Experimental Data Analysis

in ©MATLAB

Lecture "Data interpretation":

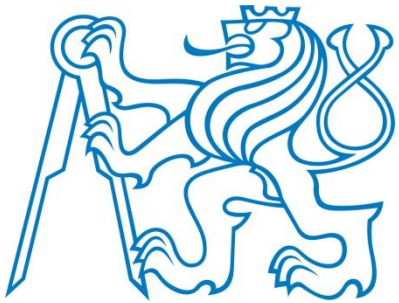
I. Introduction to research projects

II. Basic statistical analysis

III. Example of scientific project in biomedicine

IV. How to write the results of the project

Jan Rusz
Czech Technical University in Prague

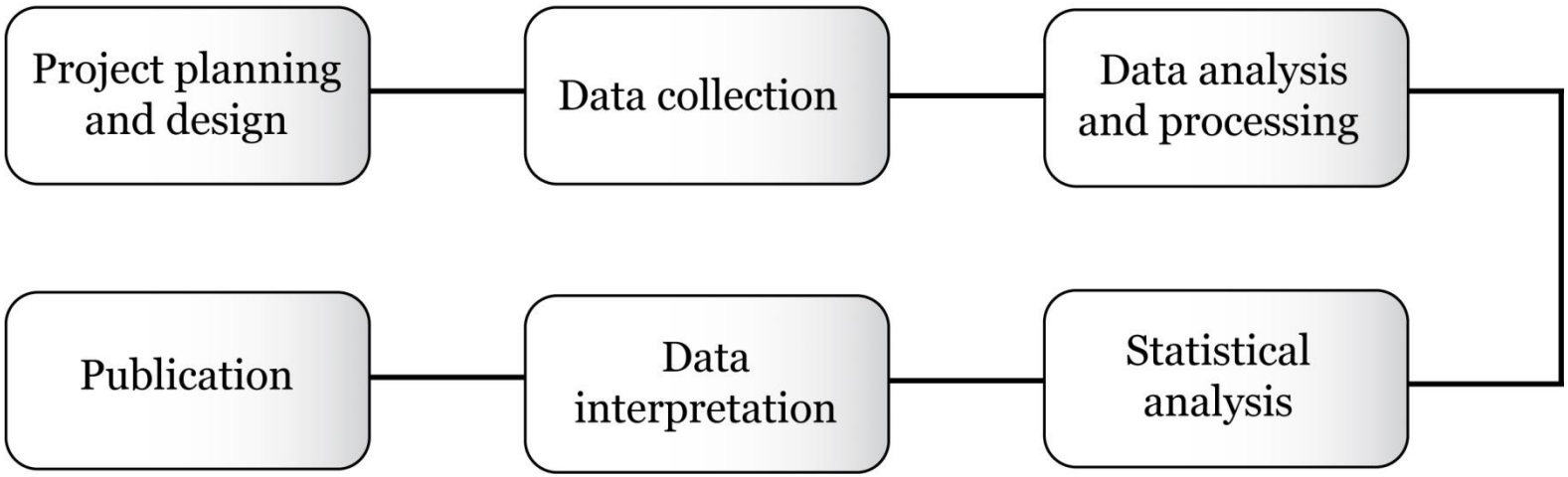


I. Introduction

Why the statistical analysis of data is important?

- Essential for solving different research and developmental projects
- Statistics represents language for quantitative interpretation of data
- Typical questions in biomedicine are of statistical nature:
  - „Can I use gait velocity to define type of disease?“
  - „Is this new treatment better than previous one?“
- Current development in biomedicine is driven by publications of new observations, which are dependent on data interpretation
- On the basis of these observation, we can advance in development of novel technologies, medical procedures, etc.
- Current computational power allows us quantitative data processing using various statistical methods
- Currently, it is impossible to publish original paper without statistical methods (with an exception of “case study”)

General picture of research project



Project planning and design

- We cannot study the entire population of interest
- The aim of the research is to generalize knowledge obtained from chosen population to overall population
- We have to define aim, purpose and type of the study
  - *aim 1*: to find persons in high risk of development of neurodegenerative disease (for example Parkinson's disease)
  - *purpose 1*: to perform preventive treatment
  - *type 1*: „cross-sectional" comparisons
  - *aim 2*: to find side-effects of treatment
  - *purpose 2*: therapy optimization
  - *type 2*: "longitudinal" observation
- We have to know well state of the knowledge („state-of-the-art)"
  => risk of incremental research
- Before start of the project, it is necessary to formulate hypotheses that should be critically assessed
  - *hypothesis 1*: speech disorder will be early sign of neurological disease
  - *hypothesis 2*: high doses of antiparkinsonian medication will cause development of stuttering
- Major mistake is to formulate hypotheses after data collection

Data collection

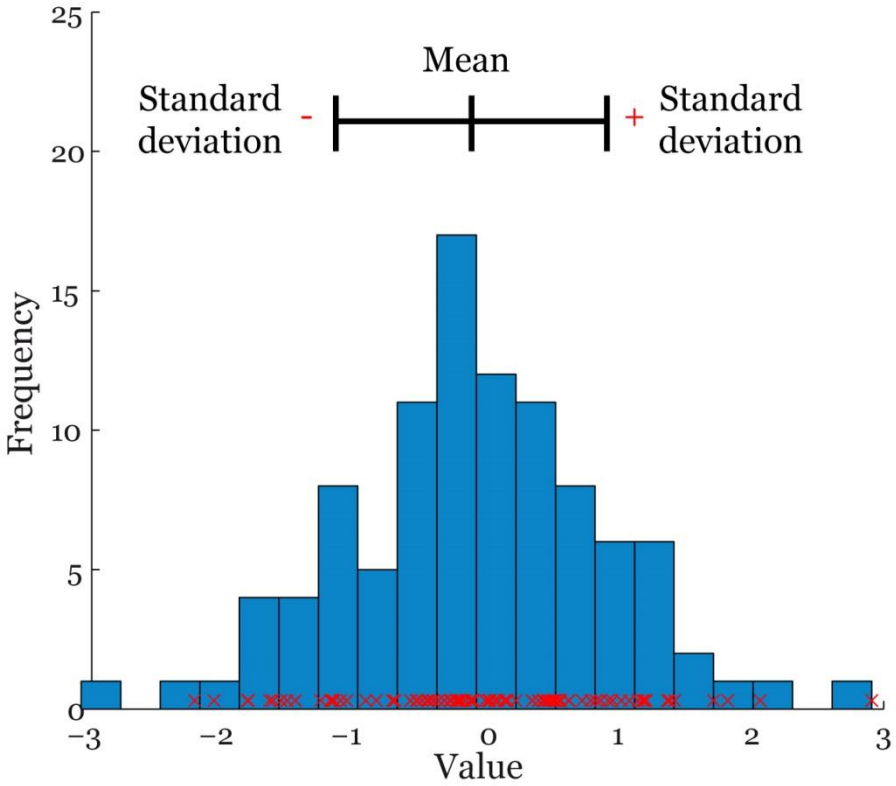
- We define population of interest
- We define „inclusion/exclusion“ criteria
- It is beneficial to plan “pilot study” using small sample of studied population in order to verify quality of data and exclude:
    - formal errors
    - random errors (inattention)
    - systematic errors (bad hypotheses, inaccurate set-up of measurement device)
- Possibility to perform power analysis and estimate of required sample size to find statistical significance

Data analysis and processing

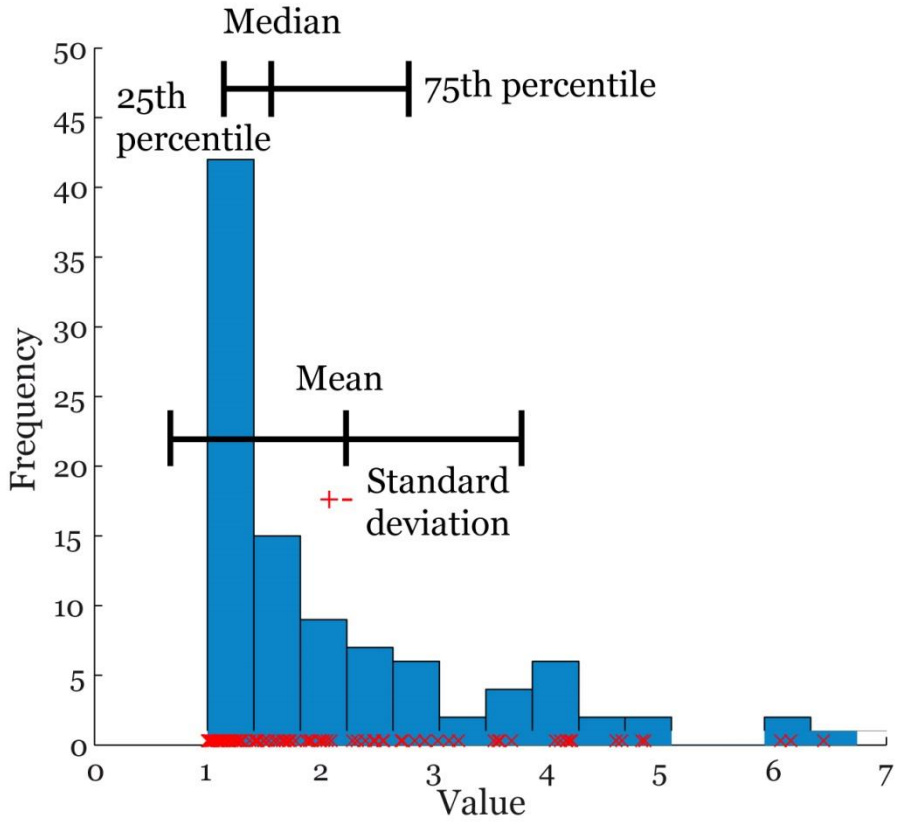
- Signal preprocessing and filtration
- Signal segmentation using DSP methods
  - Signal analysis in time domain (thresholding, signal envelope, zero crossing, etc.)
  - Signal analysis in frequency domain (spectrum, autocorrelation)
- Signal segmentation using statistical methods (mean, median, standard deviation, clustering, etc.)
- Development of features to evaluate objects of interests (moments, segment length, spectrum kurtosis, etc.)

II. Basic statistical analysis

Histogram of normal data



Histogram of non-normal data



Mean: $\mu(x) = \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$

Standard deviation: $\sigma(x) = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}$

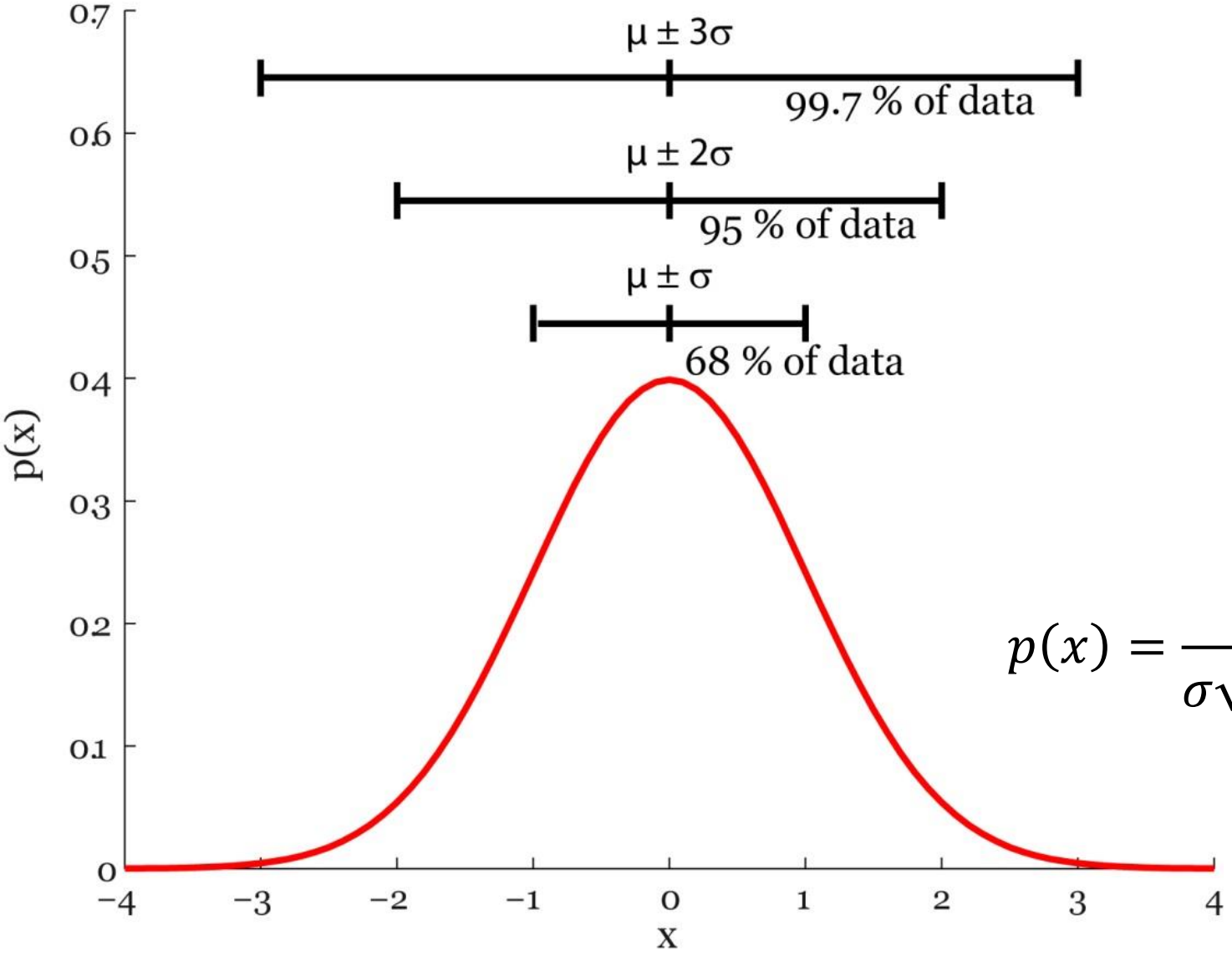
Median: is the value separating the higher half of a data sample, from the lower half.

