

Experimental Design and Data Analysis

Lecture 1

Eduard Belitser

VU Amsterdam

Lecture Overview

- 1 course organization
- 2 experimental design
- 3 recap of statistical concepts

Course organisation

- **Prerequisites:** basic statistics course, R knowledge
- **Lectures and practical sessions** 10 lectures + 3 practical sessions (TA's)
- **Assignments:** 3 regular assignments, made by **groups of 3 students** (sign up in canvas)
- **Final assignment** is to be made in the last week of the course
- **Grade** is based on assignments and final assignment
- **All relevant information is on canvas:** schedule, lecture slides, assignments (in due time), R manual(s) and suggestions additional literature
- **R** is an open software, widely adopted in the academic community, it is a programming language (object oriented), a statistical package.
- **RStudio** is a powerful user interface for R.

Experimental design

Experimental design, randomization

- Statistics allows to generalize from **data** to a **true state of nature**.
Statistical inference requires assumptions and mathematical modeling.
- The data should be obtained by a carefully **designed (chance) experiment** (or at least it must be possible to think about the data in this way).
- Any good design involves a chance element: “**experimental units**” are assigned to “**treatments**” by chance, or by **randomization**. The purpose is to exclude other possible explanations of an observed difference.
- We need probability to quantify the **randomization**. In practice, randomization is implemented with a **random number generator**. In R:

```
> x=rep(c("A","B"),each=5); x
[1] "A" "A" "A" "A" "A" "B" "B" "B" "B" "B"
> sample(x) # create a sequence of 5 A's and 5 B's in random order
[1] "A" "B" "A" "B" "B" "A" "B" "A" "A" "B"
> rbinom(10,1,0.5) # toss a fair coin 10 times
[1] 1 0 1 1 1 0 1 0 0 0
> rbinom(10,1,0.5) # again toss a fair coin 10 times
[1] 1 0 0 0 0 1 0 1 1 0
> rbinom(5,1,0.8) # toss a biased coin (success probability=0.8) 5 times
[1] 1 1 0 1 1
```

Examples, observational studies

EXAMPLE To compare two fertilisers we prepare 20 plots of land, apply the first fertilisers to 10 **randomly** chosen plots and the second one to the remaining plots. We plant a crop and measure the total yield from each plot.

EXAMPLE To compare two web designs we **randomly** select 50 subjects and measure the time needed to find some information. All 50 subjects perform this task with both designs, but for each subject the order of the two designs is based on **tossing a coin**.

EXAMPLE If an experiment involves subjects, then it could be wrong to assign “task A” to the first 10 subjects who arrive and “task B” to the last 10. (There may be a reason for arriving early.) Instead assign the tasks **at random**. Then an observed difference is due to the task (or chance).

Data obtained by registering an ongoing phenomenon, without randomization or applying other controls, is called **observational**.

EXAMPLE The incidence of lung cancer among 500 smokers is observed to be higher than among 500 non-smokers. Does this finding generalize to the full population? Does this show that smoking causes lung cancer?

Recap of basic statistical concepts: data summaries

(if needed consult the book Elementary Statistics, by Mario Triola)

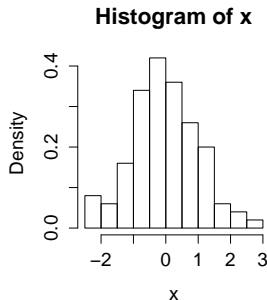
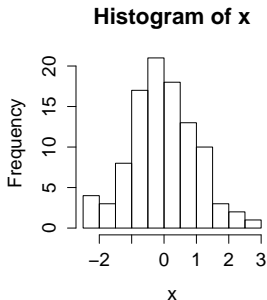
Histogram

The **histogram** of a sample of observed values x_1, x_2, \dots, x_N is a barplot, where the area of the bar over a **cell** (also called bin) C corresponds to the fraction

$$\frac{\text{number of observations in cell } C}{\text{sample size}} = \frac{\#\{1 \leq i \leq N : x_i \in C\}}{N}.$$

```
> x=rnorm(100); par(mfrow=c(1,2)) # two plots next to each other  
> hist(x) # frequencies on y-axis  
> hist(x,prob=T) # probabilities on y-axis
```

Why are the Y-axes different?

