LAB 04

PH142 Fall 2025

Announcements

- **Lab04:** due 9/19 at 11:59pm
- **Quiz03:** due 9/19 at 11:59pm
- **Data Project Part I:** due 9/26 at 11:59pm

*Your group must meet with your assigned GSI before the due date.

Next week in lab: Midterm I Review Session

Week 4 Lecture Review

Exposure group	Disease	No disease	Row total
Exposed	A	B	A+B
Not Exposed	C	D	C+D
Column total	A+C	B+D	A+B+C+D

Two-Way Tables

Marginal Distribution: The distribution of a single categorical variable in the entire population

- Use the totals in the margins of the table to calculate proportions
- Percent of Exposed individuals= (A+B)/(A+B+C+D)
- Percent of Unexposed individuals = (C+D) / (A+B+C+D)

Conditional Distribution: The distribution of one categorical variable within the other

- Use a single row or column to calculate proportions
- Percent of Diseased given exposure = A / (A+B)
- Percent of Exposed given disease = A / (A+C)

Week 4 Lecture Review

Sensitivity, Specificity, PPV, NPV

	Disorder	No Disorder
Positive Test Result	True Positive (TP)	False Positive (FP)
Negative Test Result	False Negative (FN)	True Negative (TN)

Sensitivity = TP/(TP+FN)
Specificity = TN/(TN+FP)
PPV = TP/(TP+FP)
NPV = TN/(FN+TN)

Week 4 Lecture Review

Observation vs Experimentation

A study is observational if the researcher **observes** what happens and does not control who is treated or exposed.

- Does not control for confounding

A study is experimental if the investigator is **experimenting** (or intervening) by controlling who is treated or exposed.

- In an experimental study, the exposure is assigned by a randomization mechanism that is controlled by the investigator

Week 4 Lecture Review

Internal and External Validity

Internal Validity

- The extent to which the observed results represent the truth in the population we are studying and, thus, are not due to methodological errors

External Validity

- The extent to which you can generalize the findings of the study to other situations, people, settings, and measures
- Representativeness/generalizability

Week 4 Lecture Review

Population and Sampling

Population of Interest

- Target Population - entire group of individuals about which we want estimates to apply *problem in PPDAC*
- Study Population - part of the population which we can select individuals & collect information to draw conclusions about the entire population
- Study Sample - individuals who have been sampled from the study population; the group that you collect data from

Week 4 Lecture Review

Simple Random Sampling

Simple Random Sample (SRS): A sample chosen by chance, where each individual in the dataset has an equal chance of being selected

Functions in R:

- slice_sample(n = 100), selects n rows at random
- slice_sample(prop = 0.05), selects a random proportion of rows
- set.seed(#), makes results reproducible by taking the same sample

Example: CS_100 <- CS_data %>% slice_sample(n = 100)

LAB 04 Walkthrough

Lab Submission

- Follow the directions on the LAB04 file
- Submit using the **Terminal Tab** (next to the console in the bottom left pane)
  - Copy and paste the given line into the terminal
  - Follow prompts (NOTE: the terminal will **not** show your password being typed out!)
- **CHECK IN GRADESCOPE THAT ALL YOUR TESTS PASSED**