

PLOTS

Normal quantile plots / Normal Q-Q plots

quantile-quantile

Q-Qs don't have to be normal, we can use another distribution too...

X_i	i	$X_i = \text{sort}(X_i)$	Proportion below i $q_i = (i - 1/2)/n$	Theoretical quantile z for normal distribution $N(\bar{x}, s)$
10.38	1	5.80 (min)	0.05	$z_1 \mid P(z_1) \leq 0.05$
11.68	2	7.50	0.15	
5.80	3	:	0.25	
7.50	4		0.35	
10.68	5		0.45	
11.20	6		0.55	
6.99	7		0.65	
10.68	8		0.75	
18	9	11.68	0.85	
13.60	10	13.60 (max)	0.95	

$$\bar{x} = 9.97$$

$$s = 2.21$$

$$n = 10$$

Center value
of the quantile
where data lies

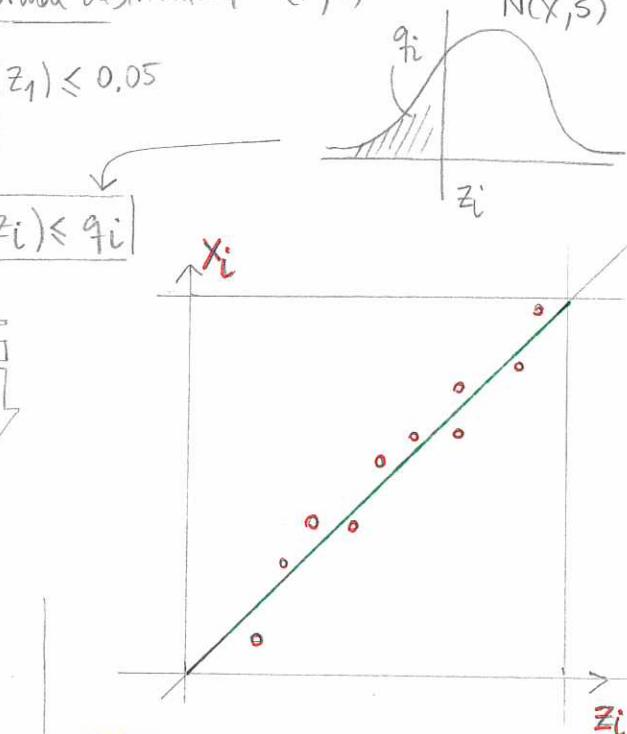
0) Get data x_i

1) Get sample values and sort: x_i, \bar{x}, s, n

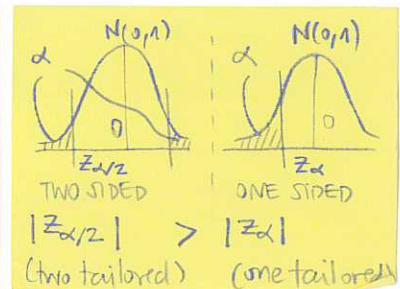
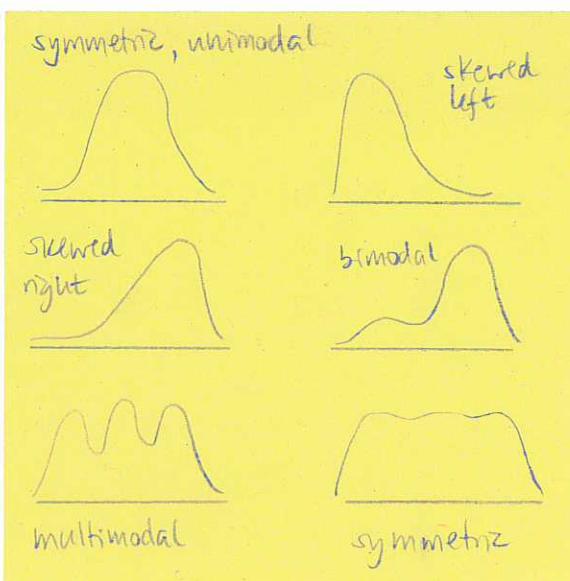
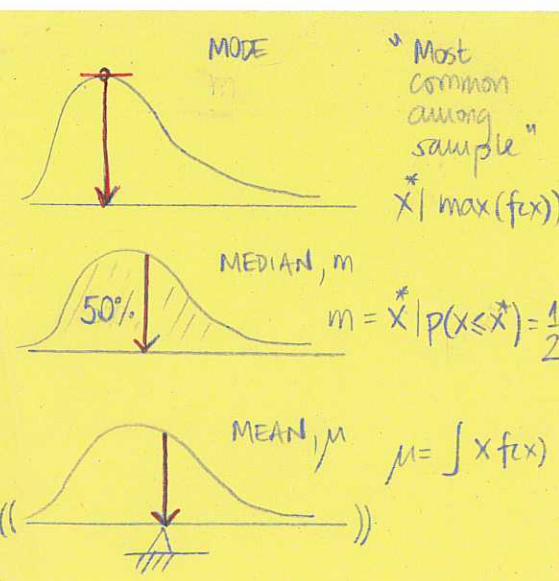
2) Compute center value of quantile where each x_i lies: $q_i = (i - 1/2)/n$

3) Compute $z_i \mid P(z_i) \leq q_i$ for each i , with the normal distribution $N(\mu = \bar{x}, \sigma = s)$

4) Plot (z_i, x_i)

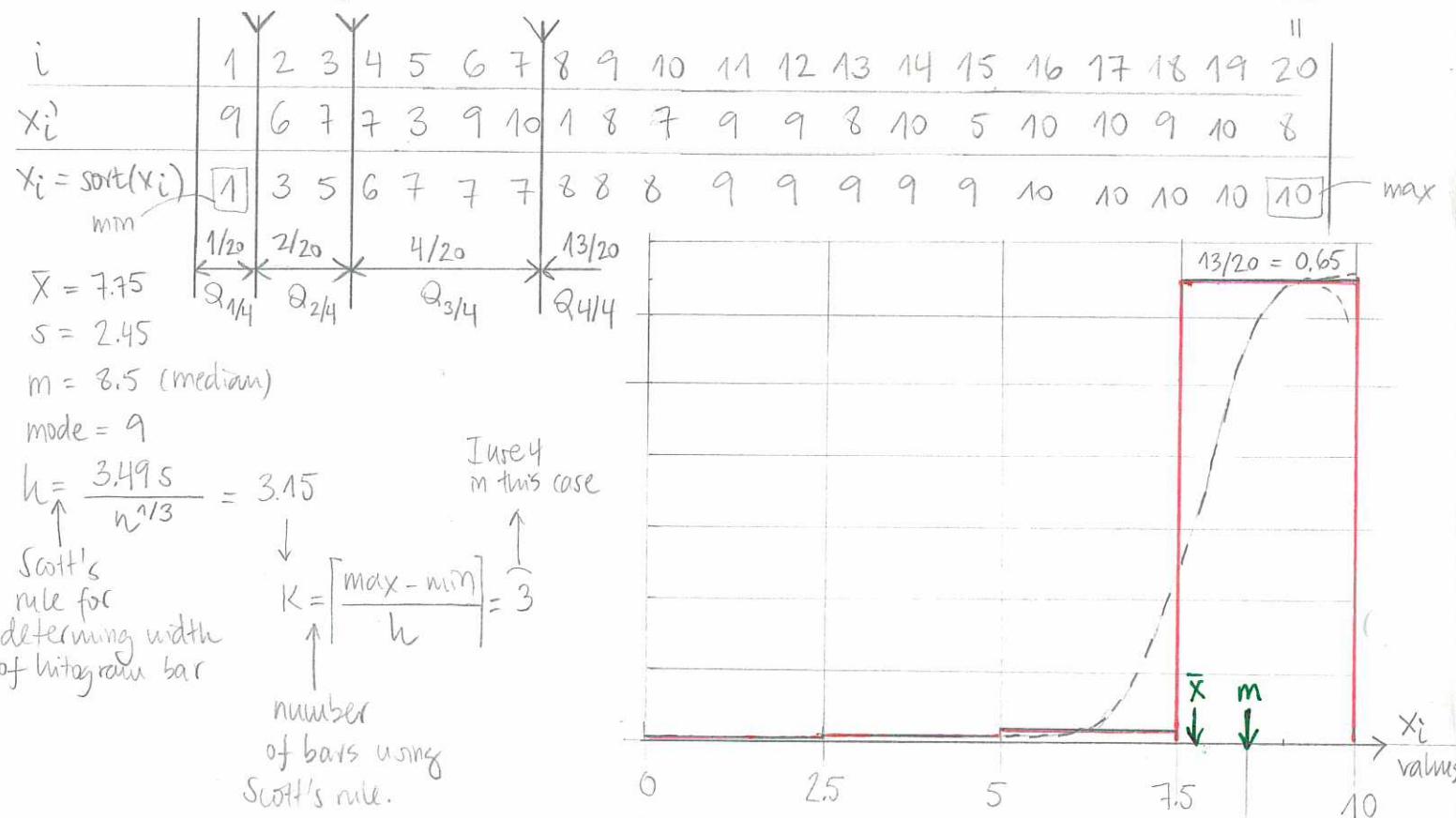


- If x_i were perfectly normal distributed, (x_i, z_i) would lie on the line $z = x$.
- Normal Q-Q plots are used to check normal distribution of samples, essential for ANOVAs!



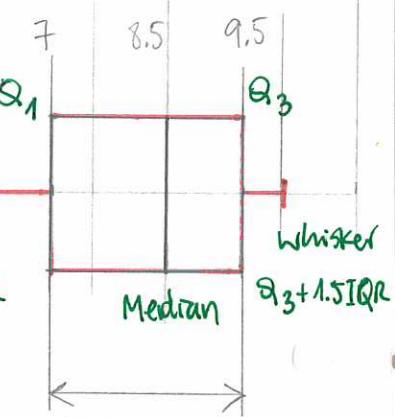
Box plots / Box-whisker plots and Histograms

(bigote)



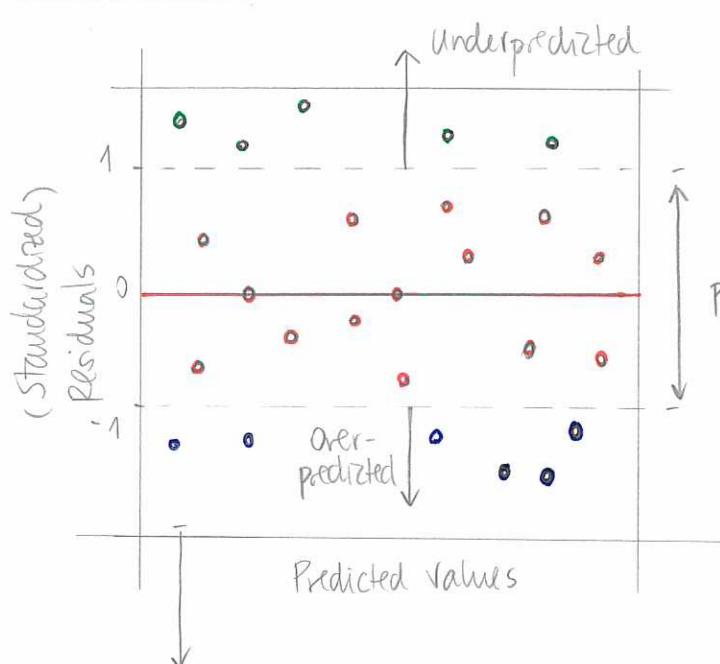
- Minimum: $\text{min}(x_i)$
 - Maximum: $\max(x_i)$
 - First quartile (Q_1): the 25% smallest x_i are $\leq Q_1$
 - Median (Q_2), m : the 50% smallest x_i are $\leq m = Q_2$
 - Third quartile (Q_3): the 75% smallest x_i are $\leq Q_3$
 - Interquartile range, $IQR = Q_3 - Q_1$: 50% of the x_i are here
 - Whiskers
 - min: $\min x_i \mid x_i \leq Q_1 - 1.5 \text{IQR} \mid = 7 - 1.5 \times 2.5 = 3.25 \rightarrow x_i = 5 \text{ (i=3)}$
 - max: $\max x_i \mid x_i \geq Q_3 + 1.5 \text{IQR} \mid = 9.5 + 1.5 \times 2.5 = 13.25 \rightarrow x_i = 10 \text{ (i=17)}$
 - Outliers: all values out of the whiskers
- from samples x_i :
- $\rightarrow 3.75$
- \downarrow
- $\rightarrow x_i = 5 \text{ (i=3)}$
- \downarrow
- $\rightarrow x_i = 10 \text{ (i=17)}$
- 20

→ Box plots help detect outliers and give an idea of how normally distributed data is



$$\text{IQR} : \text{interquartile range} = 9.5 - 7 = 2.5$$

Residual plots



$$\left\{ \begin{array}{l} \text{Regression model: } \hat{Y} = \hat{\beta}_0 + \sum \hat{\beta}_i X_i \\ \text{Residuals: } \varepsilon_i = Y_i - \hat{Y}_i \rightarrow \frac{\varepsilon_i}{s} \\ \text{standardization} \end{array} \right.$$

estimate of std. deviation: \hat{s}

Residuals should be stochastic errors if our linear model is good enough, i.e., random and unpredictable errors

- (Good model, with good residuals plots, since there is no pattern, no predictive information

— Problematic residual plots: those which contain patterns

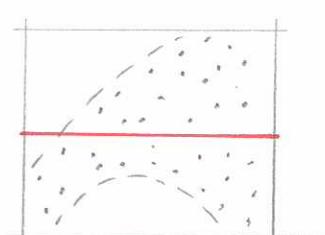
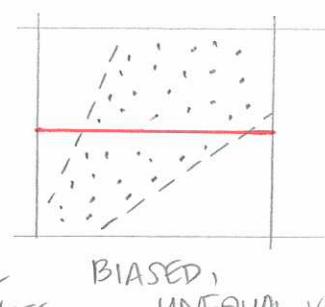
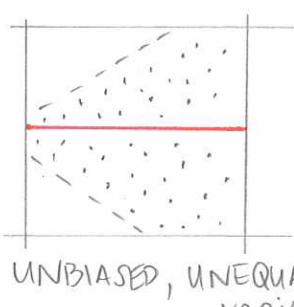
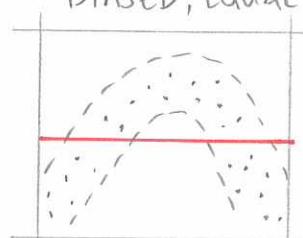
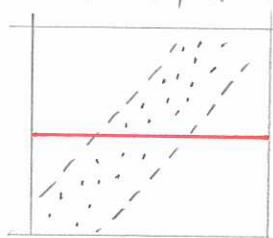
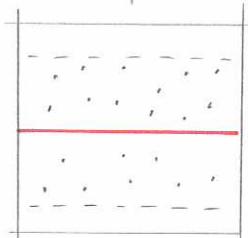
• Causes

- Missing variable
- Missing high order term of a variable in the model (could explain curvature) (X_i^2)
- Missing interaction between terms already in the model ($X_i \cdot X_j$)

• Important points:

- If residuals can be predicted with a variable \rightarrow this variable shouldn't be in model
- Plot residuals against variables
- One residual ε_i shouldn't predict next $\varepsilon_{i+1} \rightarrow$ include time variable into model, that could explain correlation between ε_i & ε_{i+1} .

UNBIASED, EQUAL VAR. BIASED, EQUAL VAR. BIASED, EQUAL VAR



UNBIASED, UNEQUAL VARIANCES

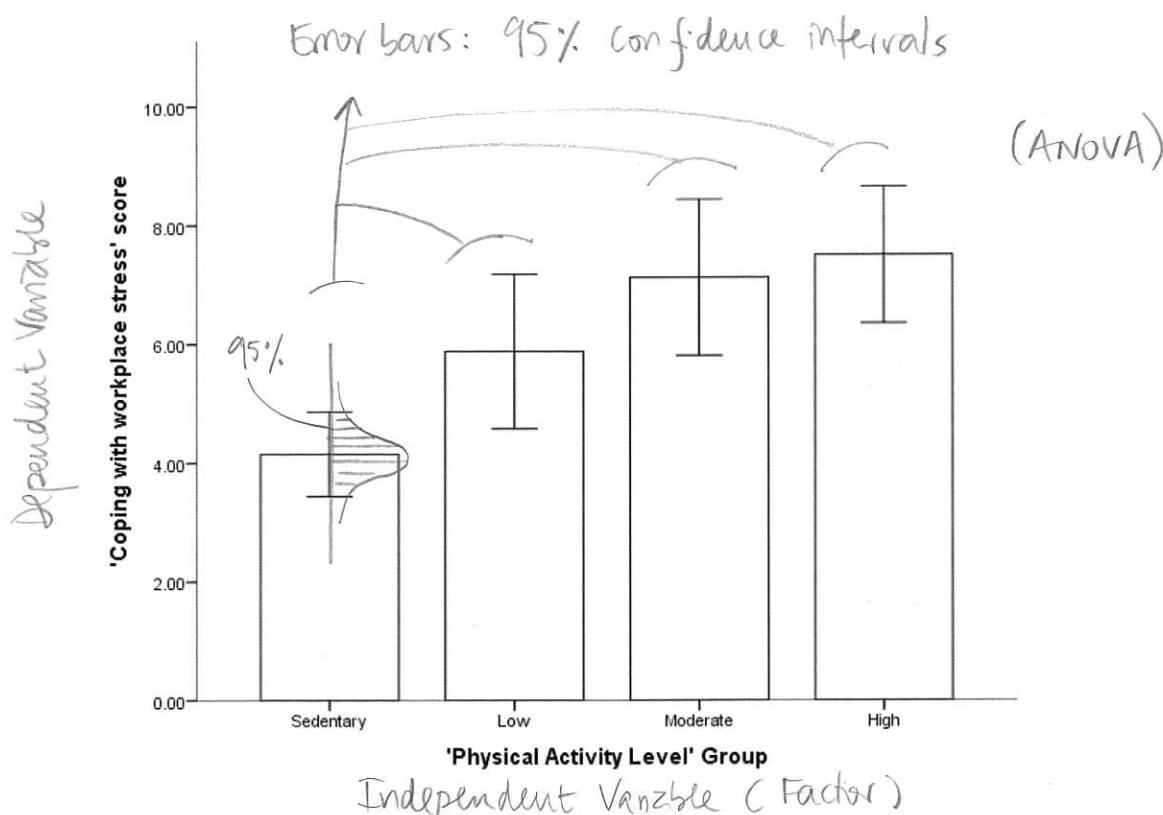
BIASED, UNEQUAL VAR.

BIASED, UNEQUAL VAR.

