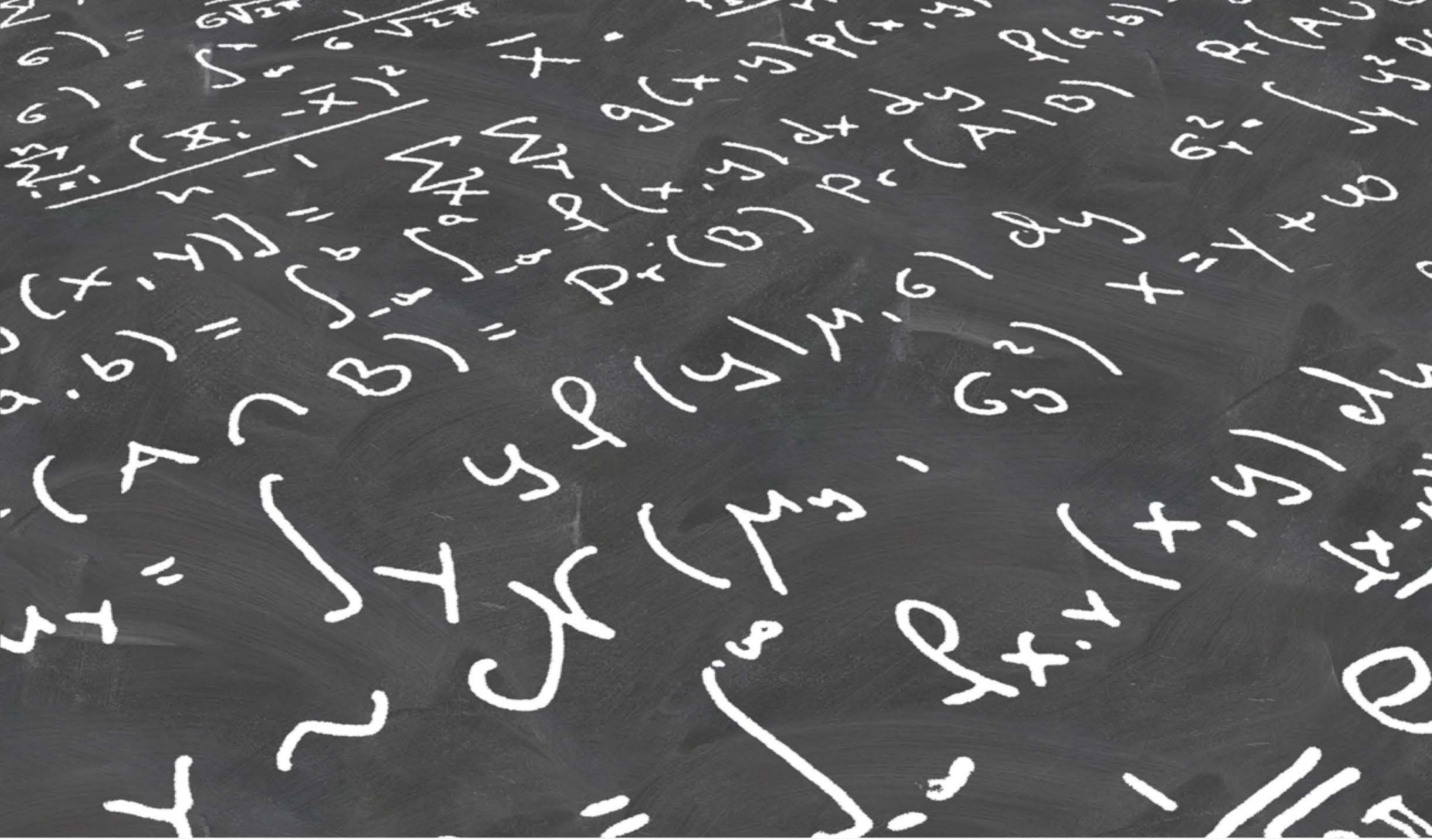


10. Hypothesis Test

Gunvor Elisabeth Kirkelund
Lars Mandrup

Slides and material provided in parts by Henrik Pedersen

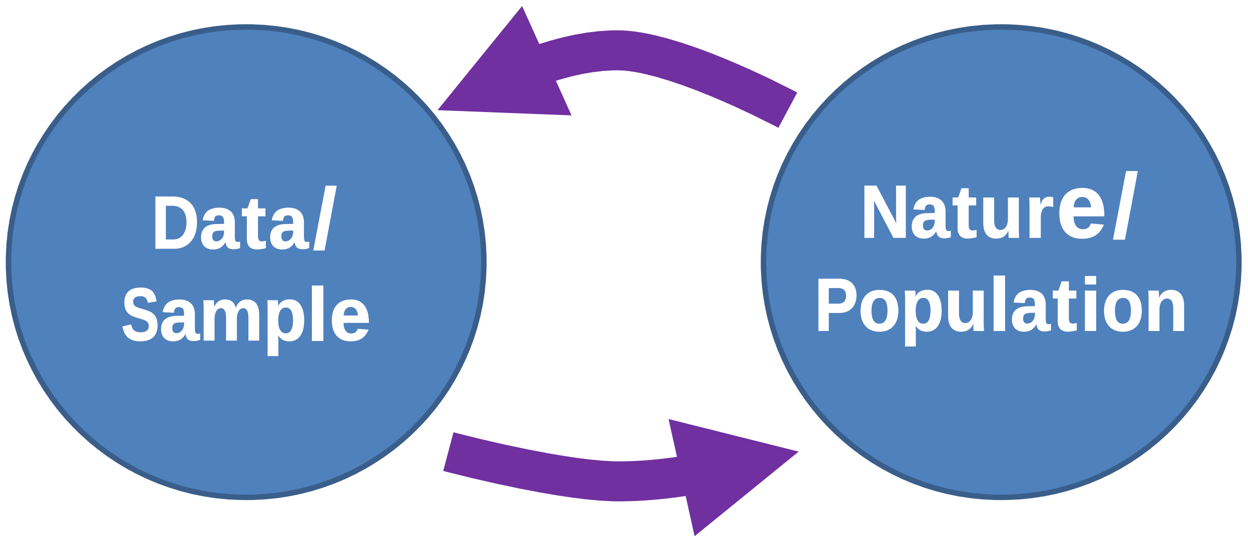
Todays Content

- Repetition from last time
- Hypothesis Test
- p-values
- Test of the mean with known variance (z-test)
- Test of the mean with unknown variance (t-test)

Introduction to Statistics

Probability theory

Given the cause (population), what should the data (sample) look like?



Statistics

Given the data (sample), what caused them (population)?

- Testing a hypotesis
- Estimating means and variances

- If we don’t know better: We assume data are normally distributed

Estimator

Estimator:

- An estimator $\hat{\theta}(X)$ is a statistic used to estimate the unknown parameter θ of a random sample X.
- An estimator is unbiased if $E\left[\hat{\theta}\right] = \theta.$

Unbiased estimators:

- The sample mean: $\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

Degrees of freedom

- The sample variance: $$\hat{s}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Degrees of freedom

Statistical Model

Statistical model:

- A random sample and its pdf, $f_X(x;\theta)$, where θ is the parameter(s) of the pdf.

- Because of the Central Limit Theorem (CLT) we often can use the normal distribution $\mathcal{N}(\mu, \sigma^2)$ with mean μ and variance σ^2 as statistical model for the sample mean $\bar{X}$

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim \mathcal{N}(\mu, \sigma^2) \quad \text{(n>30)}$$

Test Statistics

Test statistics:

- A random variable that summarized a data-set by reducing the data to one value that can be used to perform the hypothesis test.

- For a sample assumed to follow the normal distribution $\mathcal{N}(\mu, \sigma^2)$ with <u>known mean</u> μ and <u>variance</u> σ^2 we can use the z-statistics (z-score):

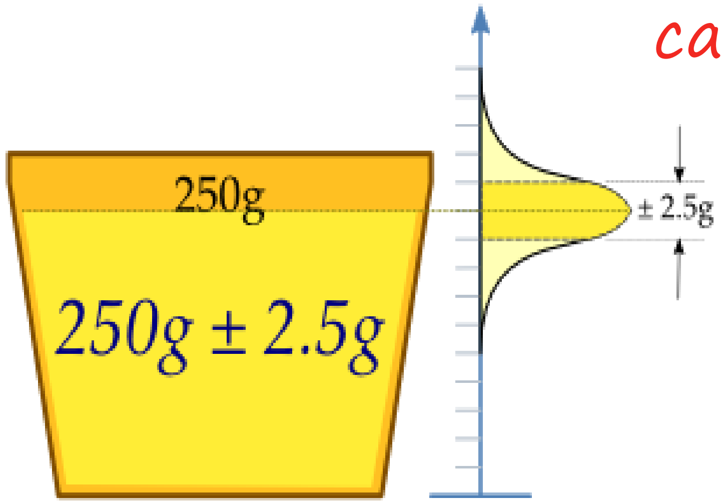
$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \sim \mathcal{N}(0,1)$$

Standard (normalized) normal distribution (μ=0 and σ^2=1)

➢ **Assignment:** Show that if $\bar{x} \sim \mathcal{N}(\mu, \sigma^2)$ then $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$ (ie. having μ=0 and σ^2=1)

Cup Example

- A machine fills cups with a liquid, the content of the cups is 250 grams of liquid.
- The machine cannot fill with exactly 250 grams, the content added to individual cups shows some variation, and is considered a random variable, X.



If the machine is adequately calibrated, X is normally distributed

$$X \sim N(\mu, \sigma^2)$$

with mean $\mu = 250$ g and standard deviation $\sigma = 2.5$ g

Cup Example

ONE sample of the population!

- To determine if the machine is adequately calibrated, a sample of $n = 25$ cups of liquid is chosen at random and the cups are weighed.
- The resulting measured masses of liquid are $X_1, X_2, \ldots, X_{25}$, a random sample from X.

Population — All cups for all times

- To get an impression of the population mean (μ), we use the average (or sample mean) as an estimate:

Sample mean is NOT the expected value (true mean)!

^ means an estimator of the true polulation value

$$\hat{\mu} = \frac{1}{25} \sum_{i=1}^{25} X_i = 250.2\ g$$

- Is the machine adequately calibrated?

