Typical set up for the experiment**:
  - **One sample** of independent test subjects, e.g. a sample of patients, a sample of customers, etc
  - Run the same experiment on each subject and collect the outcomes, e.g. give a new drug to a sample of patients and measure the effect on each individual patient; test a new web design on a sample of customers and record the time they spend on the web page, etc
  - The result contains one i.i.d. sample with **continuous numerical values**
- **Purpose**: to test if the mean of the result differs from a predefined constant
- **Data**: $x_1, \cdots, x_N$, e.g. blood pressure after taking a new drug
- **Random variable** and **assumption**: $X_1, \cdots, X_N$
  - X_i i.i.d.
  - X_i Gaussian or large N (CLT)
  - X_i standard deviation σ known
- **Parameter of interest**: μ
- **Parameter estimate**: $\bar{x}$, $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/N)$
- **Hypotheses** H_0 and H_A: given c a constant

$$H_0: \quad \mu = c$$
$$H_A: \quad \mu \neq c$$

One-sample z-test

- **Typical set up for the experiment**:
  - **One sample** of independent test subjects, e.g. a sample of patients, a sample of customers, etc
  - Run the same experiment on each subject and collect the outcomes, e.g. give a new drug to a sample of patients and measure the effect on each individual patient; test a new web design on a sample of customers and record the time they spend on the web page, etc
  - The result contains one i.i.d. sample with **continuous numerical values**
- **Purpose**: to test if the mean of the result differs from a predefined constant
- **Data**: $x_1, \cdots, x_N$, e.g. blood pressure after taking a new drug
- **Random variable** and **assumption**: $X_1, \cdots, X_N$
  - X_i i.i.d.
  - X_i Gaussian or large N (CLT)
  - X_i standard deviation σ known
- **Parameter of interest**: μ
- **Parameter estimate**: $\bar{x}$, $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/N)$
- **Hypotheses** H_0 and H_A: given c a constant

$$\begin{aligned} H_0: &\quad \mu = c \\ H_A: &\quad \mu \neq c \end{aligned}$$

Note: only two-tailed H_A is illustrated here.

One-sample z-test (cont.)

- **Test statistic**:

$$z_0 = \frac{\bar{x} - c}{\sigma/\sqrt{N}}$$

One-sample z-test (cont.)

- **Test statistic**:

$$z_0 = \frac{\bar{x} - c}{\sigma/\sqrt{N}}$$

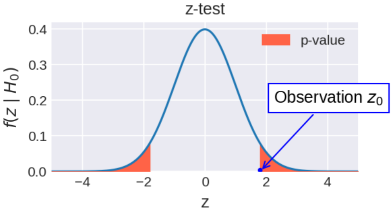
- **Null distribution**: standard normal distribution
  - PDF: $f(z \mid H_0) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$
  - Python: stats.norm.pdf(z, 0, 1)

One-sample z-test (cont.)

- **Test statistic**:

$$z_0 = \frac{\bar{x} - c}{\sigma / \sqrt{N}}$$

- **Null distribution**: standard normal distribution
  - PDF: $f(z \mid H_0) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$
  - Python: stats.norm.pdf(z, 0, 1)
- ***p*-value**
  - Definition: $p = 2 \min \left(P(Z \leq z_0 \mid H_0), P(Z \geq z_0 \mid H_0) \right)$
  - Python: $2 * \min$ (stats.norm.cdf(z_0, 0, 1), 1-stats.norm.cdf(z_0, 0, 1))



Two-sample z-test

- **Typical set up for the experiment**:

Two-sample z-test

- **Typical set up for the experiment**:
  - **Two samples** of independent test subjects, where the two samples $\mathcal{X}$ and $\mathcal{Y}$ letters with a calligraphic font are typically used to denote sets are independent from one another, e.g. two samples of independent patients, two samples of independent customers, etc

Two-sample z-test

- **Typical set up for the experiment**:
  - **Two samples** of independent test subjects, where the two samples $\mathcal{X}$ and $\mathcal{Y}$ letters with a calligraphic font are typically used to denote sets are independent from one another, e.g. two samples of independent patients, two samples of independent customers, etc
  - Run two sets of experiments A and B on the test subjects from the two samples $\mathcal{X}$ and $\mathcal{Y}$, respectively, and collect the outcomes, e.g. give different drugs to the two samples of patients and measure the effect on each individual patient; test two web designs on two samples of customers and record the time they spend on the web page, etc