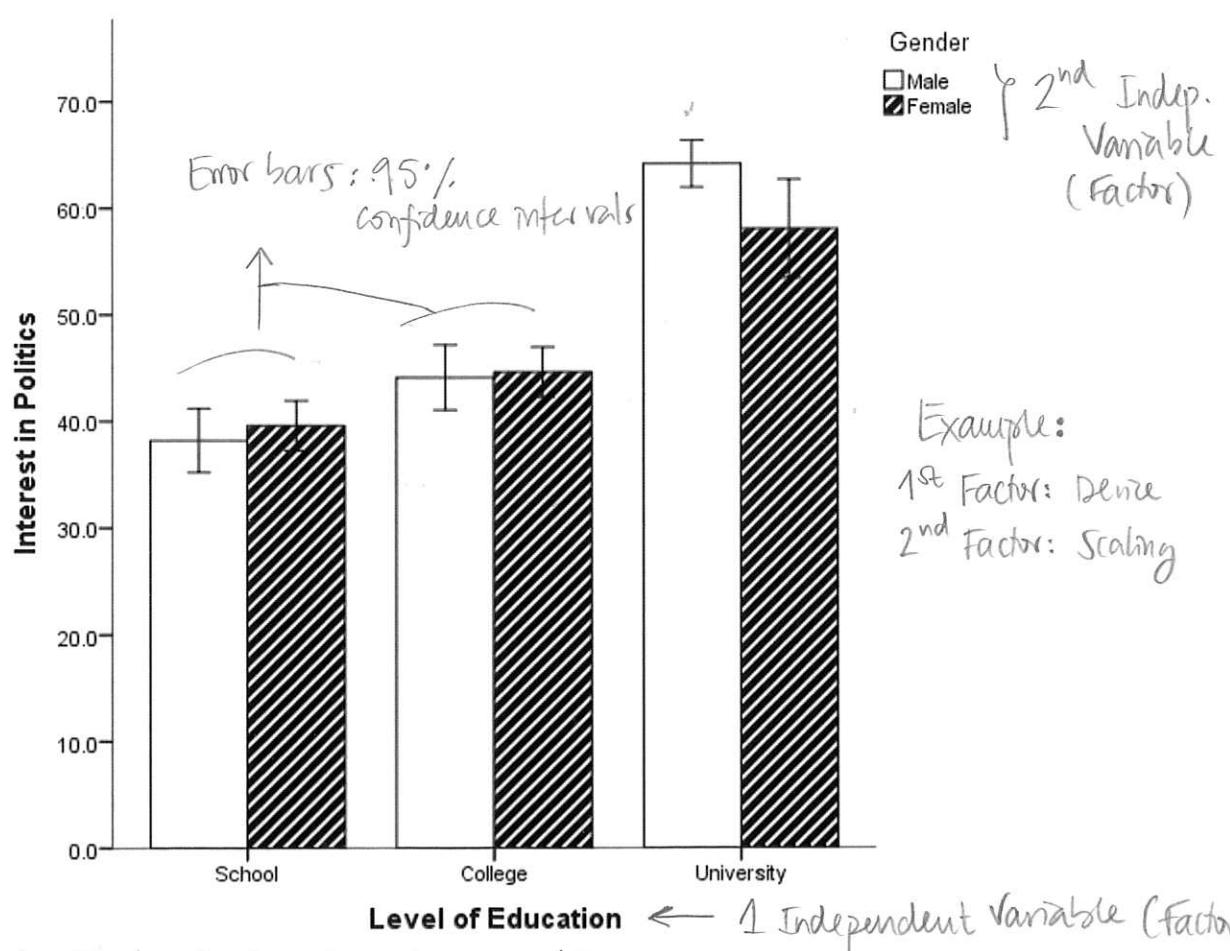
Bar Charts

- Typical for } - Independent samples t-test
 } - One-way ANOVA

Simple Bar Chart

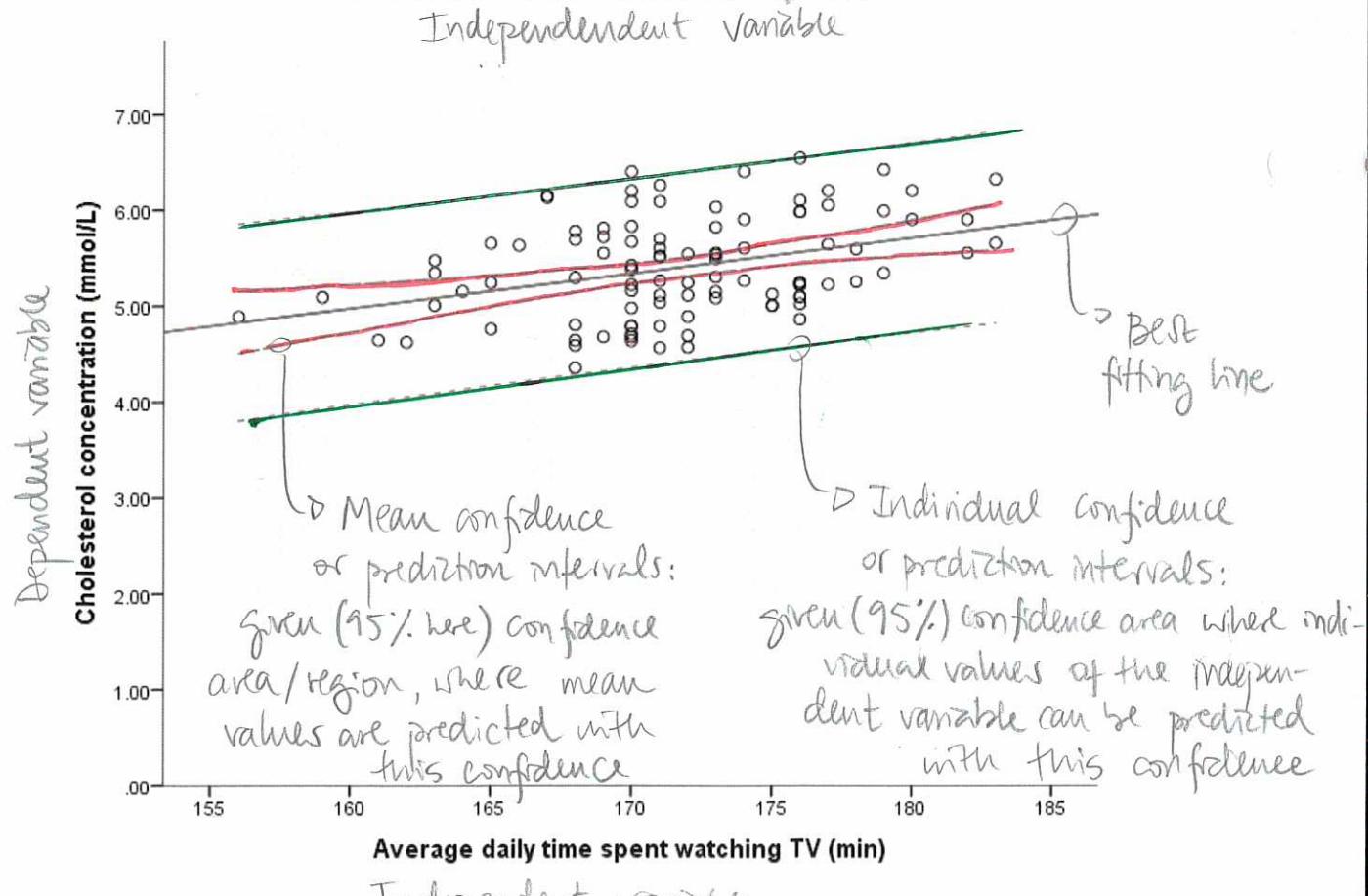
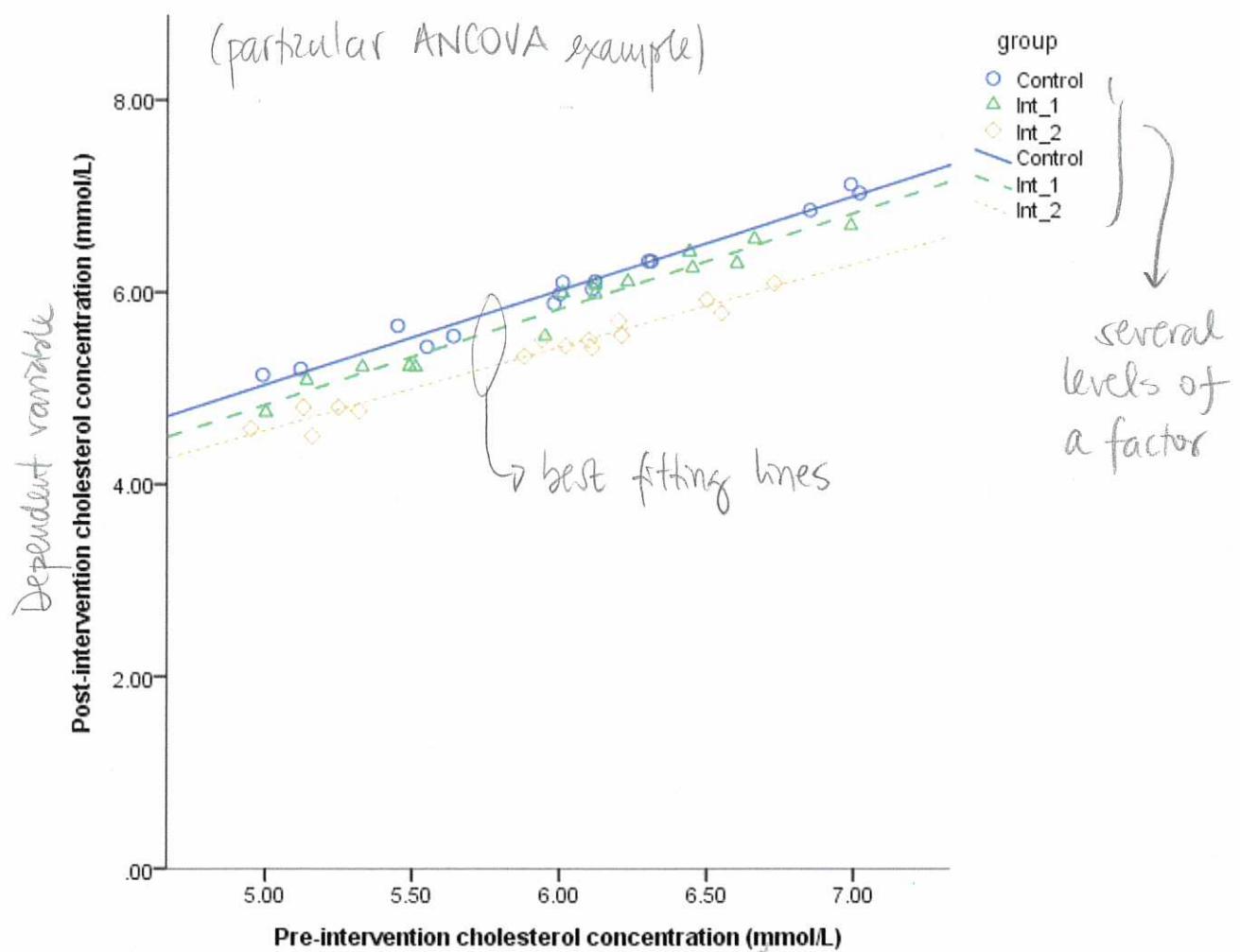


Clustered Bar Chart



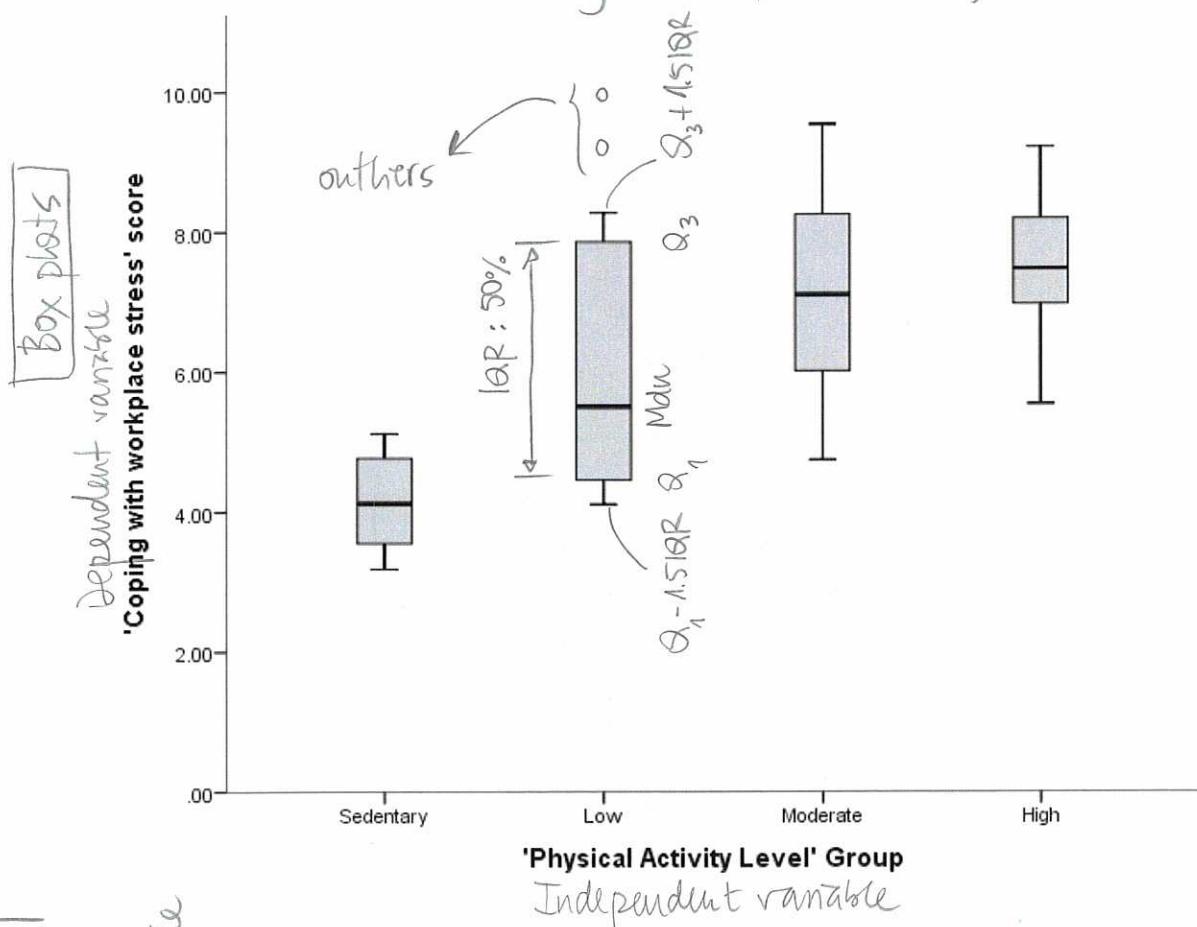
- Typical for } Factorial ANOVAs: two-way (three-way...)

Scatter Plots, AKA Dotplots → For displaying linear regressions



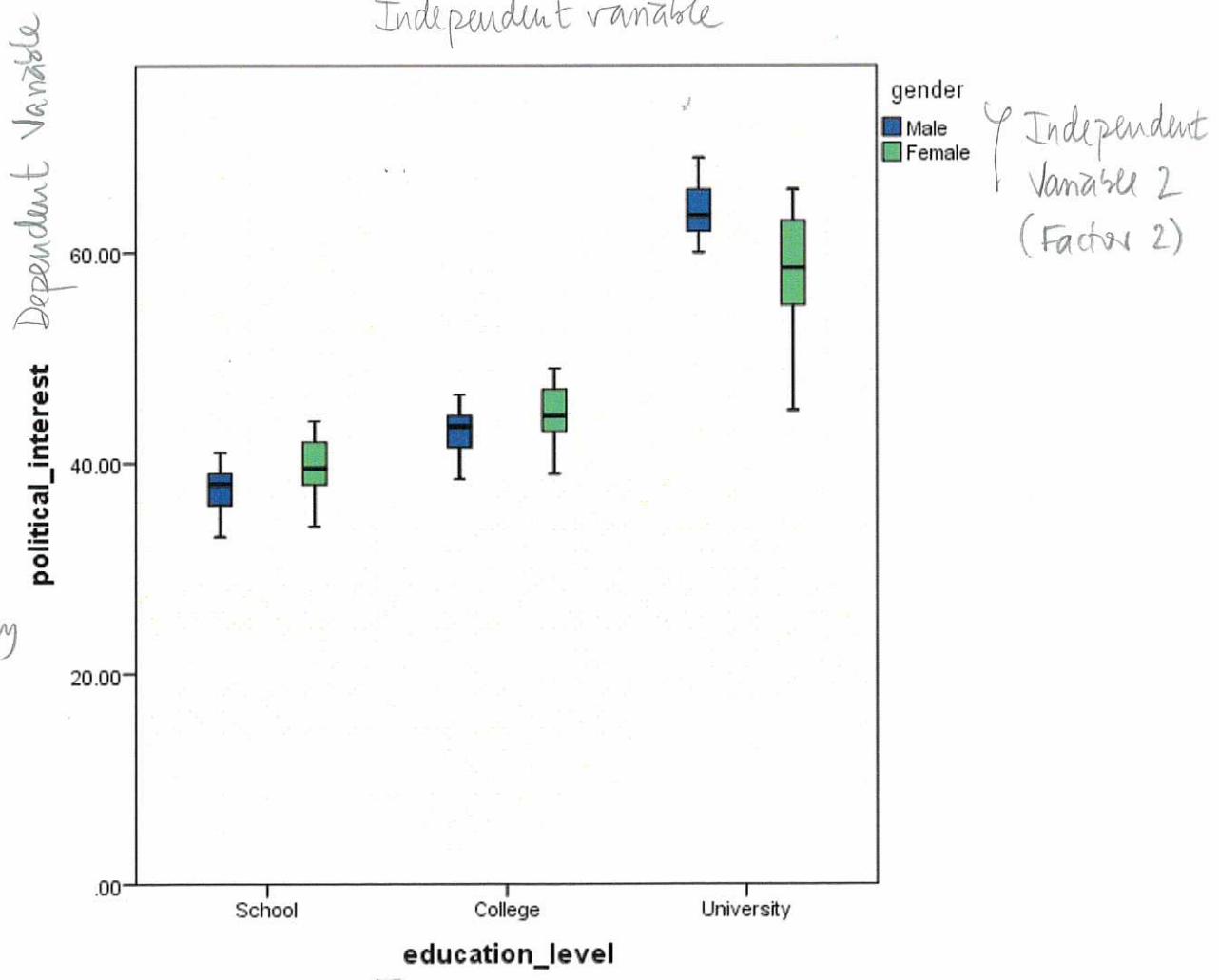
Box Plots

— Uses: One-way ANOVAs, detection of outliers

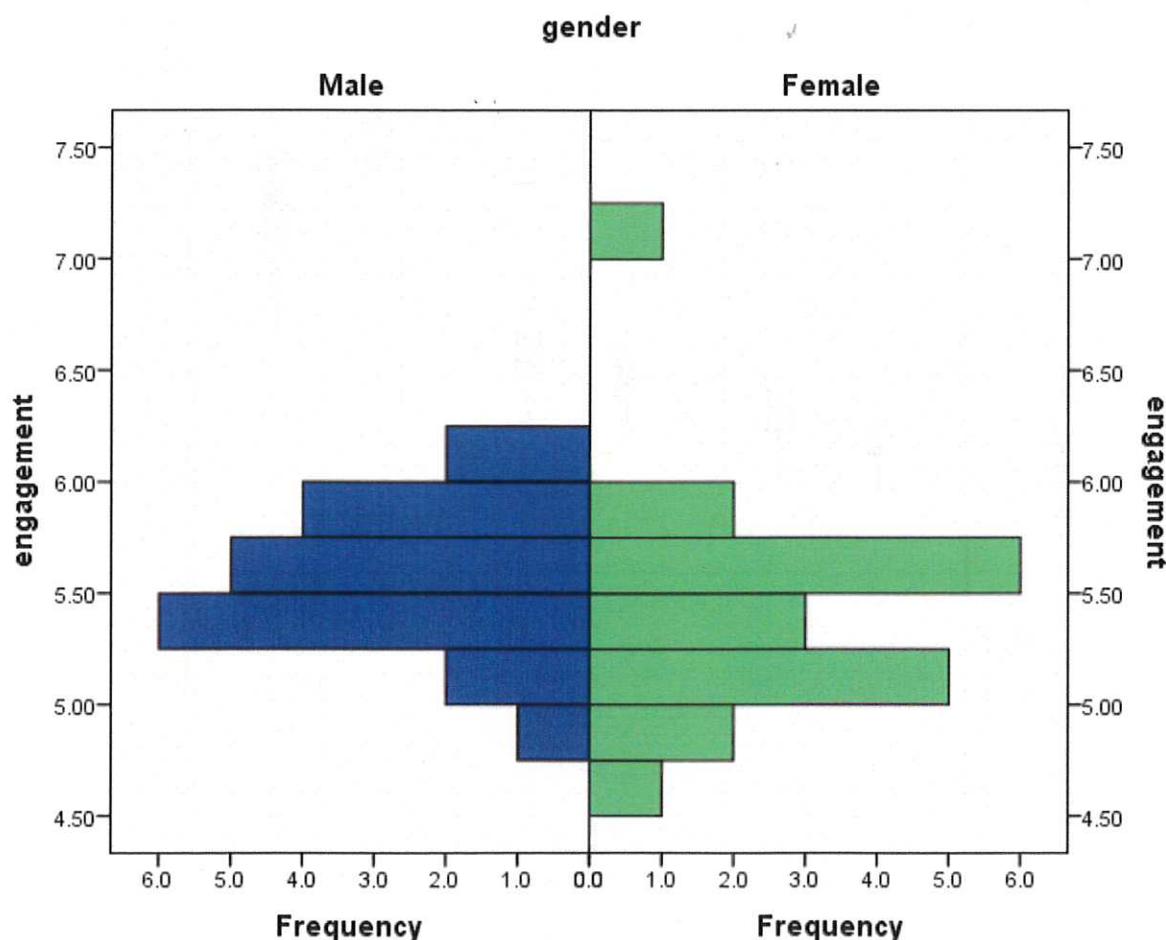
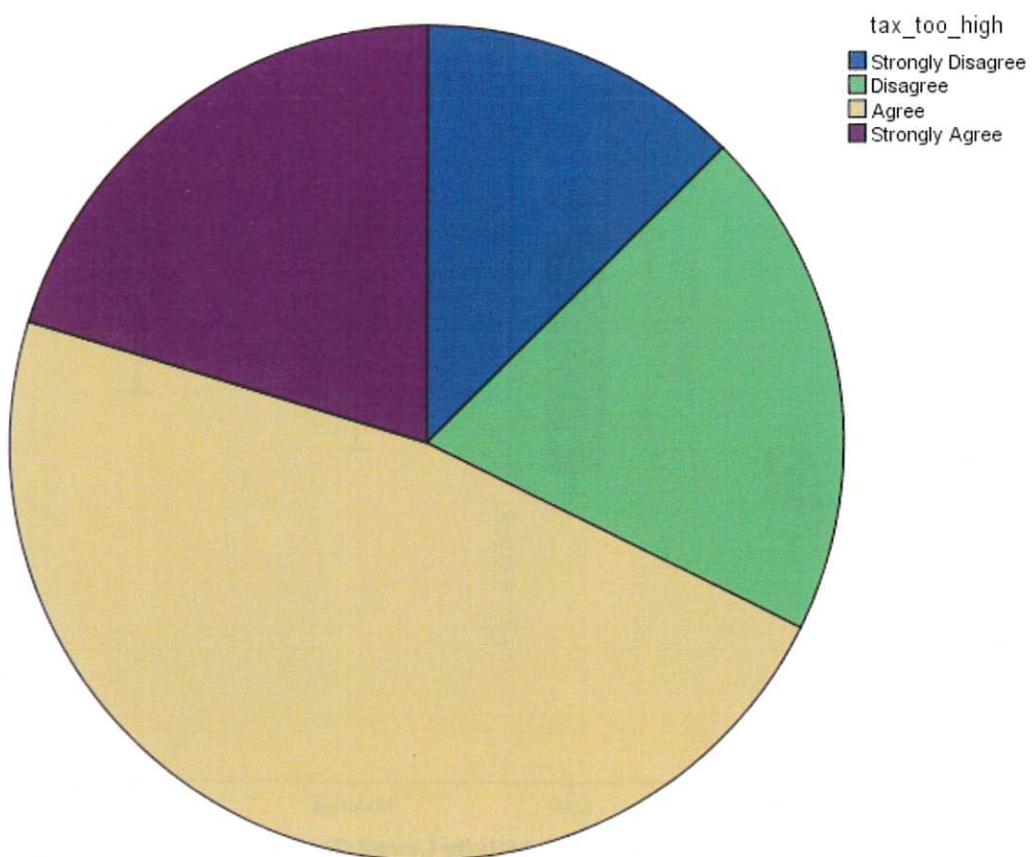


Instead Box Plots

— Use:
Factorial
ANOVAs,
ex: Two-Way
ANOVA



Pies and Pyramids



EXPERIMENTAL DESIGN

- The experimental error should be minimized by keeping factors which are not being investigated as constant as possible!
 - : If repeated observations (=replicates) under apparently identical conditions are made, then the magnitude of the experimental error can be estimated
 - ↳ Standard deviation of the replicates ↳
 - Factors to keep constant: day / time, instructions, age, gender, weather, temperature, light, ... anything we think could affect...
- Other strategies during experimental design that could minimize experimental error:
 - Randomization:
 - Divide experiments into blocks and try blocks in a randomized sequence
 - Define control groups, apart from the experimental groups:
 - People in both groups should be similar
 - Control group doesn't receive treatment
 - Blind & double blind trials: they help avoiding unwanted unconscious influencing!
 - Blind trial: when subject doesn't know real treatment
 - Double blind: when experimenter — " — in each case
 - Triple blind: when statistician doesn't know who took which treatment
- Lack of effective experimental design makes
 - { - experimental error bigger
 - real analyzed effect more difficult to detect.

