

Data Science in Medicine: Tutorial 2

Summarising and visualising data – Part 2

Semester 1, 2020-2021

- Please attempt all questions on this worksheet in advance of the tutorial, and bring with you all work, including printouts of code and other results. Tutorials cannot function properly unless you do the work in advance.
- You are welcome to bring along any questions you may have from the lectures, textbook, etc.
- Assessment is formative, meaning that tutorials do not contribute to your final grade.
- Attendance is compulsory. If you have good reasons to miss a session, you should contact your year coordinator in advance to arrange to attend a different session.

Introduction

In this tutorial you are given a simple (entirely fictitious) dataset¹ from a clinical trial and you are asked to summarise it with the use of appropriate summary statistics and visualisations. You can calculate the statistics and draw the visualisations by hand, or use R by following what you've learnt in Lab 1. Note, however, that the focus of this tutorial is not on R programming but rather on the concepts covered in the lectures.

The dataset involves a clinical trial for testing a new antidepressant drug called Joyzepam. The researchers are interested in the effectiveness of Joyzepam on patients' mood and sleeping patterns. The study involves three separate drugs to be administered: a placebo, an existing antidepressant / anti-anxiety drug called Anxifree, and the new drug Joyzepam. A collection of 18 participants with moderate to severe depression are recruited for the initial testing. Half of them are undergoing cognitive behavioural therapy (CBT) and the other half are not. Participants are randomly assigned a treatment, such that there are 3 CBT people and 3 no-therapy people assigned to each of the 3 drugs. A psychologist assesses the mood of each person after a 3 month run with each drug, and the overall improvement in each person's mood is assessed on a scale ranging from -5 to +5. Over the course of the 3 months, participants keep track of the number of minutes they sleep every night, which is averaged at the end of the 3-month period. This is compared to

¹ The dataset is an extension of the clinicaltrial.Rdata data file provided by Danielle Navarro, University of Adelaide, as part of the "Learning Statistics with R: A tutorial for psychology students and other beginners" textbook.

equivalent data before administering the drug, and the average increase in each person's sleeping is recorded (in minutes).

You can obtain the dataset `tut2_clinicalTrial.csv` from the course website. The dataset is also provided in the following table:

drug	therapy	mood.gain	sleep.gain
placebo	no.therapy	0.5	34
placebo	no.therapy	0.3	6
placebo	no.therapy	0.1	20
anxifree	no.therapy	0.6	25
anxifree	no.therapy	0.4	66
anxifree	no.therapy	0.2	24
joyzepam	no.therapy	1.4	50
joyzepam	no.therapy	1.7	47
joyzepam	no.therapy	1.3	23
placebo	CBT	0.6	8
placebo	CBT	0.9	12
placebo	CBT	0.3	5
anxifree	CBT	1.1	47
anxifree	CBT	0.8	40
anxifree	CBT	1.2	33
joyzepam	CBT	1.8	58
joyzepam	CBT	1.3	57
joyzepam	CBT	1.4	26

Part 1: Summarising a clinical trial dataset

- (1) How would you characterise each of the 4 variables? Categorical, ordinal, interval or ratio?
- (2) Calculate appropriate measures of central tendency for a variable of your choice in the dataset. How would you interpret the results?
- (3) Calculate appropriate measures of dispersion for a variable of your choice in the dataset. How would you interpret the results?

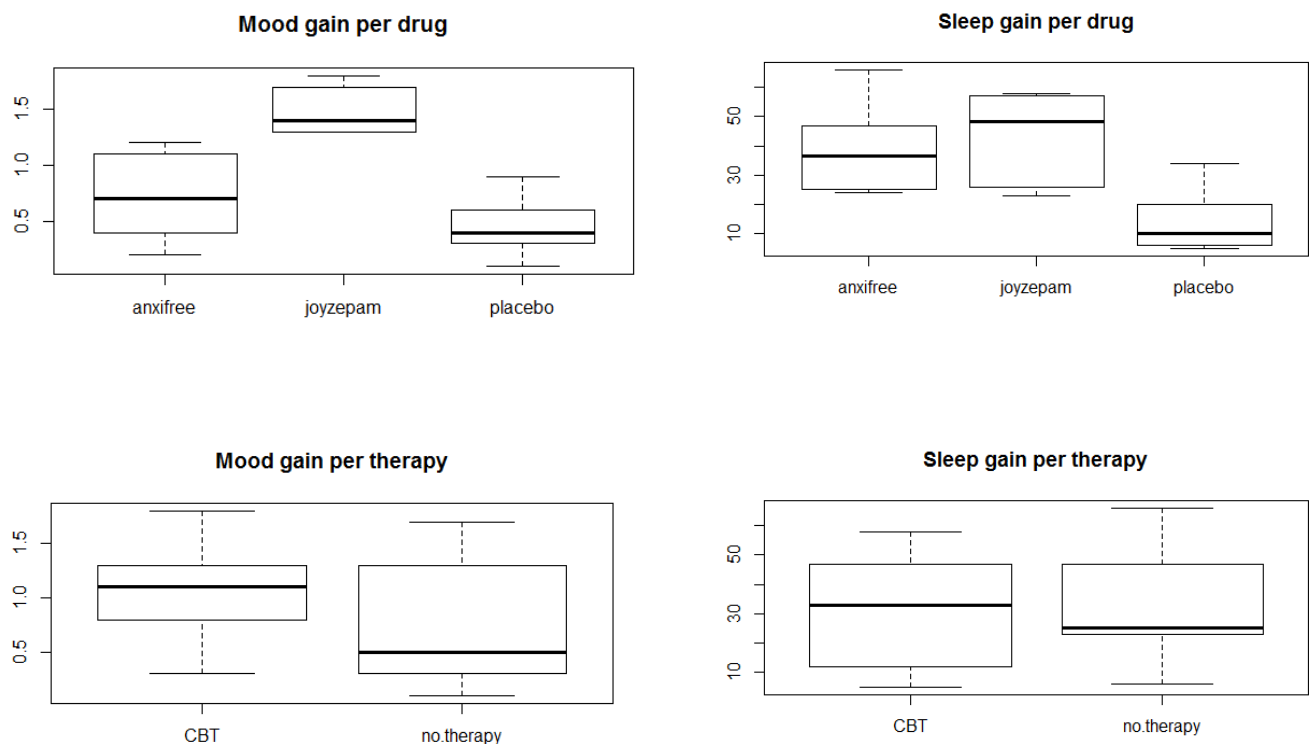
- (4) We have calculated the mean and standard deviation of mood gain and sleep gain separately for the placebo, the anxifree and the joyzepam groups. How would you interpret the results in the following table?

	Mood gain			Sleep gain		
	<i>anxifree</i>	<i>joyzepam</i>	<i>placebo</i>	<i>anxifree</i>	<i>joyzepam</i>	<i>placebo</i>
Mean	0.7167	1.483	0.450	39.17	43.50	14.17
Standard deviation	0.3920034	0.2136976	0.2810694	15.81666	15.31992	11.14301

Part 2: Visualising a clinical trial dataset

- (1) Produce appropriate visualisations for a quantitative variable of your choice in the dataset. What do these tell us about the clinical trial?

- (2) We have produced the following boxplots per group, i.e. per drug group and per therapy group. What do these tell us about the clinical trial?



- (3) Produce appropriate visualisations for a qualitative variable of your choice in the dataset. What do these tell us about the clinical trial?