Two-sample z-test

- **Typical set up for the experiment**:
  - **Two samples** of independent test subjects, where the two samples $\mathcal{X}$ and $\mathcal{Y}$ **letters with a calligraphic font are typically used to denote sets** are independent from one another, e.g. two samples of independent patients, two samples of independent customers, etc
  - Run two sets of experiments A and B on the test subjects from the two samples $\mathcal{X}$ and $\mathcal{Y}$, respectively, and collect the outcomes, e.g. give different drugs to the two samples of patients and measure the effect on each individual patient; test two web designs on two samples of customers and record the time they spend on the web page, etc
  - The result contains two i.i.d. samples with **continuous numerical values**

Two-sample z-test

- **Typical set up for the experiment**:
  - **Two samples** of independent test subjects, where the two samples $\mathcal{X}$ and $\mathcal{Y}$ letters with a calligraphic font are typically used to denote sets are independent from one another, e.g. two samples of independent patients, two samples of independent customers, etc
  - Run two sets of experiments A and B on the test subjects from the two samples $\mathcal{X}$ and $\mathcal{Y}$, respectively, and collect the outcomes, e.g. give different drugs to the two samples of patients and measure the effect on each individual patient; test two web designs on two samples of customers and record the time they spend on the web page, etc
  - The result contains two i.i.d. samples with **continuous numerical values**
- **Purpose**: to test if two alternative options have different effects by testing if the mean of the result from one sample differs from the mean of the other sample

Two-sample z-test

- **Typical set up for the experiment**:
  - **Two samples** of independent test subjects, where the two samples $\mathcal{X}$ and $\mathcal{Y}$ letters with a calligraphic font are typically used to denote sets are independent from one another, e.g. two samples of independent patients, two samples of independent customers, etc
  - Run two sets of experiments A and B on the test subjects from the two samples $\mathcal{X}$ and $\mathcal{Y}$, respectively, and collect the outcomes, e.g. give different drugs to the two samples of patients and measure the effect on each individual patient; test two web designs on two samples of customers and record the time they spend on the web page, etc
  - The result contains two i.i.d. samples with **continuous numerical values**
- **Purpose**: to test if two alternative options have different effects by testing if the mean of the result from one sample differs from the mean of the other sample
- **Data**: $x_1, \cdots, x_{N_X}$ and $y_1, \cdots, y_{N_Y}$, e.g. blood pressure measured after taking two different drugs

Two-sample z-test

- Typical set up for the experiment:
  - Two samples of independent test subjects, where the two samples $\mathcal{X}$ and $\mathcal{Y}$ letters with a calligraphic font are typically used to denote sets are independent from one another, e.g. two samples of independent patients, two samples of independent customers, etc
  - Run two sets of experiments A and B on the test subjects from the two samples $\mathcal{X}$ and $\mathcal{Y}$, respectively, and collect the outcomes, e.g. give different drugs to the two samples of patients and measure the effect on each individual patient; test two web designs on two samples of customers and record the time they spend on the web page, etc
  - The result contains two i.i.d. samples with **continuous numerical values**
- Purpose: to test if two alternative options have different effects by testing if the mean of the result from one sample differs from the mean of the other sample
- Data: $x_1, \cdots, x_{N_X}$ and $y_1, \cdots, y_{N_Y}$, e.g. blood pressure measured after taking two different drugs
- Random variable and assumption: $X_1, \cdots, X_{N_X}$, $Y_1, \cdots, Y_{N_Y}$
  - X_i and Y_j independent
  - X_i i.i.d.; Y_j i.i.d.
  - X_i Gaussian or large N_X; Y_j Gaussian or large N_Y
  - X_i and Y_j have known standard deviation σ_X and σ_Y, respectively

