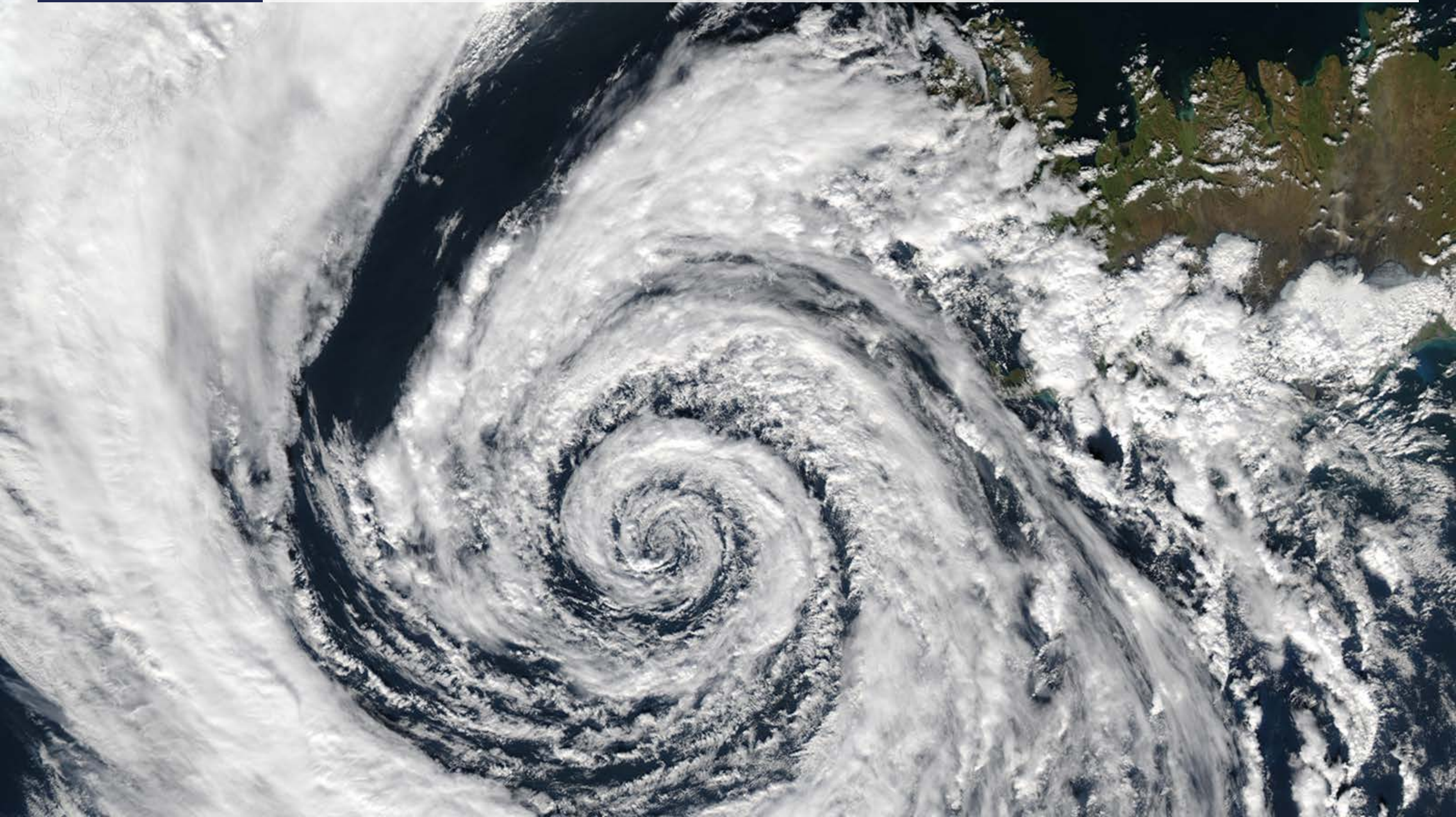
Statistics

The *t*-test

Name teacher

TU Delft



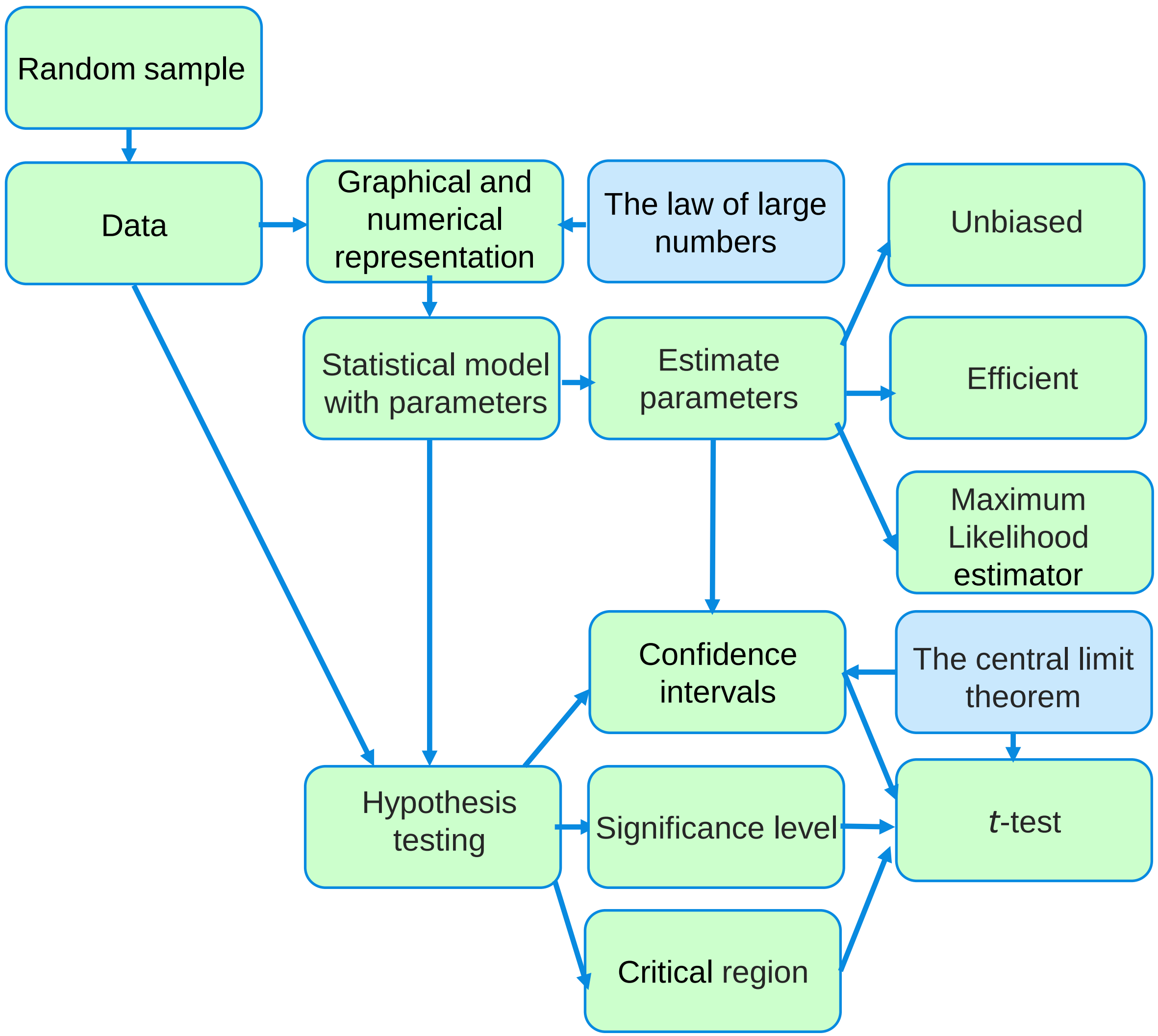
Learning objective

After this class you will be able to

- **Perform statistical hypothesis tests for the mean, under various assumptions.**
- **Perform statistical hypothesis tests for the difference between the mean of two groups .**

Book: Chapters 27 and 28

Statistics



Random sample
Data
Graphical and numerical representation
The law of large numbers
Unbiased
Statistical model with parameters
Estimate parameters
Efficient
Maximum Likelihood estimator
Confidence intervals
The central limit theorem
Hypothesis testing
Significance level
t-test
Critical region

Before this class (week 3.9 lesson 1):

Watch prelecture *'The* t-*test'*

Book: Sections 27.1 and 27.2

Programme

- The t-test for large samples
- The t-test in the regression setting
- Exercises
- Two sample t-test
- Exercises

Question from prelecture: Test $H_0 : \mu = 60$ against $H_1 : \mu < 60, \alpha = 99\%.$ The critical value for this test is