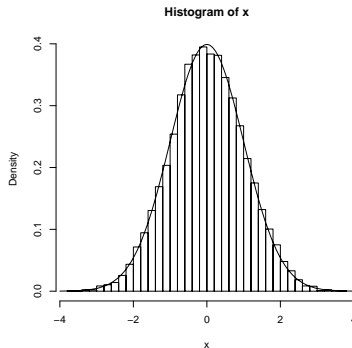


Histogram versus density

For continuous distributions, the true population density can be seen as the smoothed (or limiting as sample size $\rightarrow \infty$) histogram of the population values.

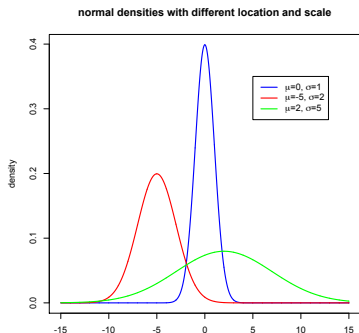
The resemblance between the true normal(0,1) density and the histogram of a sample of size 10000.

You can think of the population here as consisting of **infinitely** many values.



Location and scale, normal density

Two important characteristics of a population are **location** (or mean) μ and **scale** (or standard deviation) σ .



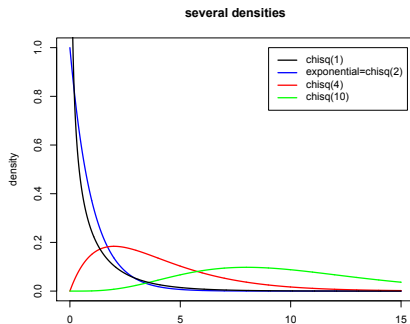
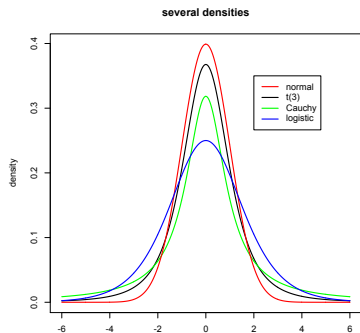
The **normal density** curve is given by

$$f_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}(x-\mu)^2/\sigma^2}.$$

The parameters μ and σ are the **location** and **scale**. Normal distributions with different μ and σ are still similar in a way.

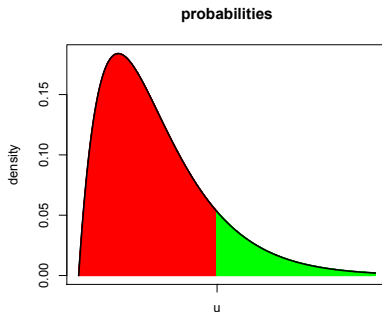
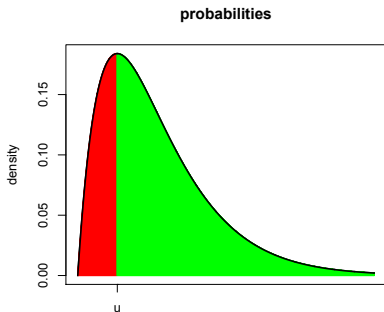
Remark. The normal curve is very specific! There are many “bell shaped” curves that are not normal.

Other symmetric and asymmetric densities



Probabilities and quantiles

If a **random variable** X is distributed according to a density curve, the probability $P(X \leq u)$ is the (**red**) area under the density curve **left** of u . Likewise, $P(X \geq u)$ is the (**green**) area under the density curve **right** of u .



For distribution P , the **quantile** of level $\alpha \in (0, 1)$ is the number q_α such that $P(X \leq q_\alpha) = \alpha$, the **upper quantile** u_α such that $P(X \geq u_\alpha) = \alpha$.

For the standard normal distribution, the quantile and upper quantile are usually denoted by ξ_α and z_α .

Data summaries

- Densities, probabilities and quantiles of many distributions can be computed in R. Commands in R: `dnorm(u,par)`, `pnorm(q,par)`, `qnorm(a,par)`, `rnorm(size,par)`, etc.
- **Numerical summaries**: sample mean, sample variance, sample median, sample standard deviation, sample α -quantile, etc. Commands in R: `mean(x)`, `var(x)`, `med(x)`, `sd(x)`, `quantile(x,a)`, `summary(x)`, `range(x)`, etc.
- **Graphical summaries**: histogram, boxplot, (normal) QQ-plot, scatter plot(s), empirical distribution function (cumulative histogram), etc. Commands in R: `hist(x)`, `boxplot(x)`, `qqnorm(x)`, `plot(x,y)`, `plot(ecdf(x))`, etc.