Two-sample z-test

- Typical set up for the experiment:
  - Two samples of independent test subjects, where the two samples $\mathcal{X}$ and $\mathcal{Y}$ letters with a calligraphic font are typically used to denote sets are independent from one another, e.g. two samples of independent patients, two samples of independent customers, etc
  - Run two sets of experiments A and B on the test subjects from the two samples $\mathcal{X}$ and $\mathcal{Y}$, respectively, and collect the outcomes, e.g. give different drugs to the two samples of patients and measure the effect on each individual patient; test two web designs on two samples of customers and record the time they spend on the web page, etc
  - The result contains two i.i.d. samples with continuous numerical values
- Purpose: to test if two alternative options have different effects by testing if the mean of the result from one sample differs from the mean of the other sample
- Data: $x_1, \cdots, x_{N_X}$ and $y_1, \cdots, y_{N_Y}$, e.g. blood pressure measured after taking two different drugs
- Random variable and assumption: $X_1, \cdots, X_{N_X}$, $Y_1, \cdots, Y_{N_Y}$
  - X_i and Y_j independent
  - X_i i.i.d.; Y_j i.i.d.
  - X_i Gaussian or large N_X; Y_j Gaussian or large N_Y
  - X_i and Y_j have known standard deviation σ_X and σ_Y, respectively
- Parameter of interest: μ_X, μ_Y

Two-sample z-test

- **Typical set up for the experiment**:
  - **Two samples** of independent test subjects, where the two samples $\mathcal{X}$ and $\mathcal{Y}$ letters with a calligraphic font are typically used to denote sets are independent from one another, e.g. two samples of independent patients, two samples of independent customers, etc
  - Run two sets of experiments A and B on the test subjects from the two samples $\mathcal{X}$ and $\mathcal{Y}$, respectively, and collect the outcomes, e.g. give different drugs to the two samples of patients and measure the effect on each individual patient; test two web designs on two samples of customers and record the time they spend on the web page, etc
  - The result contains two i.i.d. samples with **continuous numerical values**
- **Purpose**: to test if two alternative options have different effects by testing if the mean of the result from one sample differs from the mean of the other sample
- **Data**: $x_1, \cdots, x_{N_X}$ and $y_1, \cdots, y_{N_Y}$, e.g. blood pressure measured after taking two different drugs
- **Random variable** and **assumption**: $X_1, \cdots, X_{N_X}$, $Y_1, \cdots, Y_{N_Y}$
  - X_i and Y_j independent
  - X_i i.i.d.; Y_j i.i.d.
  - X_i Gaussian or large N_X; Y_j Gaussian or large N_Y
  - X_i and Y_j have known standard deviation σ_X and σ_Y, respectively
- **Parameter of interest**: μ_X, μ_Y
- **Parameter estimate**: $\bar{x}$, $\bar{y}$

Two-sample z-test

- **Typical set up for the experiment**:
  - **Two samples** of independent test subjects, where the two samples $\mathcal{X}$ and $\mathcal{Y}$ letters with a calligraphic font are typically used to denote sets are independent from one another, e.g. two samples of independent patients, two samples of independent customers, etc
  - Run two sets of experiments A and B on the test subjects from the two samples $\mathcal{X}$ and $\mathcal{Y}$, respectively, and collect the outcomes, e.g. give different drugs to the two samples of patients and measure the effect on each individual patient; test two web designs on two samples of customers and record the time they spend on the web page, etc
  - The result contains two i.i.d. samples with **continuous numerical values**
- **Purpose**: to test if two alternative options have different effects by testing if the mean of the result from one sample differs from the mean of the other sample
- **Data**: $x_1, \cdots, x_{N_X}$ and $y_1, \cdots, y_{N_Y}$, e.g. blood pressure measured after taking two different drugs
- **Random variable** and **assumption**: $X_1, \cdots, X_{N_X}$, $Y_1, \cdots, Y_{N_Y}$
  - X_i and Y_j independent
  - X_i i.i.d.; Y_j i.i.d.
  - X_i Gaussian or large N_X; Y_j Gaussian or large N_Y
  - X_i and Y_j have known standard deviation σ_X and σ_Y, respectively
- **Parameter of interest**: μ_X, μ_Y
- **Parameter estimate**: $\bar{x}$, $\bar{y}$
- **Hypotheses H_0 and H_A**: given c a constant (typically $c = 0$)

$$
\begin{aligned}
H_0 &: \quad \mu_X - \mu_Y = c \\
H_A &: \quad \mu_X - \mu_Y \neq c
\end{aligned}
$$

Two-sample z-test (cont.)