Percentile:

$n = \left\lceil \frac{P}{100} \times N \right\rceil$

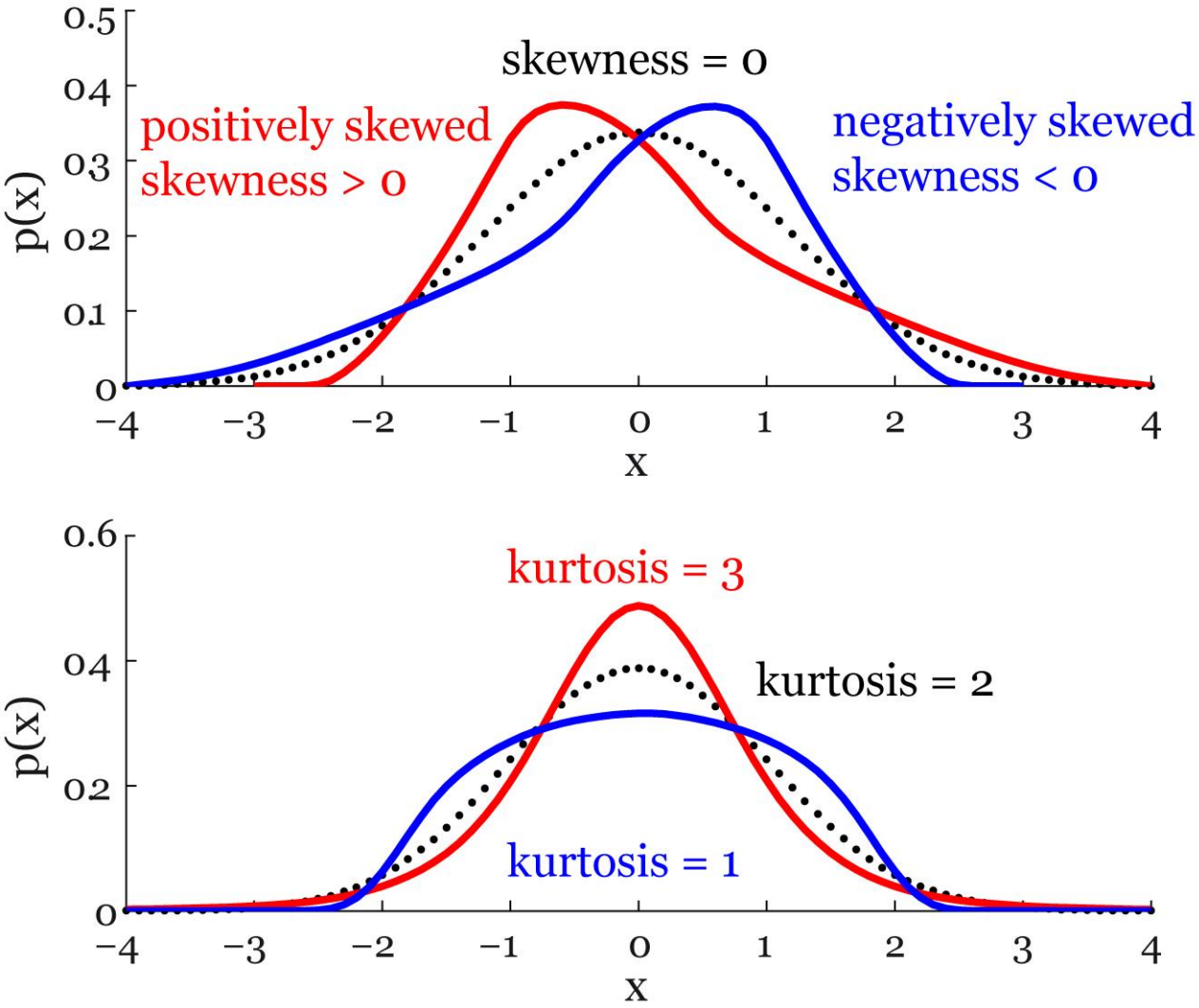
P-th percentile of the N ordered values (sorted from least to greatest)

Gaussian probability distribution

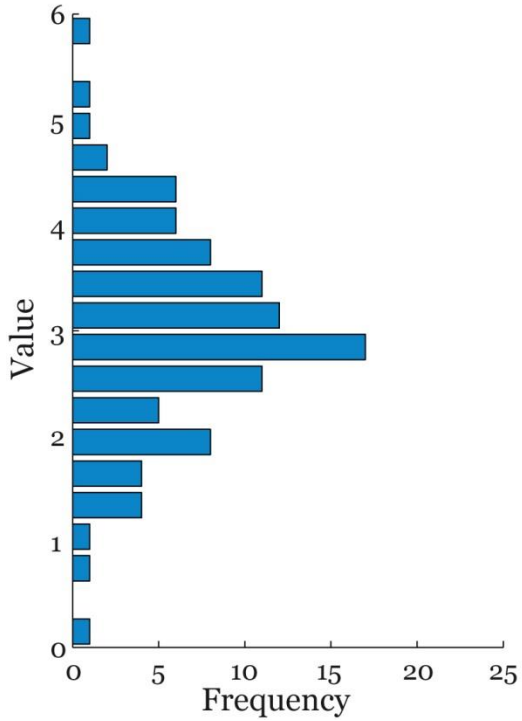


$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

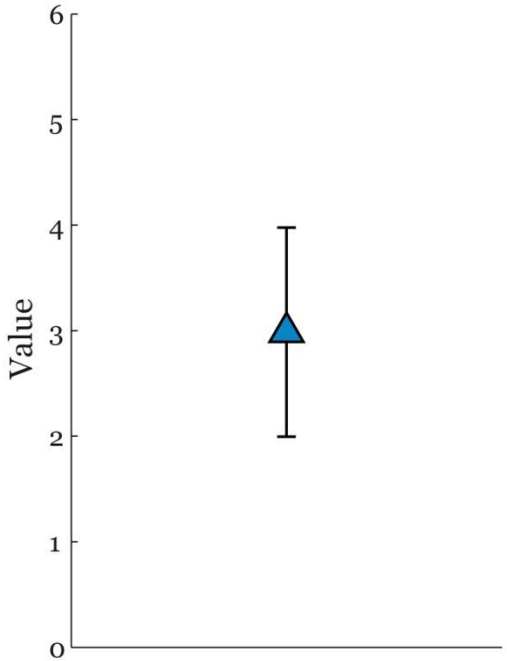
Skewness & Kurtosis



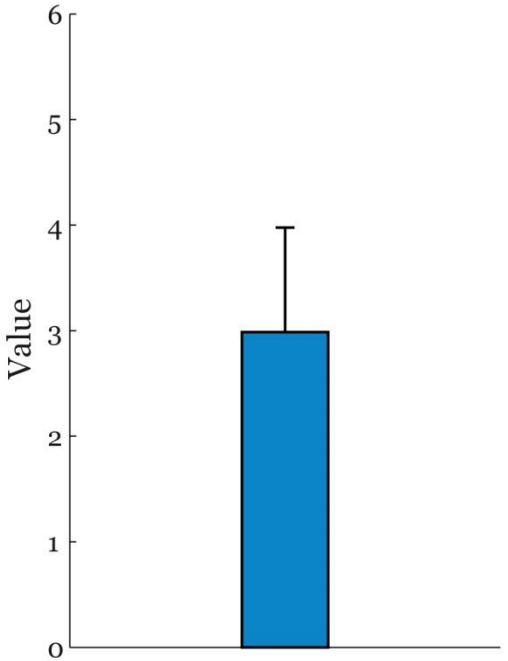
Histogram of normal data

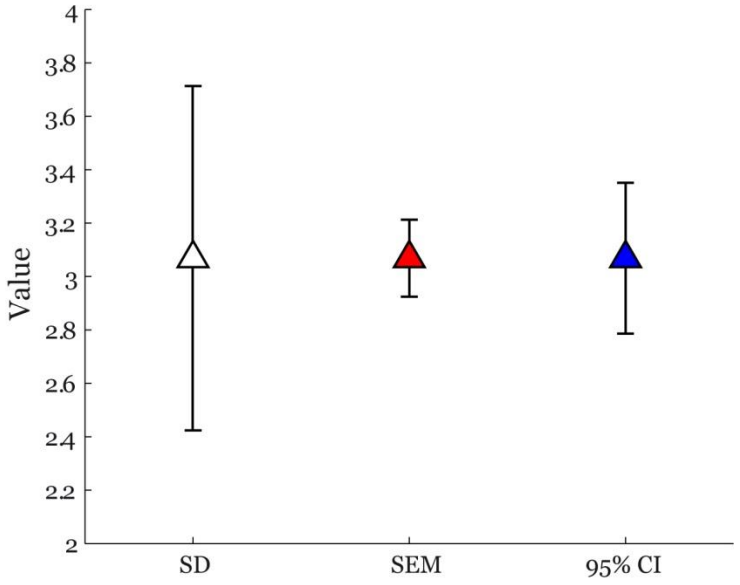


Errorbar



Bar + errorbar





Standard Deviation (SD):

$$\sigma(x) = \sqrt{\frac{\sum_{i=1}^{n}(x_{i-}\bar{x})^2}{n-1}}$$

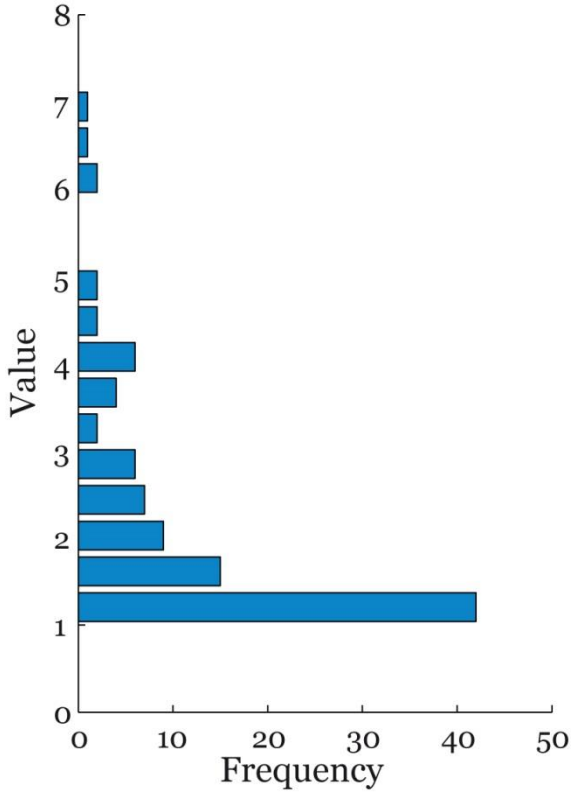
Standard Error of the Mean (SEM):

$$SEM = \frac{\sigma(x)}{\sqrt{n}}$$

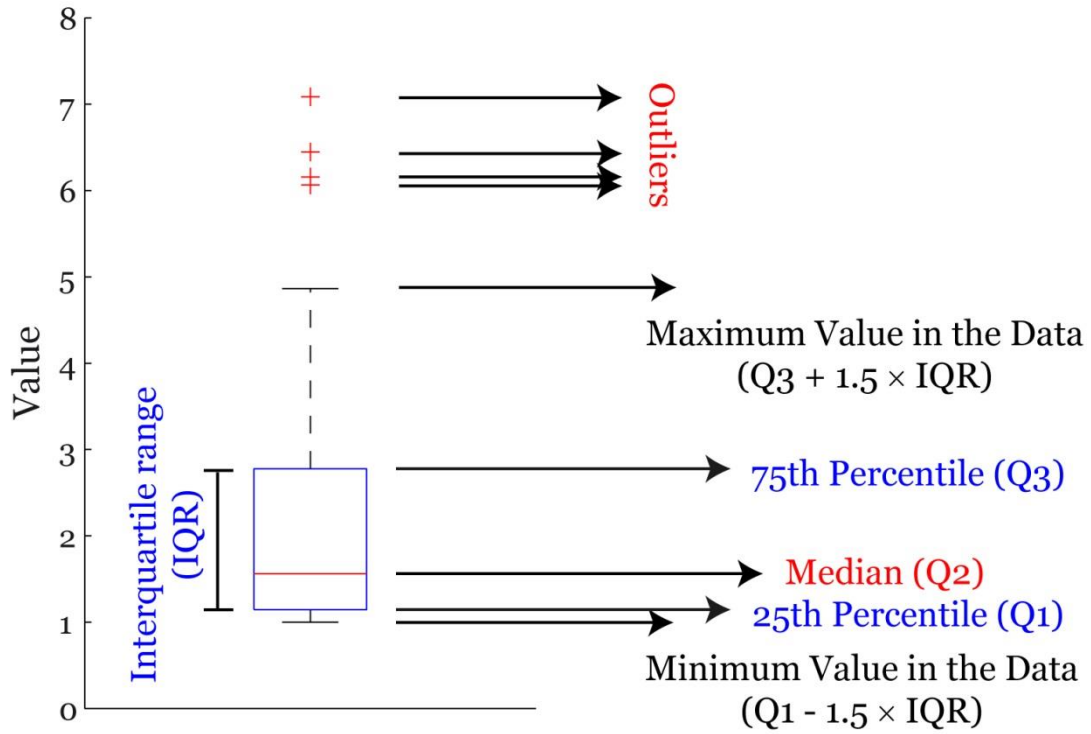
95% Confidence Interval (95% CI):