Study [Assignment 0](#).

The **boxplot** of a sample of observed values is a box with *whiskers* and (possibly) extremes. From a boxplot you can see the **scale** of the data, its **symmetry** and whether there are **extremes**.

Some numerical summaries: reminder

sample size		n	
location	<i>mean</i>	$\bar{x} = n^{-1} \sum_{i=1}^n x_i$	
	<i>median</i>	$\text{med}(x) = \begin{cases} x_{((n+1)/2)}, & \text{if } n \text{ odd} \\ (x_{(n/2)} + x_{(n/2+1)})/2, & \text{if } n \text{ even} \end{cases}$	
scale	<i>variance</i>	$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$	
	<i>standard deviation</i>	$s = \sqrt{s^2}$	

Here $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ is the ordered sample.

Interpretation of location measures:

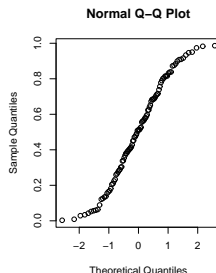
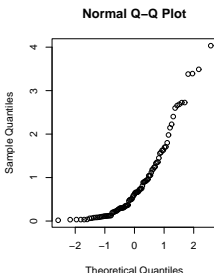
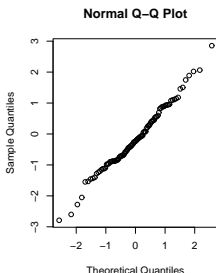
- **mean** – average value
- **median** – middle value in sorted values

Interpretation of scale measures:

- **variance** – average squared deviation from mean
- **standard deviation** – square root of variance

QQ-plots

- A **QQ-plot** can reveal whether data (approximately) follows a certain distribution P (e.g., the normal one: `qqnorm(x)`).
- It plots the **ordered data** $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(N)}$ versus the quantiles $q_{1/N}, q_{2/N}, \dots, q_{N/N}$ of the distr. P (i.e., $P(X \leq q_\alpha) = \alpha$ for $X \sim P$).
- If the x_i 's are from distribution P , then approx. a fraction i/N of the population should be smaller than the i/N -quantile $q_{i/N}$, i.e., the plot points should follow a straight line.
- If the points are approximately on a **straight line**, then the data can be assumed to be sampled from the population, possibly with different location and scale.



Recap basic stat. concepts: estimation, confidence intervals

The sample mean and its distribution

- The **sample mean** of a sample X_1, \dots, X_n of sample size n is

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i.$$

- We consider the **sampling distribution of the sample mean** \bar{X} .
- When the sample is taken from the $N(\mu, \sigma^2)$ distribution, then the sample mean \bar{X} has **exactly** the $N(\mu, \sigma^2/n)$ distribution.
- When the sample is taken from some other distribution with expectation μ and variance σ^2 , then \bar{X} has **approximately** the $N(\mu, \sigma^2/n)$ distribution (\bar{X} is **asymptotically normal**) because of the **Central Limit Theorem**.
- The **mean varies less** than the individual observations: the standard deviation σ is replaced by σ/\sqrt{n} .

Standardizing the mean

- Any normal random variable $X \sim N(\mu, \sigma^2)$ can be **standardized** into a standard $N(0, 1)$ -variable by $Z = (X - \mu)/\sigma \sim N(0, 1)$.
- Converse is also true: if $Z \sim N(0, 1)$, then $X = \mu + \sigma Z \sim N(\mu, \sigma^2)$.
- General fact: if $X \sim N(\mu_x, \sigma_x^2)$ and $Y \sim N(\mu_y, \sigma_y^2)$ are independent, then $V = aX + bY + c \sim N(a\mu_x + b\mu_y + c, a^2\sigma_x^2 + b^2\sigma_y^2)$.
- As $\bar{X} \sim N(\mu, \sigma^2/n)$ (exactly or approximately), **standardizing \bar{X}** yields

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \sim N(0, 1).$$

- In a real data set X_1, \dots, X_n , the population standard deviation σ is **unknown** and needs to be estimated by the **sample standard deviation** s .
- This uncertainty influences the distribution of the resulting statistics $\frac{\bar{X} - \mu}{s/\sqrt{n}}$.
- The random variable $T = \frac{\bar{X} - \mu}{s/\sqrt{n}}$ does not have the $N(0, 1)$ distribution.
- Instead, T has a **t -distribution with $n - 1$ degrees of freedom**.