- **Test statistic**:

$$z_0 = \frac{\bar{x} - \bar{y} - c}{\sqrt{\frac{\sigma_X^2}{N_X} + \frac{\sigma_Y^2}{N_Y}}}$$

Hint: $\bar{X} - \bar{Y} \sim \mathcal{N}\left(\mu_X - \mu_Y, \sigma_X^2/N_X + \sigma_Y^2/N_Y\right)$

Two-sample z-test (cont.)

- **Test statistic**:

$$z_0 = \frac{\bar{x} - \bar{y} - c}{\sqrt{\frac{\sigma_X^2}{N_X} + \frac{\sigma_Y^2}{N_Y}}}$$

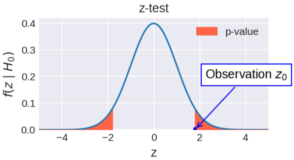
  Hint: $\bar{X} - \bar{Y} \sim \mathcal{N}\left(\mu_X - \mu_Y, \sigma_X^2/N_X + \sigma_Y^2/N_Y\right)$
- **Null distribution**: standard normal distribution
  - PDF: $f(z \mid H_0) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$
  - Python: stats.norm.pdf(z, 0, 1)

Two-sample z-test (cont.)

- **Test statistic**:

$$z_0 = \frac{\bar{x} - \bar{y} - c}{\sqrt{\frac{\sigma_X^2}{N_X} + \frac{\sigma_Y^2}{N_Y}}}$$

  Hint: $\bar{X} - \bar{Y} \sim \mathcal{N}\left(\mu_X - \mu_Y, \sigma_X^2/N_X + \sigma_Y^2/N_Y\right)$
- **Null distribution**: standard normal distribution
  - PDF: $f(z \mid H_0) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$
  - Python: stats.norm.pdf(z, 0, 1)
- ***p*-value**
  - Definition: $p = 2 \min\left(P(Z \leq z_0 \mid H_0), P(Z \geq z_0 \mid H_0)\right)$
  - Python: $2 * \min$ (stats.norm.cdf(z_0, 0, 1), 1-stats.norm.cdf(z_0, 0, 1))



One-sample t-test

One-sample t-test

- **Typical set up for the experiment** (same as the one-sample z-test):
  - One sample of independent test subjects, e.g. a sample of patients, a sample of customers, etc
  - Run the same experiment on each subject and collect the outcomes, e.g. give a new drug to a sample of patients and measure the effect on each individual patient; test a new web design on a sample of customers and record the time they spend on the web page, etc
  - The result contains one i.i.d. sample with **continuous numerical values**
- **Purpose**: to test if the mean of the result differs from a predefined constant
- **Data**: $x_1, \cdots, x_N$, e.g. blood pressure after taking a new drug
- **Random variable** and **assumption**: $X_1, \cdots, X_N$
  - X_i i.i.d.
  - X_i Gaussian or large N
  - X_i standard deviation σ **unknown**
- **Parameter of interest**: μ
- **Parameter estimate**: $\bar{x}$
- **Hypotheses H_0 and H_A**: given c a constant

$$
\begin{aligned}
H_0: &\quad \mu = c \\
H_A: &\quad \mu \neq c
\end{aligned}
$$

One-sample t-test (cont.)

- **Test statistic**:

$$t_0 = \frac{\bar{x} - c}{s/\sqrt{N}}$$

where $s = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2}$ is the sample standard deviation

One-sample t-test (cont.)

- **Test statistic**:

$$t_0 = \frac{\bar{x} - c}{s/\sqrt{N}}$$

where $s = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2}$ is the sample standard deviation

- **Null distribution**:
  - Student's-t distribution with degrees of freedom $df = N - 1$
  - Python: stats.t.pdf$(t, df = N - 1)$

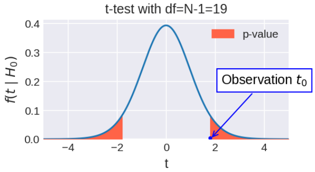
One-sample t-test (cont.)