$$95\%\ CI = SEM \times 1.96$$

Histogram of non-normal data

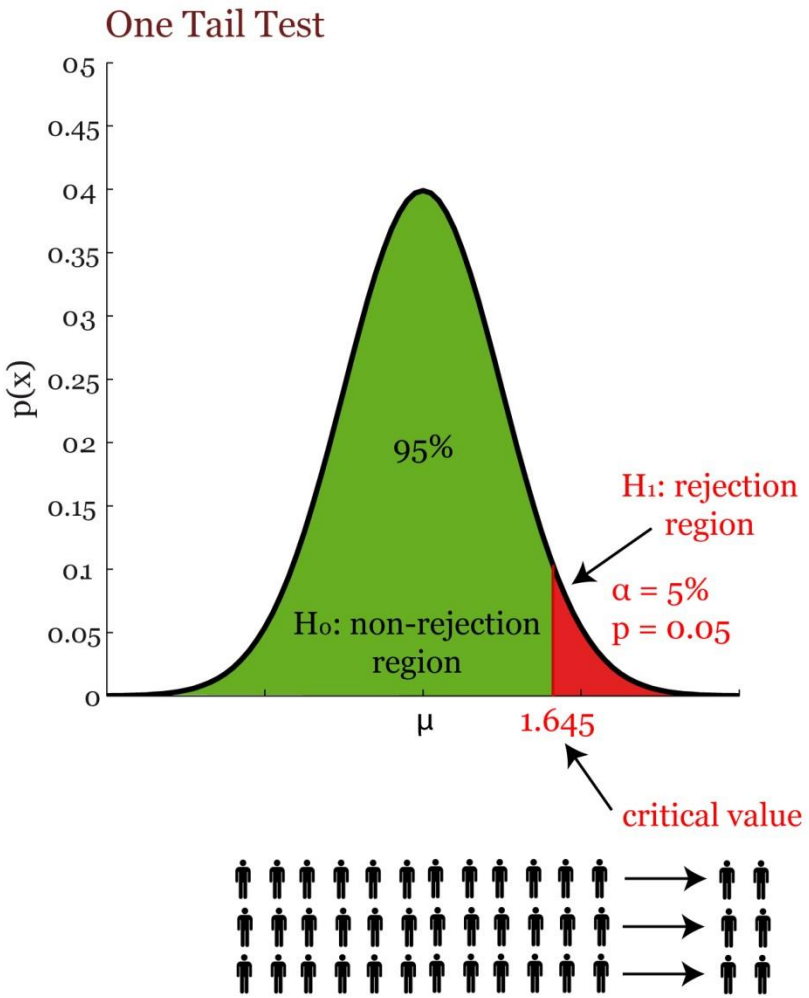


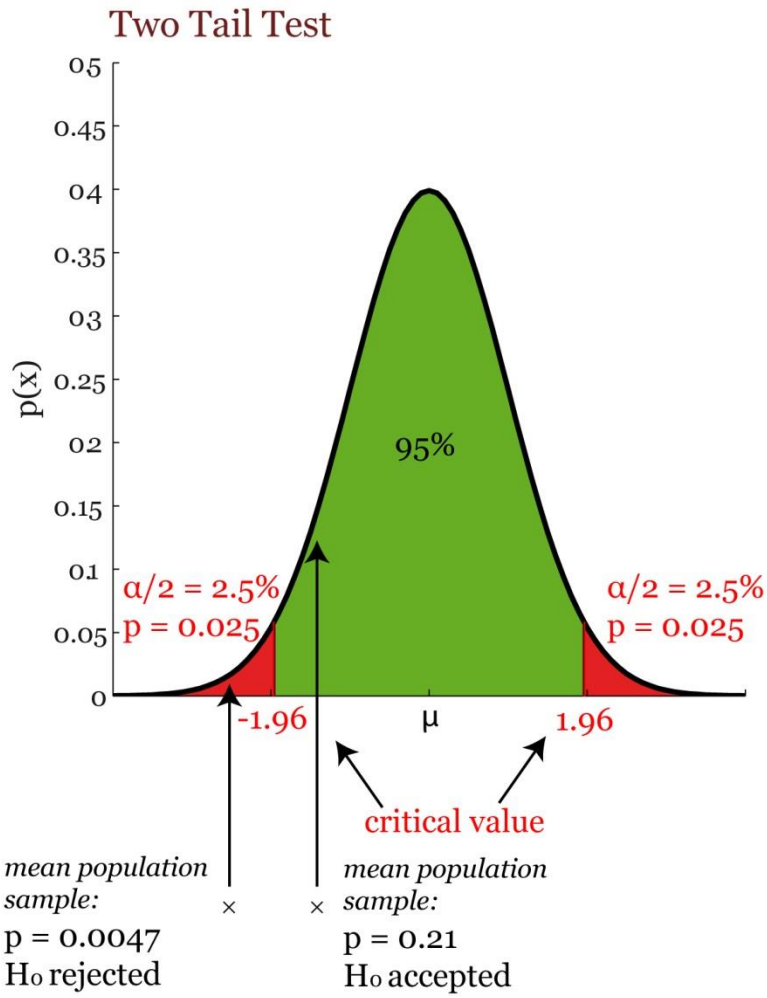
Box plot



Hypothesis testing

distribution: μ = 0, σ = 1





How to report *p* values?

The most common levels of significance reported with respect to research questions:
$p < 0.05$ * (minimal level of significance)
$p < 0.01$ **
$p < 0.001$ *** (statistically highly significant – less than one in a thousand chance of being wrong)

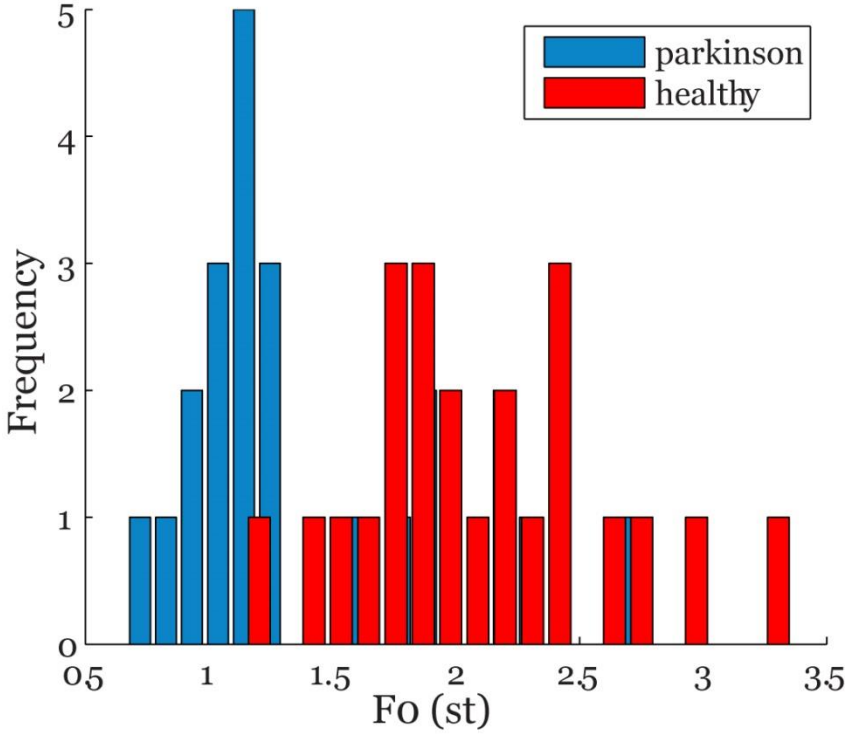
How to report them?
p <1 , 0.01>: 2 digits ($p = 0.02$, $p = 0.51$; 3 digits in special cases $p = 0.049$)
p (0.01 , 0.001>: 3 digits ($p = 0.009$, $p = 0.001$)
p (0.001 , 0>: always $p < 0.001$

Group differences
for normally distributed data (t-test)

Researcher wants to verify if the intonation patterns differ between speakers with Parkinson's disease (PD) and healthy control speakers. He collected short reading texts from 23 speakers with PD and age- and 23 sex-matched controls, extracted fundamental frequency contour and converted it into semitone scale (st).

How to report the results?

$t(45) = -4.40, p < 0.001$



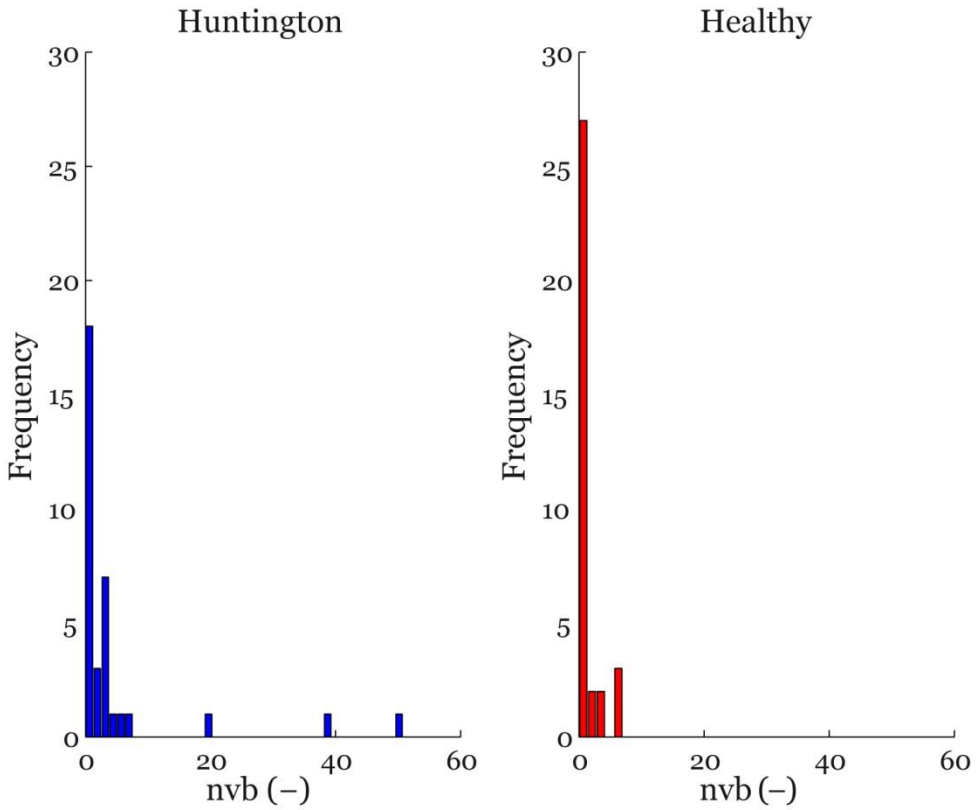
Group differences
for non-normally distributed data (Wilcoxon rank sum test)

Normal speaker is able to perform sustained vowel phonation without voice breaks that represent impaired function of vocal folds. To verify if vocal fold function differs in patients with Huntington's disease (HD), researcher collected sustained phonations from 34 speakers with HD and 34 age- and sex-matched controls and extracted analyzed number of voice breaks (nvb).

How to report the results?
$z(67) = 3{,}81,\ p < 0.001$

What about t-test?
$t(67) = 1.98,\ p = 0.051$

Failed to found significant group differences



Independent samples

Scores are separate (e.g. testing the blood pressure of group of people on active drugs against group of people taking placebo)