Hypothesis Test

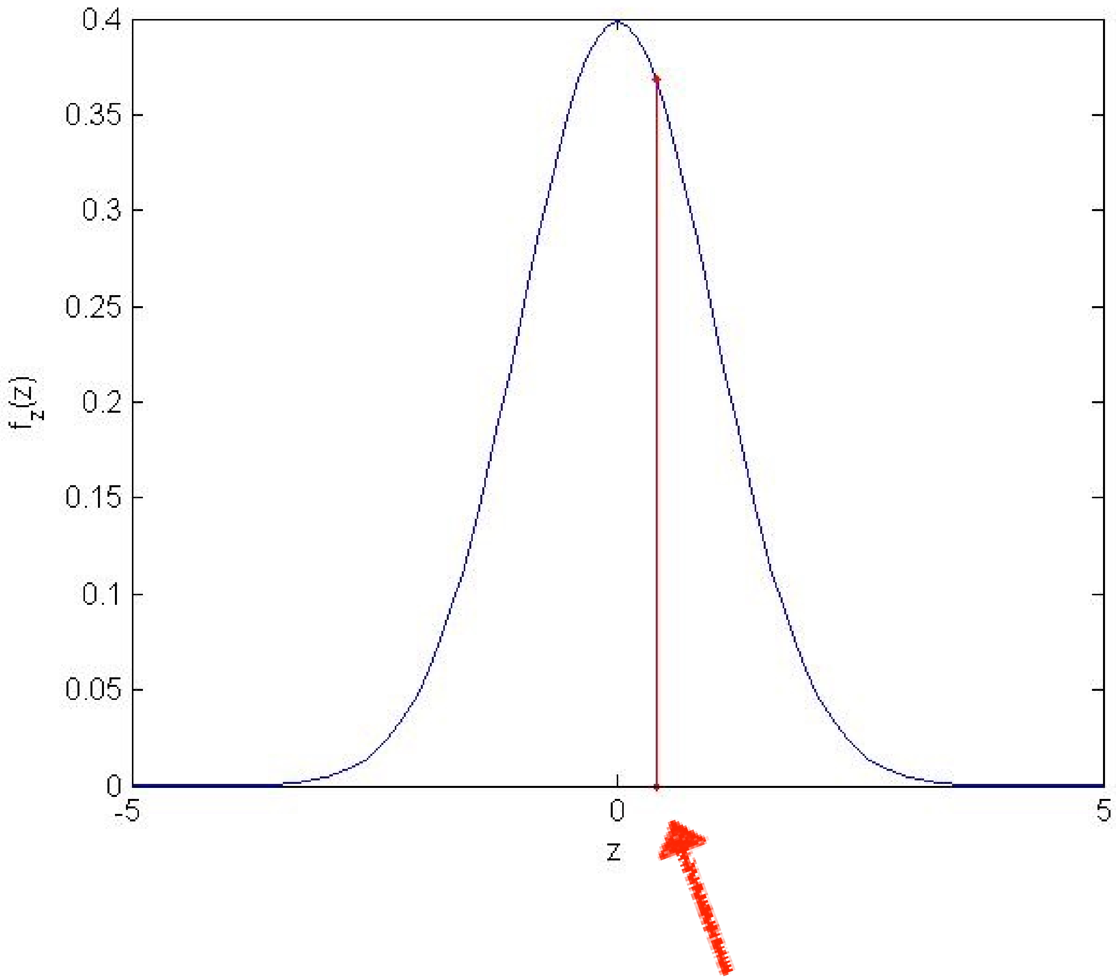
Since we know both the true mean μ and variance σ^2, we use the test statistics z

Test statistics: $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$

Does it seem plausible that z=0.4 is an observation drawn from a standard normal distribution?

Same as asking: what is the probability of observing a test size (z) that is more extreme than 0.4?

Standard normal distribution (PDF)



$$z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{250.2-250}{2.5/\sqrt{25}} = 0.4$$

Hypothesis

- **Definition – Null hypothesis (H_0)**
  - The statement being tested in a test of statistical significance is called the **null hypothesis**. The test of significance is designed to assess the strength of the evidence against the null hypothesis.
  - Usually, the null hypothesis is a statement of 'no effect‘, 'no difference‘ or ‘no relation' between the phenomena whose relation is under investigation.
- **Definition – Alternative hypothesis (H_1)**
  - The statement that is hoped or expected to be true instead of the null hypothesis is the **alternative hypothesis**
  - The alternative hypothesis, as the name suggests, is the alternative to the null hypothesis: it states that there is some 'effect/difference’ or some ‘kind of relation'.

Important!

- One cannot “prove” a null hypothesis, one can only test how close it is to being true.
- Therefore, we never say that we *accept* the null hypothesis, but that we either ***reject it*** or ***fail to reject it***.

Hypothesis

An example of a null hypothesis:

- A certain drug may reduce the chance of having a heart attack.
- Possible null hypothesis H_0: “This drug has no effect on the chances of having a heart attack".
- An alternative hypothesis H_1: “This drug has an effect on the chances of having a heart attack”.
- The test of the hypothesis consists of giving the drug to half of the people in a study group as a controlled experiment.
- If the data show a statistically significant change in the people receiving the drug, the null hypothesis is rejected.

Hypothesis

- The term "**null hypothesis**” H_0 is a general statement or default position that there is no relationship between two measured phenomena, or no association among groups.
- Rejecting or disproving the null hypothesis is a central task in the modern practice of science; the field of statistics gives precise criteria for rejecting a null hypothesis.
- The null hypothesis H_0 is generally assumed to be true until evidence indicates otherwise.
- A null hypothesis is rejected if the observed data are significantly unlikely to have occurred if the null hypothesis were true. In this case an alternative hypothesis H_1 is accepted in its place – concluding that there are grounds for believing that there *is* a relationship between two phenomena.
- If the data are consistent with the null hypothesis, then the null hypothesis is not rejected (i.e., accepted).
- **In neither case is the null hypothesis or its alternative proven**; the null hypothesis is tested with data and a decision is made based on how **likely or unlikely** the data are. This is analogous to a criminal trial, in which the defendant is assumed to be innocent (null is not rejected) until proven guilty (null is rejected) beyond a reasonable doubt (to a statistically significant degree).

Hypothesis testing

- Hypothesis testing works by collecting a randomly selected representative sample X (data) and measuring how likely the particular set of data is, assuming the null hypothesis H_0 is true: $Pr(X|H_0)$

- The data-set is usually specified via a **test statistic** which is designed to measure the extent of apparent departure from the null hypothesis – fx. z-statistic or t-statistic.

- If the data-set of a randomly selected representative sample is very unlikely relative to the null hypothesis – i.e. only rarely (usually in less than either 5% or 1% (the significance level α)) will be observed – we rejects the null hypothesis concluding it (probably) is false.

- If the data do not contradict the null hypothesis, then only a weak conclusion can be made: namely, that the observed data set provides no strong evidence against the null hypothesis. In this case, because the null hypothesis could be true or false, it is interpreted as there is no evidence to support changing from a currently useful regime (the null hypothesis) to a different one.

Example of Hypothesis

- **Example 1 – cup filling example**
  - If the machine is adequately calibrated, the true population mean should be 250 grams. Hence, the null hypothesis is

$$H_0\!: \mu = 250$$

  - If we are not concerned about the direction of a possible deviation from $\mu = 250$, the alternative hypothesis is

$$H_1\!: \mu \neq 250$$

Significance Level

- **Definition 9 – Significance level**

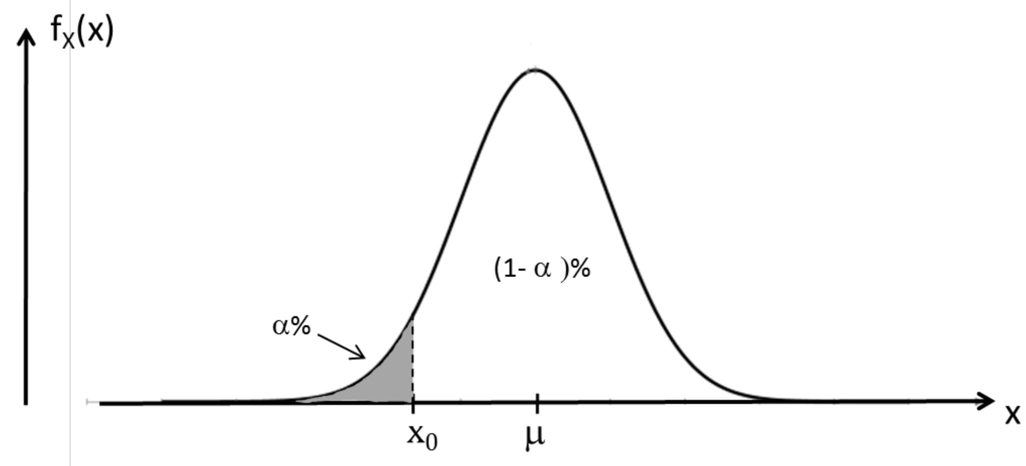
  - **The statistical significance level α is the lower limit we will accept for the probability of getting a more extreme result assuming the null hypothesis H_0 is true.**

  - The significance level is the cutoff level to reject the null hypothesis: If the probability of a randomly selected representative sample X under the assumption that the null hypothesis H_0 is true, is less than the significance level, we will reject the null hypothesis H_0.