POWER ANALYSIS (in R, package 'pwr')

- Power analysis consists in obtaining one of these four values (usually N), give 3 of them are known:

- Effect size, ES
- Significance, P-value, α
- Power, $1-\beta$
- Sample size, N

- Types of power analyses:

$$\text{POWER} = 1 - \beta = \\ = P(\text{choose } H_1 | H_1 \text{ is true})$$

	H_0 is TRUE H_1 is FALSE	H_0 is FALSE H_1 is TRUE
Choose H_0 Reject H_1	Correct Decision $1 - \alpha$	Incorrect Decision Type II Error $P(\text{Type II}) = \beta$
Reject H_0 Choose H_1	Incorrect Decision Type I Error $P(\text{Type I}) = \alpha$	Correct Decision $1 - \beta$

	Get	Given
MOST IMPORTANT	A priori	N
	Post-hoc	β
	Criterion	α
	Sensitivity	ES
		α, β, ES
		α, N, ES
		β, N, ES
		α, β, N

Usually done during design, using values of pilot studies/research
It's good doing it after all experiments to estimate β .

- Comments on the 4 values:

- Significance, P-value, α :

- $P(\text{choose } H_1 | H_0 \text{ is true})$: probability of finding a significance when there's none
- False positive
- $P(\text{type I error})$
- usually set to $\alpha = 0.05$
- It tells whether the effect found is generalizable to all population using that sample.

- Power, $1-\beta$:

- $P(\text{choose } H_1 | H_1 \text{ is true})$: probability of finding a true significance
- True positive
- β : false negative, probability of not finding significance when there is $P(\text{type II error})$.
- $1-\beta$ is usually set to 0.8.

- Sample size, N : usually what we want to determine

- Effect Size, ES:

- It can be guessed with $\left\{ \begin{array}{l} \text{- Pilot studies; compute out of data} \\ \text{- Research of results in similar studies} \\ \text{- Educated guess, out of experience, knowledge} \end{array} \right\} \leftarrow \text{DO!}$

- Effect size is the magnitude of the effect that is significant;
One possible computation: standardized difference of means:

$$ES = \frac{\bar{X}_A - \bar{X}_B}{S_{AB}}, \quad S_{AB} = \sqrt{\frac{(n_A-1)S_A^2 + (n_B-1)S_B^2}{n_A + n_B - 2}}, \quad S_A, S_B : \text{standard deviations of groups A, B, where in each a different level of the factor which effect is being analyzed is used}$$

↳ pooled standard deviation

But, depending on the type of statistical analysis, other values are defined and used - the effect size is also classified in SMALL, MEDIUM, LARGE

Statistic	SMALL	MEDIUM	LARGE
Means - Cohen's d	0.2	0.5	0.8
ANOVA - f (not F!)	0.1	0.25	0.4
ANOVA - η^2	0.01	0.06	0.14
Regression - f	0.02	0.15	0.35
Correlation - r^2	0.01	0.06	0.14
...			

Student's t-Test

① One-sample t-Test

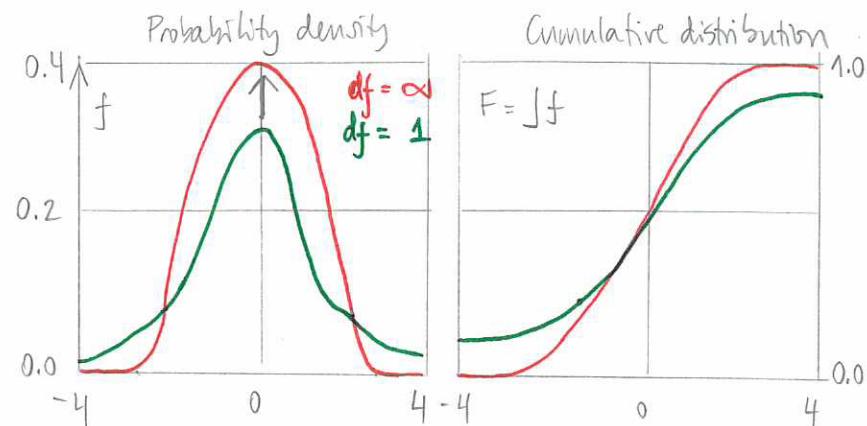
$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}, \quad df = n-1$$

\bar{X} : sample mean

s : sample standard deviation

n : sample size

μ_0 : POPULATION MEAN of the NULL HYPOTHESIS



df: degrees of freedom : depends on sample size

$df \rightarrow \infty$: t distribution becomes $N(0, 1)$

$$t_{\alpha, df}$$

Procedure :

- 1) Select NULL HYPOTHESIS: we check POPULATION MEAN of the selected hypothesis (μ_0) with our sample mean and standard deviation

2) Compute t value

3) See p-value in t distribution :

if p-value $\leq \alpha$: NULL HYPOTHESIS

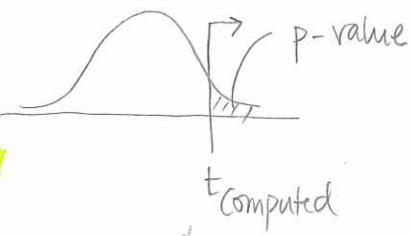
$$\Downarrow$$

is WRONG !

$$P(t_{df} > t_{\text{computed}}) < \alpha$$

\Leftrightarrow Ho wrong (\Leftarrow signif. diff.)

threshold values:
(select according to application)



$$P(t_{df} > t_{\text{computed}}) < \alpha$$

1% : 0.01

Statistical significance level

② Independent Two-Sample Test

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_{12} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \quad s_{12} = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$$

\bar{X}_i : sample mean of group i

s_i : sample standard deviation of group i

$i=1, 2$: groups 1 & 2 \rightarrow for each group: $df = n_i - 1$

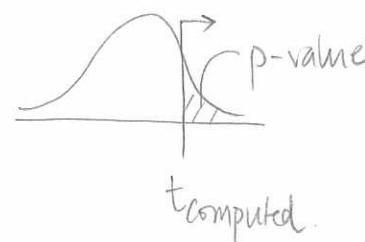
$$df = n_1 + n_2 - 2$$

n_i : sample size of group i

Called:
 $t_{\alpha/2, df}$

→ Procedure:

- 0) Define "before" & "after" groups ($i=1, 2$) (\rightarrow no null hypothesis!?)
- 1) Compute t value
- 2) Look up p-value from t distribution:



If $p\text{-value} \leq \alpha$: both groups do not belong to same distribution \Rightarrow "treatment" has effect!

$$P(t_{df} > t_{\text{computed}}) < \alpha$$



H₀ wrong



significantly diff. groups

threshold value: 5%: 0.05

(select

according
to application)

1%: 0.01

} Statistical significance level

③ Considerations for Student's Tests

- Assumptions: depend on population and sampling

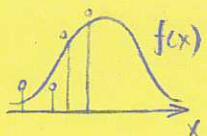
- Normal distribution
- Same variance (population)
- Independent sampling between groups (random?)

- Why not use t-Tests repeatedly in groups pairwise?

- ↳ t-Tests are thought to be used to compare 2 means
- ↳ Comparing 2+ means pairwise leads to false positives!
- ↳ use ANOVA instead!

$$\mu = \int x f(x) dx$$

$$\hookrightarrow \bar{x} = \frac{1}{n} \sum x_i$$



$$\sigma^2 = \int (x - \mu)^2 f(x) dx$$

$$\hookrightarrow s_n^2 = \frac{1}{n} \sum (x_i - \bar{x})^2$$

$$\hookrightarrow s^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2 \quad (\text{Bessel correction})$$

μ, σ : population data ; \bar{x}, s : sampled data

ANOVA: Analysis of Variance

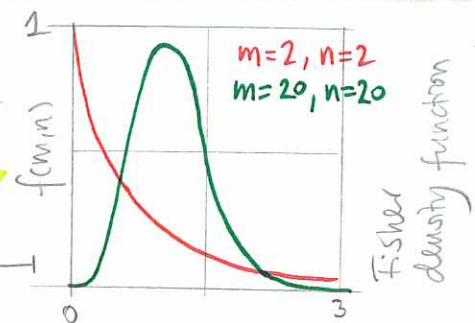
(- When t-Test not possible)

- Groups > 2

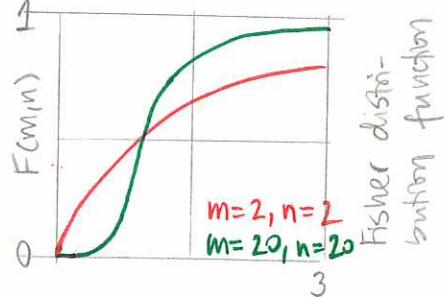
- Factors / Independent variables { 1: One-way
2: Two-way

- Variables / Dependent vars { 1: ANOVA
2: MANOVA (multivariate)

Probability of variation ratios (F) between groups and between observations (within groups)



Fisher function definition



m=2, n=2
m=20, n=20

Fisher distribution function

One-way ANOVA (independent samples)

I	Group 1	Group 2	Group 3	K groups : 3
N = $\sum_{i=1}^I n_i$	$n_1 = 5$	$n_2 = 6$	$n_3 = 4$	$n_i = 1, 2, \dots, K$ sample size of group i
TOTAL SAMPLE SIZE	$x_{11} = 7$ 8 6 5 $x_{15} = 2$	$x_{21} = 5$ 4 3 5 $x_{26} = 2$	$x_{31} = 1$ 2 3 $x_{34} = 2$	3 in this case
15	$\bar{x}_1 = 5.6$	$\bar{x}_2 = 4.16$	$\bar{x}_3 = 2$	\bar{x}_i : sample mean of group i
	$s_1 = 2.30$	$s_2 = 1.47$	$s_3 = 0.82$	s_i : sample standard deviation of group i

IMPORTANT:

- "Between subjects design": each group gets only one treatment
- "Within subjects design": all subjects try all treatments \Rightarrow REPEATED MEASURES!

\bar{x} : mean of all observations:

$$\bar{x} = \sum_{i=1}^K \frac{n_i}{N} \bar{x}_i = \frac{1}{N} (n_1 \bar{x}_1 + n_2 \bar{x}_2 + n_3 \bar{x}_3) = 4.06$$

in our case

- Within groups variance: "weighted total variance"

$$s_w^2 = \sum_{i=1}^K \frac{(n_i - 1)}{(N - K)} \cdot s_i^2 = \frac{(n_1 - 1)}{(N - K)} s_1^2 + \frac{(n_2 - 1)}{(N - K)} s_2^2 + \frac{(n_3 - 1)}{(N - K)} s_3^2 = 3,51$$

$\hookrightarrow = (n_1 - 1) + (n_2 - 1) + (n_3 - 1) = \sum_{i=1}^I (n_i - 1) = N - K$

variances weighted with dfs.

- Between groups variance: "total variance as if groups' means were observations"

$$s_b^2 = \sum_{i=1}^I n_i \frac{(\bar{x}_i - \bar{x})^2}{K-1} = 14.45$$

$$F = \frac{14.45}{3.51} = 4.12 \rightarrow P(F_{2,12} > 4.12) = 0.026$$

Hypothesis building | Necessary when detecting statistical significances between groups (ANOVA, t-Tests)

- H_0 : Population means of all groups are equal
 H_a : Population means are not equal \neq they are unequal!

→ ANOVA Table: it's basically S_b^2 & S_w^2

Source of variation	Sums of squares (SS)	df	Mean Square
BETWEEN Groups (Treatment)	$SSG = \sum_{i=1}^K n_i (\bar{X}_i - \bar{X})^2 = (K-1) S_b^2$ Variability BETWEEN group means	K-1	$MSG = \frac{SSG}{K-1} = S_b^2$
WITHIN Error	$SSE = \sum_{i=1}^K \sum_{j=1}^{n_i} (x_{ij} - \bar{X}_i)^2 = \sum_{i=1}^K (n_i - 1) S_i^2 = (N-K) S_w^2$ Variability WITHIN group means	N-K	$MSE = \frac{SSE}{N-K} = S_w^2$
Total	$SST = \sum_{i=1}^K \sum_{j=1}^{n_i} (x_{ij} - \bar{X})^2 = SSG + SSE$	N-1	

F Statistic computation

$$F = \frac{MSG}{MSE} = \frac{S_b^2}{S_w^2}$$

= variability due to treatment effect & variability due to chance

Degrees of freedom:

(S_b^2) K-1 (groups)

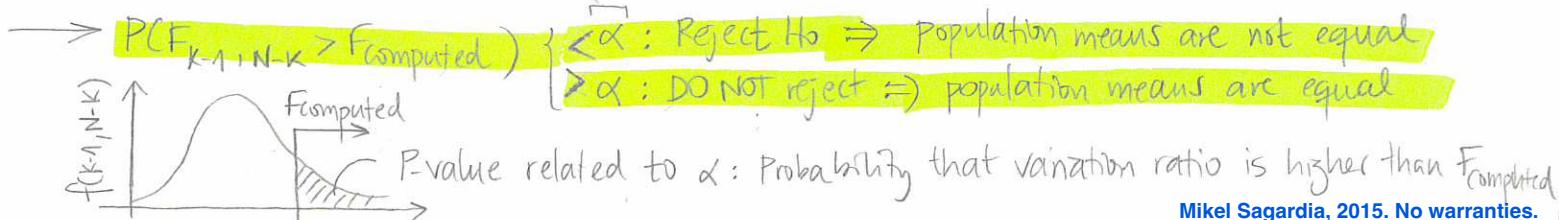
(S_w^2) N-K (observations)

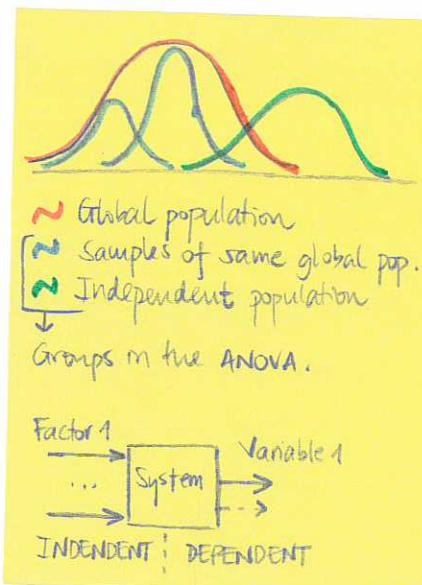
S_b^2 : variation of group means (\bar{X}_i) around overall mean \bar{X} ,
summed all group variation values (K-1)

S_w^2 : variation of observations (x_{ij}) around their group mean (\bar{X}_i),
summed all observation variation values (N-k)

- $F \uparrow \Leftrightarrow S_b^2 \gg S_w^2$: variations between groups are higher than variations between observations
 \Rightarrow populations from which group data was drawn are significantly different
 \Rightarrow groups belong to differentiated distributions
 \Rightarrow conditions / treatments have a significant effect
 \Rightarrow we reject null hypothesis (H_0) of equal means in favour of H_a
 $(\alpha \sim 0.05 - 0.001)$

- $F \downarrow \Leftrightarrow S_b^2 \ll S_w^2$: H_0 of equal means cannot be rejected





ANOVA & Hypothesis testing

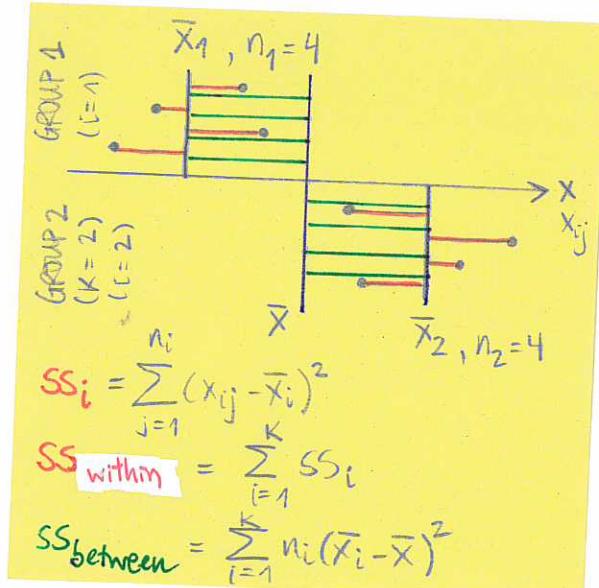
- The influence of factors F_i (independent variables) upon response variables (V_i , dependent) is analyzed.

- H_0 : V_i means are the same for all F_i combinations (groups)
 \Leftrightarrow are groups are a sub-population \rightsquigarrow of the general population \rightsquigarrow

If H_0 is rejected, we conclude a factor has significance to create indep pop \Rightarrow

ANOVA Assumptions to be checked

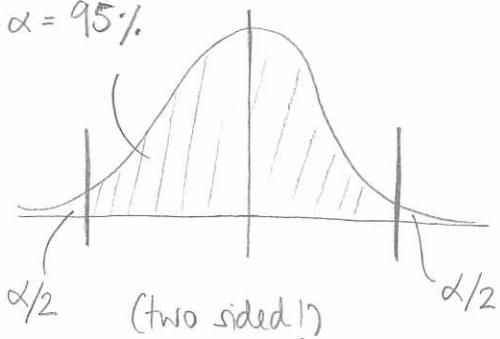
- (i) Random sampling
- (ii) Within each group, dependent variable (response) normally distributed
 → Normal Q-Q plots for checking
- (iii) While group means can be different, group standard deviations must be the same!
 → Check that $\frac{\max(s_i)}{\min(s_i)} \leq 2$.



- Statistical significances (typical values): α
 - put always better p-value
 - p-value doesn't indicate the importance of the observed effect, only the consistency of H_0
- (typical significance levels)

- Confidence intervals: typical CI: 95% → which is the interval where 95% of the values lie?

$$1-\alpha = 95\%$$



$$x_i \rightarrow \left\{ \begin{array}{l} \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \\ s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \end{array} \right.$$

if H_0 is true, then $s^2 = S_W^2$??

df
degrees of freedom

• Standard error: $e = \frac{s}{\sqrt{n}}$

• $1-\frac{\alpha}{2}$ quantile of t distribution: $t_{CI} = t_{\frac{1-\alpha}{2}, n-1}$

Confidence interval of $1-\alpha$

$$[\bar{x} - e \cdot t_{CI}, \bar{x} + e \cdot t_{CI}]$$

- What to do with SIGNIFICANT ANOVA results?

- H_0 of equal means rejected, at least one population mean different

We could use the normal distribution if $n \geq 30$

ONLY if ANOVA SIGNIFICANT!

• ANOVA doesn't say how different!

doesn't say which is different!

→ For checking pair-wise, a modified t-Test could be used for detecting significances

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S_W \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \quad S_W = \sqrt{\sum_{i=1}^K \left(\frac{n_i - 1}{N-K} \right) s_i^2} : \text{For all groups, not only 2!}$$

→ Fisher's least Significant Difference (LSD): same idea as before?

$$|\bar{x}_1 - \bar{x}_2| > LSD = t_{\alpha, N-K} \cdot S_W \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

significance level α , given: 0.05, 0.01

$\sqrt{2/n}$ if groups same size!

The difference between 2 group means ($|\bar{x}_1 - \bar{x}_2|$) is significant (α) if it's bigger than the LSD for a given α significance level".

→ Example for LSD: 50 participants divided in 5 groups with 10 each watch a movie where car accidents occur.

They are asked for the velocity of the cars when cars

H_0 : velocity means don't change when different words chosen

H_a : velocity means do change depending on chosen words (what we want to prove with reduction to absurd)

- group 1: make contact
- 2: hit
- 3: bump
- 4: collide
- 5: smash

1	2	3	4	5
Contact	Hit	Bump	Collide	Smash
21	23	35	44	39
20	30	35	40	44
1	1	1	1	1
26	35	21	44	55
\bar{x}	30	35	38	41
				46

ANOVA

Source	df	SS	MS	F	$P_e(F)$
Groups	4	1460	365	4.56	0.00361
Error	45	3600	80		
Total	49	5060	102		

$$LSD = t_{\alpha/2, N-K} \cdot S_{W2} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = t_{\alpha/2, N-K} \sqrt{\frac{80.2}{10}} = 4 t_{\alpha/2, N-K}$$

< 0.01,
highly signif.

We can either compute a LSD for a certain α (0.05, 0.01) $| \bar{x}_a - \bar{x}_b | > LSD = 4 t_{\alpha/2, 45} = 8.04$: significant mean diff. between 2 groups.
or obtain the p-value (α) for a given difference of means $| \bar{x}_a - \bar{x}_b |$

$\alpha = 0.01$

	1	2	3	4	5
	Contact	Hit	Bump	Collide	Smash
\bar{x}_i	30	35	38	41	46
1 Contact	30	0	5 ns	8 ns	11 **
2 Hit	35	0	3 ns	6 ns	11 **
3 Bump	36		0	3 ns	8 ns
4 Collide	41			0	5 ns
5 Smash	46				0

→ Fisher's LSD can be corrected to decrease false significance with the Fisher-Hayter procedure: MLSD: $MLSD = \frac{q_{\alpha, N-K}}{\sqrt{\frac{1}{n}}} S_{W2}$

$$q_{\alpha, N-K} \sqrt{\frac{1}{n}} S_{W2}$$

tables ← | Studentized range distribution for } range = $k-1$
df = $N-k$

Repeated Measures One-Way ANOVA (Related samples)

- "Within subjects design" used: all users try all conditions/treatments \Rightarrow samples are **independent!**
- Typical for experiments where
 - time effect is the analyzed factor
 - different conditions tested

$k=3$ time points or conditions

Subjects	T_1/C_1	T_2/C_2	T_3/C_3	Subject Means	x_j
1	45	50	55	50	x_j
2	42	42	45	43	
3	36			40	
4	39			38	
5.	51			55	
6	44	45.3	49.7	49.7	
\bar{x}_i	42.8	45.3	49.7		$\bar{x} = 45.9$

Subjects are added as a block, as if they were a new condition!