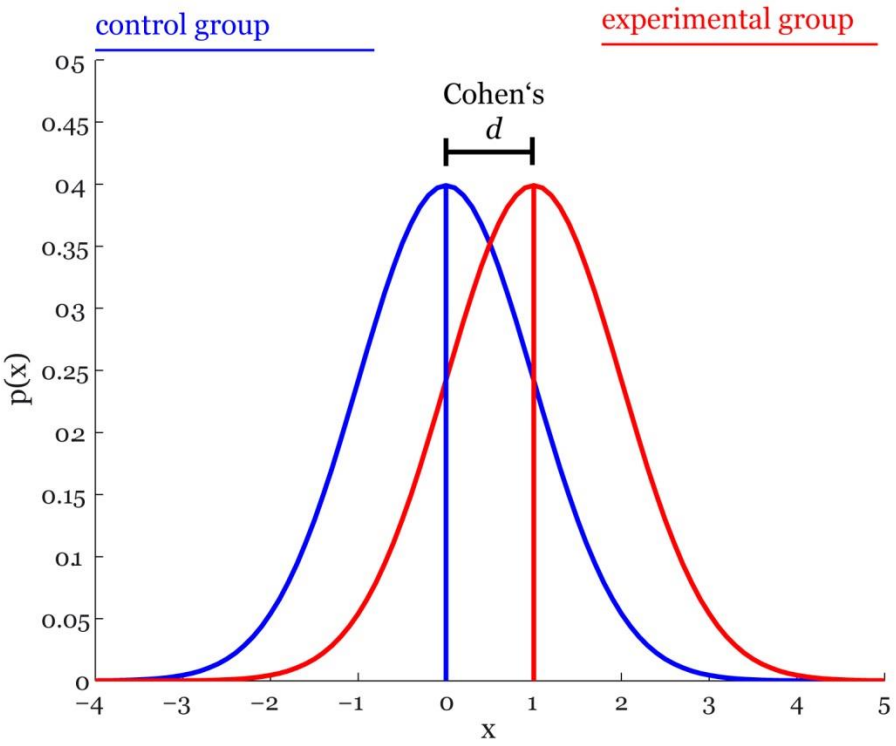
Paired samples

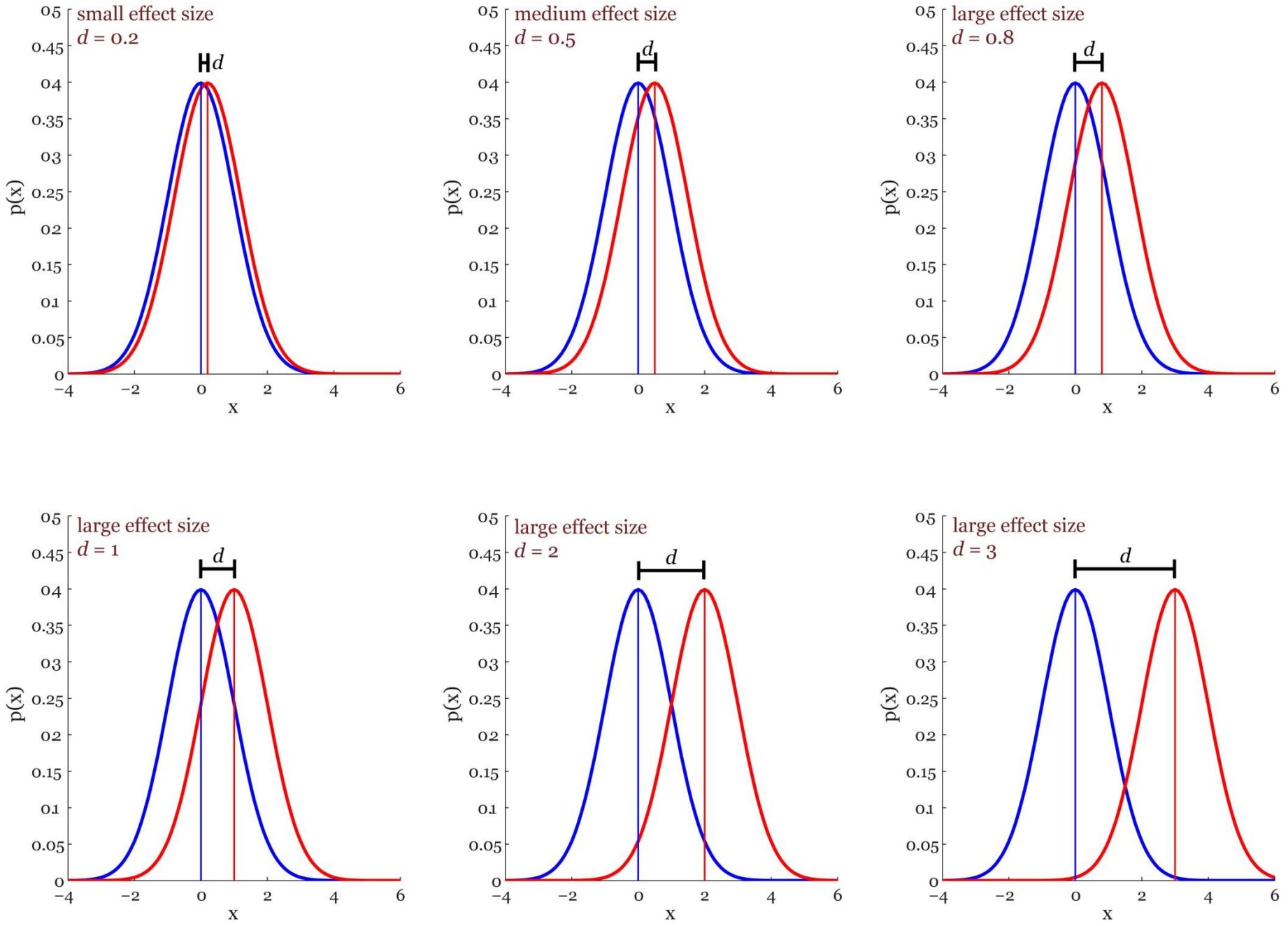
Scores are linked (e.g. measuring the blood pressures of the same people before and after they receive a dose)

Cohen‘s effect size:

a quantitative measure of the strength of a phenomenon, independent of the measured units.

$$d = \frac{\mu_c - \mu_e}{\sigma} \qquad \sigma = \sqrt{\frac{\sigma_c^2 + \sigma_e^2}{2}}$$

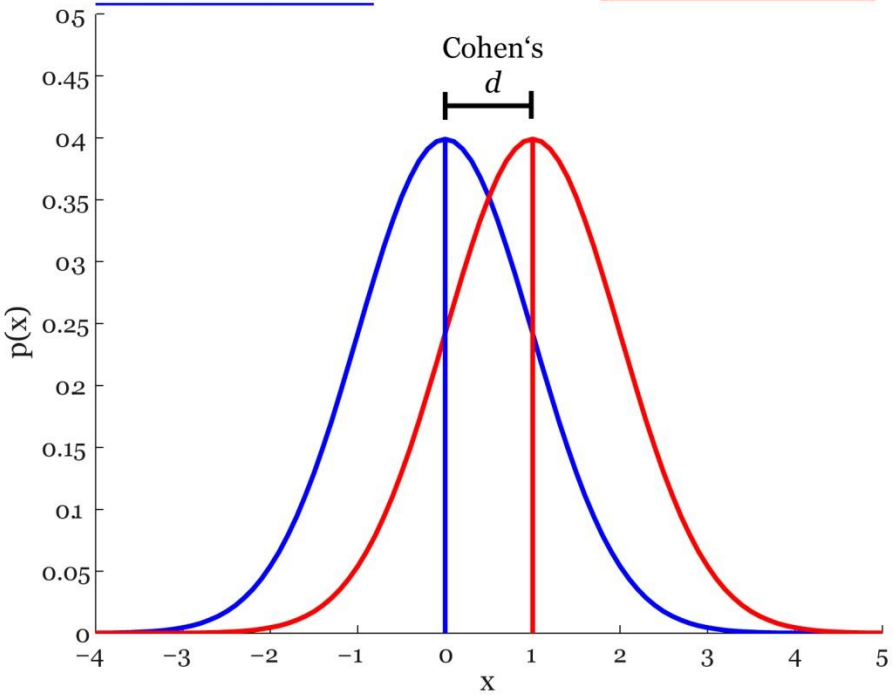



small effect size
d = 0.2
d
medium effect size
d = 0.5
d
large effect size
d = 0.8
d
large effect size
d = 1
d
large effect size
d = 2
d
large effect size
d = 3
d
p(x)
x

Cohen‘s effect size:

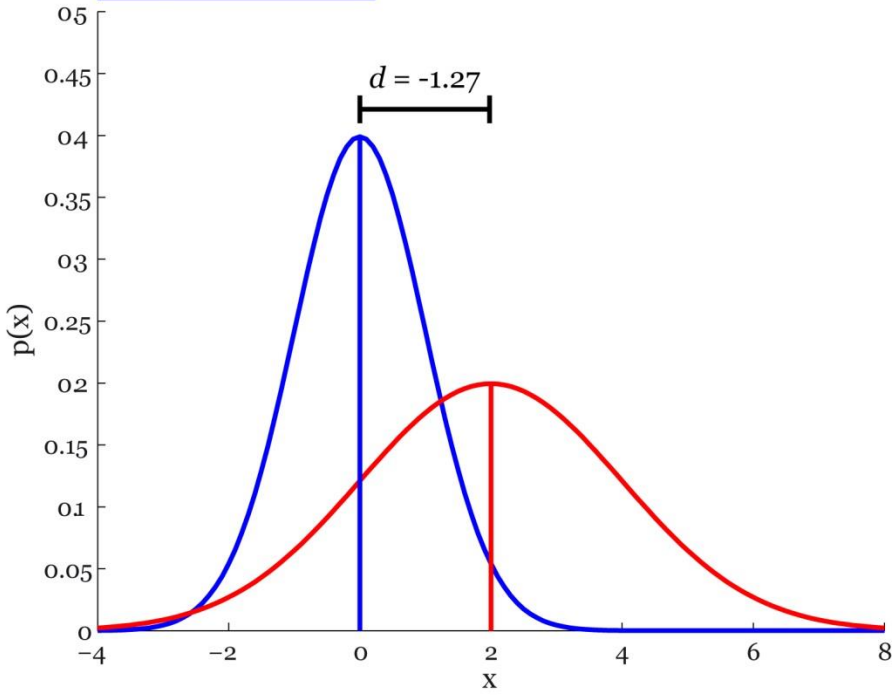
a quantitative measure of the strength of a phenomenon, independent of the measured units.

$$d = \frac{\mu_c - \mu_e}{\sigma} \qquad \sigma = \sqrt{\frac{\sigma_c^2 + \sigma_e^2}{2}}$$



Effect size estimate for unequal sample size: $d = \frac{\mu_c - \mu_e}{\sigma}$ $\sigma = \sqrt{\frac{(n_c - 1)\sigma_c^2 + (n_e - 1)\sigma_e^2}{n_c + n_e + 2}}$

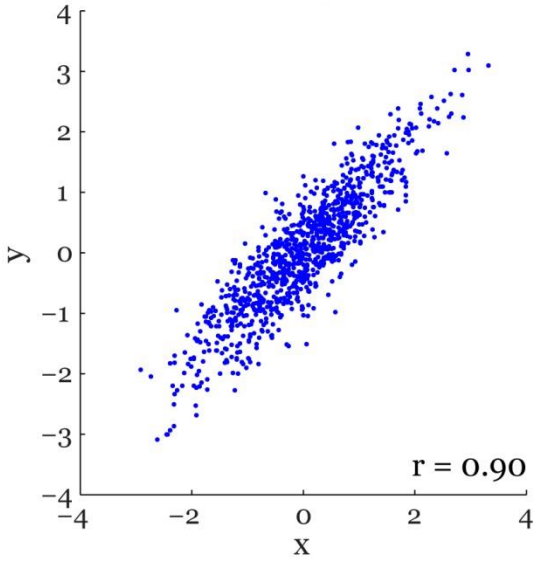
$$\sigma = \sqrt{\frac{1^2 + 2^2}{2}} = 1.58 \qquad d = (0 - 2)/1.58 = -1.27$$



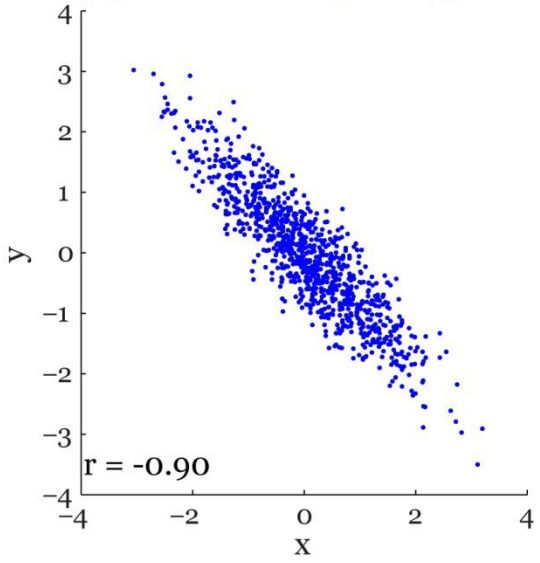
Non-parametric effect size estimate for Mann-Whitney U test: $r = \frac{z}{\sqrt{n_c + n_e}}$