- **Test statistic**:

$$t_0 = \frac{\bar{x} - c}{s/\sqrt{N}}$$

where $s = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2}$ is the sample standard deviation

- **Null distribution**:
  - Student's-t distribution with degrees of freedom $df = N - 1$
  - Python: stats.t.pdf$(t, df = N - 1)$
- ***p*-value**:
  - Definition: $p = 2\min\left(P(T \leq t_0 \mid H_0), P(T \geq t_0 \mid H_0)\right)$
  - Python: $2 * \min$ (stats.t.cdf$(t_0, df = N - 1)$, 1-stats.t.cdf$(t_0, df = N - 1)$)



Two-sample t-test

Two-sample t-test

- **Typical set up for the experiment** (same as the two-sample z-test):
  - Two samples of independent test subjects, where the two samples $\mathcal{X}$ and $\mathcal{Y}$ are independent from one another, e.g. two samples of independent patients, two samples of independent customers, etc
  - Run two sets of experiments A and B on the test subjects from the two samples $\mathcal{X}$ and $\mathcal{Y}$, respectively, and collect the outcomes, e.g. give different drugs to the two samples of patients and measure the effect on each individual patient; test two web designs on two samples of customers and record the time they spend on the web page, etc
  - The result contains two i.i.d. samples with **continuous numerical values**
- **Purpose**: to test if two alternative options have different effects by testing if the mean of the result from one sample differs from the mean of the other sample
- **Data**: $x_1, \cdots, x_{N_X}$ and $y_1, \cdots, y_{N_Y}$, e.g. blood pressure measured after taking two different drugs
- **Random variable** and **assumption**: $X_1, \cdots, X_{N_X}$, $Y_1, \cdots, Y_{N_Y}$
  - X_i and Y_j independent
  - X_i i.i.d.; Y_j i.i.d.
  - X_i Gaussian or large N_X; Y_j Gaussian or large N_Y
  - X_i and Y_j have **unknown** standard deviation σ_X and σ_Y, respectively
- **Parameter of interest**: μ_X, μ_Y
- **Parameter estimate**: $\bar{x}$, $\bar{y}$
- **Hypotheses** H_0 and H_A: given c a constant

$$
\begin{aligned}
H_0 : \quad & \mu_X - \mu_Y = c \\
H_A : \quad & \mu_X - \mu_Y \neq c
\end{aligned}
$$

Two-sample t-test (cont.)

- **Test statistic**:

$$t_0 = \frac{\bar{x} - \bar{y} - c}{\sqrt{\frac{s_X^2}{N_X} + \frac{s_Y^2}{N_Y}}}$$

with degrees of freedom $df = \frac{(s_X^2/N_X + s_Y^2/N_Y)^2}{(\frac{s_X^2}{N_X})^2/(N_X - 1) + (\frac{s_Y^2}{N_Y})^2/(N_Y - 1)}$

Two-sample t-test (cont.)

- **Test statistic**:

$$t_0 = \frac{\bar{x} - \bar{y} - c}{\sqrt{\frac{s_X^2}{N_X} + \frac{s_Y^2}{N_Y}}}$$

with degrees of freedom $df = \frac{(s_X^2/N_X + s_Y^2/N_Y)^2}{(\frac{s_X^2}{N_X})^2/(N_X - 1) + (\frac{s_Y^2}{N_Y})^2/(N_Y - 1)}$

- **Null distribution**:
  - Student's-t distribution with degrees of freedom df
  - Python: stats.t.pdf($t, df = df$)

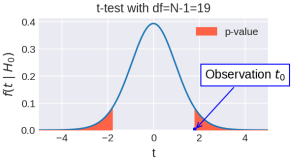
Two-sample t-test (cont.)