Estimation – the concepts

- Suppose we assume that our population of interest has a certain distribution with an unknown parameter, e.g., its mean μ or a fraction p .
- A **point estimate** for the unknown parameter is a function of **only** the observed data (X_1, \dots, X_n) , seen as a random variable.
- We denote estimators by a hat: $\hat{\mu}$, \hat{p} , etc.
- Examples of point estimates: $\hat{\mu} = \bar{X}$, the sample proportion \hat{p} .
- A **confidence interval** (CI) of level $1 - \alpha$ for the unknown parameter is a **random interval** based **only** on the observed data (X_1, \dots, X_n) that contains the true value of the parameter with probability at least $1 - \alpha$.

Estimating the mean, CI

- Recall that $\bar{X} \sim N(\mu, \sigma^2/n)$ for X_1, \dots, X_n from $N(\mu, \sigma^2)$ distribution.
- The **upper quantile** z_α of the $N(0, 1)$ -distribution is such z_α that $P(Z \geq z_\alpha) = \alpha$ for $Z \sim N(0, 1)$, (in R: $z_\alpha = \text{qnorm}(1-\alpha)$). Then

$$\begin{aligned} 1 - \alpha &= P(|Z| \leq z_{\alpha/2}) = P\left(\frac{|\bar{X} - \mu|}{\sigma/\sqrt{n}} \leq z_{\alpha/2}\right) \\ &= P\left(\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right). \end{aligned}$$

- In other words, $\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} = [\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}]$ is the **confidence interval** of μ of level $1 - \alpha$.
- If the standard deviation σ is **unknown**, we estimate it by s and the confidence interval is based on a t -distribution and the **upper t -quantile** $t_\alpha = \text{qt}(1-\alpha, df=n-1)$ (i.e., $P(T \geq t_\alpha) = \alpha$ for $T \sim t_{n-1}$).
- The t -confidence interval of level $1 - \alpha$ for μ then becomes

$$\bar{X} \pm t_{\alpha/2} \frac{s}{\sqrt{n}} = \left[\bar{X} - t_{\alpha/2} \frac{s}{\sqrt{n}}, \bar{X} + t_{\alpha/2} \frac{s}{\sqrt{n}} \right].$$

The t -CI's are (nearly) always used, since σ is almost never known in practice.

Recap basic stat. concepts: hypothesis testing

Hypothesis testing: the concepts

- In **hypothesis testing**, we have two claims, the **null hypothesis** H_0 and the **alternative hypothesis** H_1 , which do not overlap.
- The **claim of interest** is usually represented by H_1 .
- A test has two possible outcomes:
 - the strong outcome: H_0 is rejected, H_1 is assumed to be true;
 - the weak outcome: H_0 is not rejected.
- A **statistical test** chooses between two possibilities: H_0 and H_1 .
- In order to perform the test, one needs a **test statistic** $T = T(X)$, which summarizes the data $X = (X_1, \dots, X_n)$ in a relevant way.
- The H_0 is rejected if the value of the test statistic is too extreme to what is expected under the H_0 : reject H_0 if $T(X) \in K$, for **critical region** K .
- In general, to perform a test, we need to know the **distribution of $T(X)$ under H_0** , required to determine when to reject, and when not to.
- The test statistic is **not unique**. We can choose different test statistics, leading to different tests for the same hypothesis H_0 .