Correlation values

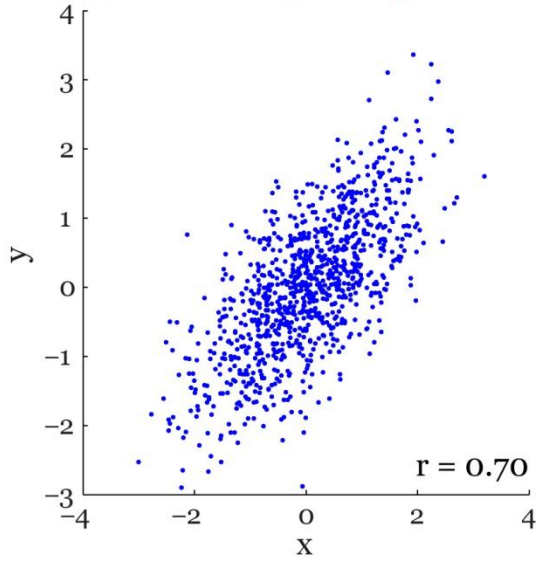
r = <-1, 1>



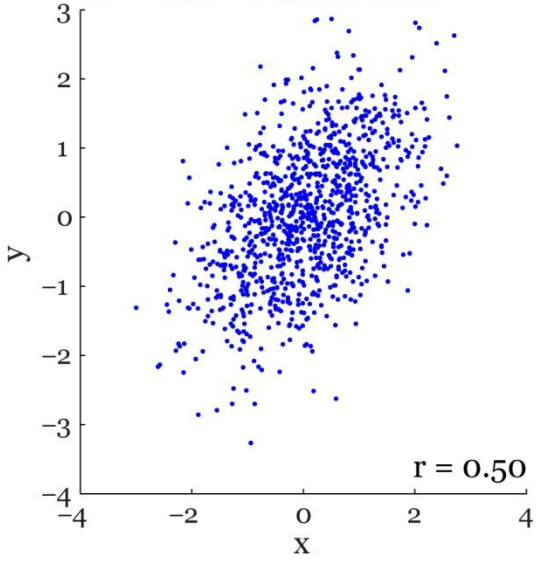
|r| = 0.80—1.0 very strong



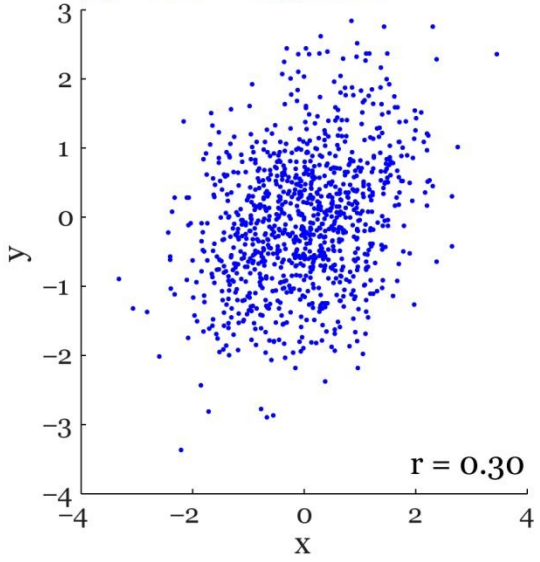
|r| = 0.60—0.79 strong



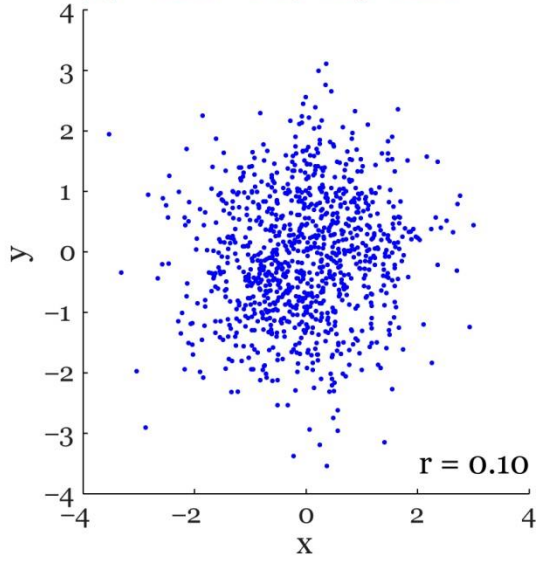
|r| = 0.40—0.59 moderate



|r| = 0.20—0.39 weak

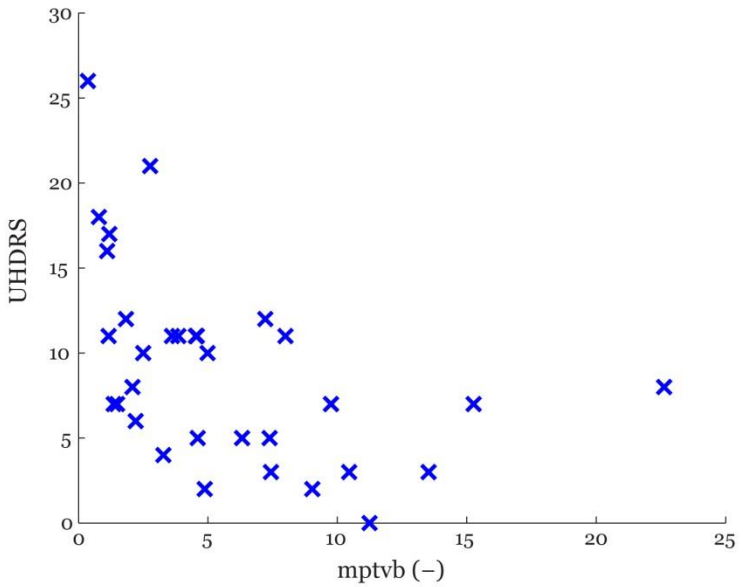


|r| = 0.00—0.19 very weak



Correlations

Normal speaker is able to perform sustained vowel phonation for several seconds without voice breaks that represent impaired function of vocal folds. To verify if vocal fold function disability in patients with Huntington's disease (HD) corresponds to overall motor disability, researcher collected sustained phonations from 32 speakers with HD and performed analysis of maximum phonation time until voice breaks (mptvb). He also assessed every patient using clinical motor scale of Unified Huntington's Disease Raring Scale (UHDRS).



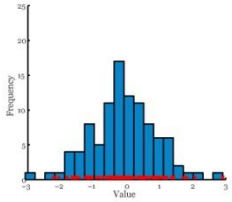
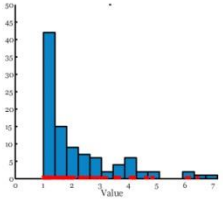
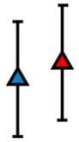
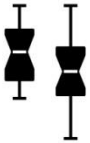
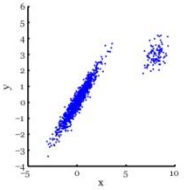
Pearson:

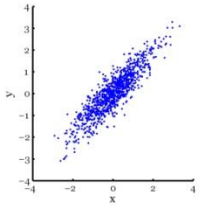
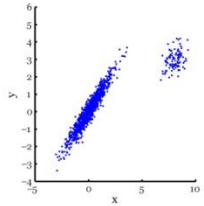
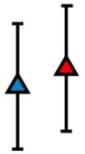
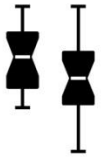
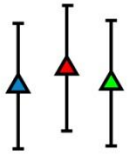
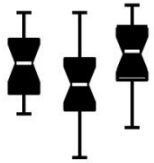
$r = -0.47$, $p = 0.01$

Spearman:

$r = -0.61$, $p < 0.001$

Spearman correlation is more powerful due to non-normally distributed data

	Parametric		Non-parametric	
Expected distribution:	Normal		No limits	
Expected variance:	Homogeneous		No limits	
Typical scale:	Interval	$y = x + d$ dates (years), temperature (°C), IQ scale	Nominal	Non-comparable YES/NO, colours, gender, phone numbers
	Ratio	$y = k \cdot x + d$ velocities, lengths, age...	Ordinal	Can be sorted in terms of *„greater“* or *„less“* education
Central measures:	Mean (SD)		Median (IQR)	
Advantages:	More powerful		Less vulnerable to outliers	
Disadvantages:	More samples needed		Less powerful	

	Parametric	Non-parametric
Correlations:	Pearson *corr* 	Spearman *corr* 
Independent 2 groups:	Independent t-test *ttest2* 	Mann-Whitney U test *ranksum* 
Paired 2 „levels“:	Paired t-test *ttest* 	Wilcoxon signed-rank test *signrank*
Independent > 2 groups:	Analysis of variance *anova1* 	Kruskal-Wallis test *kruskalwallis* 
Paired > 2 „levels“:	Repeated measures analysis of variance *ranova*	Friedman test *friedman* 

III. Research project

ACOUSTIC ANALYSIS OF WORD STRESS PATTERNS IN PARKINSON'S DISEASE

Why to conduct such research?

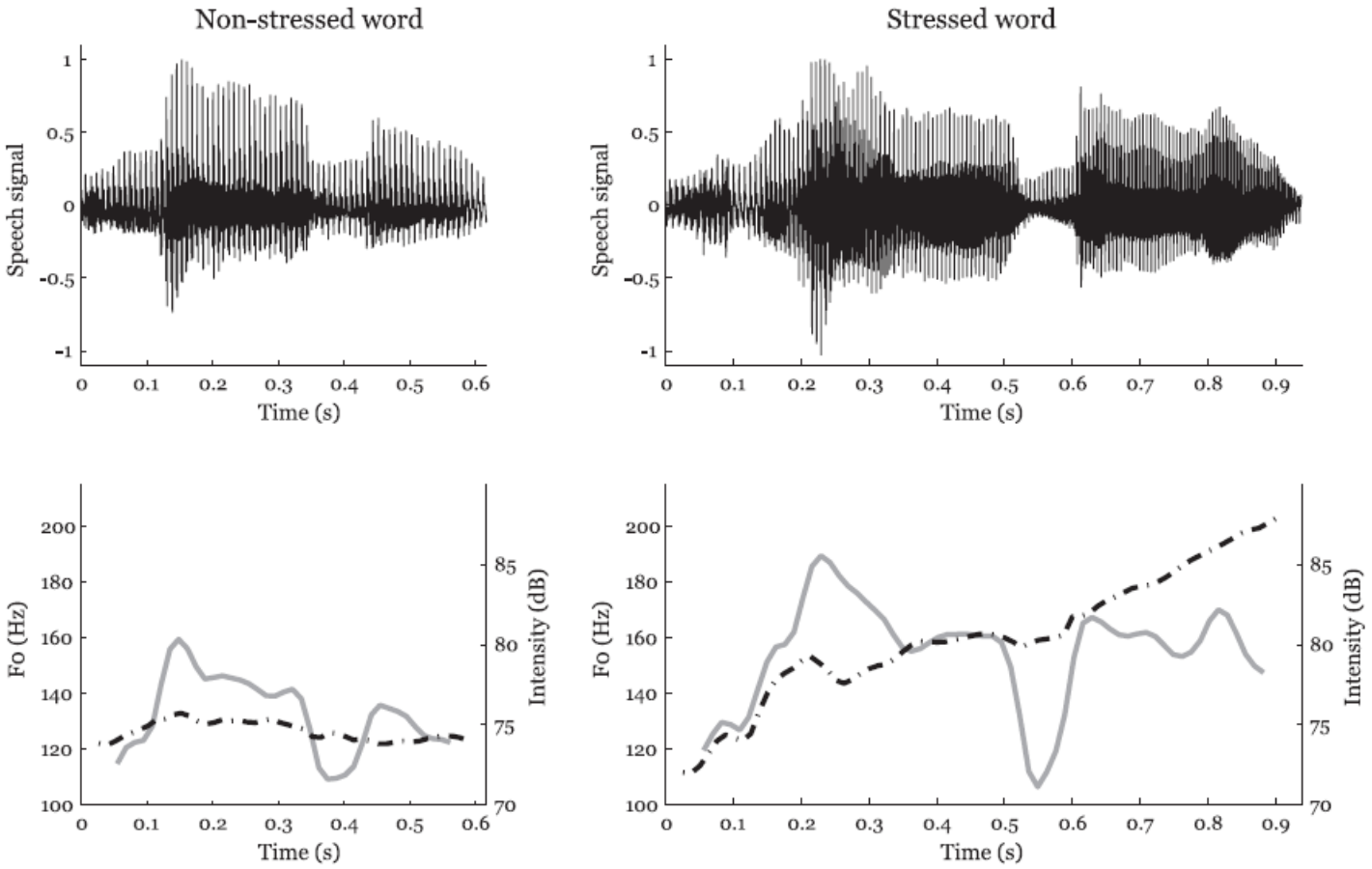
- Patients with PD have limited ability to express contrastive stress
- Impaired stress production is one the most common affected dimensions of speech in PD
- Very limited research regarding stress production in PD

General motivation

- Early and accurate diagnosis
- Monitoring progression of disease
- Monitoring treatment efficacy
- Feedback for speech therapy

Principle of stress expression

- Word prolongation (Time)
- Increase of pitch (Fo)
- Increase of loudness (Intensity)



Hypothesis: what happen during stress expression?

Parameter	Helthy controls	Parkinson's disease
Word length	↑↑↑	↑
Pitch	↑↑↑	↑
Loudness	↑↑↑	↑

Methods: data

Participants:

- 20 males with diagnosis of Parkinson's disease in the early stage of disease
- 18 healthy males as a control group

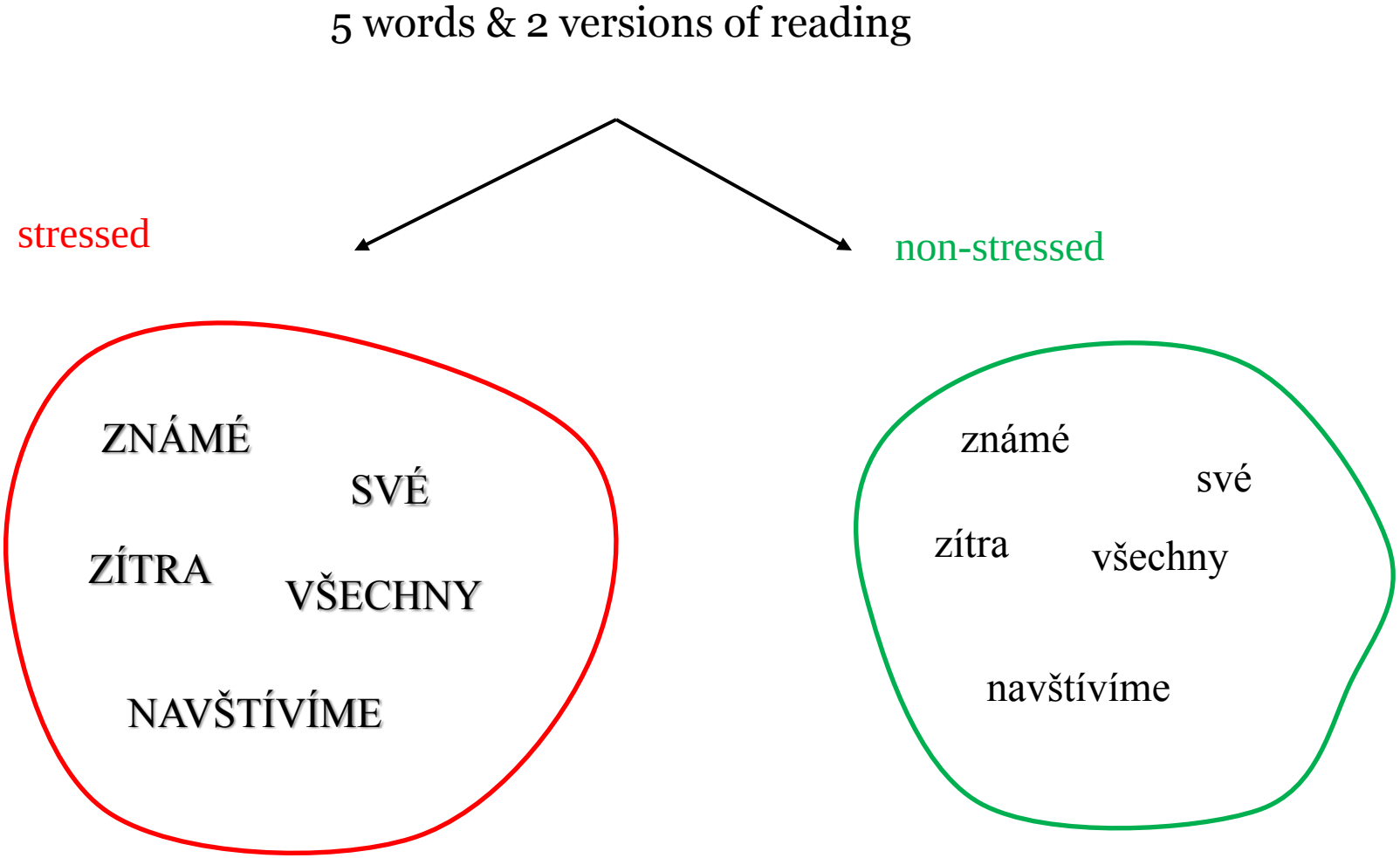
Speech task:

- The aim was to unnaturally emphasize certain keyword marked in the text

Dnes jsme to již nestihli,
možná ZÍTRA navštívíme všechny své známé.