- **Test statistic**:

$$t_0 = \frac{\bar{x} - \bar{y} - c}{\sqrt{\frac{s_X^2}{N_X} + \frac{s_Y^2}{N_Y}}}$$

with degrees of freedom $df = \frac{(s_X^2/N_X + s_Y^2/N_Y)^2}{(\frac{s_X^2}{N_X})^2/(N_X-1) + (\frac{s_Y^2}{N_Y})^2/(N_Y-1)}$

- **Null distribution**:
  - Student's-t distribution with degrees of freedom df
  - Python: stats.t.pdf($t, df = df$)
- ***p*-value**:
  - Definition: $p = 2\min\left(P(T \leq t_0 \mid H_0), P(T \geq t_0 \mid H_0)\right)$
  - Python: $2 *$ min (stats.t.cdf(t_0, df=df), 1-stats.t.cdf(t_0, df=df))



Paired t-test

Paired t-test

- **Typical set up for the experiment**:

Paired t-test

- **Typical set up for the experiment**:
  - One sample of independent test subjects, e.g. one sample of independent patients

Paired t-test

- **Typical set up for the experiment**:
  - One sample of independent test subjects, e.g. one sample of independent patients
  - Run two sets of experiments A and B on all subjects from the sample and collect the outcomes, e.g. measure the blood pressure of the patients **before** giving them a new drug (experiment A); measure the blood pressure of the patients **after** giving them the new drug (experiment B)

Paired t-test

- **Typical set up for the experiment**:
  - One sample of independent test subjects, e.g. one sample of independent patients
  - Run two sets of experiments A and B on all subjects from the sample and collect the outcomes, e.g. measure the blood pressure of the patients **before** giving them a new drug (experiment A); measure the blood pressure of the patients **after** giving them the new drug (experiment B)
  - The result contains two samples with **continuous numerical values**

Paired t-test

- **Typical set up for the experiment**:
  - One sample of independent test subjects, e.g. one sample of independent patients
  - Run two sets of experiments A and B on all subjects from the sample and collect the outcomes, e.g. measure the blood pressure of the patients **before** giving them a new drug (experiment A); measure the blood pressure of the patients **after** giving them the new drug (experiment B)
  - The result contains two samples with **continuous numerical values**
- **Purpose**: to test if two alternative options have different effects by testing if the mean of the difference between two results differs from a predefined constant

Paired t-test

- **Typical set up for the experiment**:
  - One sample of independent test subjects, e.g. one sample of independent patients
  - Run two sets of experiments A and B on all subjects from the sample and collect the outcomes, e.g. measure the blood pressure of the patients **before** giving them a new drug (experiment A); measure the blood pressure of the patients **after** giving them the new drug (experiment B)
  - The result contains two samples with **continuous numerical values**
- **Purpose**: to test if two alternative options have different effects by testing if the mean of the difference between two results differs from a predefined constant
- **Data**: $x_1, \cdots, x_N$, $y_1, \cdots, y_N$

Paired t-test

- **Typical set up for the experiment**:
  - One sample of independent test subjects, e.g. one sample of independent patients
  - Run two sets of experiments A and B on all subjects from the sample and collect the outcomes, e.g. measure the blood pressure of the patients **before** giving them a new drug (experiment A); measure the blood pressure of the patients **after** giving them the new drug (experiment B)
  - The result contains two samples with **continuous numerical values**
- **Purpose**: to test if two alternative options have different effects by testing if the mean of the difference between two results differs from a predefined constant
- **Data**: $x_1, \cdots, x_N$, $y_1, \cdots, y_N$
- **Random variable** and **assumption**: $X_1, \cdots, X_N$, $Y_1, \cdots, Y_N$
  - $X_i - Y_i$ i.i.d.
  - $X_i - Y_i \sim \mathcal{N}\left(\mu_{X-Y}, \sigma^2_{X-Y}\right)$ or large N (CLT)
  - standard deviation unknown