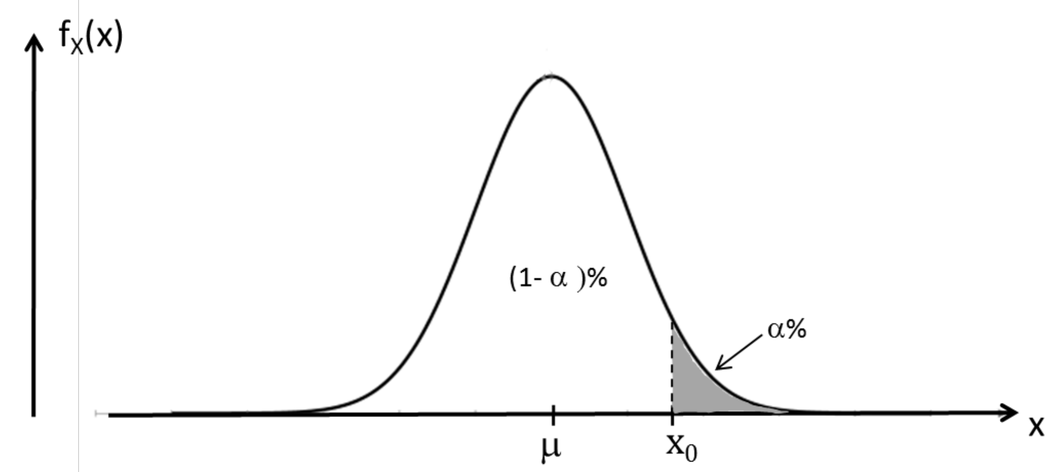
  - The most common used significance level is $\alpha = 0{,}05$ (5%)

Significance Level

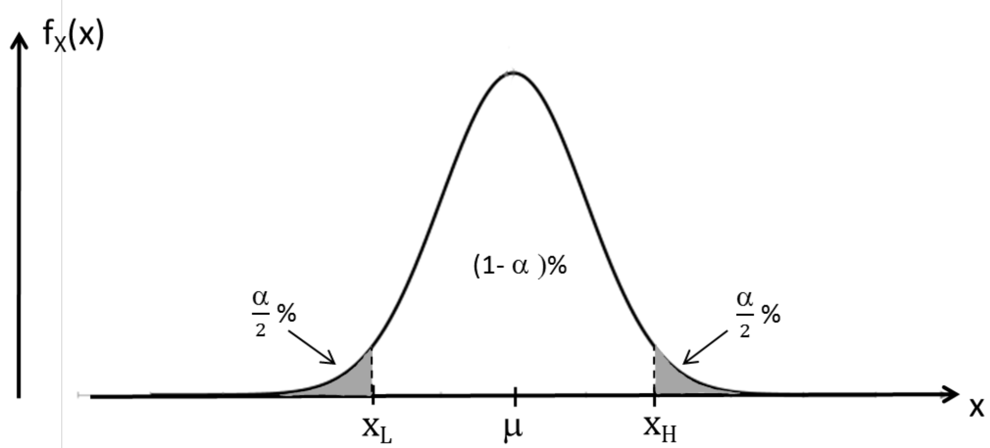
Left-tailed



Right-tailed



Two-tailed



p-value

The p value is found with matlab or in a table.

- **Definition 10 – p-value**
  - **The p-value is the probability of getting a result equal to or more extreme than the observed test-sample X under the assumption of a null hypothesis H_0:**

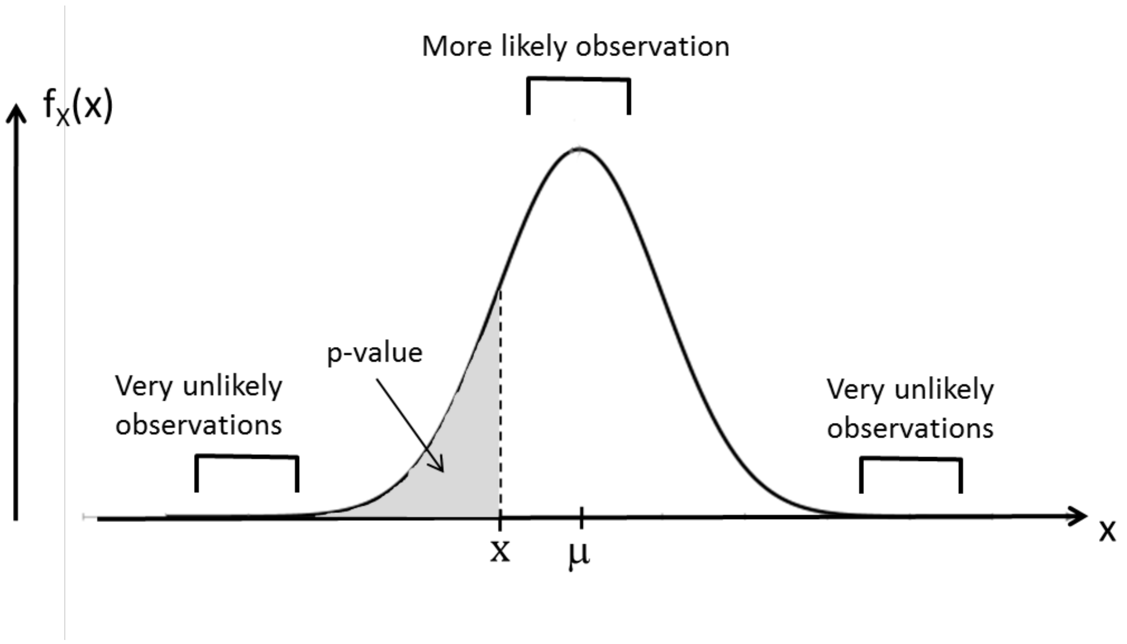
$$\boldsymbol{p - value = Pr(Worse\ result\ than\ X|H_0)}$$

- If $\boldsymbol{x}$ denotes the observed quantity, the p-value is:
  - $\boldsymbol{Pr(X \geq x|H_0)}$ for a right-tailed event
  - $\boldsymbol{Pr(X \leq x|H_0)}$ for a left-tailed event
  - $2 \cdot min\{\boldsymbol{Pr(X \leq x|H_0)}, \boldsymbol{Pr(X \geq x|H_0)}\}$ for a two-tailed event

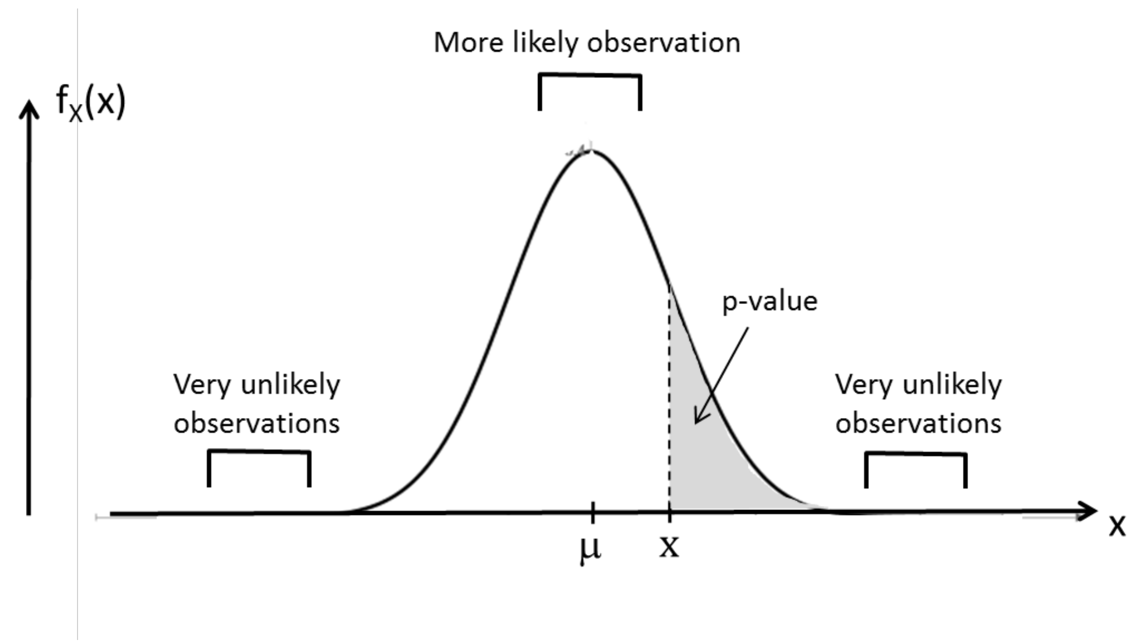
- By comparing the p-value for the test with the significance level α we can deside whether the null hypothesis H_0 should be rejected (p< α) or not (p>α).