Příbuzné jsme již navštívili,
možná zítra navštívíme všechny své ZNÁMÉ.

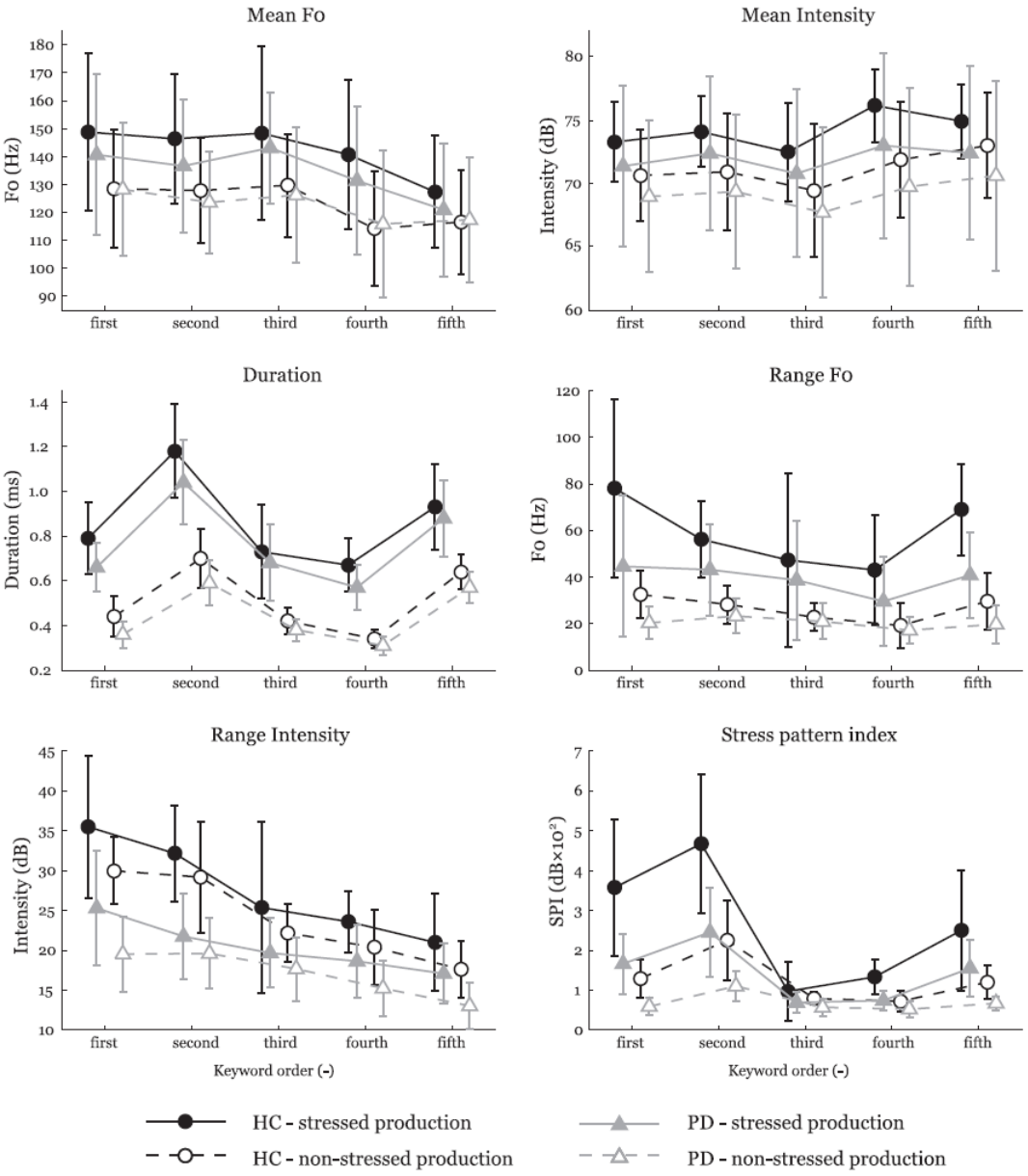
Methods: speech data



Methods: variables

- Mean value of Fo
- Range of Fo (min-max)
- Mean value of intensity of loudness
- Range of intensity of loudness
- Length of word
- Stress pattern index (combining effects of all three fundamental aspects including pitch, loudness, and length)

Results



- 2 groups (PD, HC)
- 6 variables
- 5 words
- 2 conditions (stressed, non-stressed)
  +1 condition (difference between stressed and non-stressed)

For each PD vs. HC comparison:
6 × 5 × 3 = **90 tests**

PD = Parkinson's disease
HC = Healthy controls

Results

MEAN F0		STRESSED	NORMAL	ABS. STRESS
ZNAME	PN x HC	NS	NS	NS
NAVSTIVIME	PN x HC	NS	NS	NS
SVE	PN x HC	NS	NS	NS
VSECHNY	PN x HC	NS	NS	NS
ZITRA	PN x HC	NS	NS	NS

MEAN INTENSITY		STRESSED	NORMAL	ABS. STRESS
ZNAME	PN x HC	NS	NS	NS
NAVSTIVIME	PN x HC	NS	NS	NS
SVE	PN x HC	NS	NS	NS
VSECHNY	PN x HC	NS	NS	NS
ZITRA	PN x HC	NS	NS	NS

DURATION		STRESSED	NORMAL	ABS. STRESS
ZNAME	PN x HC	NS	p < 0.01	NS
NAVSTIVIME	PN x HC	p < 0.05	p < 0.01	NS
SVE	PN x HC	p < 0.05	p < 0.05	p < 0.05
VSECHNY	PN x HC	NS	p < 0.05	NS
ZITRA	PN x HC	p < 0.01	p < 0.01	NS

RANGE INTENSITY		STRESSED	NORMAL	ABS. STRESS
ZNAME	PN x HC	p < 0.05	p < 0.001	NS
NAVSTIVIME	PN x HC	p < 0.001	p < 0.001	NS
SVE	PN x HC	p < 0.001	p < 0.01	NS
VSECHNY	PN x HC	p < 0.05	p < 0.01	NS
ZITRA	PN x HC	p < 0.001	p < 0.001	NS

RANGE F0		STRESSED	NORMAL	ABS. STRESS
ZNAME	PN x HC	p < 0.001	p < 0.01	p < 0.01
NAVSTIVIME	PN x HC	p < 0.05	NS	NS
SVE	PN x HC	NS	NS	NS
VSECHNY	PN x HC	NS	NS	NS
ZITRA	PN x HC	p < 0.01	p < 0.001	NS

STRESS PATTERN INDEX		STRESSED	NORMAL	ABS. STRESS
ZNAME	PN x HC	p < 0.05	p < 0.001	NS
NAVSTIVIME	PN x HC	p < 0.001	p < 0.001	p < 0.05
SVE	PN x HC	p < 0.001	p < 0.05	p < 0.01
VSECHNY	PN x HC	NS	p < 0.01	NS
ZITRA	PN x HC	p < 0.001	p < 0.001	p < 0.01

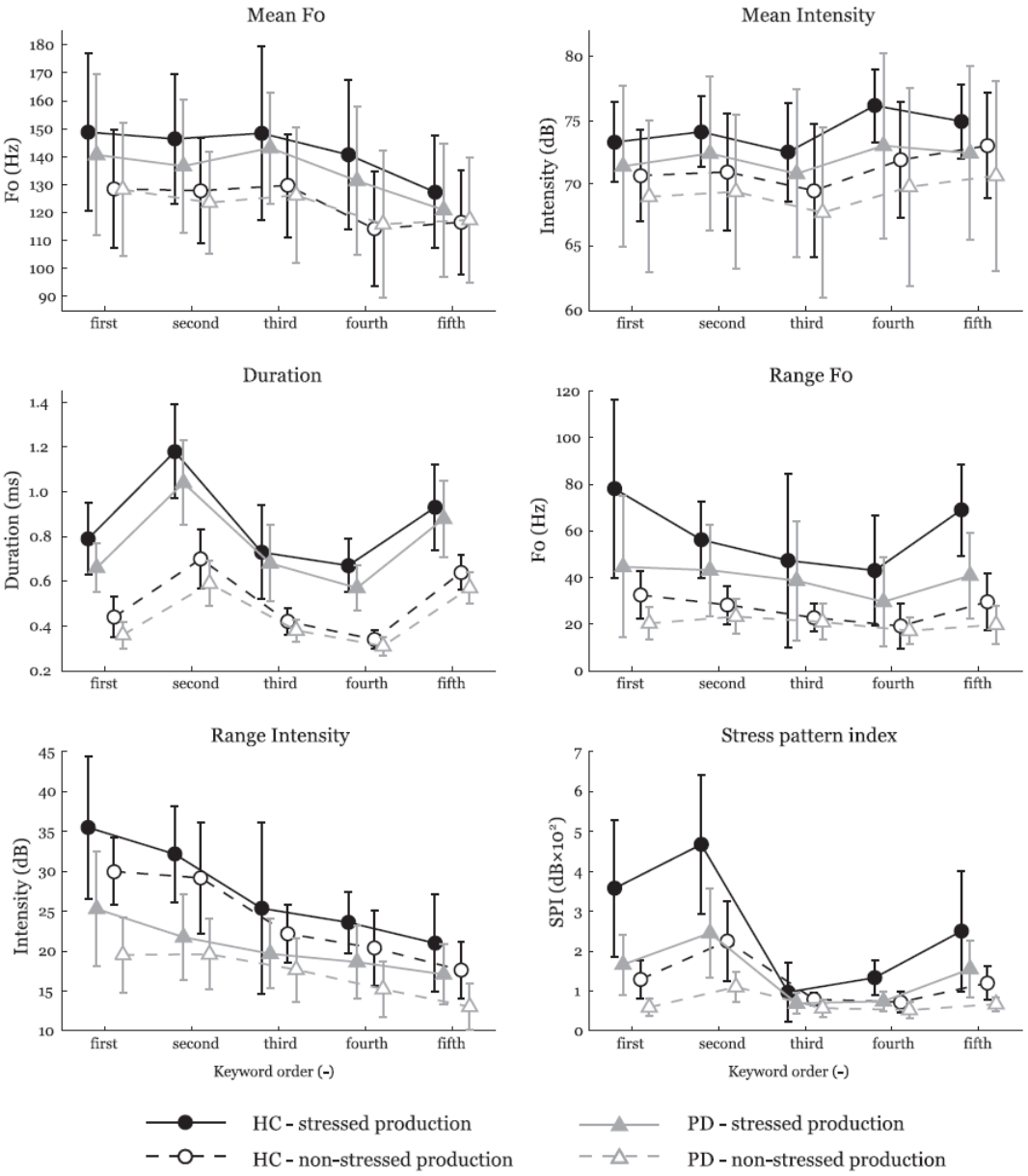
- Very hard to interpret the results
- Type I Error
- Corrected: $p = 0.05/90 \Rightarrow p < 0.00056$
- => no result is significant

NS = not significant

Solution: Three-way ANOVA

- 2 × 2 × 5 ANOVA
- 3 factors
  - SPEAKERS GROUP (PD, HC)
  - STRESS CONDITION (stressed, non-stressed)
  - KEYWORD ORDER (first, second, third, fourth, fifth)

Three-way ANOVA: Interpretation



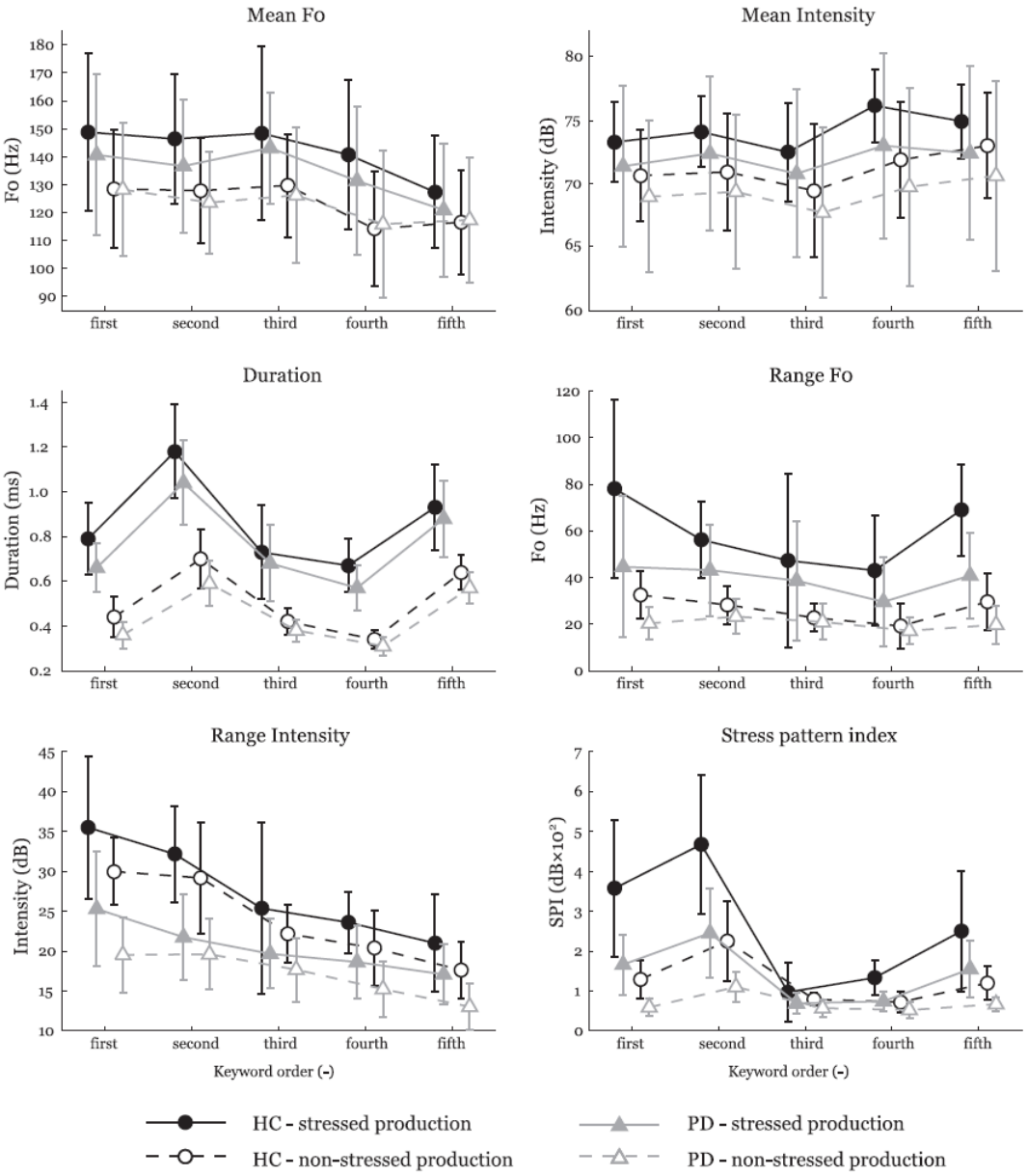
SPEAKERS GROUP

H_A: PD patients will manifest affected speech

Three-way ANOVA: Results

	mean F0	mean Intensity	duration	range F0	range Intensity	SPI
Speakers group	**n.s.**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**
Stress condition	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Keyword order	<0.001	n.s.	<0.001	<0.001	<0.001	<0.001
Speakers group x Stress condition	n.s.	n.s.	n.s.	0.0014	n.s.	<0.001
Keyword order x Speakers group	n.s.	n.s.	n.s.	n.s.	<0.001	<0.001
Keyword order x Stress condition	n.s.	n.s.	<0.001	n.s.	n.s.	<0.001