$n_i = n$ for all $i = 1 \dots k$: all groups/conditions have same number of subjects...

→ Repeated Measures ANOVA table:

	SS	df	$MS = \frac{SS}{df}$	F
Conditions or Groups	$SSG = \sum_{i=1}^k n_i (\bar{x}_i - \bar{x})^2 = (k-1) S_b^2$	$k-1$	$MSG = \frac{SSG}{k-1}$	$MSG / MSE!$
Subjects	$SS_{\text{Subjects}} = k \cdot \sum_{j=1}^n (x_j - \bar{x})^2$	$n-1$	$MS_{\text{Subjects}} = \frac{SS_{\text{Subjects}}}{n-1}$	$MS_{\text{Subj}} / MSE!$
Error (is modified, decreased!)	$SSE' = \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 = \sum_{i=1}^k (n_i - 1) S_e^2 = (N-k) S_e^2$ SSW	$N-k$	$MSE' = \frac{SSE'}{(n-1)(k-1)}$	
Total	$SST = \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2$	$N-1$		

- Subjects & Total line are often omitted!

- Basic idea: error is corrected by subtracting the variances introduced by subjects, which are the same for all conditions! \Rightarrow Error $\downarrow \Rightarrow F \uparrow \Rightarrow$ more probable significance!

This is the used F-statistic with corrected error (subject variability subtracted!).

→ For Repeated Measures ANOVA, Sphericity must be checked: if it is not fulfilled, df_s are corrected (decreased)

• Sphericity occurs when the variances of the differences between all combinations of related groups are equal.

→ it is tested with Mauchly's Test ← widely used, but criticized for:

↓
if sphericity violated ($p < 0.05$)

- overdetecting sphericity in large samples
- underdetecting " in small "

statistic ϵ is proposed for correcting df_s
using 3 different correction values (choose one)

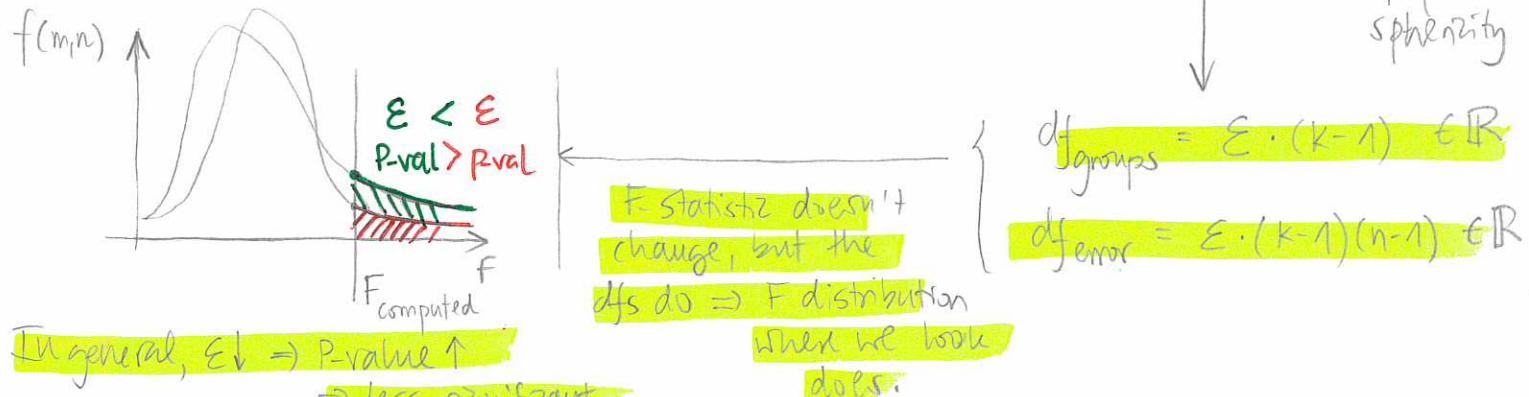
1) Lower bound: $\epsilon = \frac{1}{K-1}$ (don't use it)

2) Greenhouse-Geisser: use if $\epsilon < 0.75$

3) Huynh-Feldt: use if $\epsilon > 0.75$

$$\epsilon \in \left[\frac{1}{K-1}, 1 \right]$$

perfect sphericity



Example:

Within Subjects Effect	Mauchly's W	χ^2 Approx.	df	P-value, Sig.	(K-1)?	ϵ
Condition (ex: Time)	0.434	3.343	2	0.188	0.636	0.760

↓
W-Statistic is computed in the Mauchly's test and checked with the χ^2

↑
Mauchly's Test sphericity violated if P-value < 0.05

	df group	df error	$F = \frac{\text{SSG}/\text{df group}}{\text{SSE}/\text{df error}}$	P-value, Sig.
Sphericity assumed	2	10	12.534	$P(F_{2,10} > 12.534) = 0.002$
Greenhouse-Geisser	1.277	6.384	12.534	$P(F_{1.277, 6.384} > 12.534) = 0.009$
Huynh-Feldt	1.520	7.602	12.534	$P(F_{1.520, 7.602} > 12.534) = 0.005$
Lower bound	1.0	5.0	12.534	$P(F_{1.0, 5.0} > 12.534) = 0.017$

SPHERICITY IN ANOVA REPEATED MEASURES

Assumption for ANOVA.

$$\text{COV}(A) = \begin{bmatrix} S_1^2 & S_{12} & S_{13} & S_{14} \\ S_{21} & S_2^2 & S_{23} & S_{24} \\ S_{31} & S_{32} & S_3^2 & S_{34} \\ S_{41} & S_{42} & S_{43} & S_4^2 \end{bmatrix}$$

A₁ A₂ A₃ A₄

Covariance matrix of FACTOR A with Levels 1 - 4 ($K=4$)
(let's consider One-way ANOVA)

- Equal variances requires:

$$S_1^2 \approx S_2^2 \approx S_3^2 \approx S_4^2$$

$$\hookrightarrow 3 \text{ or } 4 \times \min(s_i^2) \geq \max(s_j^2)$$

- Matrix is symmetric: $s_{ij} = s_{ji} \quad \forall i, j$
always

- If independent measures (between, not RM): $s_{ij} = 0 \quad \forall i, j \leq K, i \neq j$

→ if not, s_{ij} is the shared/overlapping variance between two levels, and in extension to the "equal variances" requirement, sphericity must be fulfilled!

- Compound symmetry:

- It is met when all s_{ij} are similar.
- Sufficient condition for sphericity, BUT NOT NECESSARY.

- Sphericity:

- It is met when the variances of ^{all} the differences between levels are similar:

$$\underline{s_{x-y}^2} = s_x^2 + s_y^2 - 2 \cdot s_{xy} : \text{all must be similar} \quad \forall x, y \text{ levels } x \neq y$$

difference between levels x & y

Example:

	A ₁	A ₂	A ₃	A ₄
Participant 1	8	9	12	4
2	6	11	16	3
:				

Differences b/w. Levels				
	A ₁ -A ₂	A ₁ -A ₃	A ₁ -A ₄	A ₂ -A ₃
1	-1	-4	4	
2	-5	-10	3	
:				

All have to be similar $\leftarrow s_{1-2}^2 = s_1^2 + s_2^2 - 2s_{12}$
(3~4 factors between min & max)

- What happens if SPHERICITY is not met?

a) Use MANOVA, since it doesn't assume sphericity

- But: MANOVA is not as powerfull as ANOVA RM always

Rule of thumb: if $n > 10 + k$

$$\varepsilon < 0.7 \text{ in ANOVA RM correction}$$

→ use MANOVA

b) Use ε correction for dfs {

- Greenhouse-Geisser [GG], if $\varepsilon < 0.75$
- Huynh-Feldt [HF], if $\varepsilon > 0.75$

- About ε corrections :

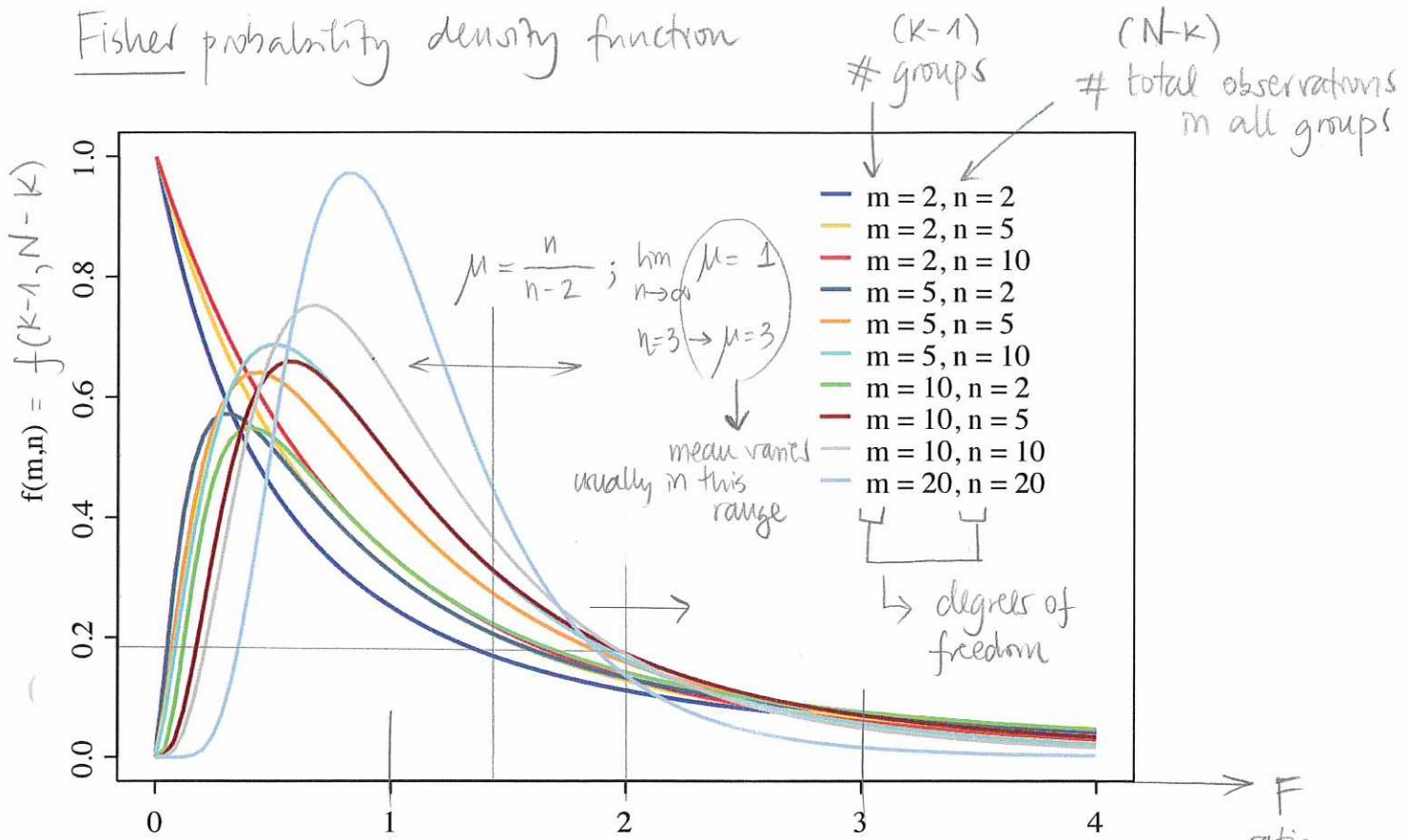
- 1) First check sphericity with Mauchly's test, but also try to observe covariances, because the Mauchly's test isn't allmighty...
 - 2) See GG & HF corrections for ε and decide which to take
 - if $\varepsilon \approx 1 \Rightarrow$ ASSUME SPHERICITY
 - $\varepsilon \approx 0.75 \left\{ \begin{array}{l} \text{• GG underestimates } \varepsilon \text{ if } \varepsilon \approx 1 \Rightarrow \text{use it if } \varepsilon < 0.75 \\ \text{• HF overestimates } \varepsilon \Rightarrow \text{use it if } \varepsilon > 0.75 \end{array} \right.$
 - Lower bound for ε : $1/(k-1)$, that means:

Levels	lowest ε possible
2	1
3	0.5
4	0.3
5	0.25
...	

 with 2 levels
 you have always
 sphericity - it makes sense;
- $K=2 \uparrow \begin{bmatrix} S_1^2 & S_{12} \\ S_{21} & S_2^2 \end{bmatrix} \rightarrow S_{12}$ → "is similar to itself"

→ R: use ezANOVA package for ANOVA RM with sphericity corrections (Mauchly's, GG, HF).

Fisher probability density function



$$F = \frac{S_b^2}{S_w^2} : \text{variation of group means } (\bar{x}_i) \text{ around overall mean } (\bar{x})$$

S_b^2 : variation of group means (\bar{x}_i) around overall mean (\bar{x})

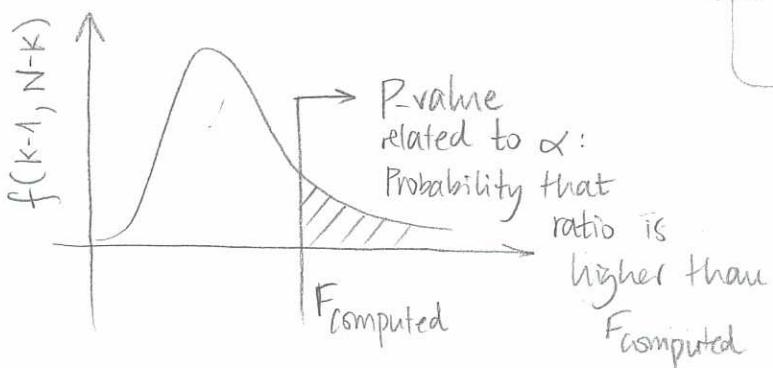
S_w^2 : Variation of observations (x_i) around group mean (\bar{x}_i)

(= F_{computed})

- The Fisher probability density function measures the probability that a variation ratio between group means and within group (F_{computed}) occurs

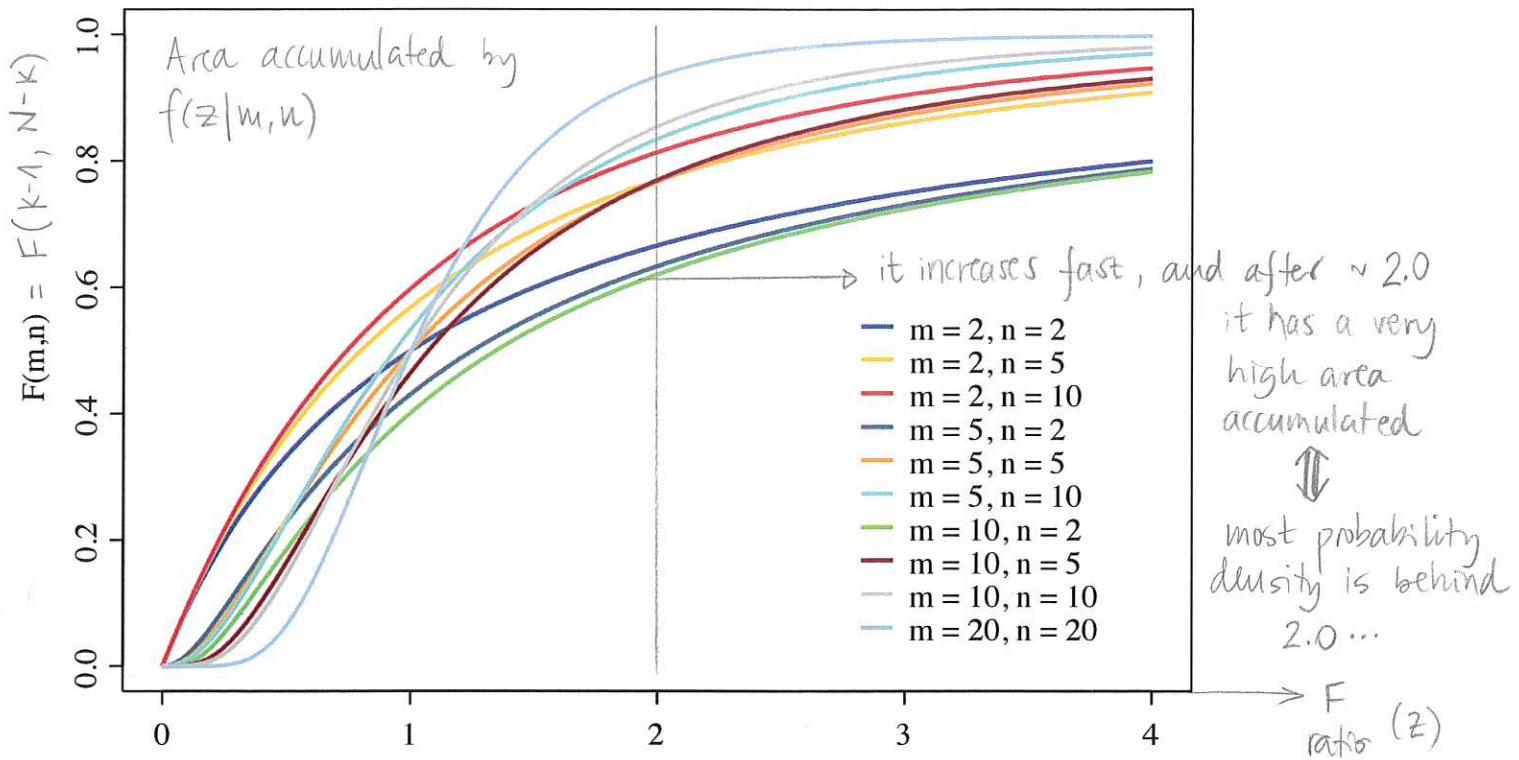
$P(F_{k-1, N-k} > F_{\text{computed}}) < \alpha$: Reject $H_0 \Rightarrow$ population means are not equal

$> \alpha$: DO NOT reject $H_0 \Rightarrow$ population means are equal



"F tends to be larger if the alternative hypothesis is true"

$$\text{Fisher distribution function : } F_{m,n} = \frac{\chi_m^2/m}{\chi_n^2/n} = \int_z f(z|m,n) dz$$



$$F(z|m,n) = \int_z f(x|m,n) dx$$

TWO-WAY ANOVA

- Factorial ANOVAs are the extension of one-way ANOVA to cope with several factors or independent variables for the same dependent variable.
- Two factors: two-way ANOVA
Three n : three-way ANOVA
...
→ for each, Repeated Measures could be applied as in one-way...
- Same assumptions as for one-way: in each group
 - normality
 - equal variances
 - randomly chosen sample
- Example: "How does studying geometry improve spatial visualization skills?"
→ skills tested with / without geometry classes for boys & girls

X_{ijk} : one value

i = levels of factor B
j = levels of factor A
K = sample number/id

Factor B: Education: Geometry YES / NO (i=1,2)

		B1	B2	Overall mean (means of rows)
		YES	NO (Control)	
Females		6, 8, 5, 16, 7, 8, 4, 9, 6, 5 ($n_{1.} = 10$)	6, 3, 4, 2, 4, 6, 4, 3, 5, 3 ($n_{2.} = 10$)	$\bar{X}_{1.} = 5.2, n_{1.} = 20$
Factor A: Gender (= 1, 2)	A1	$\bar{X}_{11} = \frac{1}{n_{1.}} \sum_{k=1}^{n_{1.}} X_{11k} = 6.4$	$\bar{X}_{12} = \frac{1}{n_{2.}} \sum_{k=1}^{n_{2.}} X_{12k} = 4.0$	$\bar{X}_{1.} = 5.2, n_{1.} = 20$
	Males	7, 9, 8, 7, 9, 10, 6 6, 10, 8 ($n_{21} = 10$)	5, 3, 2, 4, 4, 3, 3, 1 5, 4 ($n_{22} = 10$)	$\bar{X}_{2.} = 5.7, n_{2.} = 20$
Overall mean (means of columns)		$\bar{X}_{.1} = \frac{1}{n_{1.}} = 7.2$	$\bar{X}_{.2} = \frac{1}{n_{2.}} = 3.7$	$\bar{X}_{..} = 5.45$ (mean of all values)
		means \bar{X}_i		

— ANOVA table 1/11

Source	SS, sum of squares	df, degrees of freedom
Between		
main		
Factor A (rows)	$SS_A = \sum_i n_{i.} (\bar{X}_{i.} - \bar{X}_{..})^2$	$df_A = \text{levels}_A - 1$
Factor B (columns)	$SS_B = \sum_j n_{.j} (\bar{X}_{.j} - \bar{X}_{..})^2$	$df_B = \text{levels}_B - 1$
Interaction AxB	$SS_{AXB} = \sum_i \sum_j n_{ij} (\bar{X}_{ij} - \bar{X}_{..})^2 - SS_A - SS_B$	$df_{AXB} = df_A \cdot df_B$
Within / Error	$SS_W = \sum_i \sum_j \sum_k (x_{ijk} - \bar{X}_{ij})^2$	$df_{error} = N - (\text{levels}_A) \cdot (\text{levels}_B)$
Total	$SS_{TOT} = SS_W + SS_A + SS_B + SS_{AXB}$	$df_{total} = N - 1$

- ANOVA table 11/11

	Sum of Squares	df	MS, mean square	Same as M one-way ANOVA: detect significance in main effects or interaction
Between				
Factor A (main)	SS_A	df_A	$MS_A = SS_A / df_A$	$F_A = MS_A / MS_W$
Factor B (main)	SS_B	df_B	$MS_B = SS_B / df_B$	$F_B = MS_B / MS_W$
Interaction	SS_{AXB}	df_{AXB}	$MS_{AXB} = SS_{AXB} / df_{AXB}$	$F_{AXB} = MS_{AXB} / MS_W$
Within / Error	SS_W	df_{error}	$MS_W = SS_W / df_{error}$	
Total	SS_T	df_{total}		

- Basically, 3 one-way ANOVAs are done:
 - 2 for main effects of each factor
 - 1 for interaction effects; an interaction implies that differences in one of the factors depend on differences in another factor.
- ↓
VERY IMPORTANT
- If interaction is high: be carefull with main effect interpretation...

- What were our hypothesis?

