Statistics for Life Sciences (YSC2255)

Semester 2 - AY21/22

Module coordinator: Nathan Harmston Email: nathan.harmston@yale-nus.edu.sg

This course provides students with a conceptual and practical understanding of the application of statistical methods to the Life Sciences. The topics covered include probability, common probability distributions used in modelling biological data, experimental design, linear models and hypothesis testing. This course will involve programming in R and the analysis of data.

Please note, that the order and exact content of the course may change ...

Week 1	**Data and Probability**
	Summary statistics
	The laws of probability
Week 2	**Data and Probability**
	Visualising data
	Probability distributions
Week 3	**Hypothesis testing – categorical data**
	What is a p-value? – what is a hypothesis test
	Binomial test / Fishers exact test / Chi-squared test
	Testing for the difference of two proportions (theory and simulation)
Week 4	**Hypothesis testing – numerical data**
	T-test
	Wilcoxon signed-rank
	Testing for difference of two means (theory and simulation)
	Confidence intervals
Week 5	**Most common statistical techniques are special cases of linear models**
	Introduction to linear modelling
Week 6	**Most common statistical techniques are special cases of linear models**
	ANOVA
	Likelihood tests
Week 7	**Most common statistical techniques are special cases of linear models**
	Multiple regression
	Logistic regression
Week 8	**Experimental design**
Week 10	**Survival analysis**
Week 9	**Other useful tests / techniques**
Week 10	**Multiple hypothesis testing and power calculations**
Week 11	**Bayesian Statistics** - *in a week*
Week 12	**Lies, damned lies and statistics**
Week 13	**Probabilistic modelling for Life Sciences**

Course textbook

We will using **Intuitive Biostatistics: A Nonmathematical Guide to Statistical Thinking** as a basis for a number of exercises and discussions during the course

Assessment

	N	%	Σ
Problem sets	6	5	30
Project	1	30	30
Exam	1	30	30
Participation		10	10
			100

Problem Sets

- All of the questions will involve programming in R
- Coursework to be submitted as a RMarkdown / R script and the associated output files (html)
- Collaboration on problem sets
  - Allowed, but work independently on each problem **before** discussing it
  - Write solutions on your own
  - Acknowledge sources and collaborators

Office hours

- There will be office hours ..... will be first come, first served! I may end up working with two or more of you at the same time!
- These are there to help you with material from the whole course

Absence Policy

Part of the grade for this course is based on participation. Students are expected to attend all classes, and to notify me in advance if you will be absent. It is your responsibility to get class notes from your peers and be prepared to rejoin the class after your absence. You will not be penalized for absences if you receive a Medical Certificate or AD Note.

Late Submission Policy

Deadlines are hard for problem sets and projects unless there are extenuating circumstances or you have spoken to me prior to the deadline.

Contact

Email or canvas .... Nathan or Prof. Nathan

I do not check/respond to e-mails on Saturdays - I will endeavour to respond to emails / messages on canvas within one day.