

Statistics for Biologists

(What, when and where to use it)

(*Minimal use of equations)

Free Registration: <http://doodle.com/poll/v8uxyxqs8rhx6ddnsh5z64pz/admin#table>

Schedule

	Date	Time	Topics	Place
Lecture 1	August 2016, Monday 29	2:00 PM - 3:00 PM	Need of statistics in basic research, Statistics in drug development, Types of clinical study, Phases of Randomized Controlled Trials, Blinding and placebos, How to select samples, Bias, Confounding factors, Replicates, Optimal Sample Size-one sample case, Confidence Intervals, (t-table, z-table) Issues @ http://www.nature.com/authors/policies/reporting.pdf	Faculty Club, 16.6.16
Lecture 2	September 2016, Monday 5	2:00 PM - 3:00 PM	Making sense of distributions and graphs from journal publications (box-and-whisker plot, STD, SEM, median, IQR) Hypothesis Testing, p-value, Type I/II error, Power of study, Level of significance, Effect Size, one/two tail, Sample size calculation in two sample case	Auditorium 29.01.30
Lecture 3	September 2016, Monday 12	2:00 PM - 3:00 PM	Data Normalization (z-score), How to compare two or more groups, ANOVA (one/two/n-way), degree of freedom, t-test, Wilcoxon–Mann–Whitney test	Victor Haderup auditorium
Lecture 4	September 2016, Monday 19	2:00 PM - 3:00 PM	Theory of Correlation, Causality & association. Pearson's, Spearmans, Kendalls r.	Adolf Hannover auditorium
Lecture 5	September 2016, Monday 26	2:00 PM - 3:00 PM	Multiparametric analysis, Linear modeling (Logistic Regression), Odds ratio, prevalence, incidence, relative risk. Sensitivity, Specificity, TP, FP, PPV, NPV, Area under ROC curve.	Adolf Hannover auditorium
Lecture 6	October 2016, Monday 3	2:00 PM - 3:00 PM	Guidelines for reporting statistics in publications (www.Nature.com : Points of Significance) Detailed Issues @ http://www.nature.com/authors/policies/reporting.pdf	Adolf Hannover auditorium
Lecture 7	October 2016, Monday 10	2:00 PM - 3:00 PM	k-means clustering, Kruskal-Wallis test, PCA, Kaplan-Meier Analysis, COX hazard model, Introduction to machine learning and big data in drug discovery	Adolf Hannover auditorium
Lecture 8	October 2016, Monday 17	2:00 PM - 3:00 PM	Hands -on activities and demonstrations (Statistical Toolbox in Matlab)	Adolf Hannover auditorium

Lecture 1

- 1) What are populations and samples, statistics and parameters?
 - a. How to set null hypothesis before any of your experiment?
 - b. Types of variables and scales of measurement
 - c. How many samples are enough for my experiment? i.e. understanding sample size (n)
- 2) Clinical study design (Basic introduction with few examples from epidemiology study)
 - a. Randomized controlled trial
 - i. Blind trial
 - ii. Non-blind trial
 - b. Cohort study
 - c. Case-control study
 - d. Cross-sectional study
 - e. Longitudinal study-> Follow-up-> Temporal studies
 - f. What are efficacy and placebo effect of a drug and ways to measure it?

Lecture 2

- 3) Making sense of distributions and graphs
 - a. Examples from journals publications Nature, Cell, Science etc.
 - b. General trends in statistics to representing your results
 - c. Understanding the Mean, Median, and Mode of a Skewed Distribution
 - d. How to calculate the Variance and Standard Deviation, Interquartile range (IQR)
- 4) Why is the normal distribution so important?
 - a. How to test if your data or features or measurement is normal: Chi-square test?
 - b. What is Box plot? What it tells you about your data?