• Main effects

$$H_0: \mu_{A1} = \mu_{A2}$$

$$H_0: \mu_{B1} = \mu_{B2}$$

H_1 : means are not equal

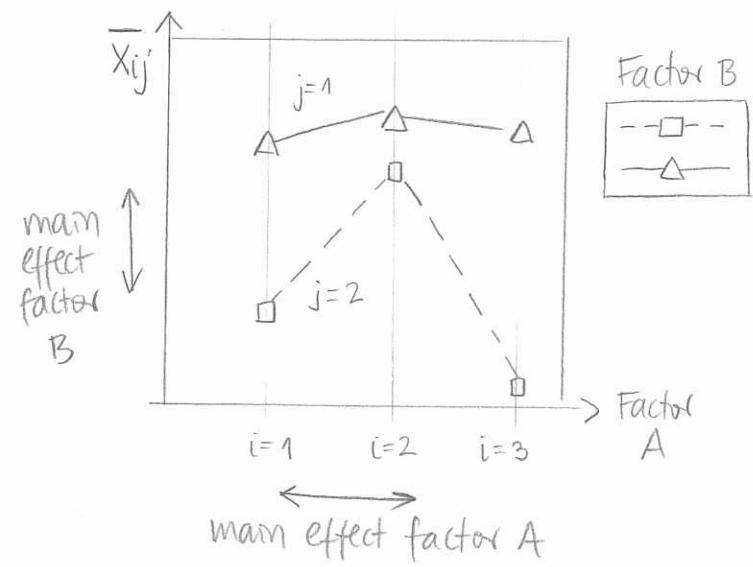
• Interaction

H_0 : there is no interaction effect in the population:

$$\mu_{A1B1} - \mu_{A2B1} = \mu_{A1B2} - \mu_{A2B2}$$

H_1 : not both differences are equal (all)

- Plots:



- Means are plotted (\bar{X}_{ij})
- Factor with most levels has the horizontal axis (easier interpretation)
- { Parallel lines: NO INTERACTION
- { Crossing lines: INTERACTION: check if it's significant
- A main effect is significant if the mean for one level of the factor is sufficiently different from the mean for another level of the factor

ANOVA: Between vs Within (Repeated Measures) Design

```
> aov.1 <- aov(ttc ~ condition * difficulty, data = Peg_Total)
> aov.2 <- aov(ttc ~ condition * difficulty + Error(subject/(condition * difficulty)), data = Peg_Total)
> summary(aov.1)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
condition	2	1983	991.7	3.863	0.02233 *
difficulty	1	1400	1400.5	5.455	0.02034 *
condition:difficulty	2	2654	1327.0	5.168	0.00634 **
Residuals	240	61618	256.7		

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