p-value

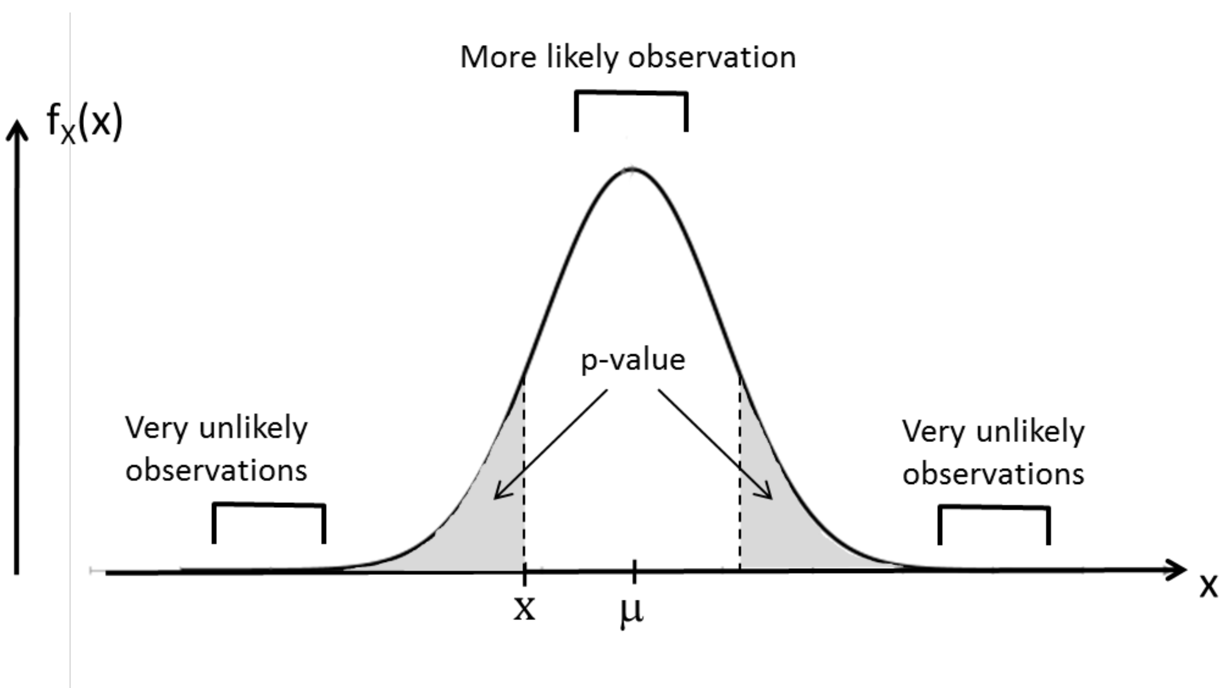
Left-tailed $p = Pr(X \leq x|H_0)$



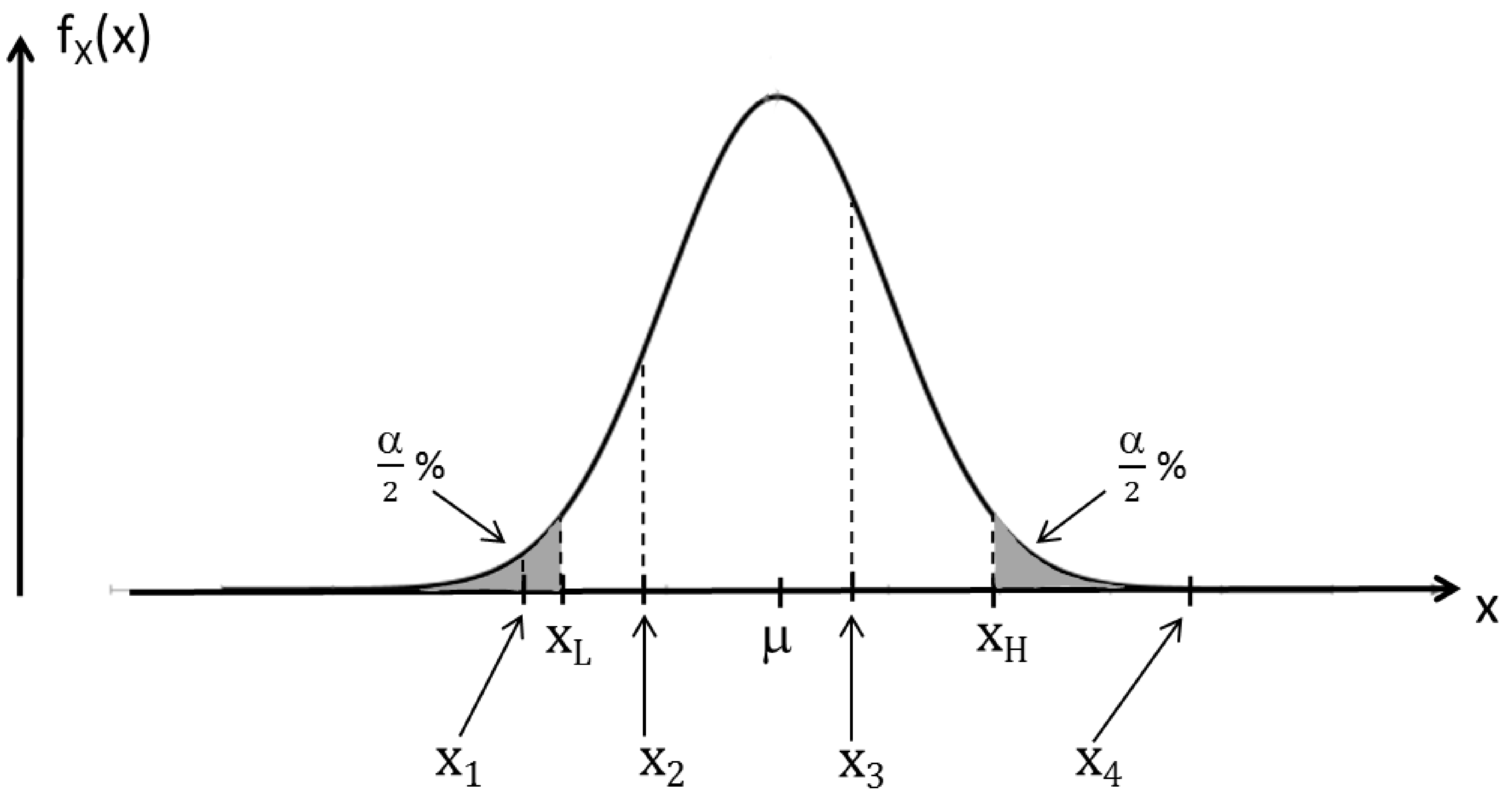
Right-tailed $p = Pr(X \geq x|H_0)$



Two-tailed $p = 2 \cdot min\{Pr(X \leq x|H_0), Pr(X \geq x|H_0)\}$



Hypothesis testing



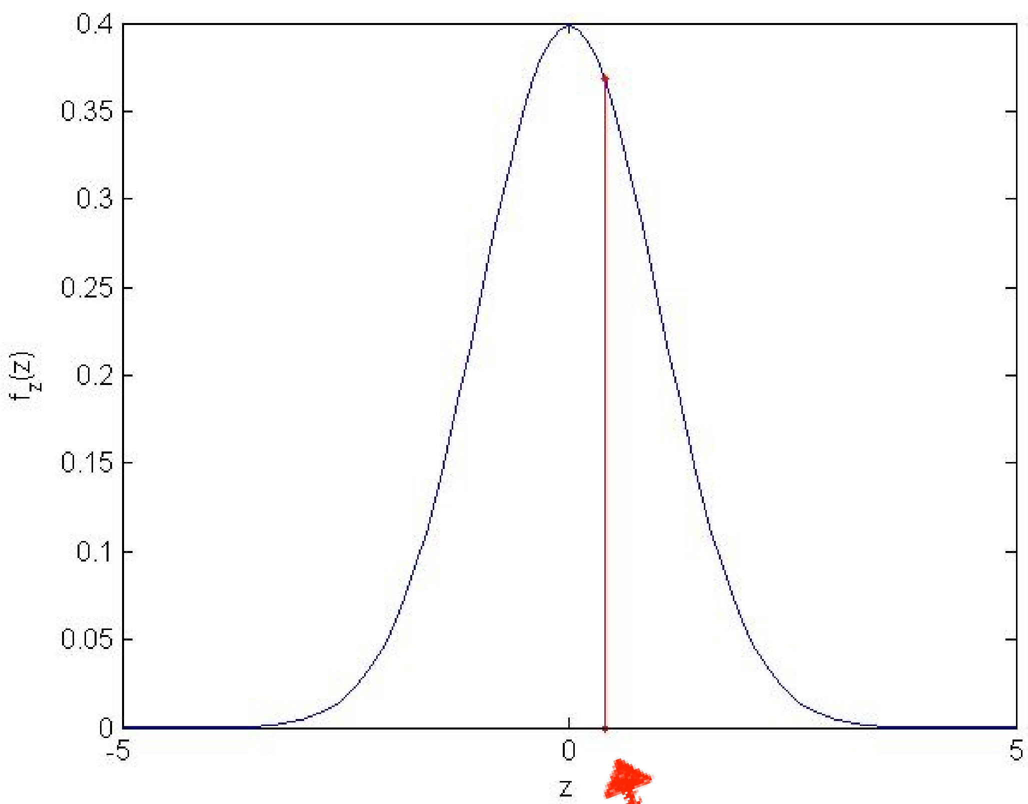
Observations (test-samples): x_1 and x_4 → Reject H_0 ($p<\alpha$)

x_2 and x_3 → Failed to reject H_0 ($p>\alpha$)

Cup Example

Here, we fail to reject the null hypothesis (H_0: $\mu = 250$), because the p-value is larger than $\alpha = 0.05$.

Standard normal distribution (PDF)



p-value:

$$
\begin{aligned}
&Pr(Z > z \cup Z < -z) \\
&= \Pr(Z > z) + \Pr(Z < -z) \\
&= 1 - \Pr(Z \leq z) + 1 - \Pr(Z \leq z) \\
&= 2 \cdot (1 - \Pr(Z \leq z)) \\
&= 2 \cdot (1 - \Pr(Z \leq 0.4)) \\
&= 2 \cdot \big(1 - \Phi(0.4)\big) \\
&= 2 \cdot (1 - 0.6554) \\
&= 0.6892
\end{aligned}
$$

`normcdf(0.4)`

Compare with α= 0.05

Test statistics: $z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \dfrac{250.2 - 250}{2.5/\sqrt{25}} = 0.4$

Tests and Types of Errors

- There will sometimes be wrong conclusions in hypothesis testing
- We can classify the errors in hypothesis testing as:

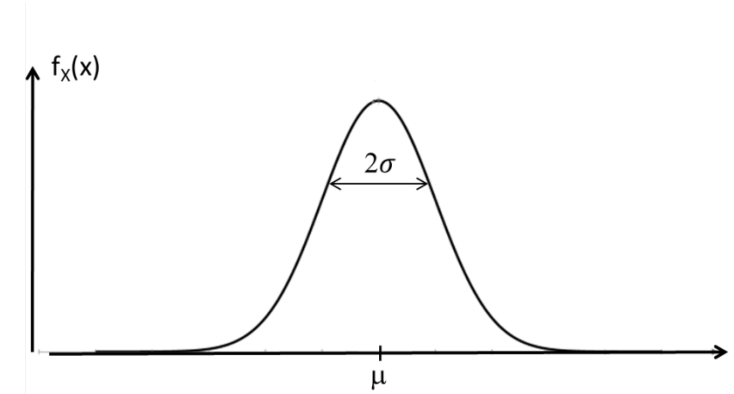
Table of error types		**Null hypothesis H_0**	
		True	**False**
Hypothesis test result	**Reject**	Type I Error (False positive)	Correct inference (True positive)
	Fail to reject	Correct inference (True negative)	Type II Error (False negative)

- The Type I Error rate is the significance level α
- Decreasing the Type I Error rate (α) will increase the Type II Error rate

The Normal Distribution

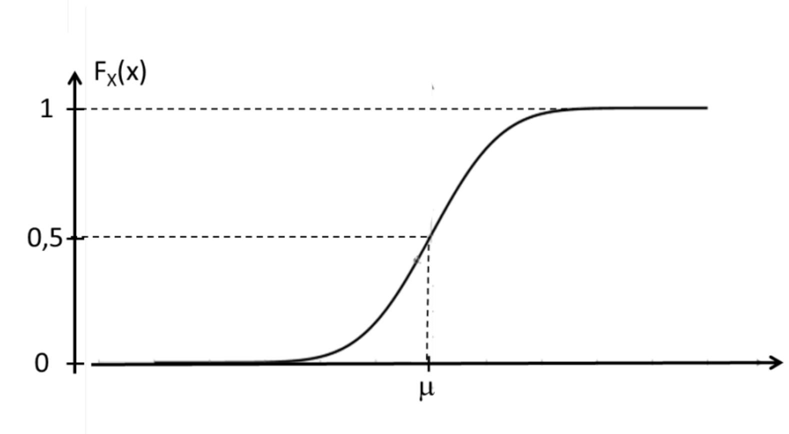
- Let X be a normally distributed random variable with mean μ and variance σ^2:

$$X \sim \mathcal{N}(\mu, \sigma^2)$$



- Probability density function (pdf):