Hypothesis testing: p -values

- 3 ways to test, say, $H_0 : \mu = \mu_0$, with test statistics $T(X)$ and level α :
 - by checking whether $T(X) \in K_\alpha$: $|T(X)| \geq |t_{\alpha/2}|$ or not;
 - by comparing the p -value to α : $P(|T(X)| \geq |t|) \leq \alpha$ or not;
 - by checking whether μ_0 is in the $(1 - \alpha)$ -CI (for μ) or not.
- By using p -values is the most common way. The value of the test statistic $T(X)$ is converted into a p -value.
E.g., $p = P(|T(X)| \geq |t|)$ for $T(x) = t$ and $T(X) \sim t_{n-1}$.
- The p -value of a test is the probability that an experiment **in the situation that H_0 is true** will deliver the data actually observed. A small p -value indicates that the observed data would be unlikely if H_0 were true.
- When the p -value is below the chosen **significance level** α (e.g., 0.05), reject H_0 (**strong** outcome), otherwise do not reject H_0 (**weak** outcome).
- If H_0 is rejected, the data are said to be **statistically significant** at level α .
- By construction, **under H_0 , the p -value is like a uniform draw from $[0, 1]$.**

Example: the t -test

- The t -test is for testing the population mean μ of a **normal population**.

1. $H_0 : \mu \leq \mu_0$ vs $H_1 : \mu > \mu_0$ (`t.test(data,mu= μ_0 ,alt="g")`)
2. $H_0 : \mu \geq \mu_0$ vs $H_1 : \mu < \mu_0$ (`t.test(data,mu= μ_0 ,alt="l")`)
3. $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$ (`t.test(data,mu= μ_0)`)

- In all 3 cases, at the border of H_0 and H_1 (i.e. for $\mu = \mu_0$), the

test statistic $T = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$ **has t -distribution** with $n - 1$ degrees of freedom.

- The **p -value** for observed value $T(x) = t$ of the test statistic is

1. $p = P(T \geq t)$ under H_0 ;
2. $p = P(T \leq t)$ under H_0 ;
3. $p = P(|T| \geq |t|) = 2 \times \min(P(T \geq t), P(T \leq t))$ under H_0 .

- For testing, say, situation 3, $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$, we reject H_0 if

either $|T(x)| > |t_{\alpha/2}|$,
or $p = P(|T| \geq |t|) < \alpha$ under H_0 ,
or μ_0 does not belong to the CI $\bar{X} \pm t_{\alpha/2} \frac{s}{\sqrt{n}}$.

Hypothesis testing: types of errors, power of the test

- Statistical tests are typically not perfect, but make two types of errors:
 - Error of the first kind (type I error): rejecting H_0 while it is true.
 - Error of the second kind (type II error): not rejecting H_0 while it is false.
- Tests are constructed to have small $P(\text{type I error})$ (typically, $< 5\%$).
- $P(\text{type II error})$ depends (among others) on the amount of data.
- The probability of (correctly, when H_0 is not true) rejecting H_0 is called the **power** of the test. Under H_1 , $\text{power} = 1 - P(\text{type II error})$.
- Different test statistics can yield different statistical power of the test.
- Higher sample sizes yield higher power.
- Tests with high statistical power are preferred, while keeping the **level** of the test (probability of type I error, often taken to be 5%) **fixed**.

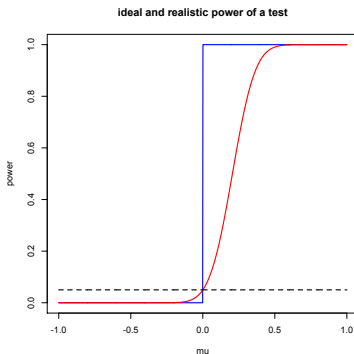
The power of a test is specified for each possibility under H_1 . E.g., if $H_0 : \mu \leq 0$ then the power can be calculated in each $\mu > 0$. A *good* test (that is, a test based on a *good* test statistic) has high power in all positive μ -values, relative to other tests.

When H_0 is true, $\text{power} = P(\text{type I error})$, which must be less than α .

Ideal test and realistic test

The **ideal test** would make no errors in the sense that

- never falsely reject (no error of type I)
- always reject when H_1 is true (no error of type II)



The power of the **ideal test** and a **realistic test** for $H_0 : \mu \leq 0$. The dashed line is the level of the test, here 0.05 (the probability of type I error).

Practical significance

- **Statistical significance** is about generalization: an observed effect is not due to chance, it should be observed again if a new experiment were performed. Data can be statistically significant even though the deviation from H_0 is very small!
- In practice, this boils down to **practical significance** which is about the relation between the size of the effect and the available information.

EXAMPLE Suppose that a coin has probabilities $1/2 - 10^{-10}$ and $1/2 + 10^{-10}$ to land HEAD or TAIL.