```
> summary(aov.2)
```

Error: subject

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Residuals	40	15910	397.7		

Error: subject:condition

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
condition	2	1983	991.7	4.193	0.0185 *
Residuals	80	18920	236.5		

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Error: subject:difficulty

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
difficulty	1	1400	1400.5	6.815	0.0127 *
Residuals	40	8220	205.5		

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Error: subject:condition:difficulty

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
condition:difficulty	2	2654	1327.0	5.717	0.00478 **
Residuals	80	18568	232.1		

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

1983,37 (manually computed)

$SS_A = SS_{\text{condition}} = SS_{\text{Between groups of factor A}}$

$$= n_{V_1} (\bar{Y}_{V_1..} - \bar{Y}_{..})^2 + n_{V_2} (\bar{Y}_{V_2..} - \bar{Y}_{..})^2 + n_{FF} (\bar{Y}_{FF..} - \bar{Y}_{..})^2$$

$$= \sum_{k=1}^{K_A} n_{A_k} (\bar{Y}_{\text{Factor } A \text{ cell}} - \bar{Y}_{\text{Total}})^2$$

(61618,2 ✓)

$SS_{\text{Residual}} = SS_{\text{Error}} = SS_{\text{Within groups of factor B}}$

$$= \sum_{i=1}^n (Y_{V_1e,i} - \bar{Y}_{V_1e..})^2 + \sum_{i=1}^n (Y_{V_2e,i} - \bar{Y}_{V_2e..})^2 + \sum_{i=1}^n (Y_{FFe,i} - \bar{Y}_{FFe..})^2 +$$

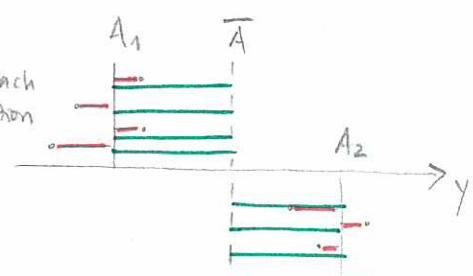
$$+ \sum_{i=1}^n (Y_{V_1d,i} - \bar{Y}_{V_1d..})^2 + \sum_{i=1}^n (Y_{V_2d,i} - \bar{Y}_{V_2d..})^2 + \sum_{i=1}^n (Y_{FFd,i} - \bar{Y}_{FFd..})^2$$

$n_{AK,..} = n + n = 82$
2 diff levels for each condition

$$K_A = 2, K_B = 3, n = 41$$

degrees of freedom:
possible contrasts for
each variable

$$df_A = K_A - 1 \quad df_B = K_B - 1$$



$$\begin{aligned} 1971,66 &= 1971,66 \\ 431,01 &= 431,01 \\ 501,52 &= 501,52 \\ 1947,24 &= 1947,24 \\ 361,21 &= 361,21 \\ 56406,16 &= 56406,16 \end{aligned}$$

the σ is also very big!!! ?!

$$1400,496 \checkmark$$

$$SS_B = n_{B_{K_1}} (\bar{Y}_{v,e} - \bar{Y}_{..})^2 + n_{B_{K_1}} (\bar{Y}_{v,d} - \bar{Y}_{..})^2 = \sum_{i=1}^{K_B} n_{B_{K_1}} (\bar{Y}_{v,i} - \bar{Y}_{..})^2 \quad 41 \times 3 = 123$$

2653,902 \checkmark

$$SS_{AB} = n_{B_{K_1}} (\bar{Y}_{v,e} - \bar{Y}_{..})^2 + n_{B_{K_1}} (\bar{Y}_{v,d} - \bar{Y}_{..})^2 + n_{B_{K_1}} (\bar{Y}_{FF,e} - \bar{Y}_{..})^2 +$$

$$+ n_{B_{K_1}} (\bar{Y}_{v,d} - \bar{Y}_{..})^2 + n_{B_{K_1}} (\bar{Y}_{FF,d} - \bar{Y}_{..})^2$$

$$- SS_A - SS_B$$

- When we have a repeated measures each subject tries all combinations of factors

↓ Each subject has his/her inherent abilities / characteristics responses to the treatment
 ↓ We can explain part of the error with this pattern \Rightarrow the error is smaller when repeated measures performed!!!

$$SS_{\text{Subjects Total}} = (A) \sum_{i=1}^n (\bar{Y}_{v,e,i} - \bar{Y}_{..})^2 + (B) \sum_{i=1}^n (\bar{Y}_{v,d,i} - \bar{Y}_{..})^2 + (C) \sum_{i=1}^n (\bar{Y}_{FF,i} - \bar{Y}_{..})^2 \quad [80]$$

2 difficulties

$$(B) \sum_{i=1}^n (\bar{Y}_{v,e,i} - \bar{Y}_{..})^2 + \sum_{i=1}^n (\bar{Y}_{v,d,i} - \bar{Y}_{..})^2 \quad [40]$$

3 conditions

$$(AxB) \sum_{i=1}^n (\bar{Y}_{v,..,i} - \bar{Y}_{..})^2 \quad 2274,635$$

3x2 conditions x difficulties

Mean of subject i for condition = \checkmark

and all difficulties = e, d.

② I don't know if this computation is correct...
 Should be checked...

- Subject means are computed for Repeated Measures (RM) ANOVA

- The Error/Residual SS is decreased with the Subjects SS!

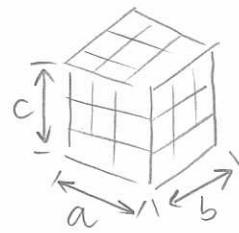
② I don't understand how the Residuals are partitioned into A, B, A:B and how each group's SS_{subjects} is computed completely — But I understand the main idea ...

THREE-WAY ANOVA

: Extension to ONE-WAY and TWO-WAY ANOVAs

- Model:

$\left\{ \begin{array}{l} \cdot \text{factor A with levels } i = 1, \dots, a \\ \cdot \text{factor B with levels } j = 1, \dots, b \\ \cdot \text{factor C - n - } k = 1, \dots, c \end{array} \right\} \rightarrow \text{observations within a cell}$
 $X_{ijkl}, l = 1, \dots, n_{ijk}$



▷ 3 factors / indep variables ; abc groups/cells

if $n_{ijk} = n$: balanced.

- ANOVA table built analogously as in two-way:

$$N = \sum_i \sum_j \sum_k n_{ijk}$$

Source	SS	df	MS	F	P-value
<i>Between</i>					
MAIN	A	$a - 1$	MS_A	MS_A / MS_W	When dividing to 2-way analyses don't forget adjusting the df's and the F distributions... the F distributions
	B	$b - 1$	MS_B	MS_B / MS_W	
	C	$c - 1$	MS_C	MS_C / MS_W	
INTERAC TION	AB	$(a-1)(b-1)$	MS_{AB}	MS_{AB} / MS_W	... the F distributions
	AC	$(a-1)(c-1)$	MS_{AC}	MS_{AC} / MS_W	
	BC	$(b-1)(c-1)$	MS_{BC}	MS_{BC} / MS_W	
	ABC	$(a-1)(b-1)(c-1)$	MS_{ABC}	MS_{ABC} / MS_W	
Error / Within	SS_W	$N - abc$	MS_W		
Total	SS_T				

- General guidelines:

ABC	Significant	NOT significant	all 3 significant	NOT significant
AB		Significant	NOT significant	NOT significant
AC		Significant	NOT significant	NOT significant
BC		Significant	Significant	NOT significant
A			Significant	Significant
B				Significant
C				NOT significant
- Do not interpret main effects (A, B, C)	- Do not interpret main effects (A, B, C)	- Interpret main effect A	- Interpret main effects A, B.	
- Do not interpret 2-way interactions (AB, AC, BC)	- Interpret 2-way interactions (AB, AC, BC)	- Interpret interaction BC		
- Divide into 2-way analyses for levels of a factor: ex: C=1 → AB, C=2 → AB, ...	- Plot AB ignoring C to interpret it; repeat for AC / B BC / A	- Plot AB ignoring C to interpret it; repeat for AC / B BC / A		

OTHER TESTS / TOPICS

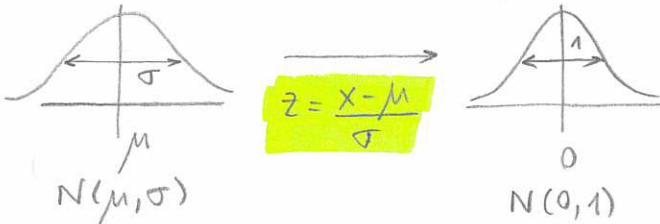
- Branches of statistics analysis:

- DESCRIPTIVE: compute properties of samples (\bar{x} , s, ...)
- EXPLORATIVE: detect patterns (plot; understand)
- INFERENCE: generalization of models from models to population (ANOVA, regression, ...)

- Dependent vs Independent measurements

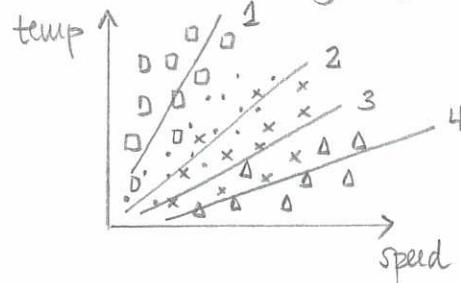
INDEPENDENT	DEPENDENT
BETWEEN - subjects - design Sample of each group is different	WITHIN - subjects - design Groups share same sample, measurements are dependent
a group is a combination of values for independent variables	What's interesting is whether for EACH person there's a change from one group of combination to the other
a variance BETWEEN groups is of interest = systematic fraction of variance of all subjects	The difference between all subjects within one group is not interesting = it is the error
The variance/change within the groups is not interesting	The difference between all subjects within one group is not interesting

- Standardization for obtaining z-value:



- Some other tests:

* ANCOVA: Analysis of covariance



- Test the hypothesis that different groups have the same regression lines
- Example: measure chirping speed vs temperature of 4 species of crickets

* Logistic regression:

$$Y \sim bX \quad (\text{simple})$$

$$Y \sim b_1 X_1 + b_2 X_2 + \dots \quad (\text{multiple})$$

→ nominal variable: Yes/No, A/B/C, ...

Skalenniveaus

- Nominalskala

Geschlecht: M/W

Farbe: Rot/Grau/Blau ...

- Ordinalskala

Schulnoten: "sehr gut", "gut", ...

→ Wie nominal, aber mit Reihenfolge

- Kardinalskala

• Intervallskala

Zeit: $t \in \mathbb{R}$

• Verhältnisskala

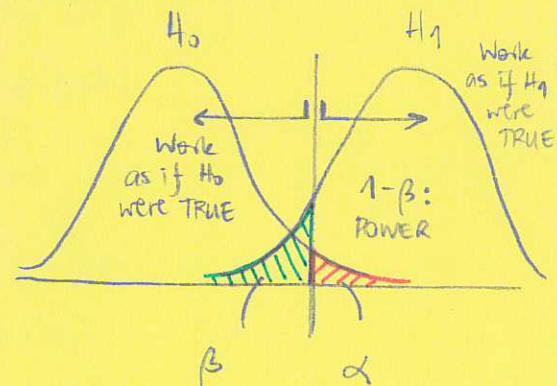
Alter $y \in \mathbb{R} \setminus \{0\}$

Temperatur $T \in \mathbb{R} \setminus \{0\}$

Nullpunkt

Unterschied: Nullpunkt. Dank des Nullpunkts kann man alle Werte absolut vergleichen!

		Sample DECISION	
		Take H_0	Take H_1
Population TRUTH	H_0 is True	✓	α ERROR (Type I)
	H_1 is True	β ERROR (Type II)	✓



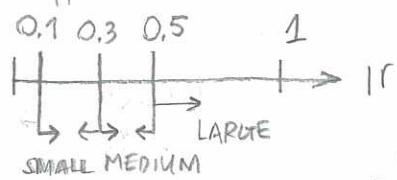
The statistical power ($1 - \beta$) is the probability of taking H_1 when it is true = the probability that our theory is true.

EFFECT SIZE

- Apart from detecting significant differences between groups/treatments, we need to determine how big these differences are: this is the effect size and there are many methods.

- * Regressions
 - Pearson's correlation coefficient, r :
 - Coefficient of determination, r^2 :

Variable	Type	Analysis
Depend. (Measured)	Continuous	Regression
Indep. (Forced)	Nominal	ANOVA



→ "change behavior"

$$r = 0.21 \rightarrow r^2 = 0.0441$$

→ 4.41% of the variance of either variable is shared with the other variable

- * ANOVAs, or similar

- Eta square, η^2 :

$$\eta^2 = \frac{SS_{\text{Treatment}}}{SS_{\text{Total}}}$$

- Analogous to r^2
- Ratio of variance explained in the dependent variable by an independent variable
- It's biased, as r^2 : additional variables $\eta^2 \uparrow$

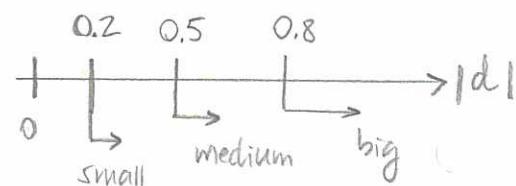
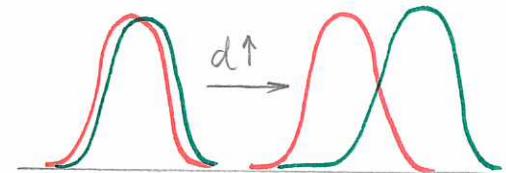
- Cohen's d:

$$d = \frac{\bar{x}_1 - \bar{x}_2}{s_{12}}$$

$$s_{12} = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}}$$

$$s_1^2 = \frac{1}{n_1-1} \sum_{i=1}^{n_1} (x_{1i} - \bar{x}_1)^2$$

Pooled standard deviation



- Hedges' g: similar to Cohen's d, but with a correction

- Confidence intervals (look in ANOVA summary):

$$\bar{x} \pm \underbrace{SE \cdot t_{CI}}_{\text{CI of } 95\%}$$

$$\left. \begin{array}{l} SE, \\ \text{Standard Error} \\ \text{Error} \\ t_{CI} = t_{1-\frac{\alpha}{2}, n-1} \end{array} \right\} \text{Standard Error} = \frac{s}{\sqrt{n}}$$

$$\left. \begin{array}{l} \text{example,} \\ \text{CI of } 95\% \end{array} \right\} 95\% = 1 - \alpha \Rightarrow$$

$$\Rightarrow \alpha = 0.05 \Rightarrow$$

$$\underbrace{t_{0.975, n-1}}_{\text{The value of the t distribution with this parameters}}$$

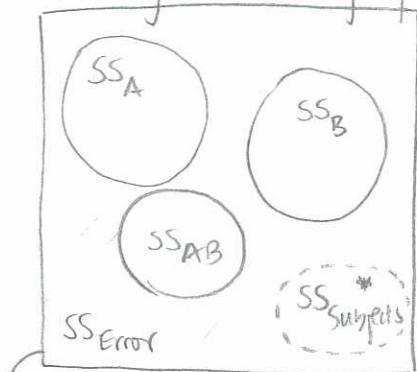
The value of the t distribution with this parameters

TYPES OF SUMS OF SQUARES IN ANOVAS

(From notes by Rudolf Cardinal)

① UNCORRELATED INDEP. VARS.

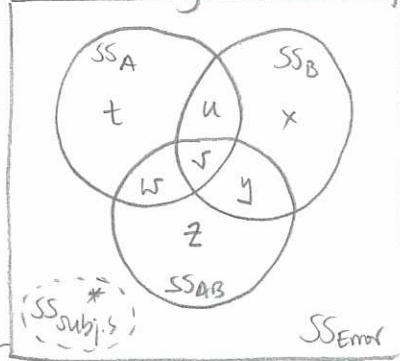
(Orthogonal sums of squares)



$$SS_{\text{TOTAL}} = SS_A + SS_B + SS_{AB} + SS_{\text{Error}}$$

② CORRELATED DEP. VARS.

(non-orthogonal sums of squares)



$$SS_{\text{TOTAL}} \neq SS_A + SS_B + SS_{AB} + SS_{\text{Error}}$$

* In within-design
 SS_{Error} is decreased with the variance introduced by same subjects when computing F.

$$\rightarrow \left[F_A = \frac{SS_A / df_A}{SS_{\text{Error}} / df_{\text{Error}}} \right] = \frac{\text{Treatment}}{\text{Error}}$$

$$F_B = \frac{SS_B / df_B}{SS_{\text{Error}} / df_{\text{Error}}} \dots$$

: "how big is the variance introduced by the treatment compared to the error variance in the system?"

$F \uparrow \rightarrow$ Effect of treatment \rightarrow treatment significant

① If independent variables are not correlated sums of squares don't overlap and the sum of all is the total ss.

② If independent variables are correlated, sums of squares overlap and there are 3 types or ways of computing them:

Type I (default R)

- "Sequential"
- Order of variables affects result!
- If A first: $SS_A = t + u + v + w$
- $SS_B = x + y$
- $SS_{AB} = z$
- Do not use it.

Type III (default SPSS)

- "Marginal" or "orthogonal"
- Order does not affect
- $SS_A = t$
- $SS_B = x$
- $SS_{AB} = z$
- More conservative than others
- Try to use this

Type II

- "Hierarchical"
- Order does not affect
- $SS_A = t + w$
- $SS_B = x + y$
- $SS_{AB} = z$

- When there's no interaction, it's the most powerful
- If there's interaction, there's no guarantee that's more powerful than type III

* Possible ANOVA calls in R:

`aov(dprvar ~ predictors) → summary()`

`lm(dprvar ~ predictors) → Anova()`

`Anova()`

Type II & III supported!

`lme((dprvar ~ predictors) → anova())`

`[ezANOVA(...)]`

very easy to use!

* ezANOVA vs aov()

`summary(aov(dprvar ~ U*V + Error(S/U*V), data = myFrame))`

`ezANOVA(myFrame, dr = .(dprvar), wid = .(S), within = .(U,V),
detailed = TRUE, type = "III")`

- `aov()` can handle also type III SSs, but function `drop1()` must be used for that
 - "error" model is singular": when data of a subject is (partly) missing (for all conditions)
 → in those cases, `lmer()` can be used!

- It's possible to force type III SSs in general R functions (`aov()`) → alternative to giving option in additional packages
 Type at beginning of script:

`options(contrasts = c("contr.sum", "contr.poly"))`

Rudolf Cardinal - ANOVA		BS: between subjects
R Calls for (and SPSS equiv.)		WS: within subjects
- 1 BS	↓	aov, <u>ezANOVA</u>
- 2 BS	↓	
- 3 BS	↓	
- 1 WS	↓	aov, ezANOVA
- 2 WS	↓	
- 3 WS	↓	
- 2 BS, 1 WS	↓	
- 1 BS, 2 WS	↓	
- 1 BS covariate (ANCOVA)		
- 1 BS covariate, 1 BS factor		
- 1 BS covariate, 2 BS factor		
- 2+ BS covariates [1+ BS factor]		
- 1 WS covariate		
...		

NON-PARAMETRIC TESTS

- They use no parametrized distributions
- Sometimes less powerful than their parametric equivalents
 - ↳ use them only when assumptions for parametric tests not met:
 - When no normality
 - When ordinal data: ranks → Example: Subjective rankings

Other options

- Remove outliers
- Skewed distributions can be transformed into normal applying logarithmic transformation

Parametric	Non-Parametric Equivalent	Assumptions & Explanation / Example
Paired-Samples t-Test	Wilcoxon signed-rank, W	<ul style="list-style-type: none"> - Population doesn't have to be normally distributed - Sample randomly chosen - Treatment values (e.g. before & after) from same sample - Values are measured in an ordinal scale or they're transformed to a ranking. - Distribution of differences of treatment values should be symmetric around median - After ranking values are defined, differences (+ and -) are summed for the computation of the statistic W^+ & W^- - Hypothesis testing as usual (but same median instead of mean)
Subject	1 2 3 4 5	
Test before train.	0.77 0.49 0.66 0.28 0.38	
Test score after training	0.40 0.72 0.00 0.36 0.55	
Difference	0.37 -0.23 0.66 -0.07 -0.17	
Rank	1 2 3 4 5	
Abs. Difference	0.08 0.17 0.23 0.37 0.66	
Sign	- - - + +	
Sum +/- ranks: $W^- = 1+2+3 = 6$	$W^+ = 4+5 = 9$	
$S = \min(W^-, W^+) = 6$, $n_{\text{difference}} = 5$	$\rightarrow p\text{-value}_{\text{Computed}}$	
Independent-Samples t-Test	Mann-Whitney Test, U .	<ul style="list-style-type: none"> - Use it when no normality ↳ But distributions same shape! - Data samples independent - Data at least in one ordinal scale - Hypothesis testing as usual after computing statistic. (but same median instead of mean)
- n -	Wilcoxon rank sum, W : same as Mann-Whitney Test	Example of Wilcoxon rank sum:

Class (years)	Reading score	Position	Rank
2nd (8)	0.0	1	1
2nd (8)	0.36	2	2
2nd (8)	0.40	3	3
4th (10)	0.55	4	(4+5)/2 = 4.5
2nd (8)	0.55	5	(4+5)/2 = 4.5
4th (10)	0.57	6	6
4th (10)	0.70	7	7
4th (10)	0.72	8	(8+9)/2 = 8.5