A) -2.787

B) -2.485

C) -2.492

D) -2.797

Question from prelecture: Test $H_0 : \mu = 60$ against $H_1 : \mu < 60, \alpha = 99\%$. The value of the test statistic is -2. The conclusion of the test is

A) reject the null hypothesis

B) do not reject the null hypothesis

C) can not be given

Question from prelecture: Test $H_0 : \mu = 60$ against $H_1 : \mu \neq 60, \alpha = 95\%.$ The value of the test statistic is -2. The critical value for and the conclusion of the test are

A) $t_{24,0.025}$; reject the null hypothesis

B) $t_{24,0.025}$; do not reject the null hypothesis

C) $t_{24,0.05}$; do reject the null hypothesis

D) $t_{24,0.05}$; do not reject the null hypothesis

Normal samples

Definition:

Let $X_1, X_2, \cdots, X_n$ be a sample from a $N(\mu, \sigma^2)$ distribution. To test the null hypothesis $H_0 : \mu = \mu_0$, define the *t*-test statistic T as

$$T = \frac{\bar{X}_n - \mu_0}{S_n/\sqrt{n}}$$

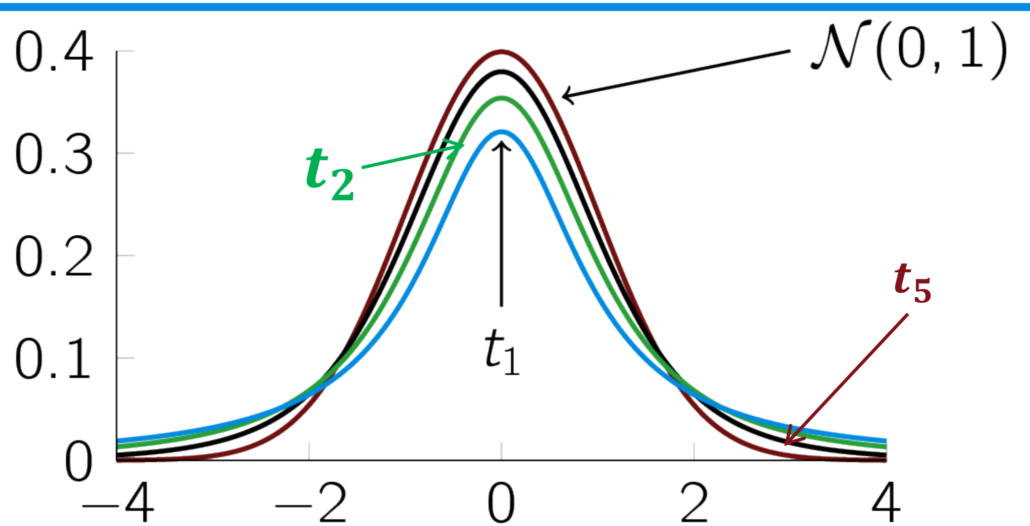
Then the distribution of this statistic under $H_0 : \mu = \mu_0$ is

$$T \sim t(n-1)$$

Normal samples

How to perform a *t*-test for samples from normal data with unknown variance at significance level α :

1. Formulate the hypotheses
2. Compute the value of the *t*-test statistic
3. Compare this value with the critical values $t_{n-1,\alpha/2}$ or $t_{n-1,\alpha}$ depending on performing a two-sided or one-sided test
4. Decide whether to reject the null hypothesis



Large samples

Let $X_1, X_2, \cdots, X_n$ be a sample from a $N(\mu, \sigma^2)$ distribution. For large n, the distribution of the *t*-test statistic can be approximated by the standard Normal distribution.

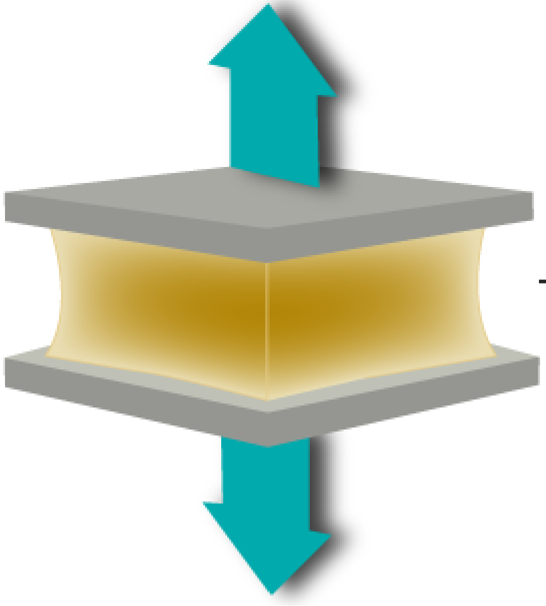
How to perform a *t*-test for large samples from normal data with unknown variance at significance level α :