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



- There is no closed expression for the cdf.
  We use a lookup table for the standardized x:

$$z = \frac{x-\mu}{\sigma} \sim \mathcal{N}(0,1)$$

- The cdf is found as:

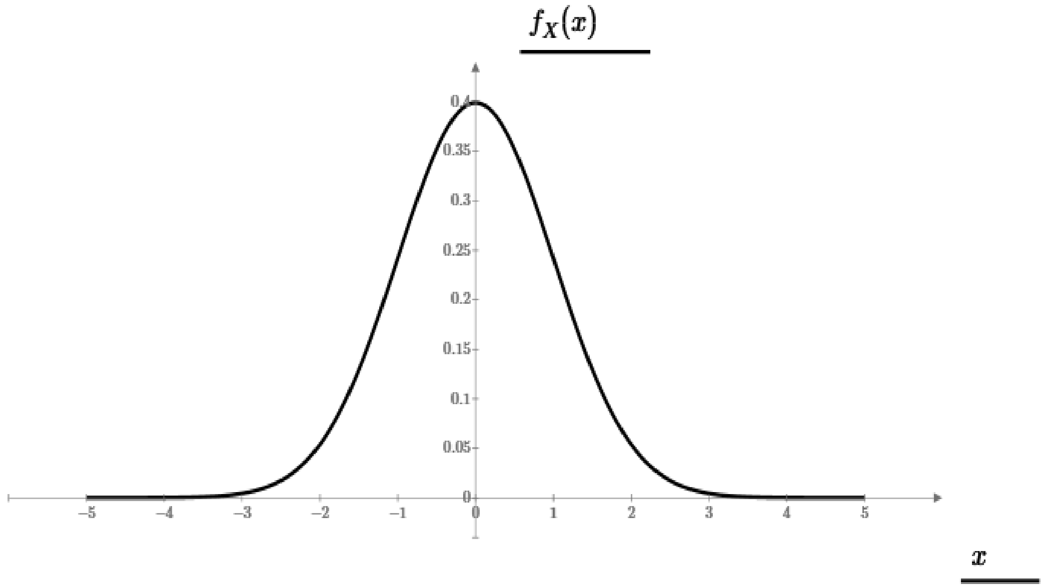
$$F_X(x) = Pr(X \le x) = Pr\left(\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}\right) = \Phi(z) = 1 - Q(z)$$

- $\Phi(z)$ is the cdf of the standardized normal random variable Z used in most tabels
- $Q(z) = 1 - \Phi(z) = 1 - F_X(x) = Pr(X \ge x)$ is the tabelled quantity in app. D in ”Random Signals”

The Normal Distribution

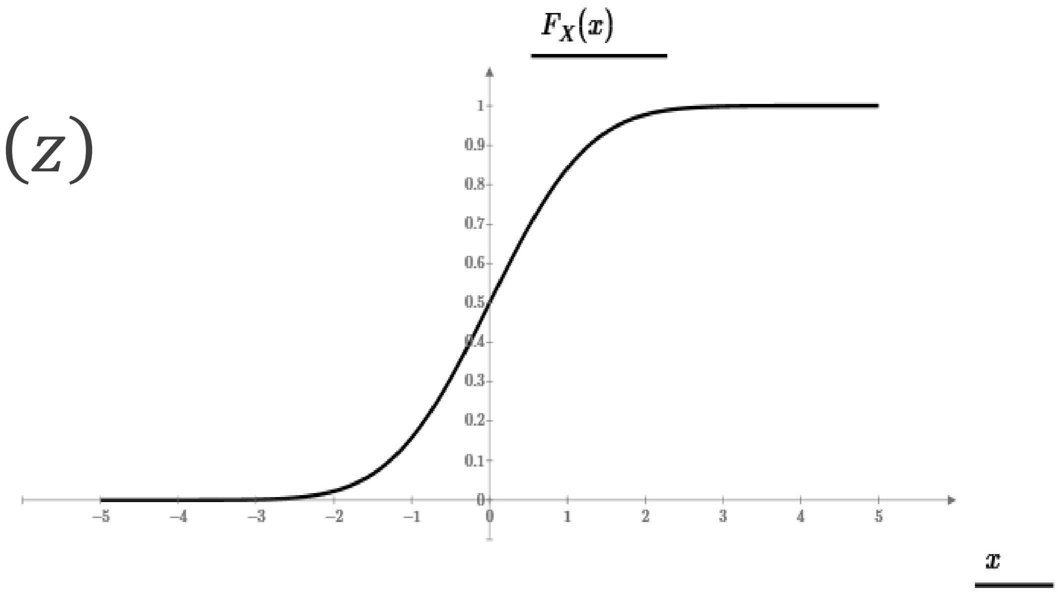
- Note that due to symmetry of the pdf:

$$Pr(Z \le -z) = Pr(Z \ge z) = 1 - Pr(Z \le z)$$



- Which imply that:

$$\Phi(-z) = 1 - \Phi(z) \quad \text{and} \quad \mathrm{Q}(-z) = 1 - \mathrm{Q}(z)$$



- Also due to symmetry:

$$Pr(Z \le 0) = \Phi(0) = Q(0) = \frac{1}{2}$$

The Normal Distribution in Matlab

- Calculating probabilities of a normally distributed random variable

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

- In Matlab: $f_X(x) =$ `normpdf(x,μ,σ)` $=$ σ·`normpdf(x)+μ` density function

  $F_X(x) = \Pr(X \leq x) =$ `normcdf(x,μ,σ)` cumulative function

- where μ is the mean and σ is the standard deviation ($\sigma = \sqrt{\sigma^2}$).

- If X is standard normally distributed $\sim\mathcal{N}(0,1)$ (ie. μ=0, σ=1), you can skip the arguments μ and σ:

```
normpdf(x) = normpdf(x,0,1)
normcdf(x)= normcdf(x,0,1)= Φ(x)
```

Simplified Calculation of the p-value from a z-Statistic

- Test size: $z \sim \mathcal{N}(0,1)$
- p-value (two-tailed event): $\boxed{pval = 2 \cdot (1 - \Phi(|z|))}$
  - If z is negative:

$$\begin{aligned} pval &= 2 \cdot min\{\boldsymbol{Pr}(\boldsymbol{Z} \leq \boldsymbol{z}), \boldsymbol{Pr}(\boldsymbol{Z} \geq \boldsymbol{z})\} \\ &= 2 \cdot \boldsymbol{Pr}(\boldsymbol{Z} \leq \boldsymbol{z}) \\ &= 2 \cdot \Phi(z) = 2 \cdot \big(1 - \Phi(-z)\big) = 2 \cdot (1 - \Phi(|z|)) \end{aligned}$$

  - If z is positive:

$$\begin{aligned} pval &= 2 \cdot min\{\boldsymbol{Pr}(\boldsymbol{Z} \leq \boldsymbol{z}), \boldsymbol{Pr}(\boldsymbol{Z} \geq \boldsymbol{z})\} \\ &= 2 \cdot \boldsymbol{Pr}(\boldsymbol{Z} \geq \boldsymbol{z}) = 2 \cdot (1 - \boldsymbol{Pr}(\boldsymbol{Z} \leq \boldsymbol{z})) \\ &= 2 \cdot \big(1 - \Phi(z)\big) = 2 \cdot (1 - \Phi(|z|)) \end{aligned}$$

Confidence Interval

- **Conficence Interval**
  - The $1-\alpha$ confidence interval is an interval $[\theta_-; \theta_+]$ such that the probability that the true value of the unknown parameter θ lies within the interval is $1-\alpha$:

$$Pr(\theta_- \leq \theta \leq \theta_+) = 1-\alpha$$

- For a two-tailed event with significance level $\alpha = 0{,}05$ the 95% confidence interval for the mean $[\mu_-; \mu_+]$ is given by:

$$Pr(\mu_- \leq \mu \leq \mu_+) = 1-\alpha = 0{,}95$$

- where the interval endpoints are:

$\Phi^{-1}(0{,}975)$

$$\mu_- = \bar{x} - 1{,}96\frac{\sigma}{\sqrt{n}} \qquad \text{and} \qquad \mu_+ = \bar{x} + 1{,}96\frac{\sigma}{\sqrt{n}}$$

Confidence Interval

- In the cup-filling example, the 95% confidence interval is

$$[\mu_-; \mu_+] = \left[\bar{x} - 1.96 \cdot \frac{\sigma}{\sqrt{n}};\ \bar{x} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}\right]$$

$$= \left[250.2 - 1.96 \cdot \frac{2.5}{\sqrt{25}};\ 250.2 + 1.96 \cdot \frac{2.5}{\sqrt{25}}\right]$$

$$= [250.2 - 0.98;\ 250.2 + 0.98] = [249.22;\ 251.18]$$

ie. μ=250 can't be rejected

Test catalog for the mean (known variance)

- **Statistical model:**
- $X_1, X_2, \dots, X_n$ are i.i.d. samples of a random variable *X* with mean μ and variance σ^2.
- Parameter estimate:

$$\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu, \sigma^2/n)$$

- Where the observation is $\bar{x}$ = 'the average of *n* samples drawn from *X*'s distribution'.
- NOTE: The statistical model is only true if *n* is sufficiently large ($n \geq 30$) <u>or</u> if the samples are drawn from a normal population with mean μ and variance σ^2.

- **Hypothesis test (two-tailed):**
- $H_0: \mu = \mu_0$
- $H_1: \mu \neq \mu_0$
- Test size: $z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$
- Approximate p-value: $2 \cdot |1 - \Phi(|z|)|$

- **95% confidence interval:**
- $\mu_- = \bar{x} - 1.96 \cdot \sigma/\sqrt{n}$
- $\mu_+ = \bar{x} + 1.96 \cdot \sigma/\sqrt{n}$

t-Score

When the variance is unknown!