2nd (8)	0.72	9	(8+9)/2 = 8.5
4th (10)	0.84	10	10

$$\Rightarrow \begin{array}{|c|c|} \hline \text{Class (years)} & \text{Sum of ranks} \\ \hline \text{2nd (8)} & 19 \\ \text{4th (10)} & 36 \\ \hline \end{array}$$

$W = \min(\text{group}_1, \text{group}_2) = \min(19, 36) = 19$
 $n_{\text{group}_1} = 5$
 $n_{\text{group}_2} = 5$

p-value computed

Parametric

Indep.-Samples
One-Way
ANOVA

Non-Parametric Equivalent

Kruskall-Wallis Test, H

Assumptions & Explanation / Example

- Extension of the Mann-Whitney U test to 3 or more groups treated independently
- Alternative to one-way ANOVA when [data not normally distributed]
- Data at least on an ordinal scale
- Distributions of samples should have same shape, although not necessarily normally distributed
- Hypothesis testing as always, but same median assumption instead of mean
- Statistic H is computed using ranks and it is checked with the χ^2 for the p-value

$$H = (N-1) \frac{SSG}{SST}$$

N: # samples, subjects, ...
SSG: Sum Squares Between Groups
SST: Sum Squares Total
K: # groups/treatments
 $\chi^2(df=K-1)$ → p-value computed

Computed like
in the ANOVA,
but using the
ranks instead of
the real values

Repeated Measures
One-way
ANOVA

Friedman Test

- Alternative when assumptions of normality or equality of variance not met.

- Extension of the Wilcoxon signed-rank
- Data measured in at least one ordinal scale.
- Hypothesis testing: equivalent, but with same median

- If the test gives a significant result, pairwise comparisons can be made using Wilcoxon signed-rank after Bonferroni correction

Example (Friedman Test)

	j = 1..K			Each point ranked ↓	j = 1 j = 2 j = 3			
	Start	1 Month	1 Year		Start	1 Mon	1 Year	
Point 1	17.4	13.6	13.2	Mean rank if two same values	1	3	2	1
	2	15.7	10.1		2	3	2	1
	3	12.9	10.3		3	3	1.5	1.5
	4	9.8	9.2		4	3	2	1
	5	13.4	11.1		5	3	2	1
	6	18.7	20.4		6	1	3	2

Sum of ranks: R_j	16	12.5	7.5
R_j^2	256	156.25	56.25

Formula of the statistic:

$$\begin{aligned} M &= \left(\frac{12}{n \cdot K(K+1)} \cdot \sum_{j=1}^K R_j^2 \right) - 3n(K+1) \\ &= \frac{12}{6 \cdot 3(3+1)} \cdot (256 + 156.25 + 56.25) - 3 \cdot 6(3+1) \\ &= \dots \end{aligned}$$

→ M, n, K] → tables are checked for p-value or M_{crit} for critical level α is computed

- If K & n exceed tables (i.e., high numbers) ⇒ $M \sim \chi^2(df=K-1)$ ⇒ compute p-value

ASSUMPTIONS for DEPENDENCE TESTS

visual inspection

T-Tests, ANOVA make these assumptions (although they can be robust if violation ...).

1) Each sample is an independent random sample

2) The distribution of the response variable follows a normal distribution check (2)

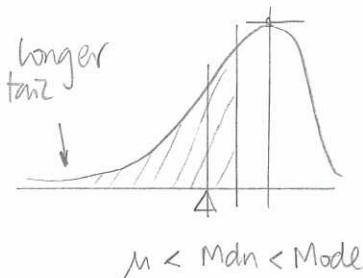
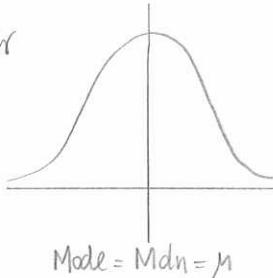
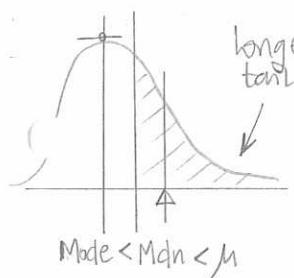
3) The population variances are equal across responses for the group level, that is called homogeneity of variance or homoscedasticity.

check (3)

- **Levere's Test**
- **Bartlett's Test**
- Box-plots
- Plot residuals vs fitted values

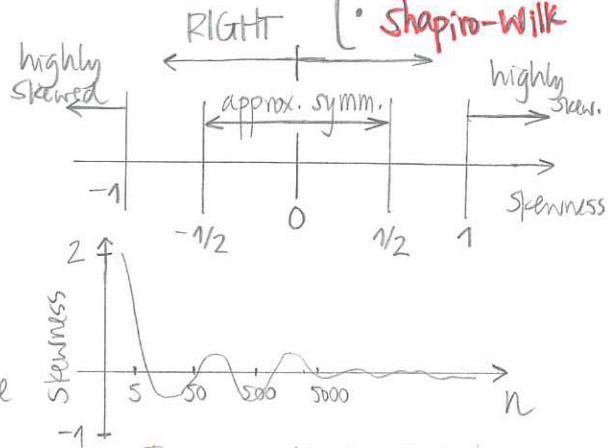
visual inspection

SKEWNESS: measure of asymmetry



Skewed left =
= negatively skewed
Skewness > 0

Skewed right =
= positively skewed
Skewness < 0



The sample size (n) has a big effect on the skewness, so be carefull!!

- Computation of skewness (g_1 for populations, G_1 for samples)

$$g_1 = \frac{m_3}{(m_2)^{3/2}}$$

$$m_2 = \frac{\sum (x - \bar{x})^2}{n} = \sigma^2$$

$$m_3 = \frac{\sum (x - \bar{x})^3}{n} \quad (\text{3rd moment})$$

(for populations)

CORRECTION FOR SAMPLES

$$G_1 = \frac{\sqrt{n(n-1)}}{n-2} g_1$$

if $n \uparrow$, $G_1 \approx g_1$

- Standard error of skewness, SES:

$$\text{SES} = \sqrt{\frac{6n(n-1)}{(n-2)(n+1)(n+3)}}$$

- Two tailed test for Skewness $\neq 0$ at $\alpha = 0.05$ significance level:

$Z_{G_1} = G_1 / \text{SES} \longrightarrow \begin{cases} \cdot Z_{G_1} < -2 : \text{population very likely to be skewed negatively (LEFT)} \\ \text{BUT YOU DON'T KNOW HOW MUCH!} \\ \cdot Z_{G_1} \in [-2, 2] : \text{no conclusion could be reached: it could be skewed} \\ \cdot Z_{G_1} > 2 : \text{population very likely skewed positively (RIGHT) or sym.} \\ \text{BUT YOU DON'T KNOW HOW MUCH!} \end{cases}$

- Confidence interval (95%) for skewness

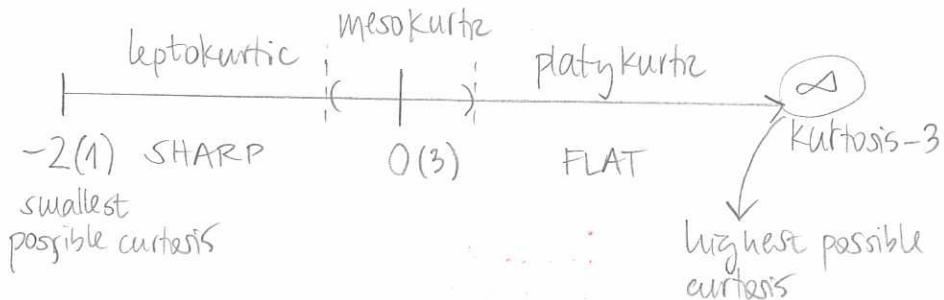
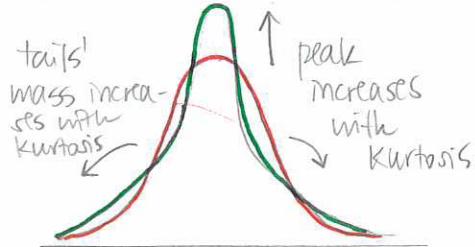
$$CI = G_1 \pm 2 \text{SES} \quad \xrightarrow{\text{example}} CI = \underbrace{-0.1098}_{G_1} \pm 2 \cdot 0.2414 = [-0.5926, 0.3730]$$

Skewness is with 95% prob. in this region

KURTOSIS: measure of height and sharpness ; excess of kurtosis is often reported:

kurtosis - 3

→ kurtosis of normal distribution



→ Computation of excess of kurtosis (g_2 for populations, G_2 for samples):

$$g_2 = \frac{m_4}{m_2^2} - 3, \quad m_4 = \frac{\sum (x-\bar{x})^4}{n}, \quad m_2 = \frac{\sum (x-\bar{x})^2}{n} = \sigma^2 \quad \left. \begin{array}{c} \text{CORRECTION} \\ \text{FOR SAMPLES} \end{array} \right\} \quad G_2 = \frac{n-1}{(n-2)(n-3)} [(n+1)g_2 + 6]$$

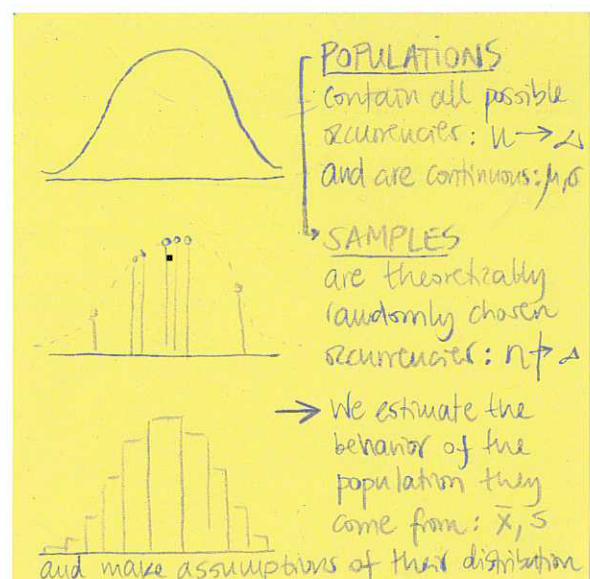
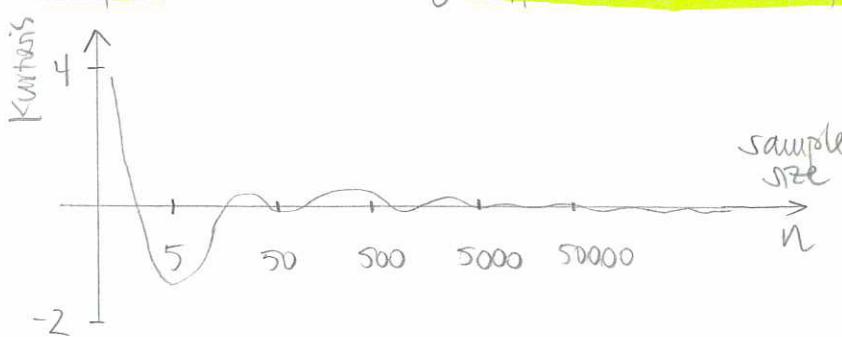
- Standard error of kurtosis, SEK:

$$SEK = 2 \cdot SES \cdot \sqrt{\frac{n^2 - 1}{(n-3)(n+5)}}$$

- Two tailed test of excess kurtosis $g_2 \neq 0$ at approximately $\alpha = 0.05$

$$Z_{g_2} = G_2 / SEK \rightarrow \begin{cases} \cdot Z_{g_2} < -2 : \text{population very likely to have negative excess kurtosis (kurtosis} < 3) \\ \cdot \text{BUT YOU DON'T KNOW HOW MUCH!} \\ \cdot Z_{g_2} \in [-2, 2] : \text{no conclusion can be reached} \\ \cdot Z_{g_2} > 2 : \text{population very likely to have a positive excess kurtosis (kurtosis} > 3) \end{cases}$$

- Sample size n has a big effect on kurtosis, be careful! (as with skewness)



NORMALITY CHECKS

① D'Agostino-Pearson / D'Agostino's K² omnibus test

$$\text{DP} = \underbrace{z_{g_1}^2}_{\substack{\text{DP-} \\ \text{statistic}}} + \underbrace{z_{g_2}^2}_{\substack{\text{skewness} \\ \text{statistic}}} \sim \chi^2(\text{df} = 2)$$

it follows
a χ^2 distribution
with $\text{df} = 2$.

With small sizes, D'Agostino-Pearson has tendency to reject normality
Use it with $n > 20$

- Example: $\text{DP} = z_{g_1}^2 + z_{g_2}^2 = (0.45)^2 + (0.44)^2 = 0.3961$

$$\Pr(\chi^2(\text{df} = 2) > 0.3961 = \text{DP}) = 0.8203 \rightarrow \text{you cannot reject the assumption of normality.}$$

② Shapiro-Wilk

W = computed using the covariance matrix...

W_{critical} is computed for n & $\alpha (= 0.05)$: if $W \leq W_{\text{critical}}$: H_0 rejected,
NO NORMAL DISTRIBUTION

HOMOGENEITY OF VARIANCE CHECKS

① Levene's test

$$W = W(N, K, X, \dots) \xrightarrow{\substack{\text{groups} \\ \text{total observations}}} \text{computed}$$

$\Pr(F(\alpha, K-1, N-K) > W_{\text{computed}}) < \alpha (0.05)$: significative probability of homogeneity of variance

② Bartlett's test

$$\chi^2 = \text{computed with variances} \quad \alpha = 0.05$$

\rightarrow if $\chi^2 > \chi^2_{n-1, \alpha}$: H_0 is rejected: variances not equal \Rightarrow no homogen. of variances
 H_0 : all n population variances are equal

HYPOTHESES

H_0 in ANOVA: all samples have equal means

- We define two hypotheses

- { · H_0 : null hypothesis ~ "no effect, no improvement, no changes..." ↗
- H_1 or H_a : alternative hypothesis ~ " H_0 is false" OR a hypothesis that not necessarily is complementary of H_0

→ Procedure: consider H_0 true and try to show with reduction to absurd that was a wrong consideration.

- Usually H_1 is the one we want to show by testing H_0 : so we pick H_0 & H_1 to be mutually exclusive & exhaustive

- Type I & II errors:

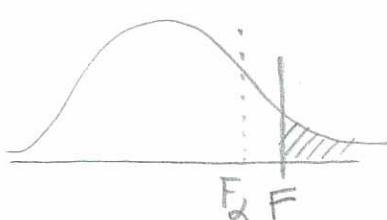
- $P(\text{choose } H_1 \mid H_0 \text{ is true}) = P(\text{type I error}) = \alpha$] → level of significance, set to $\alpha \sim 0.05$
- $P(\text{choose } H_0 \mid H_1 \text{ is true}) = P(\text{type II error}) = \beta$] → it can be improved by increasing sample size, but usually we set n and α
- $P(\text{choose } H_1 \mid H_1 \text{ is true}) = \text{Power} = 1 - \beta$

		H_0 is TRUE	H_0 is FALSE	PROBABILITIES of decisions to be correct / incorrect.
Decision	Truth	H_1 is FALSE	H_1 is TRUE	
	Choose H_0	Correct decision $1 - \alpha$	Incorrect decision Type II error β	
Reject H_1	Correct decision $1 - \beta$	Incorrect decision Type I error α		
Reject H_0	Correct decision $1 - \beta$	Incorrect decision Type I error α		
Choose H_1				

+ That's what we usually try to demonstrate/show: α is the probability of error.

- General procedure:

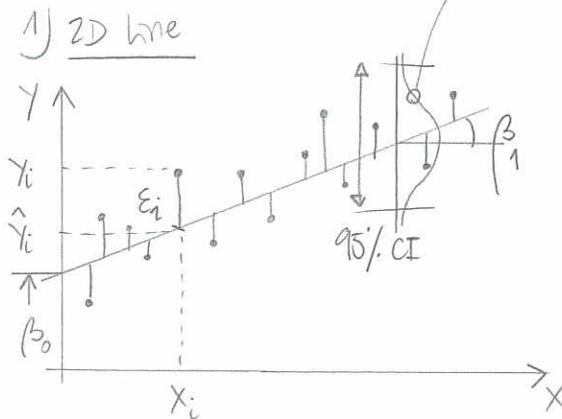
- 1) choose H_0 & H_1 ; design experiment: independent/dependent variables, n , ...
- 2) For each ANOVA/test: compute statistic of the sample: F_{comp}
- 3) Compute P-value for the statistic:



$\Pr(F_{df} > F_{\text{comp}}) \begin{cases} \leq \alpha : \text{Reject } H_0 \text{ at level } \alpha \\ > \alpha : \text{DO NOT reject } H_0 \text{ at level } \alpha \end{cases}$

LINEAR REGRESSION

MATHEMATICAL MODEL



the linear regression predicts the mean

the standard deviation of the error is minimized

$$Y = \beta_0 + x \beta_1 + \epsilon \leftrightarrow \hat{Y} = \hat{\beta}_0 + x \hat{\beta}_1$$

PREDICTION/APPROXIMATION
Best fitting line, with error.

- We look at the minimal error possible for determining the best fitting line: $\beta_i | \epsilon \stackrel{!}{=} \min$

- Error: $\epsilon_i^2 = (Y_i - \beta_0 - x_i \beta_1)^2 \rightarrow$ total error: $SSE = \sum_{i=1}^n \epsilon_i^2 = \text{RSS ; RESIDUAL STANDARD ERROR}$

- Minimum error: $\frac{\partial E}{\partial \beta_i} = 0 : \begin{cases} \frac{\partial}{\partial \beta_0} \left[\sum_{i=1}^n (Y_i - \beta_0 - x_i \beta_1)^2 \right] = 0 \\ \frac{\partial}{\partial \beta_1} \left[\sum_{i=1}^n (Y_i - \beta_0 - x_i \beta_1)^2 \right] = 0 \end{cases} \rightarrow \begin{cases} \sum_i (Y_i - \beta_0 - x_i \beta_1) (-1) = 0 \\ \sum_i (Y_i - \beta_0 - x_i \beta_1) (x_i) = 0 \end{cases}$

$$\rightarrow \begin{cases} \beta_0 \sum_i 1 + \beta_1 \sum_i x_i = \sum_i Y_i \\ \beta_0 \sum_i x_i + \beta_1 \sum_i x_i^2 = \sum_i x_i Y_i \end{cases} \rightarrow \begin{bmatrix} \sum 1 & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum x_i Y_i \end{bmatrix}$$

- Estimate of the σ : $\hat{\sigma}^2 = \frac{SSE}{n-2} = \frac{\sum \epsilon_i^2}{n-2} = \frac{\sum (Y_i - \beta_0 - x_i \beta_1)^2}{n-2} = \frac{\sum (Y_i - \hat{Y}_i)^2}{n-2}$

- $SST = \sum_{i=1}^n (Y_i - \bar{Y})^2$

$$\rightarrow R^2 = 1 - \frac{SSE}{SST} = \frac{RSS}{SST}$$

Correlation between X & Y: R^2
very important statistic that measures how close data are to the fitted regression line: $R^2 \uparrow \text{good! } R^2 \in [0,1]$

- $\hat{\beta}_i$: effect upon depend variable when independent variable i increases 1 unit, holding all other independent variables constant.

$$\left[\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \right], \quad \left[\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \cdot \bar{x} \right] : \text{the difference between } \beta \text{ & } \hat{\beta} \text{ is that the first is the one that hits the perfect prediction}$$

taking into account ϵ . In practice $\beta = \hat{\beta}$.

$$SE(\beta_1)^2 = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \text{ of } \epsilon \sim N(0, \sigma^2)$$

$$\rightarrow \left[\hat{\beta}_i \pm 2 \cdot SE(\hat{\beta}_i) \right] 95\% \text{ CI.}$$

There is approx. 95% chance that this interval will contain β_i for predicting y !

$$SE(\beta_0) = \sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right]$$

- Hypothesis testing:

$$\left. \begin{array}{l} H_0: \text{there is no relationship between } X \text{ & } Y \\ H_1: \text{there is relationship between } X \text{ & } Y \end{array} \right\} \Leftrightarrow \left\{ \begin{array}{l} H_0: \beta_1 = 0 \\ H_1: \beta_1 \neq 0 \end{array} \right.$$

$\Rightarrow t\text{-statistic: } t = \frac{\hat{\beta}_1 - \beta_1}{SE(\hat{\beta}_1)} = \frac{\hat{\beta}_1}{SE(\hat{\beta}_1)} \quad \left|_{df=n-2} \quad \rightarrow \text{Two-tailed p-value is computed.}\right.$

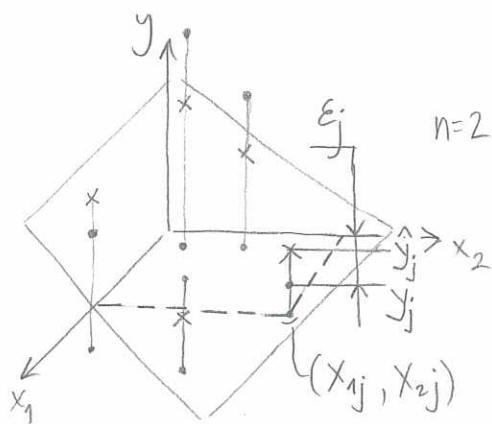
$$\beta_1 = 0$$

- Residual Standard Error: $RSE = \sqrt{\frac{RSS}{n-2}} = \sqrt{\frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{y}_i)^2} = \hat{\sigma}$

$$RSS = SSE$$

Residual sum of squares = Sum of Squares of Errors

2) 3D Plane / Multiple linear regression



$n=2$ $\left\{ \begin{array}{l} \{x_{ij}\}_{i=1,\dots,n}^{j=1,\dots,m}, \{y_j\}_{i=1,\dots,m} \\ \{x_{1j}, x_{2j}, y_j\}_{j=1,\dots,m} \end{array} \right. \rightarrow \text{sample}$

independent variables dependent variable

effect on the dependent variable

when increasing the i th

indep. variable 1 unit, holding others constant,

but usually they change together!

- Model: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon = \beta_0 + \sum_{i=1}^n \beta_i x_i + \varepsilon$

$$\hat{y} = \hat{\beta}_0 + \sum_{i=1}^n \hat{\beta}_i x_i$$

intercept: expected mean value of y when $x=0$.

- Error: $SSE = RSS = \sum_{j=1}^m \varepsilon_j^2 = \sum_{j=1}^m (y_j - \hat{y}_j)^2 = \sum_{j=1}^m (y_j - \beta_0 - \sum_{i=1}^n \beta_i x_i)^2$

$$\left. \begin{array}{l} \frac{\partial (RSS)}{\partial \beta_i} = 0 \quad \left\{ \begin{array}{l} \frac{\partial}{\partial \beta_0} \left(\sum_{j=1}^m (y_j - \beta_0 - \beta_1 x_{1j} - \beta_2 x_{2j})^2 \right) = 0 \\ \frac{\partial}{\partial \beta_1} \left(\sum_{j=1}^m (y_j - \beta_0 - \beta_1 x_{1j} - \beta_2 x_{2j})^2 \right) = 0 \\ \frac{\partial}{\partial \beta_2} \left(\sum_{j=1}^m (y_j - \beta_0 - \beta_1 x_{1j} - \beta_2 x_{2j})^2 \right) = 0 \end{array} \right|_{n=2} \\ \left. \begin{array}{l} \\ \\ \end{array} \right\} \rightarrow \end{array} \right.$$

$$\rightarrow \begin{cases} 2(-1) \cdot \sum_{j=1}^m (y_j - \beta_0 - \beta_1 x_{1j} - \beta_2 x_{2j}) = 0 \\ 2 \sum_{j=1}^m (y_j - \beta_0 - \beta_1 x_{1j} - \beta_2 x_{2j})(-x_{1j}) = 0 \\ 2 \sum_{j=1}^m (y_j - \beta_0 - \beta_1 x_{1j} - \beta_2 x_{2j})(-x_{2j}) = 0 \end{cases} \rightarrow \begin{bmatrix} \sum x_{1j}^2 & \sum x_{1j} x_{2j} & \sum x_{1j} \\ \sum x_{1j} x_{2j} & \sum x_{2j}^2 & \sum x_{2j} \\ \sum x_{1j} & \sum x_{2j} & \sum 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_0 \end{bmatrix} = \begin{bmatrix} \sum x_{1j} y_j \\ \sum x_{2j} y_j \\ \sum y_j \end{bmatrix}_{j=1 \dots m}$$

- Example table:

"Effect on sales of advertising budget for TV, radio, newspapers?"

$$Y = \beta_0 + \beta_1 \cdot \overbrace{x_1}^{\text{TV}} + \beta_2 \cdot \overbrace{x_2}^{\text{radio}} + \beta_3 \cdot \overbrace{x_3}^{\text{newspapers}}$$