Paired t-test

- **Typical set up for the experiment**:
  - One sample of independent test subjects, e.g. one sample of independent patients
  - Run two sets of experiments A and B on all subjects from the sample and collect the outcomes, e.g. measure the blood pressure of the patients **before** giving them a new drug (experiment A); measure the blood pressure of the patients **after** giving them the new drug (experiment B)
  - The result contains two samples with **continuous numerical values**
- **Purpose**: to test if two alternative options have different effects by testing if the mean of the difference between two results differs from a predefined constant
- **Data**: $x_1, \cdots, x_N$, $y_1, \cdots, y_N$
- **Random variable** and **assumption**: $X_1, \cdots, X_N$, $Y_1, \cdots, Y_N$
  - $X_i - Y_i$ i.i.d.
  - $X_i - Y_i \sim \mathcal{N}\left(\mu_{X-Y}, \sigma^2_{X-Y}\right)$ or large N (CLT)
  - standard deviation unknown
- **Parameter of interest**: μ_{X-Y}

Paired t-test

- **Typical set up for the experiment**:
  - One sample of independent test subjects, e.g. one sample of independent patients
  - Run two sets of experiments A and B on all subjects from the sample and collect the outcomes, e.g. measure the blood pressure of the patients **before** giving them a new drug (experiment A); measure the blood pressure of the patients **after** giving them the new drug (experiment B)
  - The result contains two samples with **continuous numerical values**
- **Purpose**: to test if two alternative options have different effects by testing if the mean of the difference between two results differs from a predefined constant
- **Data**: $x_1, \cdots, x_N$, $y_1, \cdots, y_N$
- **Random variable** and **assumption**: $X_1, \cdots, X_N$, $Y_1, \cdots, Y_N$
  - $X_i - Y_i$ i.i.d.
  - $X_i - Y_i \sim \mathcal{N}\left(\mu_{X-Y}, \sigma^2_{X-Y}\right)$ or large N (CLT)
  - standard deviation unknown
- **Parameter of interest**: μ_{X-Y}
- **Parameter estimate**: $m_{X-Y} = \frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)$

Paired t-test

- **Typical set up for the experiment**:
  - One sample of independent test subjects, e.g. one sample of independent patients
  - Run two sets of experiments A and B on all subjects from the sample and collect the outcomes, e.g. measure the blood pressure of the patients **before** giving them a new drug (experiment A); measure the blood pressure of the patients **after** giving them the new drug (experiment B)
  - The result contains two samples with **continuous numerical values**
- **Purpose**: to test if two alternative options have different effects by testing if the mean of the difference between two results differs from a predefined constant
- **Data**: $x_1, \cdots, x_N$, $y_1, \cdots, y_N$
- **Random variable** and **assumption**: $X_1, \cdots, X_N$, $Y_1, \cdots, Y_N$
  - $X_i - Y_i$ i.i.d.
  - $X_i - Y_i \sim \mathcal{N}\left(\mu_{X-Y}, \sigma^2_{X-Y}\right)$ or large N (CLT)
  - standard deviation unknown
- **Parameter of interest**: μ_{X-Y}
- **Parameter estimate**: $m_{X-Y} = \frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)$
- **Hypotheses H_0 and H_A**: given c a constant

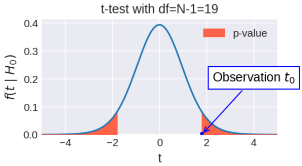
$$
\begin{aligned}
H_0: &\quad \mu_{X-Y} = c \\
H_A: &\quad \mu_{X-Y} \neq c
\end{aligned}
$$

Paired t-test

- **Test statistic**:

$$t_0 = \frac{m_{X-Y} - c}{s_{X-Y}/\sqrt{N}}$$

  where $s_{X-Y} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - y_i - m_{X-Y})^2}$
- **Null distribution**:
  - Student's-t distribution with degrees of freedom $N-1$
  - Python: stats.t.pdf($t, df = N-1$)
- ***p*-value**:
  - Definition: $p = 2\min\left(P(T \leq t_0 \mid H_0), P(T \geq t_0 \mid H_0)\right)$
  - Python: $2 * \min$ (stats.t.cdf($t_0, df = N-1$), 1-stats.t.cdf($t_0, df = N-1$))



Exercise 1

- A company claims that a new drug E they have developed can increase the average sleeping hours of people with insomnia. Design three different hypothesis tests to test this statement.

Exercise 2

- One of the tests you have designed is a two-sample test. After the experiments, you realized the test subjects being selected in the second group are parents or siblings of the first group. Would that be a problem? Can you still use the result somehow?