- Now, consider the usual *z* statistic

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

- The equivalent statistic, when replacing the standard deviation (σ) with the empirical standard deviation (s), is called a *t*-score

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

Called the student's t-distribution

- The *t*-score is *not* normally distributed; it is *t*-distributed with $v = n - 1$ degrees of freedom, which we write

$$\boxed{t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \sim t(n-1)}$$

n-1: degrees of freedom

Student's t-distribution

- **Students t-distribution: t(ν)=t(n-1):**

ν: Degrees of freedom

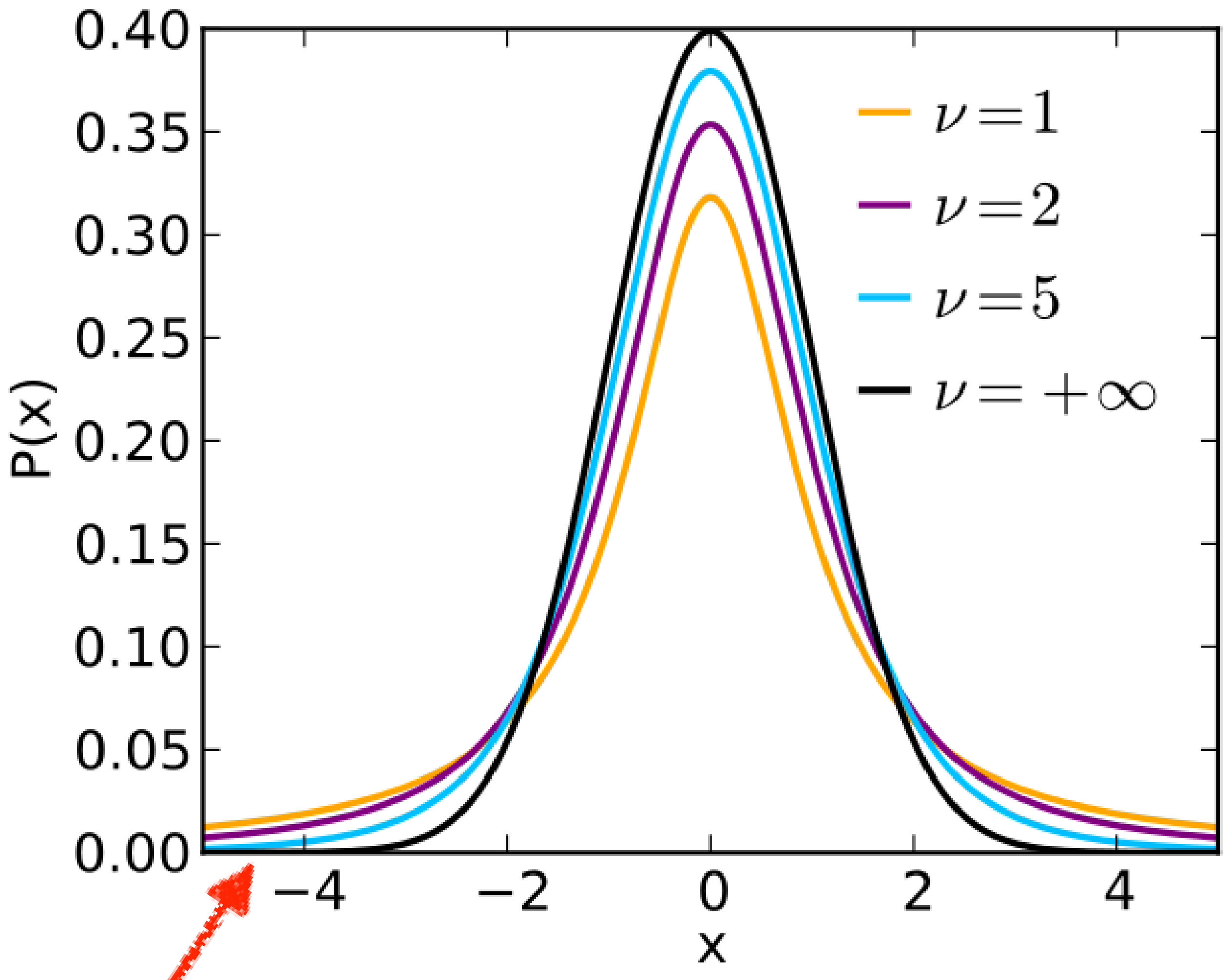
n: Number of samples

  - pdf: $f_X(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\,\Gamma(\frac{\nu}{2})}(1+\frac{x^2}{\nu})^{-\frac{\nu+1}{2}}$

    where the gamma-function:

    $$\Gamma(n) = \int_0^{\infty} y^{n-1}e^{-y}dy$$

  - Even/ symmetric: $f_X(x) = f_X(-x)$
  - Mean: $\mu_t = 0$ for $\nu>1$
  - Variance: $\sigma_t^2 = \frac{\nu}{\nu-2}$ for $\nu>2$
  - $n \to \infty$: $t(n-1) \sim \mathcal{N}(0,1)$



t(ν) is heavy (large tail) for small ν

When the variance is unknown!

Find the Mean Using the t-score

- To test the nul hypothesis H_0: $\mu = \mu_0$ using the t-statistic instead of the z-statistic, the p-value is:

$$pval = 2 \cdot (1 - t_{cdf}(|t|, n-1))$$

- where $t_{cdf}(t, n-1) = \Pr(T \le t)$ denotes the CDF of a t distribution with n-1 degrees of freedom.

- The 95% confidence interval for the mean is: $\bar{x} \pm t_0 \cdot s/\sqrt{n}$
- where t_0 is chosen such that:

$$\Pr(T \le t_0) = t_{cdf}(t_0, n-1) = 1 - \frac{\alpha}{2} = 1 - \frac{0.05}{2} = 0.975$$

- ie: $t_0 = t^{-1}_{cdf,n-1}\left(1 - \frac{\alpha}{2}\right) = t^{-1}_{cdf,n-1}(0.975)$

Depends on n-1
(degrees of freedom)

The t-Distribution in Matlab

- Calculating probabilities of a t-distributed random variable

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

- $f(t)$ = `tpdf(t,n-1)`
- $\Pr(T \leq t)$ = `tcdf(t,n-1)`

- where n is the number of samples.
- Given a probability $1 - \alpha/2$, what is the corresponding value t_0, such that $\Pr(T \leq t_0) = 1 - \alpha/2$?

two-tailed event

- `t0 = tinv(1-alpha/2,n-1)`

The Mean for a Population with Unknown Variance

```
X =[253.50
    254.70
    256.23
    244.14
    252.78
    247.71
    249.52
    253.57
    248.86
    250.86
    251.87
    247.90
    249.59
    246.69
    251.45
    252.83
    249.50
    245.41
    250.19
    250.21
    250.17
    246.27
    248.24
    251.60
    251.21];
```

Data (n=25)

Then the sample mean ($\bar{X}$) is:

```
>>MeanX=mean(X)
            MeanX = 250.2000
```

And the (unbiased) estimate of the variance (s^2) is:

```
>>VarX=var(X)
            VarX = 8.5868
```

Corresponding to an empitical standard diviation (s):

```
>>StdX=sqrt(VarX)
            StdX = 2.9303
```

The Mean for a Population with Unknown Variance

- Recall that in the cup filling machine example, the null hypothesis is:

$$H_0: \quad \mu = 250$$

- Test size:

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{250.20 - 250}{2.9303/\sqrt{25}} = 0.3413 \sim t(n-1)$$

- p-value:

$$pval = 2 \cdot \left(1 - t_{cdf}(|t|, n-1)\right) = 2 \cdot \left(1 - t_{cdf}(0.3413{,}24)\right)$$

$$= 2 \cdot (1 - 0.6321) = 0.7359 \qquad > 0.05$$

- and we fail to reject the hypothesis

The Mean for a Population with Unknown Variance

- The 95% confidence interval for the mean is $\bar{x} \pm t_0 \cdot \frac{s}{\sqrt{n}}$, so the endpoints are:

Lower bound: $$\mu_- = \bar{x} - t_0 \cdot \frac{s}{\sqrt{n}} = 250.20 - 2.0639 \cdot \frac{2.93}{\sqrt{25}} = 248.99$$

Upper bound: $$\mu_+ = \bar{x} + t_0 \cdot \frac{s}{\sqrt{n}} = 250.20 + 2.0639 \cdot \frac{2.93}{\sqrt{25}} = 251.41$$

- where

```
t0=tinv(1-alpha/2,n-1)=tinv(0.975,24)=2.0639
```

Convergence of the t-dist. towards a Std. Norm. dist.

- **z-test (Standard Normal distribution / known variance)):**
- p-value: $pval = 0.6892$
- Confidence interval: $[\mu_-; \mu_+] = [249.22; 251.18\ ]$

- **t-test (Student's t-distribution / unknown variance):**
- p-value: $pval = 0.7359$
- Confidence interval: $[\mu_-; \mu_+] = [248.99; 251.41\ ]$