Lecture 3

- 5) Standardization and z scores
 - a. What is data normalization, why is it important
 - b. Comparing results with different conditions
 - c. Standard Errors & Confidence Intervals
 - i. How to calculate it?
 - ii. How to interpret it?
- 6) How to compare two or more groups
 - a. Hypothesis Testing , Statistical Significance, Effect Size
 - b. What is p-value? What it tells you? What is ANOVA: One/Two-Way Analysis of Variance
 - c. What is Student-t test & Wilcoxon–Mann–Whitney test? Which test is suitable for your analysis?

Lecture 4

- 7) When to use correlation and what it tells us
 - a. What is Homoscedasticity?
 - b. What the Correlation Coefficient Does, and Does Not, Tell Us: Causality issue
 - c. When to use Pearson's, Spearman's, and Kendall's correlation
 - d. Statistically Significance of Correlations

Lecture 5

- 8) Regression: Linear & Logistic
 - a. What is Odds and odds ratio
 - b. How to calculate it and interpret it
- 9) Area under ROC curve:
 - a. What is Sensitivity, Specificity? What it represents in your experiment?
 - b. What are True positives (TP), False Positives (FP)? When does it matter?
 - c. What is positive predictive value (PPV & NPV)? How does it matter to your experiment?

Lecture 6

- 10) Reporting statistics in publications
 - a. Answering Statistics checklist for journal like "Nature, LANCET, NCB etc. "
 - i. How was sample size chosen to ensure adequate power to detect a pre-specific effect size?
 - ii. Describe it, if a method of randomization was used to determine how samples/animals were allocated to experimental groups and processed
 - iii. If the investigator was blinded to the group allocation during the experiment and/or when assessing the outcome, state the extent of blinding
 - iv. For every figure, are statistical tests justified as appropriate?
 - v. Do the data meet the assumptions of the tests (e.g., normal distribution)?

Lecture 7

- 11) Advance statistics
 - a. Data Clustering : k-means, fuzzy clustering
 - b. Principal Component Analysis (PCA)
 - c. Kruskal-Wallis test an non parametric alternative to the ANOVA
 - d. Survival analysis: Kaplan-Meier Analysis, COX proportional Hazards Model
 - e. Probabilistic modeling of your data
 - f. What is Machine learning, BIG data and Deep learning in brief?

Lecture 8

12) Hands -on activities and demonstrations: How to use *Matlab* (free license for all KU user) to process your data and apply above statistical tests

Reference:

- ❖ Ann Aschengrau, George R. Seage-***Essentials of Epidemiology in Public Health***-Jones & Bartlett Learning (2013)
- ❖ Helmut van Emden, ***Statistics for Terrified Biologists***, 1st Edition (2008)
- ❖ Timothy C. Urdan, ***Statistics in Plain English***-3rd Edition (2011)
- ❖ J. P. Marques de Sá, ***Applied Statistics Using SPSS, STATISTICA and MATLAB*** (2003)

Tutor:

Gopal Karemore

I received my PhD degree from Department of Computer Science at the University of Copenhagen in 2011. I'm currently employed at Novo Nordisk Center for Protein Research and Danish Stem Cell Center. I'm not a bio-statistician but I have some experience working on biomedical data and clinical studies especially with breast cancer screening. I have contributed my expertise in statistics for articles published in J Biomol Screen, J Cancer Epidemiology, J Menopause, Phys Med Biol, Breast Cancer Research, IEEE J Med Imaging, J Biomedical Optics, and Nat Struct Mol Biol.

After working with biologist and students from life science branch over five years, I feel it is important to organize the crash course on statistics to help them how to choose the appropriate statistical test for their experiment, then apply that test and interpret the results. I believe for biologist it is unimportant to understand the mathematical equation and hence during the entire course I would use minimal equations but rather give more focus to understand the concept. I hope that you will find these series of lectures to be a useful introduction to bio-statistics.

Organizing Team: Protein Imaging Platform & Danish Stem Cell Center

Claudia Lukas, Anne Grapin-Botton, Jutta Maria Bulkescher, and Gopal Karemore