1. Formulate the hypotheses
2. Compute the value of the *t*-test statistic
3. Compare this value with the critical values $z_{\alpha/2}$ or z_α depending on performing a two-sided or one-sided test
4. Decide whether to reject the null hypothesis

Exercises

Book: 27.1

27.3:



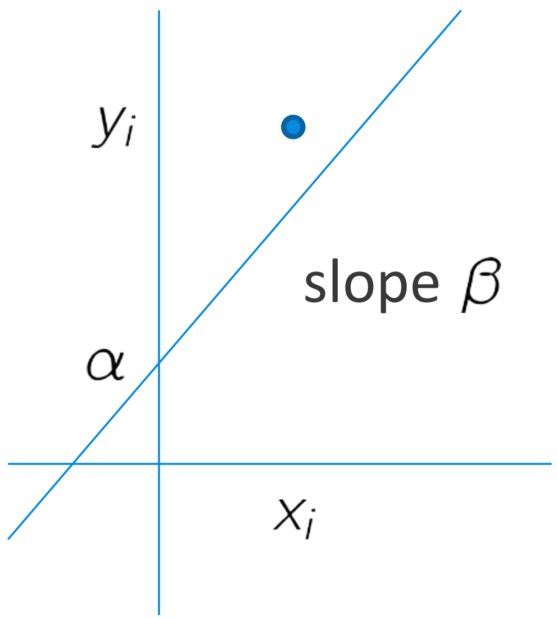
Tensile Adhesion test

The *t*-test in a regression setting: Model

Data set: $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$

Model: $Y_i = \alpha + \beta x_i + U_i$, for $i = 1, 2, \cdots, n$

The errors U_i are normally distributed with expectation 0 and constant variance.



The *t*-test in a regression setting: the slope

Hypothesis of interest: $H_0 : \beta = \beta_0$

Test statistic:

$$T_b = \frac{\hat{\beta} - \beta_0}{S_b}$$

where $\hat{\beta}$ is the least squares estimator for the slope and

$$S_b^2 = \frac{n}{n \sum x_i^2 - (\sum x_i)^2} \hat{\sigma}^2, \qquad \hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^{n} (Y_i - \hat{\alpha} - \hat{\beta} x_i)^2$$

Under $H_0 : \beta = \beta_0$ the test statistic has *t*-distribution with n-2 degrees of freedom.

The *t*-test in a regression setting: the intercept

Hypothesis of interest: $H_0 : \alpha = \alpha_0$

Test statistic:

$$T_a = \frac{\hat{\alpha} - \alpha_0}{S_a}$$

where $\hat{\alpha}$ is the least squares estimator for the intercept and

$$S_a^2 = \frac{\sum x_i^2}{n \sum x_i^2 - (\sum x_i)^2} \hat{\sigma}^2, \qquad \hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^{n} (Y_i - \hat{\alpha} - \hat{\beta} x_i)^2$$

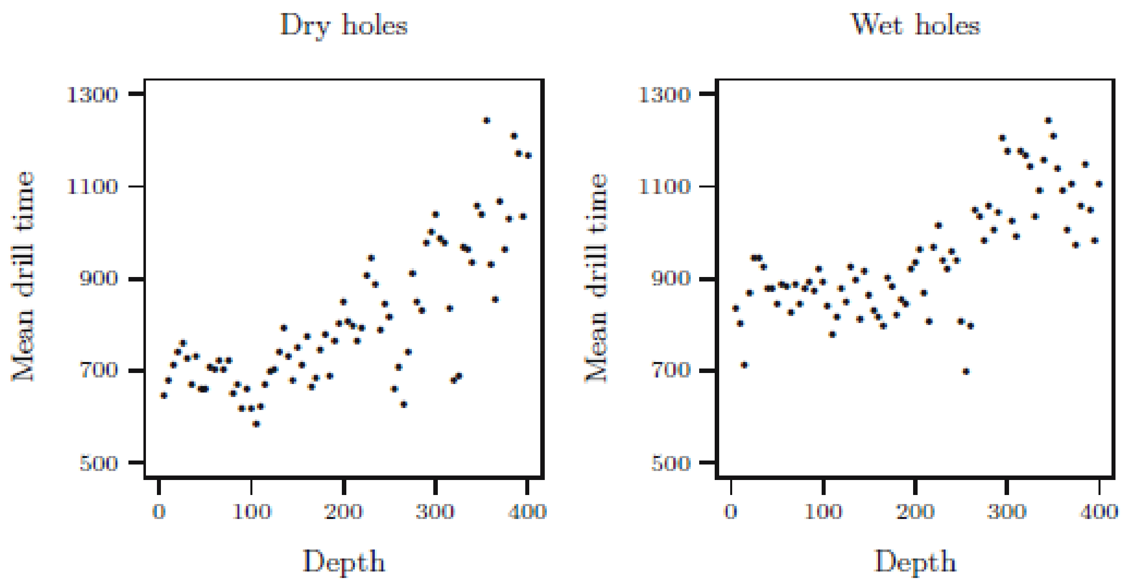
Under $H_0 : \alpha = \alpha_0$ this test statistic has also a *t*-distribution with n-2 degrees of freedom.