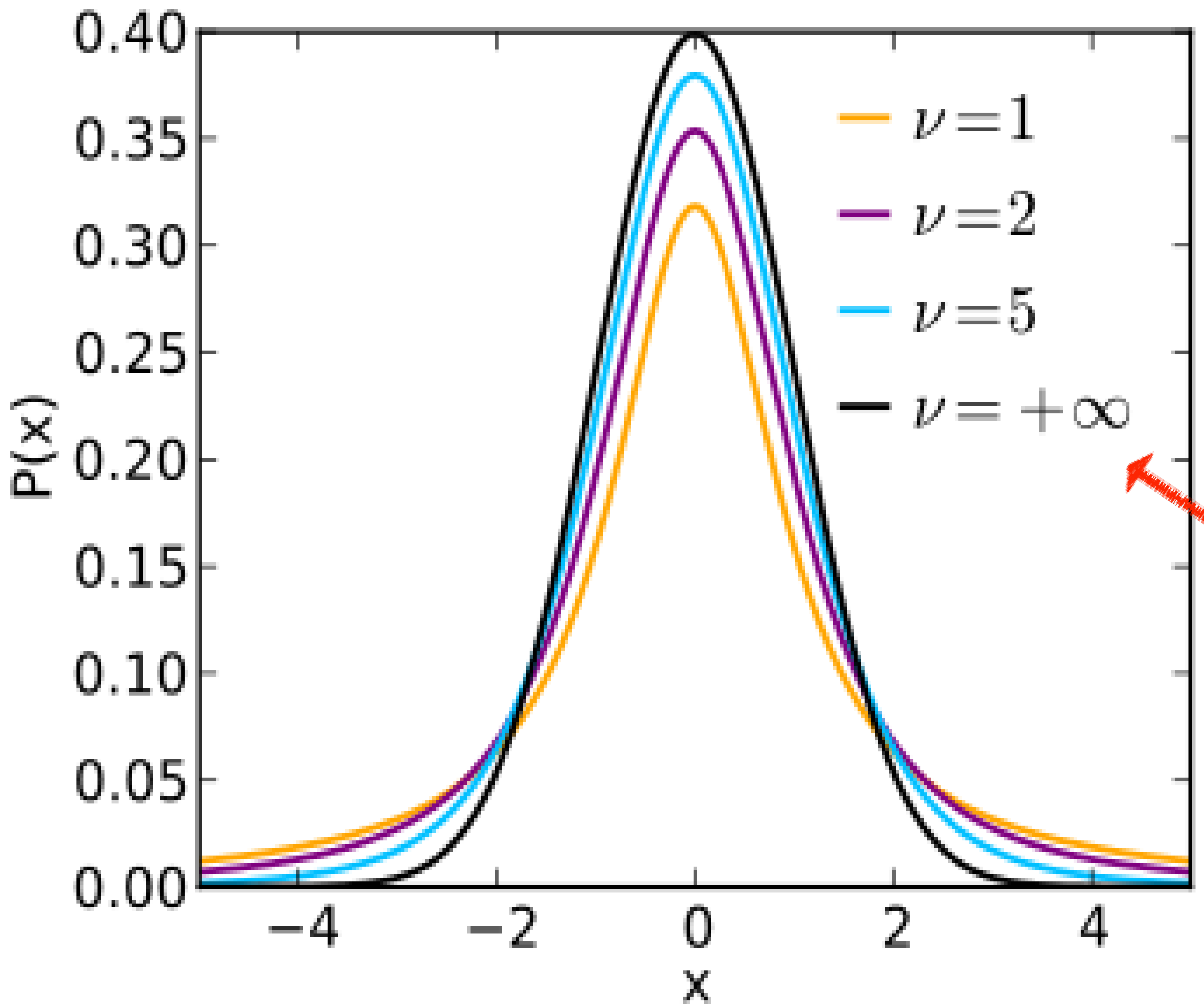
Convergence of the t-dist. towards a Std. Norm. dist.

- The confidence interval obtained with t statistic is wider than the one obtained with the z statistic.
- This results from the fact that we do not know the true standard deviation; we have to use the estimate s instead of the true value σ.
- As a result, we always have $t_0 \geq 1.96$ for a significance level of $\alpha=0.05$, which on average leads to a wider confidence interval for the t statistic.

n	2	3	5	10	30	∞
$\boldsymbol{t_0}$	12.71	4.30	2.78	2.26	2.05	1.96

Convergence of the t-dist. towards a Std. Norm. dist.

density of the t-distribution



Standard normal distribution

Checking for Normality in Sampled Data (Q-Q plots)

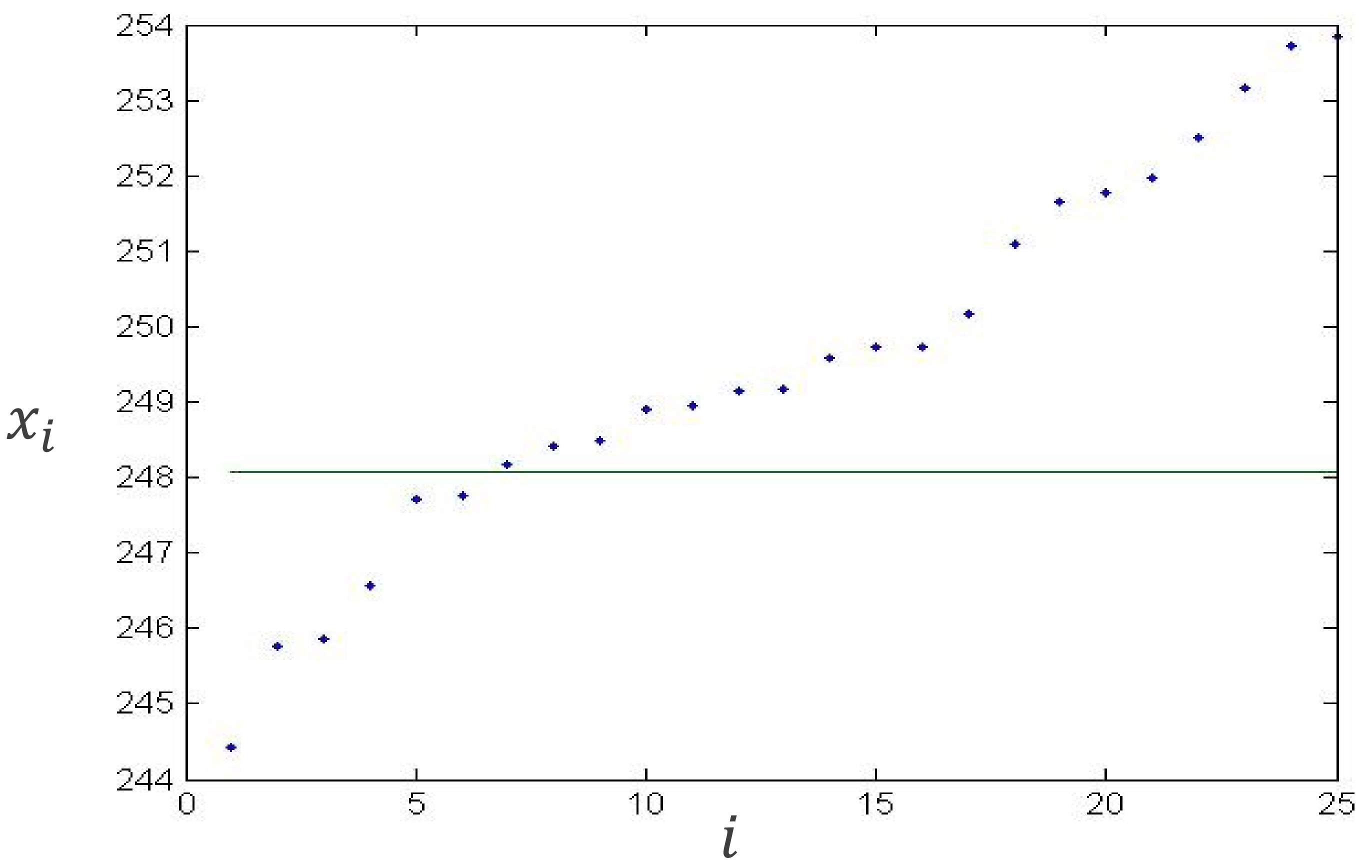
- We can quite safely use the central limit theorem (CLT) to make inference about the mean of any population (i.e., distribution), provided that the sample size is sufficiently large (say $n \geq 30$).
- However, if *n* is small the CLT does not hold anymore.
- In this case, statistical inference based on either the *z*-score or t-score only works, if the sampled data $x_1, x_2, \dots, x_n$ are themselves normally distributed.
- **Hence, we need a method to check whether the data are normally distributed.**

Quantiles *(Fraktiler)*

- The 25% quantile of the previous data

```
q25 = quantile(x,0.25)
q25 = 248.0731
```

$$\Pr(X \leq q25) = \Phi(z25) = 0.25$$
$$\Downarrow$$
$$z25 = \frac{q25 - \mu}{\sigma/\sqrt{n}} = \Phi^{-1}(0.25) = -0.675$$



Sorted data values with the estimated 25% percentile = 248.07.

Roughly 25% of the data should lie below this value.

Q-Q plot

- The quantiles of standard normally distributed data with *n* samples are roughly such that

$\sim[0;1]$

$$x_{[i]} \leftrightarrow \Phi^{-1}\left(\frac{i-0.5}{n}\right) = z_{[i]} = \frac{x_{[i]}-\mu}{\sigma}$$

