

Computer lab 6: The Relationship Between Correlation and Regression

C91AR: Advanced Statistics using R

Dr Peter E McKenna

2025-03-05

Contents

1	Packages for today	1
2	Dataset	1
3	Session outline	2
4	Group work suggestions	2

1 Packages for today

```
# load packages
pacman::p_load(tidyverse,
               corrr,
               psych,
               tidyplots,
               HistData) # new data package
```

2 Dataset

Galton's data on the heights of parents and their children, by child

This data set lists the individual observations for 934 children in 205 families on which Galton (1886) based his cross-tabulation.

I have wrangled two datasets for you to choose from today

- “father_son_height.csv”
- “mother_daughter_height.csv”

Both of these are on Canvas for download.

You are free to select either for the exercise today.

3 Session outline

- Your job is work in groups to follow the steps outlined on Monday’s lecture to
 - visualise each vectors distribution
 - visualise their relationship with one another
 - gather the relevant statistics to simulate data based on the source data
 - simulate 1000 new observations
 - plot the real against the simulated data

4 Group work suggestions

- Work with someone new today!
- Use the terminal to display Monday’s slides
- Delegate tasks as part of a team: this’ll help to speed things up.
- As always, annotate and comment your RMarkdown scripts
- Strive to keep the code tidy
- Good luck!