If we use the coin to decide who will kick-off in a soccer game, then TAIL has a slight advantage, but the difference is negligible. A statistical test based on observing 100 tosses of the coin will not reject H_0 , but a test based on observing 10^{21} coin tosses almost certainly will.

Some simulations in R

Point and interval estimation, one sample t -test

Given a random sample X_1, \dots, X_N from a population with mean μ and unknown variance σ^2 , we wish to estimate μ , construct a CI for it, and to test $H_0 : \mu = \mu_0$ against $H_1; \mu \neq \mu_0$ for some given number μ_0 , e.g., $\mu_0 = 0$.

```
> mu=0.2; x=rnorm(50,mu,1) # creating artificial data
> t.test(x,mu=0)
      One Sample t-test
data:  x
t = 2.4211, df = 49, p-value = 0.01922
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 0.05219746 0.56202370
sample estimates:
mean of x
0.3071106
```

For this (synthetic) data $X_1, \dots, X_n \sim N(0.2, 1)$, we read off from the above output the estimate $\bar{X} \approx 0.31$, the 95% CI $[0.052, 0.562]$, the p -value 0.01922. Thus, $H_0 : \mu = 0$ is rejected as 1) $|t| = 2.42 > |t_{\alpha/2}| = \text{qt}(0.975, 49) \approx 2.01$, or 2) $p\text{-value} = 0.01922 < 0.05$, or 3) $0 \notin [0.052, 0.562]$.

Standard error and confidence interval

The **standard error** $\frac{\sigma}{\sqrt{n}} \approx \frac{s}{\sqrt{n}}$ of the estimator \bar{X} is a measure of its **precision**. If the estimator is normally distributed, then

Estimate $\pm 1.96 \times \text{Std.Error}$ gives an approx. 95% CI.

The **bigger** the sample size n , the **smaller** the standard error and the confidence intervals. The estimates get more precise, as more information is available.

Generate estimates \bar{X} from standard normal samples (hence the true $\mu = 0$):

sample size	Estimate	Std.Error
10	0.3564	0.3604
50	0.2198	0.1510
100	0.1098	0.1067
1000	-0.007433	0.031466

In all cases the true value 0 is in the 95% confidence interval

Estimate $\pm 1.96 \times \text{Std.Error}$.

The margin $m = 1.96 \times \text{Std.Error}$ is based on the asymptotic normality of \bar{X} and the fact that s is a good estimator of σ . If in the CI we use the upper t -quantile $t_{0.025, n-1}$ instead of $z_{0.025} \approx 1.96$, the CI will be bigger (i.e., more “conservative”) because always $t_{\alpha, n-1} > z_{\alpha}$, but $t_{\alpha, n-1} \rightarrow z_{\alpha}$ as $n \rightarrow \infty$.

Two sample t -test

- Given two populations with means μ and ν , we wish to test $H_0 : \mu = \nu$ against $H_1; \mu \neq \nu$. We take a random sample X_1, \dots, X_M from the first population and, independently, Y_1, \dots, Y_N from the second population.
- The test is based on the difference $\bar{X}_M - \bar{Y}_N$ which is a reasonable estimate for $\mu - \nu$. If it is very different from 0 we reject H_0 .
- How different? $\bar{X}_M - \bar{Y}_N$ will not exactly be $\mu - \nu$. The estimation error depends on M and N and the standard deviations of the populations.
- The t -statistic thus divides $\bar{X}_M - \bar{Y}_N$ by an estimate $S_{M,N}$ of its standard error: with $S_{X,Y}^2 = \frac{1}{M+N-2} \left(\sum_{i=1}^M (X_i - \bar{X}_M)^2 + \sum_{j=1}^N (Y_j - \bar{Y}_N)^2 \right)$,

$$\text{under } H_0, \quad T = \frac{\bar{X}_M - \bar{Y}_N}{S_{M,N}} \sim t_{M+N-2}, \quad S_{M,N} = S_{X,Y} \sqrt{\frac{1}{M} + \frac{1}{N}}.$$

- T is compared to the critical value (a quantile from t_{M+N-2} -distribution), or the p -value (computed by using t_{M+N-2} -distribution) is compared to α .

Remark. The standard t -test assumes that the two populations are normal. If the sample sizes M and N are large, then the test performs well even without this assumption, but the test is unreliable for M, N less than 20.