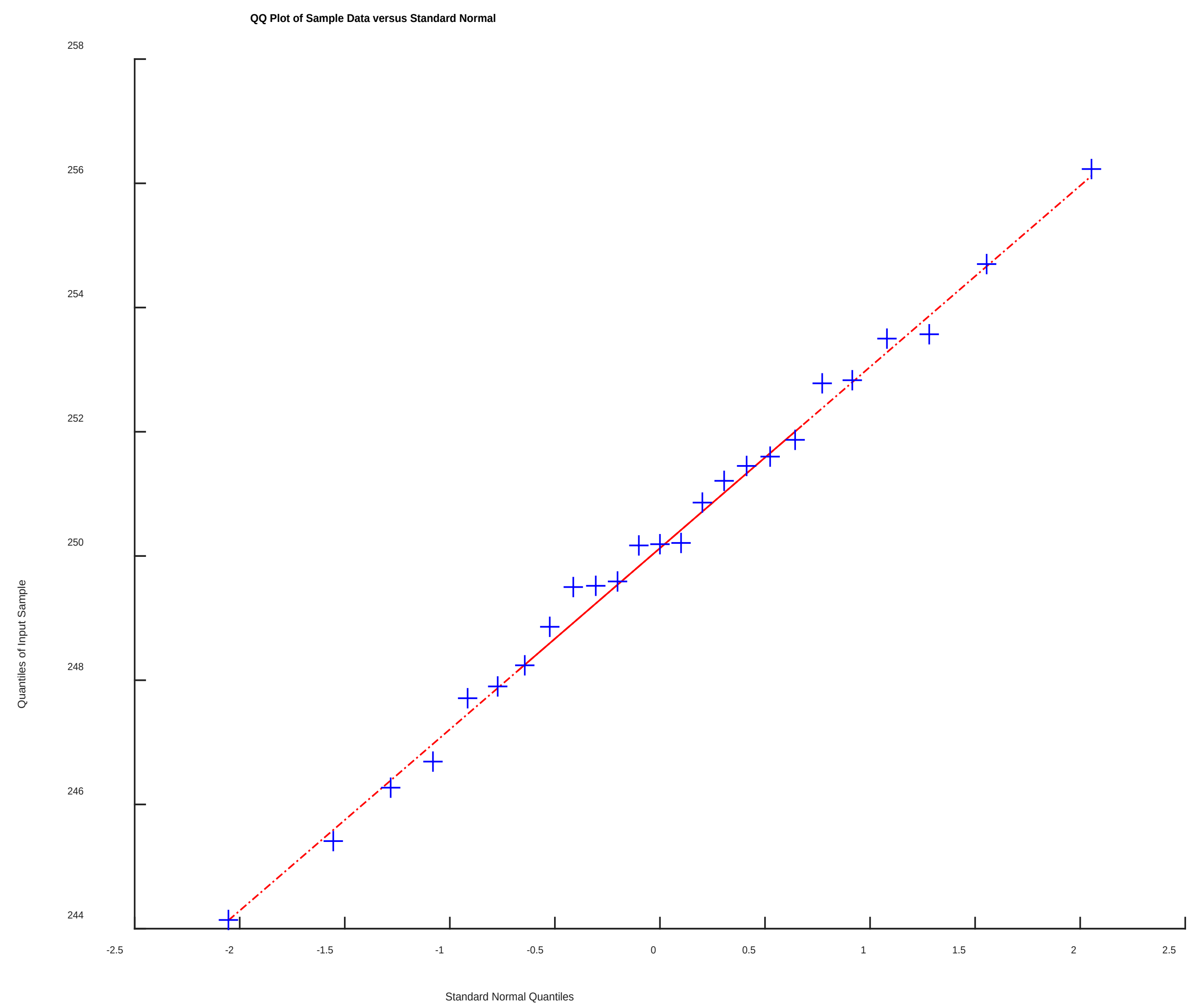
$]-\infty;\infty[$

- where $x_{[i]}$ denotes the *i*'th sample after sorting the samples $x_1, x_2, \dots, x_n$ in ascending order.
- If the data are consistent with a sample from a normal distribution, then plotting $x_{[i]}$ vs. $\Phi^{-1}\left(\frac{i-0.5}{n}\right)$ should result in a straight line.
- **This is the Q–Q plot.**

$$x_{[i]} = \sigma \cdot \Phi^{-1}_{[i]} + \mu$$

Example

$x_{[i]}$



$\Phi^{-1}(\frac{i-0.5}{n})$

Q-Q plot of the data from slide 33. The data points lie roughly on a straight line, and we conclude that the data are in fact normally distributed.

The Cavendish Experiment

- The Cavendish experiment, performed in 1797–98 by British scientist Henry Cavendish, was the first experiment to measure the force of gravity between masses in the laboratory and the first to yield accurate values for the gravitational constant.
- The value generally accepted today is 5.517.
- Cavendish's measurements were

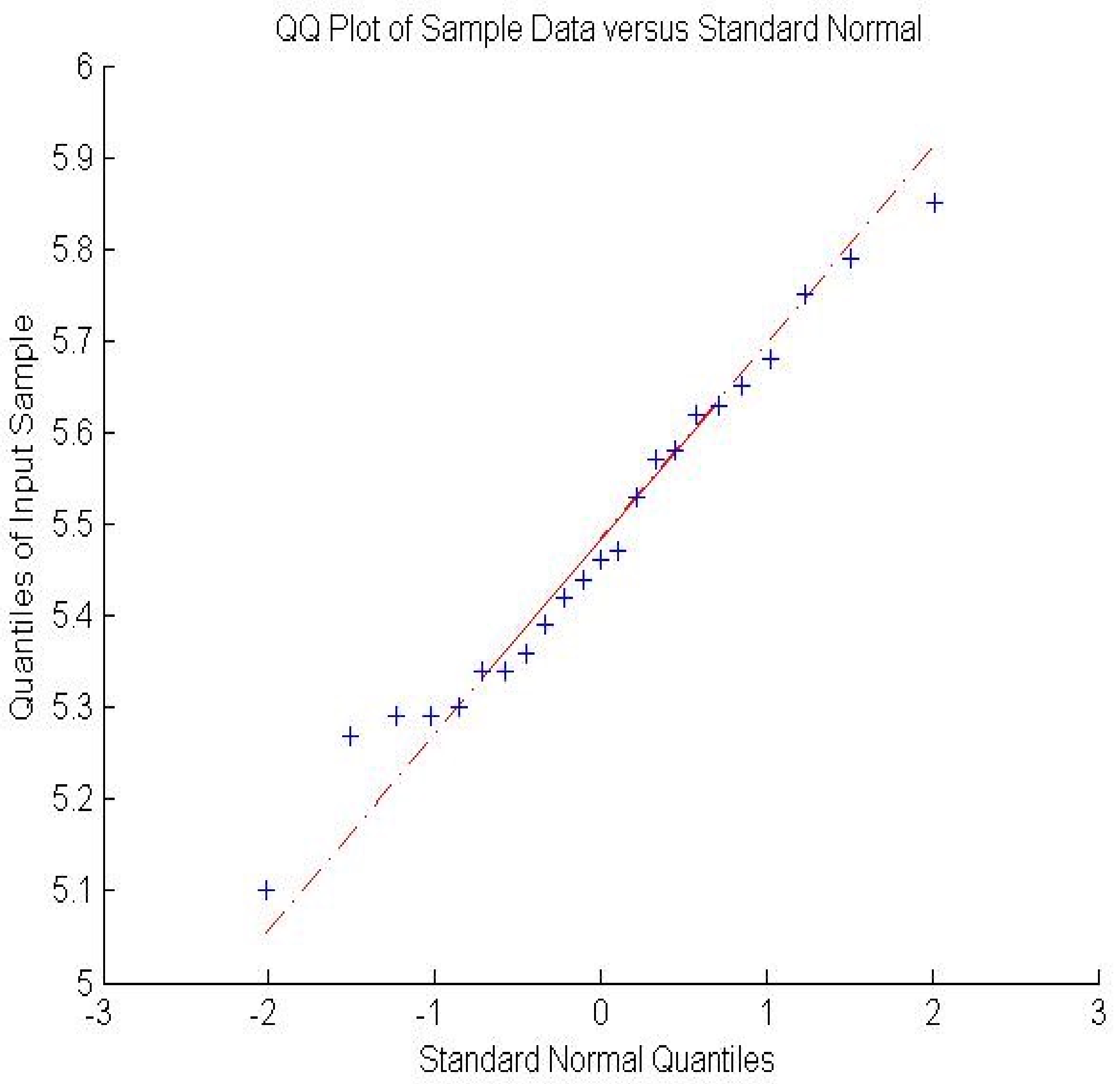
```
x = [ 5.36 5.29 5.58 5.65 5.57 5.53 5.62 5.29 ...
      5.44 5.34 5.79 5.10 5.27 5.39 5.42 5.47 ...
      5.63 5.34 5.46 5.30 5.75 5.68 5.85 ];
```

The Cavendish Experiment

- Check for normality:

    ```
    qqplot(x)
    ```

- The Q-Q plot results in a straight line, and hence we can conclude that the data are normally distributed.



The Cavendish Experiment

- The sample mean (and hence Cavendish's estimate of the gravitational constant) is

```
>> mean(x)
   5.4835
```

- with an empirical variance of

```
>> var(x)
   0.0363
```

- The null hypothesis needed to test if Cavendish's estimate corresponds to the accepted value today is

$$H_0\colon \mu = 5.517$$

The Cavendish Experiment

- Since we observe a sample mean of $\bar{x} = 5.4835$ and an empirical standard deviation of $s = \sqrt{0.0363} = 0.1904$, the test size with *n* = 23 is

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{5.4835 - 5.517}{0.1904/\sqrt{23}} = -0.8438 \sim t(n-1)$$

- P-value

$$pval = 2 \cdot \left(1 - t_{cdf}(|t|, n-1)\right) = 2 \cdot \left(1 - t_{cdf}(0.8438{,}23 - 1)\right)$$
$$= 2 \cdot (1 - 0.7961) = 0.4078 \quad > 0.05$$

- and we fail to reject the null hypothesis.
- In other words, we conclude that Cavendish's estimate of earth's gravitational constant corresponds to the accepted value today.

The Cavendish Experiment

- The 95% confidence interval for the mean is $\bar{x} \pm t_0 \cdot s/\sqrt{n}$. We have

```
t0 = tinv(1-0.05/2,n-1)= tinv(0.975,23-1)= 2.0739
```

- so the endpoints of the confidence interval are

Lower bound:

$$\mu_{-} = \bar{x} - t_0 \cdot \frac{s}{\sqrt{n}} = 5.4835 - 2.0739 \cdot \frac{0.1904}{\sqrt{23}} = 5.4012$$

Upper bound:

$$\mu_{+} = \bar{x} + t_0 \cdot \frac{s}{\sqrt{n}} = 5.4835 + 2.0739 \cdot \frac{0.1904}{\sqrt{23}} = 5.5658$$

TEST CATALOG FOR THE MEAN (UNKNOWN VARIANCE)

- **Statistical model:**
- $X_1, X_2, \ldots, X_n$ are i.i.d. samples of a random variable X with mean μ and variance σ^2.
- Parameter estimates:

$$\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu, \sigma^2/n)$$

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

- Where the observation is $\bar{x}$ = 'the average of n samples drawn from X's distribution'.
- NOTE: The statistical model is only true if n is sufficiently large ($n \geq 30$) <u>or</u> if the samples are drawn from a normal population with mean μ and variance σ^2.

- **Hypothesis test (two-tailed):**
- $H_0 : \mu = \mu_0$
- $H_1 : \mu \neq \mu_0$
- Test size: $t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}} \sim t(n-1)$
- Approximate p-value: $2 \cdot \left|1 - t_{cdf}(|t|)\right|$

- **95% confidence interval:**
- $\mu_- = \bar{x} - t_0 \cdot s/\sqrt{n}$
- $\mu_+ = \bar{x} + t_0 \cdot s/\sqrt{n}$

- where t0 = tinv(1-0.05/2,n-1)

Words and Concepts to Know

Heavy

Quantiles

Left-tailed

Null hypothesis

Test catalog

Reject

Q-Q plot

Alternative hypothesis

t-score

Fail to reject

Right-tailed

Students t-distribution

Two-tailed

Hypothesis test

Degrees of freedom