MLW / COM R & statistics short course

When: 30th November – 4th December, sessions to run daily from 9am to 1pm (except Thursday & Friday), and 2-5pm

Where: Nyika 1, Learning & Training Centre, Malawi – Liverpool – Wellcome Trust Clinical Research Programme + MS Teams

Audience: Anybody who wants to learn or refresh their basic statistics & R skills: MSc & PhD students, post-docs, …

Equipment: Please bring your own laptop, ideally with R installed (though installation will be covered in Session 1).

Cost / registration: $20 per participant for the full course; register by emailing mhenrion@mlw.mw before 22 November 2020.

Session 1: Introduction to R

Monday 30 November 2020, 9am-1pm

- Install R
- Help, documentation, tutorials
- Interfaces: R console, R GUI, R Studio, R projects and script editors (Emacs, Vim, Notepad++, …)
- Reading data, writing data
- R objects / variables
- R workspace: saving it, recovering it
- Manipulating data in R: editing, subsetting, recoding, creating variables, R objects
- R scripts, useful editors, version control

Session 2: Data visualisation in R

Monday 30 November 2020, 2pm-5pm

- Different types of data:
  nominal, ordinal, interval, ratio
- graphics & ggplot2 libraries
- Data distributions
  - Bar plots, pie charts
  - Histograms
- Covariation
  - Box & whisker, violin plots
  - Scatter plots
  - Line & time plots
  - Heat maps, contour plots
- Multi-panel figures
- A note on creating maps for visualizing spatial data

Session 3: Statistical principles & study design

Tuesday 1 December 2020, 9am-1pm

- Probability theory
- Random variables
- Discrete & continuous distributions
  - Bernoulli, binomial, Poisson, geometric
  - Normal, beta, gamma, exponential
- Simulating data (using R)
- Observational designs: cross-sectional, longitudinal, case-control
- Experimental designs: clinical trials and other intervention studies, lab & field experiments
- Power & sample size (using R)
- General design considerations
- Common & adaptive designs for clinical trials

Session 4: Parameter estimation & statistical hypothesis testing (using R)

Tuesday 1 December 2020, 2pm-5pm

- What is a statistical hypothesis test & why do it?
- Type I and II errors: significance level & power
- Estimation & confidence intervals
- Statistical hypothesis testing (parametric and non-parametric tests)
- Tests for parameter estimates
- Comparing 2 or more groups
- Comparing 2 proportions
- A note on p-values, null hypothesis significance testing and the reproducibility crisis.

Session 5: Basic regression modelling I (using R)

Wednesday 2 December 2020, 9am-1pm

- Linear model & correlation
- Relation to ANOVA
- Generalised linear model
  - Special cases: logistic, Poisson & negative binomial regression
- Estimation, model fitting
  - Maximum likelihood, least squares, optimisation techniques (brief overview only)
  - Model diagnostics
  - Note on Bayesian modelling
- Statistical tests:
  - Wald, Likelihood ratio & score tests
  - Omnibus vs. single coefficient tests
- Interaction terms and interpretation

Session 6: Basic regression modelling II (using R)

Wednesday 2 December 2020, 2pm-5pm

- Variable selection
- Model selection
- Regression and smoothing splines, Generalised Additive Models
- Missing values & multiple imputation
- Correlated data (introduction only)
  - Repeated measurements & matched data
  - Longitudinal data
  - Hierarchical & clustered data
  - Time series
  - Spatial data
- Methods for correlated data (overview only)
  - Paired tests
  - Robust estimation methods
  - Generalised Estimating Equations
  - (Generalised) Linear Mixed Models; fixed vs random factors

Session 7: Reproducible Research

Thursday 3 December 2020, 2pm-5pm

- R markdown
- GitHub
- R packages
- Licenses for (i) software, (ii) research outputs

Session 8: Ask the experts!

Friday 4 December 2020, 2pm-5pm

Drop-in session to get help & advice for specific projects.