Two sample t -test: implementing in R

We can test $H_0 : \mu = \nu$ against $H_1; \mu \neq \nu$ by the **two sample t -test**.

```
> mu=0;nu=0.5
> x=rnorm(50,mu,1); y=rnorm(50,nu,1) # creating artificial data
> t.test(x,y)

Welch Two Sample t-test

data:  x and y
t = -2.4339, df = 96.574, p-value = 0.01677
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-0.85202520 -0.08659066
sample estimates:
mean of x mean of y
0.06552453 0.53483246
```

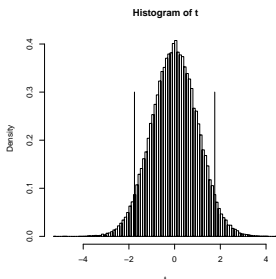
Noticed that we observed $t = -2.4339$, so that the corresponding p -value is

$$P(|T| > |t|) = 2P(T > |t|) = 2(1 - P(T \leq |t|)) \approx 0.0167.$$

This can be found in the above output, and we could also compute this directly in R as `2*(1-pt(2.4339,98))=0.01674788`.

p -value for two sample t -test

We can also evaluate this p -value by simulating from the null hypothesis.



We generate a population of T -values under H_0 by repeating the sampling. The p -value of the observed value t is approximately the fraction of this population that is bigger than $|t|$ or smaller than $-|t|$.

```
> mu=nu=0; t=numeric(100000)
> for(i in 1:100000){x=rnorm(50,mu,1);y=rnorm(50,nu,1);t[i]=t.test(x,y)[[1]]}
> sum(abs(t)>=abs(-2.4339))/length(t) ##cf. 2*(1-pt(2.4339,98))=0.01674788
[1] 0.01744
```

Different test statistics

EXAMPLE The *t*-test is for testing the population mean μ of a *normal* population, $H_0 : \mu = \mu_0$. Given a sample X_1, \dots, X_N , the test statistic is

$$T = \frac{\bar{X}_N - \mu_0}{S_N}, \quad \text{where } \bar{X}_N = \frac{1}{N} \sum_{i=1}^n X_i, \quad S_N^2 = \frac{1}{N-1} \sum_{i=1}^n (X_i - \bar{X}_N)^2.$$

When T is *very different* from 0, reject H_0 . The *critical value* for T that acts as border between rejecting and not rejecting H_0 is based on the distribution of T under H_0 . For *t*-test, this distribution is the t_{N-1} -distribution.

EXAMPLE For testing $H_0 : \mu = 0$ we can as well use the *sign test*. Given a sample X_1, \dots, X_N from the population, the test statistic for the sign test is

$$T = \#(X_i < 0).$$

When T is very different from $N/2$ we reject H_0 . For this test, the critical value comes from the $\text{Bin}(N, \frac{1}{2})$ -distribution, the distribution of number of heads in throwing N times a fair coin.

Comparing powers

Assume we have a normal sample and test $H_0 : \mu = 0$ using the t -test and the sign test. We can compare the power in $\mu = 0.5$ of the two tests by [simulation](#). Recall that always [power](#) = $P(\text{reject } H_0)$.

```
> B=1000; n=50
> psign=numeric(B)  ## will contain p-values of sign test
> pttest=numeric(B)  ## will contain p-values of t-test
> for(i in 1:B) {
+   x=rnorm(n,mean=0.5,sd=1) ## generate data under H1 with mu=0.5
+   pttest[i]=t.test(x)[[3]]                ## extract p-value
+   psign[i]=binom.test(sum(x>0),n,p=0.5)[[3]] }  ## extract p-value
> sum(psign<0.05)/B
[1] 0.746
> sum(pttest<0.05)/B
[1] 0.937
```

The power in $\mu = 0.5$ for the t -test ([0.937](#)) is higher than for the sign test ([0.746](#)) when we reject for p -values smaller than the level 0.05. [Why?](#) Because for normal data, the t -test has better performance than the sign test.

To finish

Today we dicussed

- ① course organization
- ② experimental design
- ③ recap of basic statistical concepts

Study Assignment 0 (not to be submitted), assignments to submit are to be made in groups of 3 students ([enroll in a group of 3 in canvas!](#)).

[Next time](#) bootstrap methods, one sample tests.