	Coefficient	Standard Error	t-statistic	p-value
Intercept	β_0	2.939	9.42	<0.0001
TV	β_1	0.046	32.81	<0.0001
radio	β_2	0.189	21.89	<0.0001
newspaper	β_3	-0.001	-0.18	0.8599

CORRELATIONS

Intercept	TV	radio	newsp.
β_0	1	0.0567	0.0567
β_1	1	0.3541	0.5762
β_2		1	0.2283
β_3			1

	Value
Residual Standard Error, RSS	1.69
R^2	0.897
F-Statistic	570

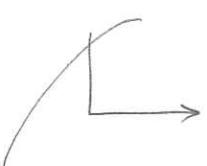
The best scenario is the one where there are no correlations between variables:

BALANCED DESIGN:

- Each β_i can be estimated and tested separately
- "Unit change in x_i causes increase of β_i on y " is TRUE

$$F = \frac{(SST - RSS)/n}{RSS/(m-n-1)} \sim F_{n, (m-n-1)}$$

$$= \frac{R^2}{(1-R^2)} \cdot \frac{n}{m-n-1}$$



$$\begin{cases} H_0: \beta_i = 0, \forall i = 0, \dots, n \\ H_1: \text{at least one } \beta_i \neq 0 \end{cases}$$

Rejection region: $F_{\alpha, n, (m-n-1)}$

Hypothesis testing & confidence

Intervals for β_i can be computed analogously to the 2D case (see example table)

- Do all the variables x_i explain y , or is only a subset useful?

→ all possible subsets (2^n) are computed for their least squares fit and the optimum subset is chosen.

- Qualitative predictors (in contrast to quantitative) $\xrightarrow{\text{€IR}}$

↳ Categorical: gender: 1 (female), 0 (male) (2 levels)
 ethnicity: asian, caucasian, afro-american (3 levels)

- Binary variables are introduced (AKA dummy variables)

Two levels: gender: $x_1 = \{ \text{male} = 0, \text{female} = 1 \}$

Three+ levels: ethnicity: $x_2 | x_3 \rightarrow$ always one dummy variable less than # levels!

x_2	x_3	
1	0	level 1: asian
0	1	level 2: caucasian
0	0	level 3: afro-american: baseline

example: ↴

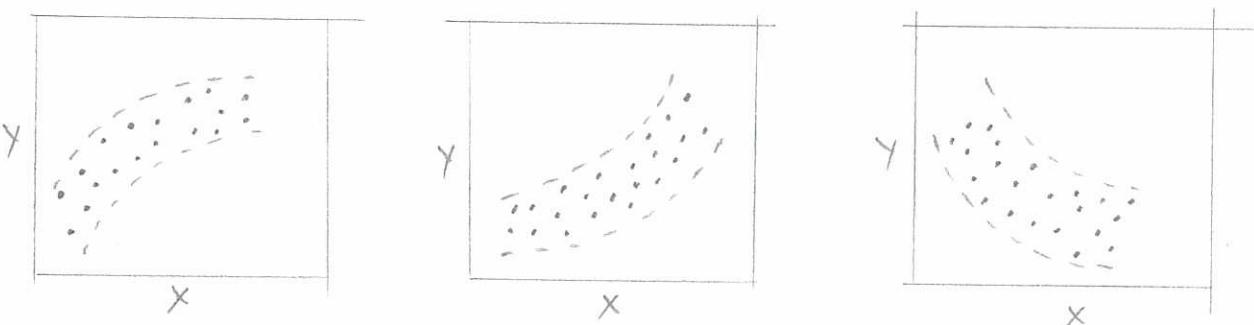
$$Y_j = \beta_0 + \beta_1 + \beta_3 : \text{if person: female, caucasian}$$

TRANSFORMATIONS: What to do if...

① X and Y are not linearly related: transform X

REMEMBER

$$\log x < \sqrt{x} < x < x^2 < e^x$$



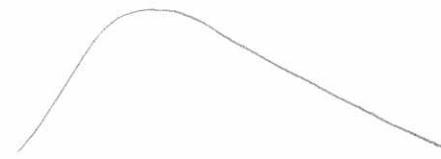
$$X \rightarrow \begin{cases} x' = \sqrt{x} \\ x' = x^a, 0 < a < 1 \\ x' = \log x \end{cases}$$

$$X \rightarrow \begin{cases} x' = x^2 \\ x' = x^a, a > 1 \\ x' = e^x \end{cases}$$

$$X \rightarrow \begin{cases} x' = 1/x \\ x' = x^a, a < 0 \\ x' = e^{-x} \end{cases}$$

② Y doesn't satisfy normal distribution of errors / same variances: transform Y

use $\lambda < 1$

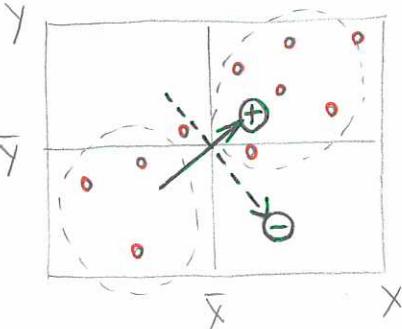


use $\lambda > 1$



$\lambda = 3$	$y' = y^3$
2	$y' = y^2$
1	$y' = y$
$1/2$	$y' = \sqrt{y}$
$1/3$	$y' = \sqrt[3]{y}$
0	$y' = \log(y)$
$-1/2$	$y' = -1/\sqrt{y}$
-1	$y' = 1/y$
-2	$y' = 1/y^2$

CORRELATION



- Covariance: value that measures the tendency of the data to go from one quadrant to the diagonally opposite

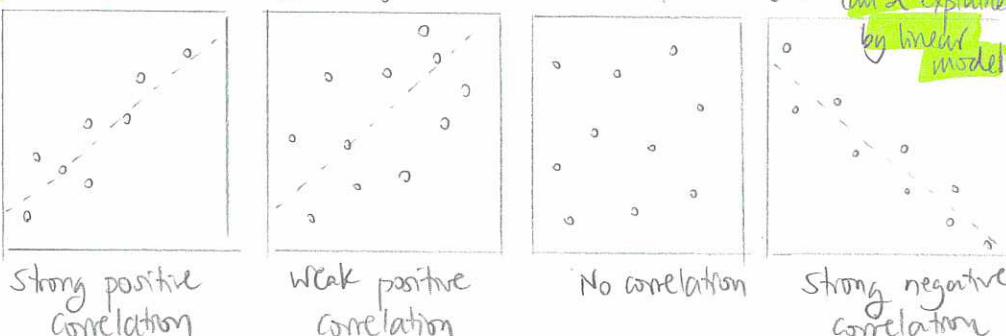
$$\text{Cov}(x, y) = s_{xy} = \frac{1}{n} \sum (x_i - \bar{x})(y_i - \bar{y})$$

↓
since the size of the value doesn't say how big the tendency is, because it depends on the scale of the data, it is normalized with the std. deviation: PEARSON'S COEFFICIENT

- Pearson's coefficient:

$$r = \frac{\text{cov}(x, y)}{s_x \cdot s_y} = \frac{\frac{1}{n} \sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n} \sum (x_i - \bar{x})^2} \cdot \sqrt{\frac{1}{n} \sum (y_i - \bar{y})^2}} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}}, \quad -1 \leq r \leq 1$$

→ $R^2 = 100 \cdot r^2$: coefficient of determination (percentage); variation that can be explained by linear model



	+	+	+	+
	-1	0	no correlation	1
-1	negative correlation	x ↑ y ↓	positive correlation	x ↑ y ↑
0				
1				

• Pearson's coefficient measures linearity: $|r| \approx 1 \Rightarrow$ points very close to same straight line
 $|r| \approx 0 \Rightarrow$ no linear connection between x & y .

• But it doesn't say how strong the slope of the straight line is!!!

• Hypothesis testing can be done to check the probability that the relationship (correlation) of the sample corresponds to the one of the population: T-test:

$$H_0: \rho = 0 \quad (\rho: \text{correlation of population}) \quad H_1: \rho \neq 0$$

$$t = \frac{r - \rho^0}{\sqrt{\frac{1 - r^2}{n-2}}}, \quad df = n-2 \longrightarrow$$

compare critical to
or compute signifi-
cance with
T-test

• Relationship to regression (regression tells about slope!)

$$Y - \bar{Y} = \frac{\text{cov}(x, y)}{s_x^2} (x - \bar{x}) \Leftrightarrow \frac{Y - \bar{Y}}{s_y} = r \cdot \frac{x - \bar{x}}{s_x} \Leftrightarrow Y = \left(r \cdot \frac{s_y}{s_x} \right) x + \left(\bar{Y} - r \cdot \frac{s_y}{s_x} \cdot \bar{x} \right)$$

slope!

intercept

- Important limitations of any correlation coefficient (VERY IMPORTANT)

- Correlation doesn't imply causality, and small correlation doesn't imply X & Y are not related
- Outliers might yield misleading correlations → ALWAYS check scatter plots!!!
- Correlation doesn't give info about the effect size of X on Y (slope)!
- Choose always the correct correlation coefficient: linear? ascending?

— Other correlation coefficients that do not measure linearity and that are used with ordinal data (ranks): *necessarily*

- Kendall's coefficient, τ :

$(x_i, y_i), i=1, \dots, N$

ordered
(ordinal)
data

	$(x_i, y_i)_1$	\dots	$(x_i, y_i)_i$	\dots	$(x_i, y_i)_n$
$(x_i, y_i)_1$					
\vdots					
$(x_i, y_i)_n$					

→ for each $(x_i, y_i)_i$ & $(x_i, y_i)_j$ pair $i \neq j$, we say the pair is:

- Concordant, if both variables increase or decrease $x_i < x_j \& y_i < y_j$ OR $x_i > x_j \& y_i > y_j$
- Discordant, if one variable decreases and the other increases $x_i > x_j \& y_i < y_j$ OR $x_i < x_j \& y_i > y_j$
- Tied, if one or both variables stay constant

$$\left[\tau = \frac{C - D}{N} = \frac{C - D}{\frac{n(n-1)}{2}} \right], \text{ where}$$

$$\left. \begin{array}{l} C = \# \text{ concordant pairs} \\ D = \# \text{ discordant pairs} \\ T = \# \text{ tied pairs} \end{array} \right\}$$

$$C + D + T = N = n(n+1) \frac{1}{2}$$

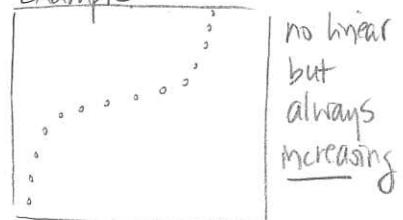
→ τ gives a measure of the degree to which a relationship is always increasing ($\tau > 0$) or always decreasing ($\tau < 0$)
- no linearity is measured

→ Usually a corrected value is used:

$$\tau_b = \frac{C - D}{\sqrt{\left(\frac{n(n-1)}{2} - tx \right) \left(\frac{n(n-1)}{2} - ty \right)}}$$

tx and ty = tied x and y values, respectively

Example



$r \approx 0, \tau = 1 !!!$

- Spearman's rank correlation coefficient, r_s :

- It is derived from Pearson's coefficient
- Used with ranked data:

IQ (x_i)	86 97 99 100 101 103 106 110 112 113
rank x_i	1 2 3 4 5 6 7 8 9 10
TV.h/week (y_i)	0 20 28 27 50 29 7 17 6 12
rank y_i	1 6 8 7 10 9 3 5 2 4
$d_i = x_i - y_i$	0 -4 -5 -3 -5 -3 4 3 7 6

$n = 10$

$$\left[r_s = 1 - \frac{6 \sum d_i^2}{n(n^2-1)} \right]$$

$$= 1 - \frac{6 \cdot 194}{10(10^2-1)} = -0,175$$

Standard deviation:

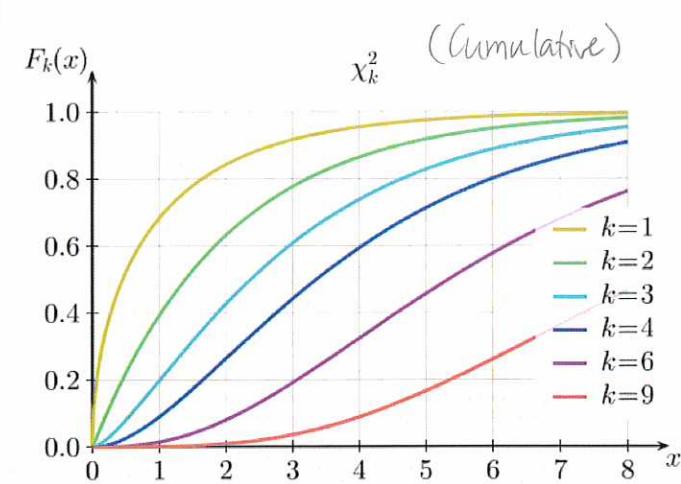
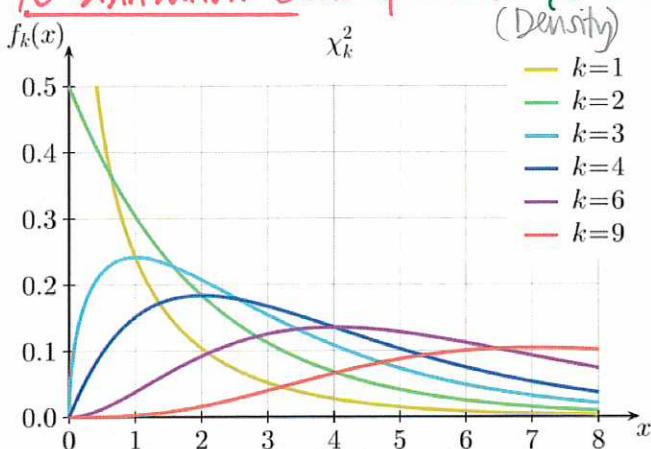
$$s_{r_s} = \sqrt{\frac{1-r_s^2}{n-2}}$$

- Hypothesis testing:

$$\left. \begin{array}{l} H_0: r_s = 0 \\ H_1: r_s \neq 0 \end{array} \right\} \rightarrow t = \frac{r_s - \rho_s}{\sqrt{\frac{1-r_s^2}{n-2}}}, df = n-2 \rightarrow T\text{-test}$$

2 Some Continuous Probability Distributions

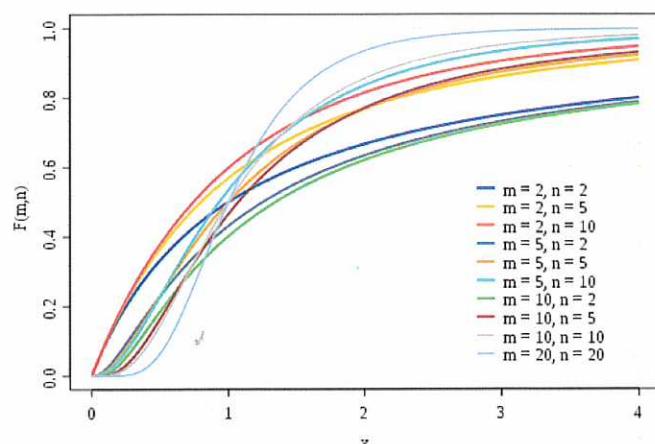
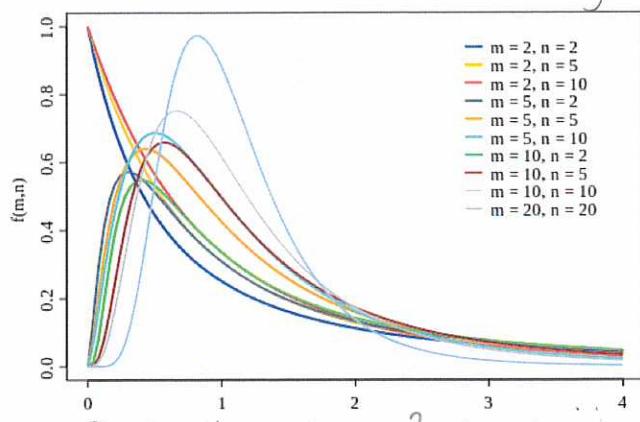
χ^2 Distribution (Chi squared)



- Uses:
 - goodness of fit of an observed distribution to a theoretical one
 - confidence interval estimation for a population standard deviation of a normal distribution from a sample std. deviation

F-Distribution (Fisher)

$F(df_1, df_2)$
(Density)

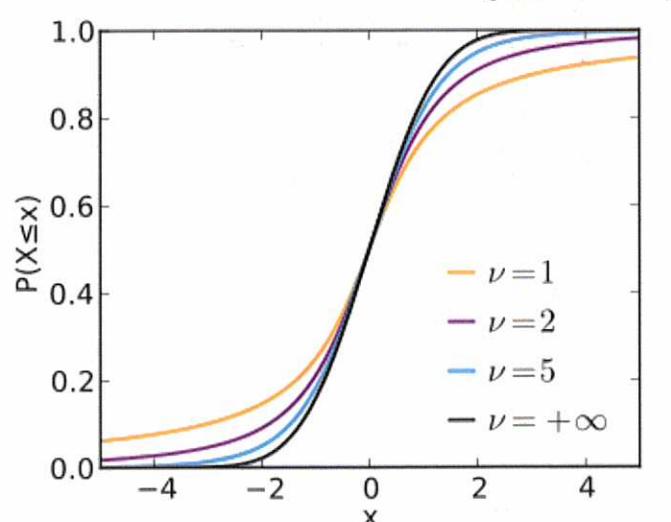
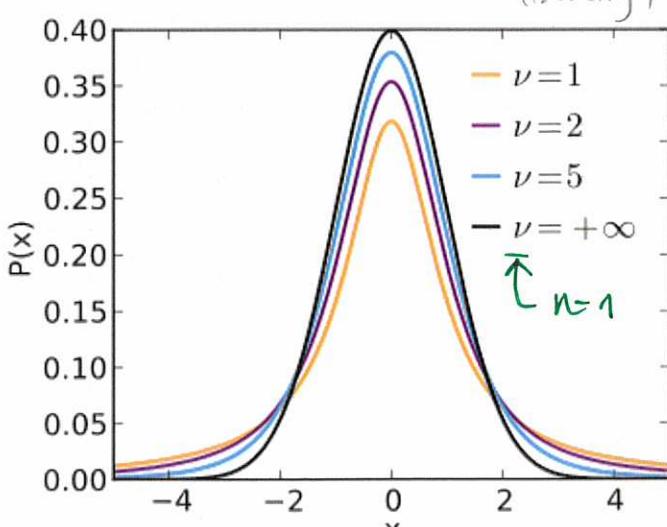


— Quotient of two χ^2 distributions

— Uses: ANOVAs

Student's t distribution

$t(n)$
(Density)

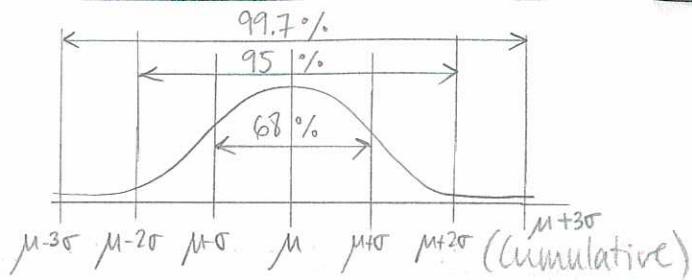
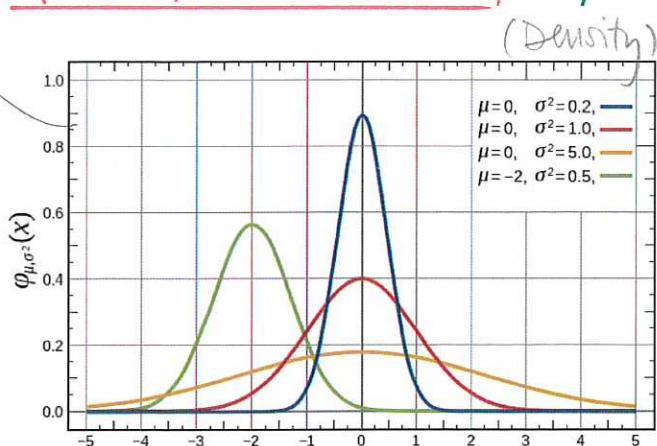


- Similar to the standard/normal distribution, but used to take into account the size of the sample (n) \Rightarrow the bigger n the more t resembles the normal distribution: $t(\infty) = N(0,1)$.
- Used in the Student t tests.

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

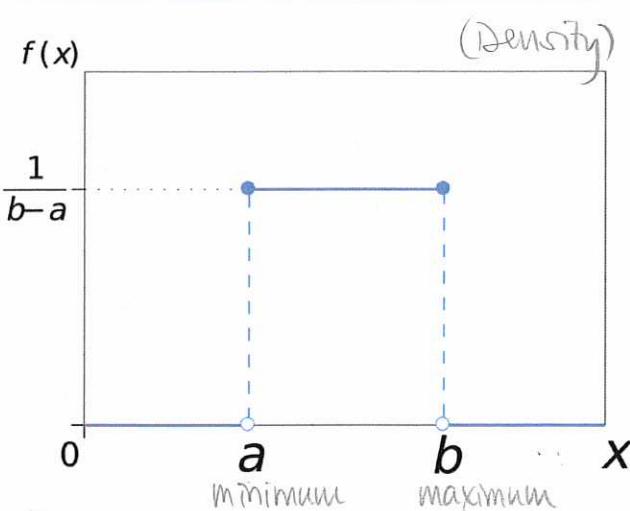
$\rightarrow (N(0, 1))$

Standard/Normal distribution $N(\mu, \sigma^2)$

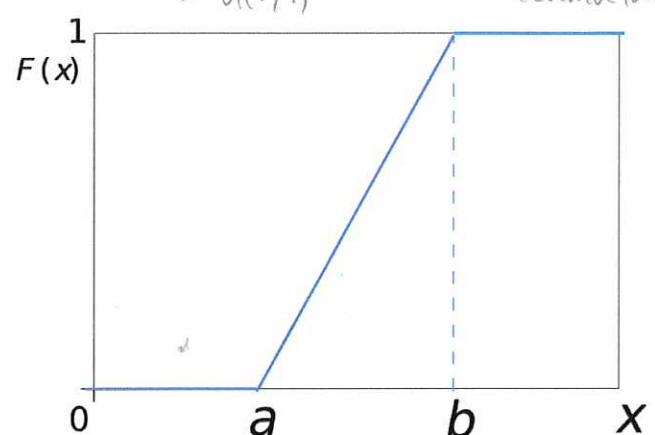


- Central limit theorem states that averages of random variables independently drawn from independent distributions are normally distributed \Rightarrow many physical quantities
- Other bell-like distributions: Cauchy, Student, logistic, ... have distributions closer to the normal!!

Continuous Uniform Distribution $U(a, b)$



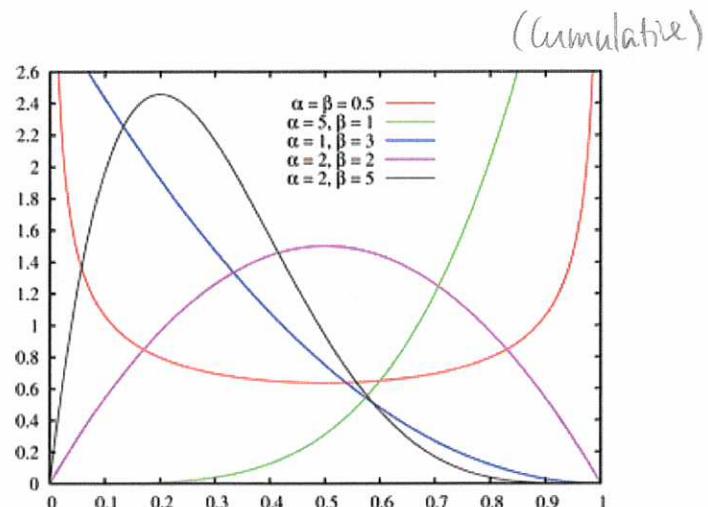
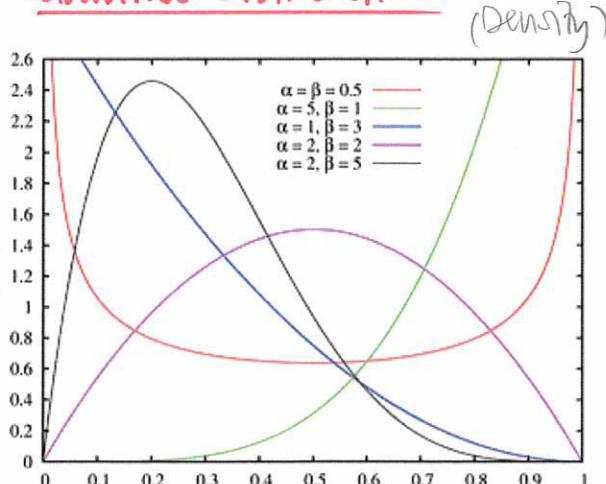
Standard $\rightarrow U(0, 1)$



- Uses: random number generation: $\text{rand}(a, b) = a + (b - a) U(0, 1)$, some errors are uniformly distributed

Beta Gamma Distribution

Suitable for random behavior of proportions

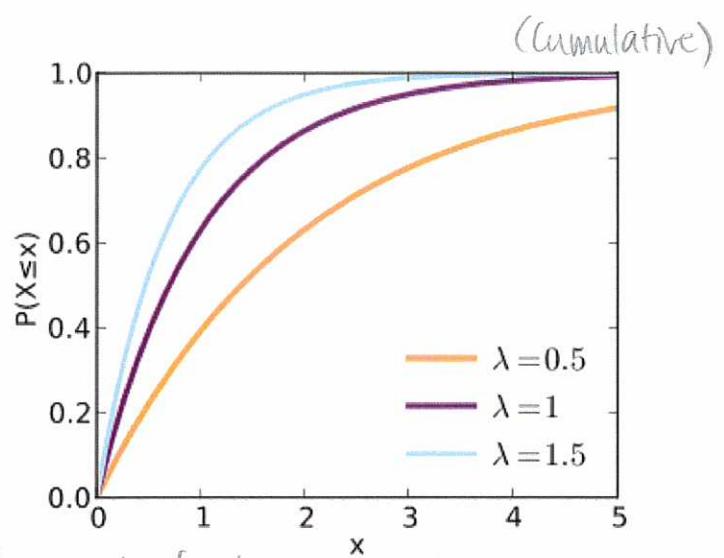
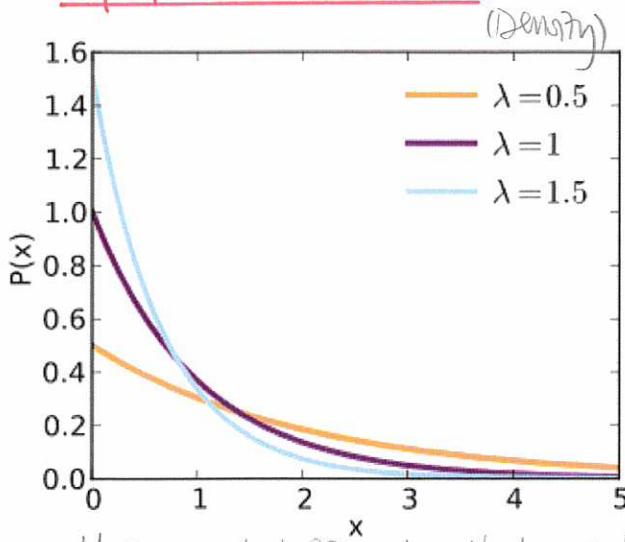


\hookrightarrow

\hookleftarrow : limited to interval $[0, 1]$

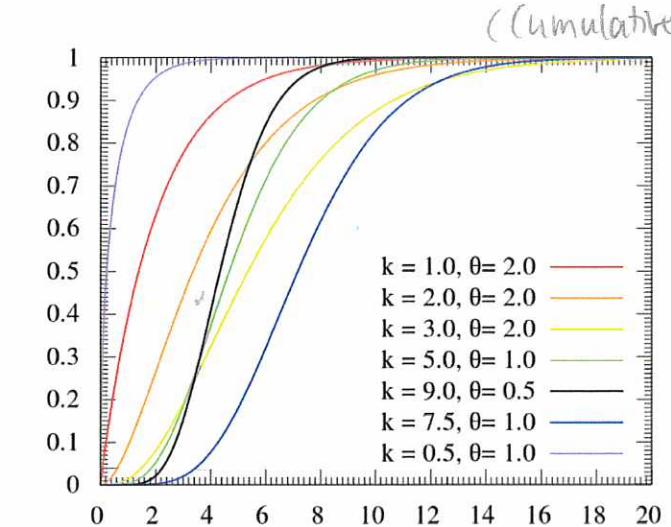
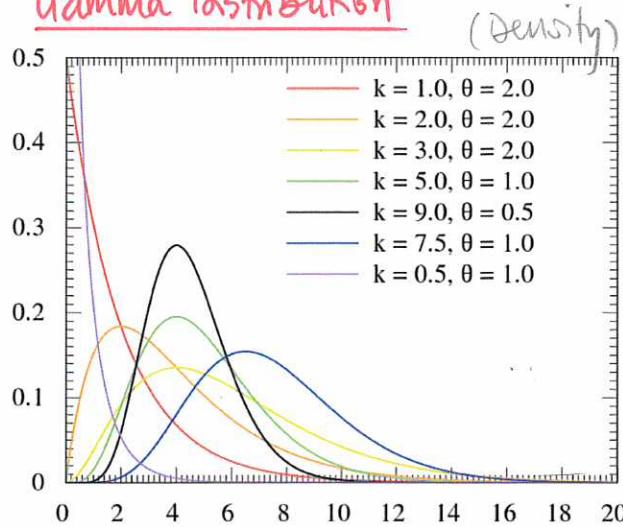
- Applied to model the behavior of random variables limited to intervals; examples: allele frequencies in population genetics, time allocation in project management, variability of soil properties, probabilities in HIV transmission.

Exponential Distribution



- Use: probability distributions for events that occur continuously and independently at a constant average rate: the time between each incidence has an exponential distribution; example: calls in call centers, earthquakes, ...

Gamma distribution



- Exponential & χ^2 are special cases of the Gamma distribution
- Use: waiting times, time until death, ...

BINOMIAL DISTRIBUTION (Discrete)

- Discrete probability distribution of the number of successes in a sequence of n independent yes/no experiments, each of which yields success with probability p

If a random number follows the binomial distribution: $X \sim B(n, p)$

Probability of getting exactly k successes given by PMF:

$$f(k|n,p) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$= \frac{n!}{k!(n-k)!} \cdot p^k (1-p)^{n-k}$$

Example:

Suppose a biased coin comes up heads with probability 0.3 when launched. What is the probability of achieving $\frac{0, 1, 2, 3, 4, 5}{K}$ or 6 heads after $\frac{6}{n}$ launches?

$$\Pr(0 \text{ heads}) = f(k=0 | n=6, p=0.3) = \Pr(X=0) = \binom{6}{0} (0.3)^0 (1-0.3)^{6-0} = 0.1176$$

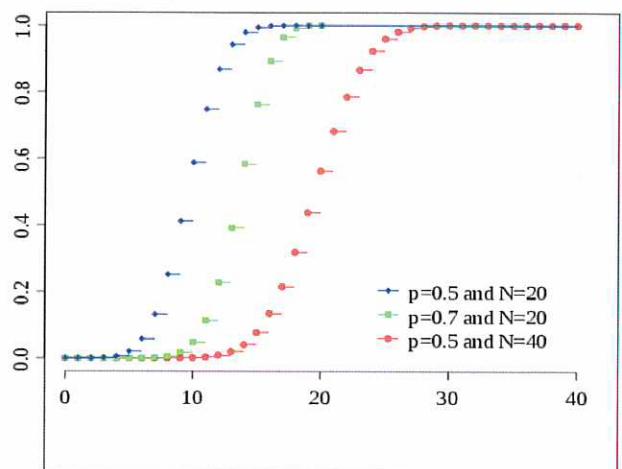
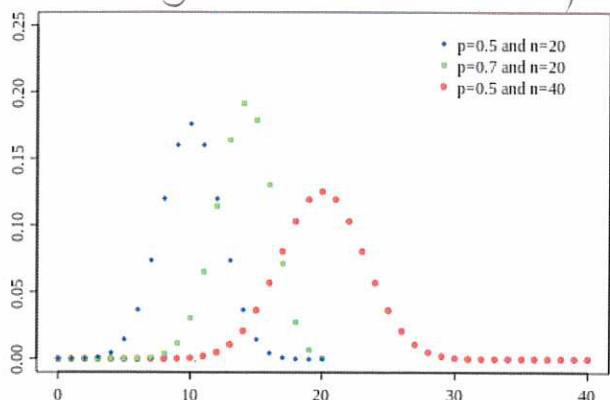
$$\Pr(1 \text{ heads}) = f(k=1 | n=6, p=0.3) = \Pr(X=1) = \binom{6}{1} (0.3)^1 (1-0.3)^{6-1} = 0.3025$$

$$\Pr(6 \text{ heads}) = f(k=6 | n=6, p=0.3) = \Pr(X=6) = \binom{6}{6} (0.3)^6 (1-0.3)^{6-6} = 0.0007$$

Normal Approximation: if $n \uparrow \rightarrow B(n, p)$ approximates to $N(np, np(1-p))$

Used in the BINOMIAL TEST

Probability Mass Function (PMF)



GOODNESS OF FIT TESTS

- The χ^2 chi-squared test is a technique for testing whether observed data are representative of a particular distribution:
 - Discrete probability models [natural numbers used]: binomial (dice), Poisson (malfunctions)
 - Continuous probability models [regions of real numbers used]: exponential, normal...
- Procedure:
 - Make table with
 - Observed frequencies for each value/interval: O_i
 - Expected frequencies (according to our model): E_i
- Example: choose 100 numbers from 0 to 9 randomly. Distribution should be discrete uniform - is it?

Number	O_i	E_i	$\frac{(O_i - E_i)^2}{E_i}$
0	11	10	0.1
1	12	10	0.4
2	8	10	0.4
3	14	10	1.6
4	7	10	0.9
5	9	10	0.1
6	9	10	0.1
7	8	10	0.4
8	14	10	1.6
9	8	10	0.4

$$\sum_{i=1}^{10} \frac{(O_i - E_i)^2}{E_i}$$

$\hookrightarrow \sum E_i = 100$: constraint

$$\sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{df=n-1}$$

77

degrees of freedom = number of classes/groups/values

number of constraints: Total frequency amount is usually fixed

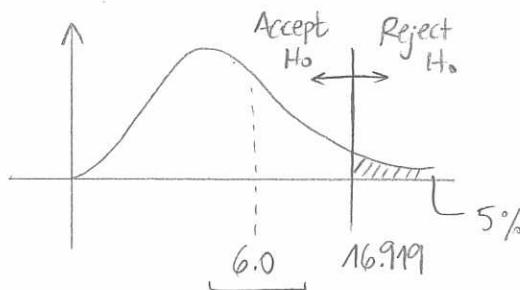
$$\chi^2 = \sum_{i=1}^{10} \frac{(O_i - E_i)^2}{E_i} = 6.0 \rightarrow \{ df = 10-1 = 9 \\ P\text{-value?}$$

H_0 : numbers are random

H_1 : numbers are not random

$\alpha = 0.05$

$$df = 10-1 = 9 \rightarrow \chi^2(9, 0.05) < 16.919$$



\hookrightarrow Numbers are random!
(at $\alpha = 0.05$ level)

- High values of χ^2 mean our values

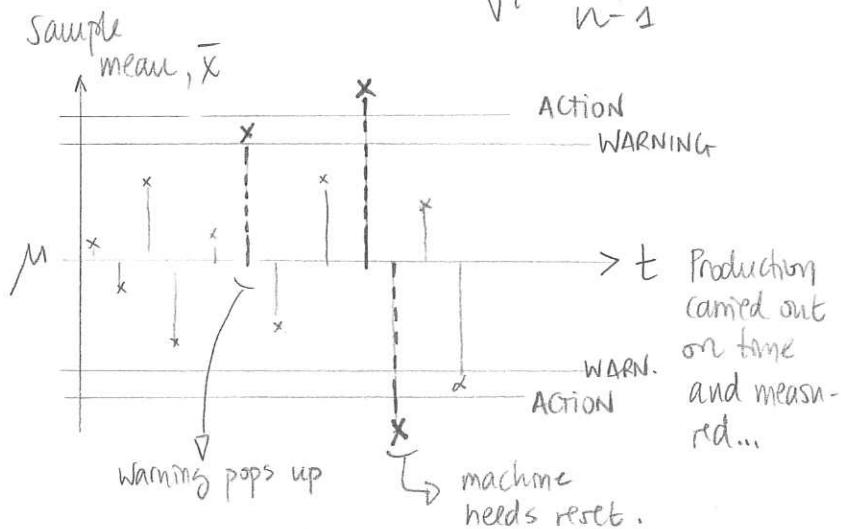
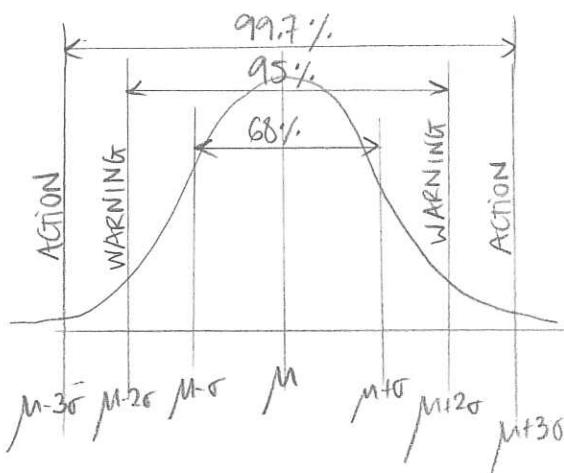
do not match the expected model at the level of $\alpha \Rightarrow$ we reject H_0

$\hookrightarrow \alpha$ gives the critical χ^2 value, under which our $\sum_i \frac{(O_i - E_i)^2}{E_i}$ must be!

STATISTICAL PROCESS CONTROL & ACCEPTANCE SAMPLING

- In production, a target value is desired to achieve, but we assume there is some kind of short term inherent variability: μ, σ .
- Confidence values are defined as
 - warning limits: ~95% of sample means lies within
 - action limits: ~99.8% of sample means lies within

- If no σ, μ available, they are computed out of big samples: $\hat{\mu} = \frac{\sum X_i}{n}$, $\hat{\sigma} = \sqrt{\frac{\sum (X_i - \hat{\mu})^2}{n-1}}$
- In the case, where the model follows a normal distribution:



$$\left\{ \begin{array}{l} \text{WARNING LIMIT: } \mu \pm \frac{2\sigma}{\sqrt{n}} \\ \text{ACTION LIMIT: } \mu \pm \frac{3\sigma}{\sqrt{n}} \end{array} \right.$$

sample size tested at the moment; $\mu \pm \sigma$ correspond to the population - if not available, compute $\hat{\mu}$ & $\hat{\sigma}$.

- Acceptance sampling: unlike statistical process control, where the purpose is to check production as it proceeds, acceptance sampling is applied to large batches (lots) of goods which have been produced
 - PROBLEM DEFINITION: decide if batch is acceptable based on testing relatively small samples
 - Know parameters:
 - Size of batch & sample (n) \oplus number of detected non-conforming from batch sample (r)
 - Operating characteristic: graph of probability of accepting batch against proportion of non-conforming (defective) in the batch
 - Binomial distribution used.
 - Example: Large batches, where samples of $n=30$ are tested according to
 - Plan A: accept batch if no non-conforming in sample ($n=30, r=0$)
 - Plan B: accept batch if no more than 1 non-conf. in sample ($n=30, r=1$)
 - Probability of accepting a batch with 2% non-conforming?
 - A: $(1-0.02)^{30} = 0.545$] !!!
 - B: $(1-0.02)^{30} + 30 \cdot 0.02 \cdot (1-0.02)^{29} = 0.978$]

$P(\text{accept})$