Exercise

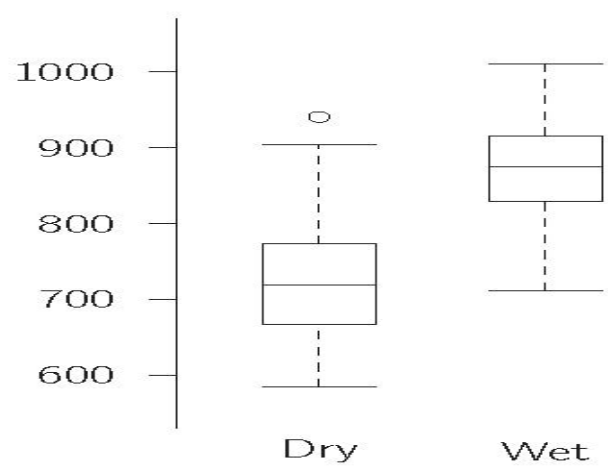
Book: 27.6

Comparing two samples

Two examples:
Mean drill time versus depth for “dry” and “wet” drilling



Boxplots:



Testing if the expectations of two samples are equal: set up

Let $X_1, X_2, \cdots, X_n$ and $Y_1, Y_2, \cdots, Y_m$ be two random samples, with expectations μ_X and μ_Y .

Hypotheses:

$H_0 : \mu_X = \mu_Y$

$H_1 : \mu_X \neq \mu_Y$ or $H_1 : \mu_X > \mu_Y$ or $H_1 : \mu_X < \mu_Y$

Two samples with equal variances: test statistic

Let $X_1, X_2, \cdots, X_n$ and $Y_1, Y_2, \cdots, Y_m$ be two random samples, with equal variances, that is $\sigma_X^2 = \sigma_Y^2$.

Definition:

The test statistic for the null hypothesis $H_0 : \mu_X = \mu_Y$ is defined by

$$T_p = \frac{\bar{X}_n - \bar{Y}_m}{S_p}$$

where

$$S_p^2 = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2}\left(\frac{1}{n} + \frac{1}{m}\right)$$

$$T_p = \frac{\bar{X}_n - \bar{Y}_m - (\mu_X - \mu_Y)}{S_p}$$

is called the pooled studentised mean difference.

Two samples with equal variances: normal case

Let $X_1, X_2, \cdots, X_n$ be a random sample from $N(\mu_1, \sigma^2)$
and let $Y_1, Y_2, \cdots, Y_m$ be a random sample from $N(\mu_2, \sigma^2)$
Then, under the null hypothesis of equal expectations,
$H_0 : \mu_1 = \mu_2$:

$$T_p = \frac{\bar{X}_n - \bar{Y}_m}{S_p} \sim t(n + m - 2)$$

Two samples with equal variances: large samples

Let $X_1, X_2, \cdots, X_n$ be a random sample from $N(\mu_1, \sigma^2)$
and let $Y_1, Y_2, \cdots, Y_m$ be a random sample from $N(\mu_2, \sigma^2)$.
If n and m are large enough, the distribution of T_p
can be approximated by the standard Normal distribution.

Two exercises

Book: 28.1 (a)

For next class (week 3.9 lecture 2):

Prepare practice exam

Book: Relevant sections of Chapters 13 - 28

After this class you are

- **Well prepared for the final exam**

Statistics

Good luck!