Three-way ANOVA: Interpretation



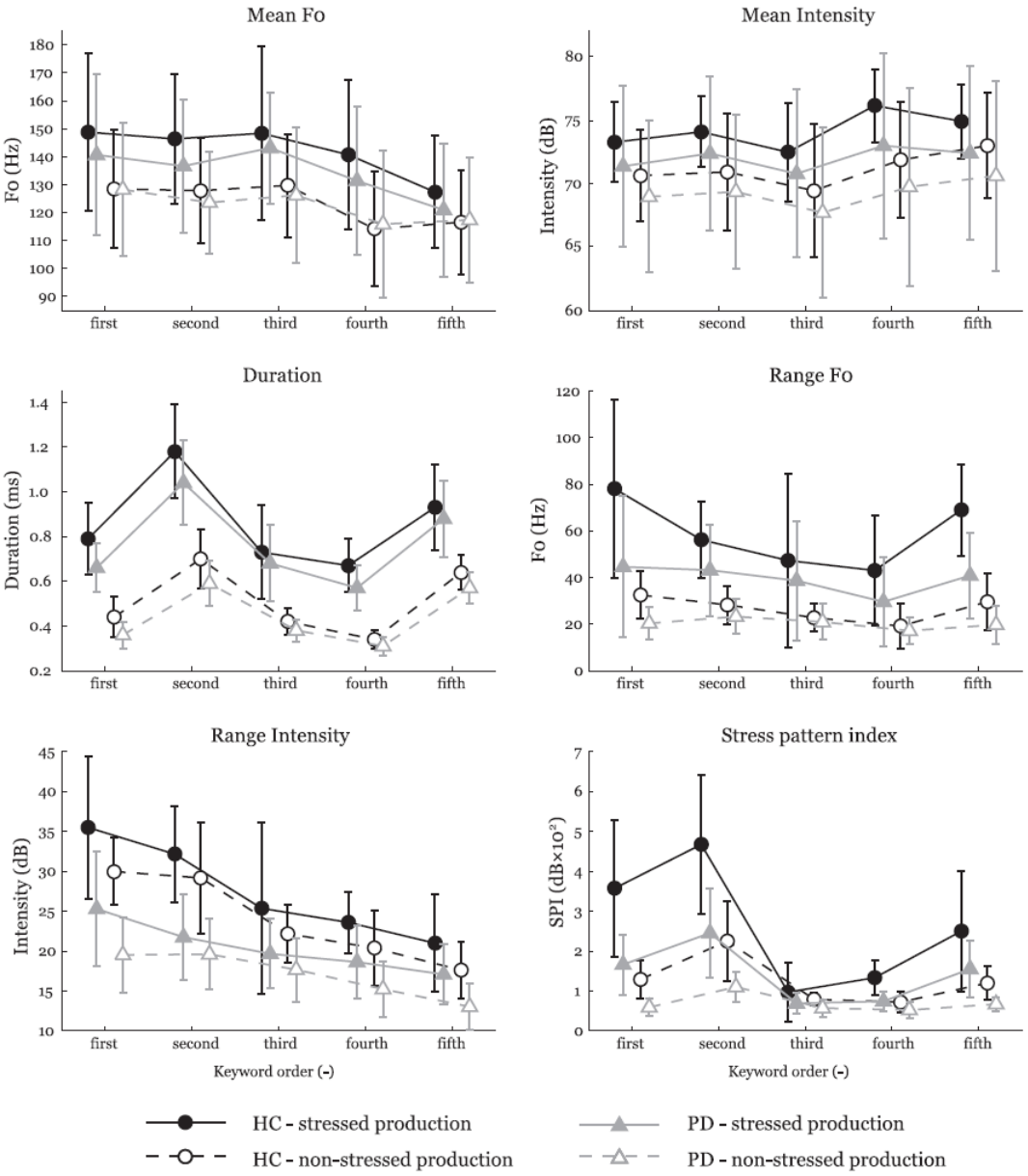
STRESS CONDITION

H_A: Variable reflects effect of the stress condition

Three-way ANOVA: Results

	mean F0	mean Intensity	duration	range F0	range Intensity	SPI
Speakers group	n.s.	<0.001	<0.001	<0.001	<0.001	<0.001
Stress condition	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**
Keyword order	<0.001	n.s.	<0.001	<0.001	<0.001	<0.001
Speakers group x Stress condition	n.s.	n.s.	n.s.	0.0014	n.s.	<0.001
Keyword order x Speakers group	n.s.	n.s.	n.s.	n.s.	<0.001	<0.001
Keyword order x Stress condition	n.s.	n.s.	<0.001	n.s.	n.s.	<0.001

Three-way ANOVA: Interpretation



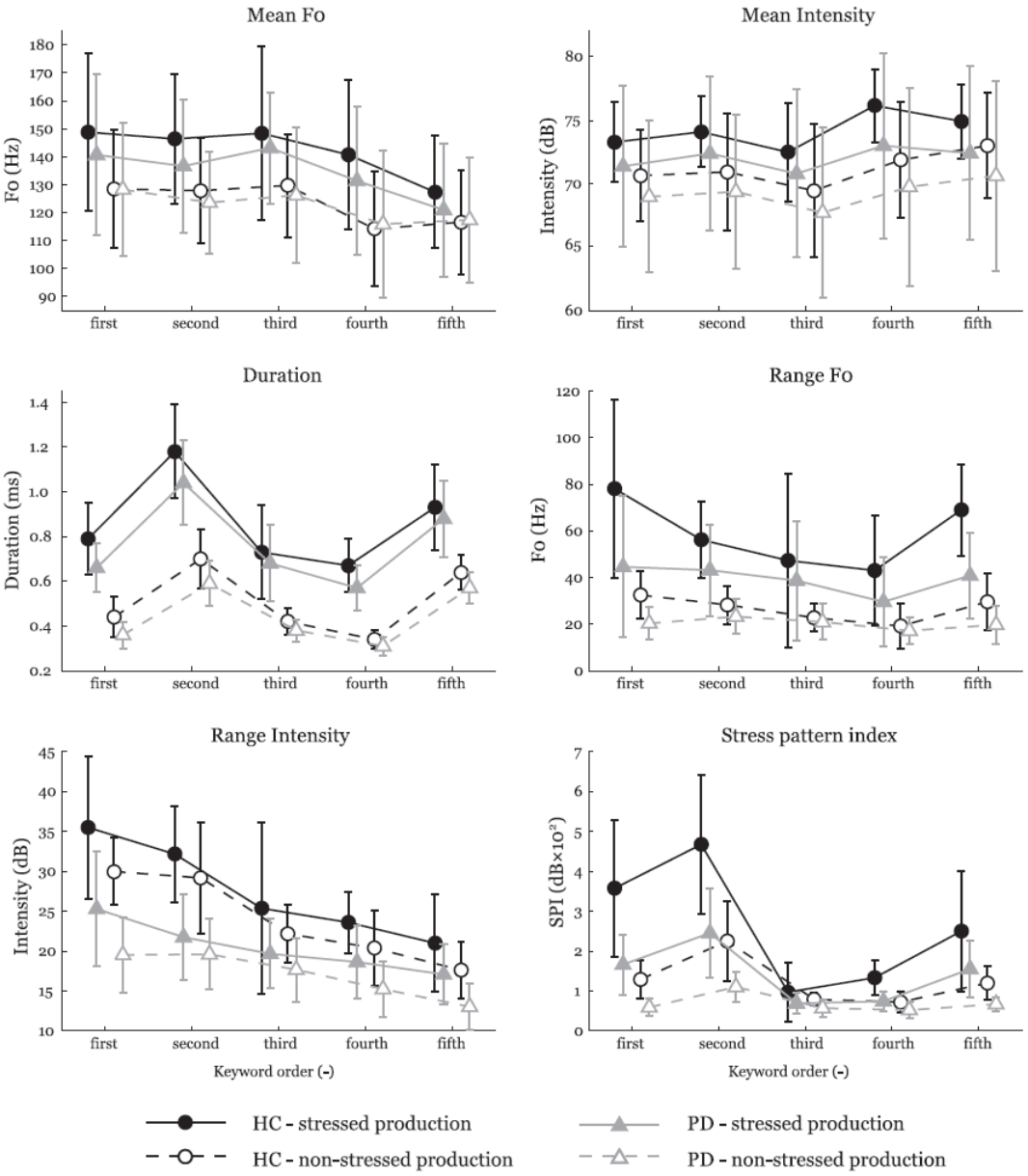
KEYWORD ORDER

H_A: Variable reflects the keyword order

Three-way ANOVA: Results

	mean F0	mean Intensity	duration	range F0	range Intensity	SPI
Speakers group	n.s.	<0.001	<0.001	<0.001	<0.001	<0.001
Stress condition	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Keyword order	**<0.001**	**n.s.**	**<0.001**	**<0.001**	**<0.001**	**<0.001**
Speakers group x Stress condition	n.s.	n.s.	n.s.	0.0014	n.s.	<0.001
Keyword order x Speakers group	n.s.	n.s.	n.s.	n.s.	<0.001	<0.001
Keyword order x Stress condition	n.s.	n.s.	<0.001	n.s.	n.s.	<0.001

Three-way ANOVA: Interpretation



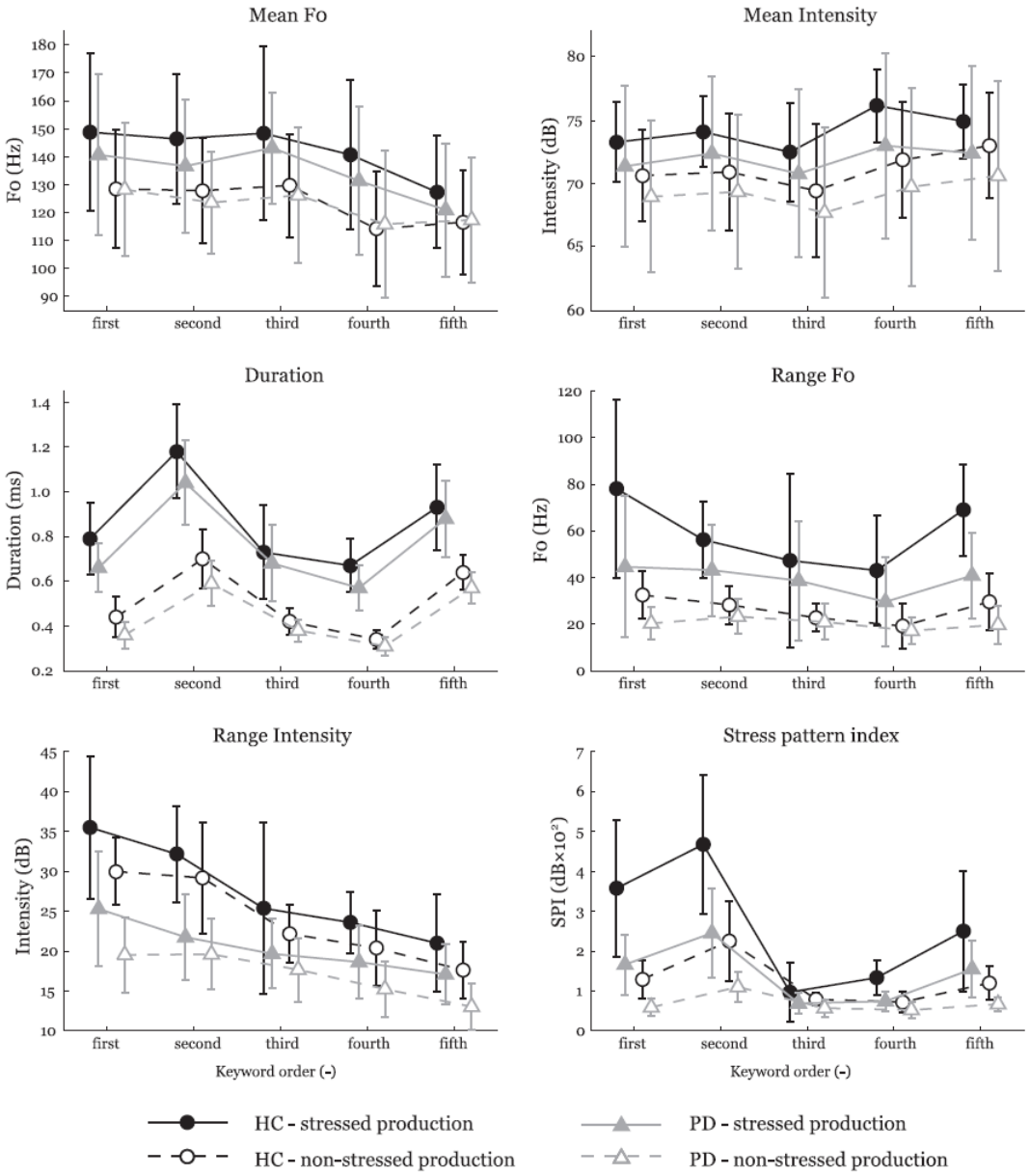
SPEAKERS GROUP × STRESS CONDITION

H_A: PD patients are not able to express stress as effectively as HC

Three-way ANOVA: Results

	mean F0	mean Intensity	duration	range F0	range Intensity	SPI
Speakers group	n.s.	<0.001	<0.001	<0.001	<0.001	<0.001
Stress condition	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Keyword order	<0.001	n.s.	<0.001	<0.001	<0.001	<0.001
Speakers group x Stress condition	**n.s.**	**n.s.**	**n.s.**	**0.0014**	**n.s.**	**<0.001**
Keyword order x Speakers group	n.s.	n.s.	n.s.	n.s.	<0.001	<0.001
Keyword order x Stress condition	n.s.	n.s.	<0.001	n.s.	n.s.	<0.001

Three-way ANOVA: Interpretation



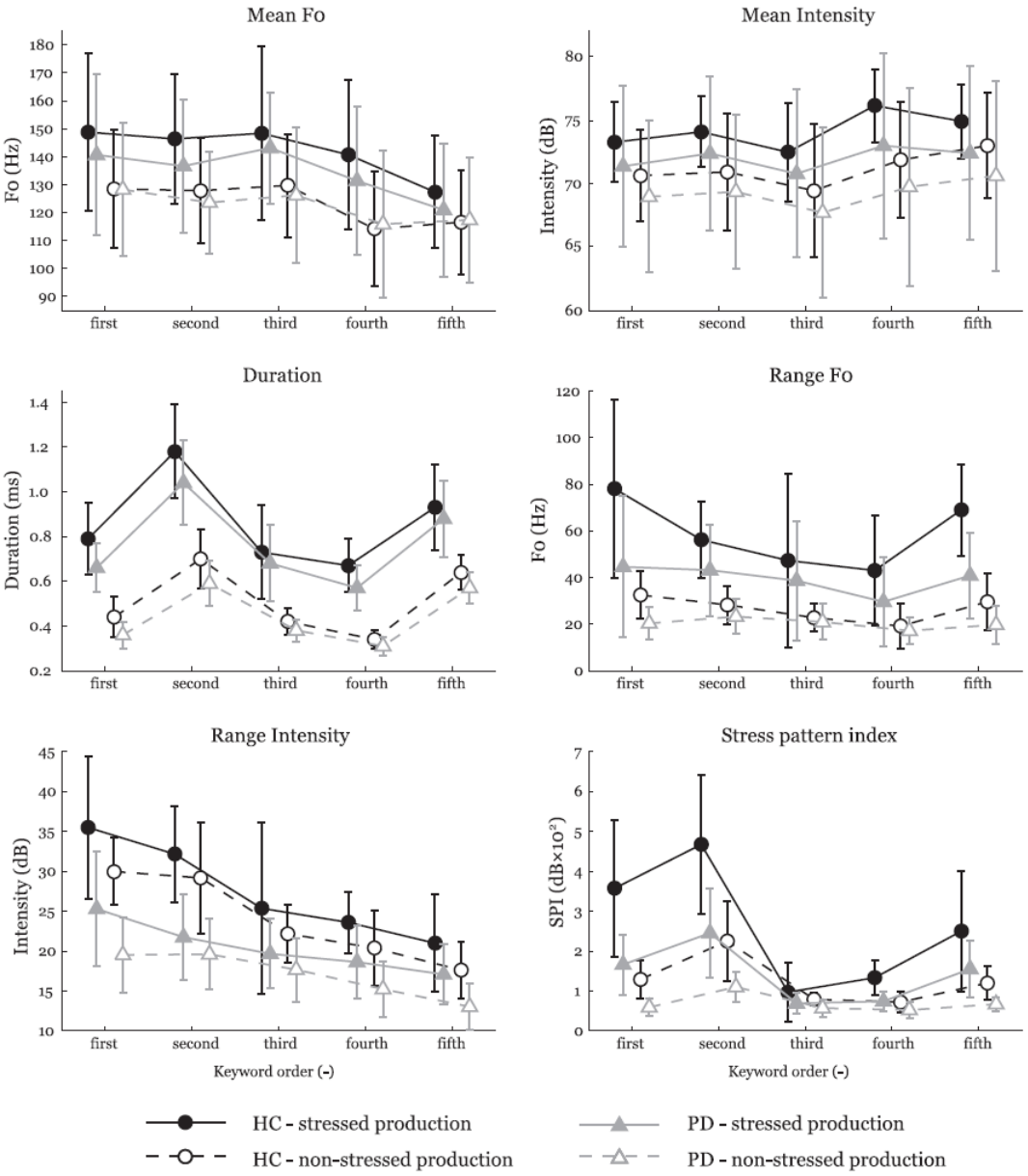
KEYWORD ORDER× SPEAKERS GROUP

H_A: Certain word is more effective in the separation of groups

Three-way ANOVA: Results

	mean F0	mean Intensity	duration	range F0	range Intensity	SPI
Speakers group	n.s.	<0.001	<0.001	<0.001	<0.001	<0.001
Stress condition	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Keyword order	<0.001	n.s.	<0.001	<0.001	<0.001	<0.001
Speakers group x Stress condition	n.s.	n.s.	n.s.	0.0014	n.s.	<0.001
Keyword order x Speakers group	**n.s.**	**n.s.**	**n.s.**	**n.s.**	**<0.001**	**<0.001**
Keyword order x Stress condition	n.s.	n.s.	<0.001	n.s.	n.s.	<0.001

Three-way ANOVA: Interpretation



KEYWORD ORDER× STRESS CONDITION

H_A: Certain word is more suitable for expression of stress condition

Three-way ANOVA: Results

	mean F0	mean Intensity	duration	range F0	range Intensity	SPI
Speakers group	n.s.	<0.001	<0.001	<0.001	<0.001	<0.001
Stress condition	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Keyword order	<0.001	n.s.	<0.001	<0.001	<0.001	<0.001
Speakers group x Stress condition	n.s.	n.s.	n.s.	0.0014	n.s.	<0.001
Keyword order x Speakers group	n.s.	n.s.	n.s.	n.s.	<0.001	<0.001
Keyword order x Stress condition	**n.s.**	**n.s.**	**<0.001**	**n.s.**	**n.s.**	**<0.001**

Three-way ANOVA: Interpretation

SPEAKERS GROUP – PD patients will manifest affected speech

STRESS CONDITION – variable reflects effect of the stress condition

KEYWORD ORDER – variable reflects the keyword order

SPEAKERS GROUP × STRESS CONDITION – PD patients are not able to express stress as effectively as HC

KEYWORD ORDER× SPEAKERS GROUP – certain word is more effective in the separation of groups

KEYWORD ORDER× STRESS CONDITION – certain word is more suitable for expression of stress condition

Matlab output to replicate

Analysis of Variance

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
group	66.628	1	66.6284	106.67	0
stress	95.182	1	95.182	152.38	0
keyword	165.214	4	41.3036	66.13	0
group*stress	8.844	1	8.8439	14.16	0.0002
group*keyword	26.217	4	6.5542	10.49	0
stress*keyword	38.71	4	9.6774	15.49	0
Error	211.123	338	0.6246		
Total	595.155	353			

Main conclusions regarding our findings

- PD patients have decreased ability to express stress (SPEAKERS GROUP × STRESS CONDITION)
- PD patients are able to consciously improve their speech performance (MAIN EFFECTS)

IV. How to “sold” the results of my project?

Before publishing your work, think about why you want to publish your work. Is it publishable?

- Have I done something new and interesting?
- Is there anything challenging in my work?
- Have I provided solution to some difficult problem?
- How novel is my approach?

- You have to known well state-of-the-art (perform quantitative literature review) in order not to perform redundant research/development, i.e. re-introduce the wheel.

Reviewers will likely have to answer following questions.

- Does the work contains sufficient new material?
- Is the paper presented concisely and is it well organized? Are the methods and experiments presented in the way that they can be replicated again?
- Are the results presented adequately?
- Is the discussion relevant, concise and well documented?
- Are the conclusions supported by the data presented? Are the limitations relevant (if applicable)?
- Is the language acceptable?
- Are figures and tables adequate and well designed? Are there some information duplicated?
- Are all references cited in the text included in the reference list?

Pay attention to structure of the paper.

- **Introduction**: What did you/others do? Why did you do it?
- **Methods**: How did you do it?
- **Results**: What did you find?
- **Discussion**: What does it all mean?
- **Conclusion**: How the work advances the field from the present state of knowledge?

I. Introduction.

- What is the problem to be solved?
- Are there any existing solutions?
- Which is the best?
- What is its main limitation?
- What do you hope to achieve?

I. Tips for Introduction.

- Never use more words than necessary. Do not make this section into history lesson. Introduce only things necessary to introduce your own work. Long introduction put readers off.
- Do not use improper citations.
- Do not cite too many references irrelevant to the work.
- Introduction have to be organized from the global to particular point of view, guiding your readers to your objectives when writing the paper.
- Do not mix introduction with results, discussion and conclusion. Always keep them separate to ensure that the paper flows logically from one section to the next.
- Hypotheses and objectives have to be clearly remarked at the end of the introduction.

II. Methods.

- Respond to the question of how the problem was studied.
- If you are proposing new method, you need to include detailed information allowing the others to reproduce the experiment.
- Do not repeat the details of already established methods.
- Methods provide critical knowledge; incomplete or incorrect methods represent major fail of the work.
- Always use standard system and nomenclature, for example international System of Units (SI).

II. Tips for Methods.

- Present methods in the logical order in which you did your research:
  - Description of the **materials and participants**.
  - Description of the **experiments** done (for example speech assessment).
  - Description of the **laboratory methods** (for example signal segmentation).
  - Description of the **statistical methods** used.
- Do not provide partial results as the part of methods.

III. Results.

- Provide results of your experiments, refers to the question “What have you found?”

III. Tips for Results.

- Ideally, present only representative results of your experiment. In other words, these results should be essential for discussion.
- Decide to present data in logical order that tells a clear story easy to understand. Generally, this will be in the same order as presented in methods section.
- You should provide clear results devoid of emotions and interpretations. You should never include references in this section. You are presenting **your results** here. Thus, you cannot refer to others, this is *discussion*.
- Do not list any methods in the results section.

IV. Discussion.

- You must explain your results to wider audience.
- Typically the most important section of your article.
- Here you get chance to sell your data.
- Your work can easily be rejected because discussion is weak.
- You have to make the discussion corresponding to the results but not reiterate the results.
- You need to **compare your results to previously observed results**.
- Do not ignore the work in disagreement with you findings, you should rather confront it and convince the readers that you are correct and better.

IV. Tips for Discussion.

- Avoid statements that go beyond what your results can support.
- Avoid unspecific expressions, quantitative descriptions are always preferred.
- Avoid sudden introduction of new terms and ideas; you have to introduce everything in the introduction.
- Speculations on possible interpretations are allowed, but these should be rooted in fact, rather than imagination.

To achieve good interpretation, think about:

- How do your results relate to original objectives outlined in introduction?
- Do the data support your hypothesis?
- Are your results consisted with what other investigators reported?
- Discuss weaknesses and discrepancies. Provide limitations. If your results were unexpected, try to explain why.
- Is there another way to interpret your results?
- What further research would be necessary to answer questions raised by your results?

IV. Conclusion.

- Shortly tells how your work advances the field from the present state of knowledge.
- It is typically separate section or last section of discussion.

IV. Tips for Conclusion.

- Do not repeat the abstract or just list experimental results.
- Provide specific or global conclusion with respect to the objectives included in the introduction.
- Provide clear scientific justification for your work and indicate uses and extensions (if applicable). Suggest future experiments or point out those that are underway.

Figures and tables

- “Figure is worth of a thousand words.” Illustrations, including both figures and tables, are the most efficient way to present your results. Your data are driving force of the paper and thus your illustrations are critical.
- Whenever you choose table or figure, no illustration should replicate the information described elsewhere in the paper.
- Figures should preferably be freestanding, and figure and table have to be self-explanatory.

Tips for figures and tables

- Avoid too many curves in one figure.
- Select appropriate axis label size.
- Include clear symbols and data sets that are easy to distinguish.
- Preferably use color only when necessary.
- Lines joining data should be used only when you present time series or consecutive sample data.
- In tables, use same number of decimal numbers for each variable.

Abstract

- Summarize to prospective reader what you did and what the important findings in your research were.
- Together with the title, it represents advertisement of your paper.
- It gives key results but minimizes experimental details.
- Typically you have less than 250 words for abstract.

Tips for Abstract

- Use structured form, i.e. Introduction, Objectives, Methods, Results, Conclusion
- Two parts that are essential:
  - What has been done?
  - What are the main findings?

Tips for References

- You have to cite all the scientific publications on which your work is based. Do not over-inflate your paper with too many references. Avoid excessive self-citations and excessive citations of publications from the same region.
- Do not provide references to unpublished observations, articles that are not peer-reviewed, and minimize citations to articles not published in English. For example, Wiki and Google are not reliable source to cite.
- Make sure that all references are provided in the same format. You can use software such as *EndNote* to format your references.

Other tips

- Guideline introducing how to write a seminar paper including tips for developing a well-received paper.

https://www.wikihow.com/Write-a-Seminar-Paper