Problem set 12

S520

Upload your answers through the Assignments tab on Canvas by 11:59 pm, Thursday 7th December. Draw graphs in R and include code.

Draw all graphs in R and include all R code. You may work with others, but you must write up your homework independently — you should not have whole sentences in common with other students or other sources.

1. (5 points.) Trosset chapter 15.7 exercise 5, parts (a), (b), and (c).

2. (10 points. From the Takehome final, Summer 2016.) The file `examanxiety.txt` on Canvas contains information on a number of variables measured on a sample of 103 students taking a math exam:

   - `Code`: a label for the individual in the sample (not scientifically interesting.)
   - `Revise`: hours spent revising for the math exam.
   - `Exam`: score on a math exam on a scale from 0 to 100.
   - `Anxiety`: "math anxiety" on a scale from 0 to 100 (100 is most anxious.)
   - `Gender`: female or male.

   Assume the data is a random sample from a larger population of students.

   (a) Is there a significant difference between average anxiety for the population of male students and the population of female students? Perform an appropriate significance test, stating hypotheses, a P-value, and a substantive conclusion.

   (b) Let anxiety be your x-variable and exam score be your y-variable. Find the regression line to predict exam score from anxiety. Carefully explain (in words or using math) what your regression line means — do not just paste R output.

   (c) Let anxiety be your x-variable and exam score be your y-variable. Which of the following regression assumptions are met? Make arguments and/or show graphs to support your answers.

      i. Linearity
      ii. Independence
      iii. Equal variance of errors (homoskedasticity)
      iv. Normality of errors

3. (5 points.) We wish to study the effects of four different rat feeds (fruit, carbs, meat, and mixed) on rat weight gain. 140 rats, each kept in a separate cage, are randomly split into four equal groups of 35. Each group is put on a different feed for a month. The amount of weight gained during the month is measured for each rat. The sample means and sample standard deviations of the results (all in grams) are:

   - Fruit: mean 83.5, standard deviation 16.9
   - Carbs: mean 92.3, standard deviation 14.6
   - Meat: mean 88.6, standard deviation 14.2
   - Mixed: mean 99.4, standard deviation 14.1

   Figure 1 (on the next page) shows normal QQ plots of the data.

   (a) What are the assumptions of the analysis of variance F-test? Are the data here approximately consistent with these assumptions?

   (b) Complete the analysis of variance table below.

Variation	Sum of squares	DF	Mean square	F	P-value
Between					
Within					
Total					

Table 1: ANOVA table for weight gain of four groups of rats.

   (c) What can you conclude from the analysis of variance? (Give a substantive conclusion, not just "reject" or "don't reject."

4. (10 points. Adapted from the S520 Takehome final, Fall 2019.) To study the relationship between video games and empathy, researchers performed a randomized experiment on 155 Italian high school students.[1]

   Each participant played a randomly selected game of one of three types:

   - "Neutral games" with no violent or sexual content (*Dream Pinball 3D* or *Q.U.B.E. 2.*)
   - Games from the *Half-Life* series: The researchers considered these games violent but not sexist.
   - Games from the *Grand Theft Auto* (GTA) series: The researchers considered these games violent and sexist, and the player's characters in these games to be misogynistic (woman-hating.)

   After playing the game, the participants were shown a photo of a victim of violence and asked a series of questions. Their answers were turned into an "empathy score" on a scale from 1 to 7.

[1] `http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0152121`. It's worth reading this comment by a statistician: `http://www.statschat.org.nz/2016/04/17/evil-within/`

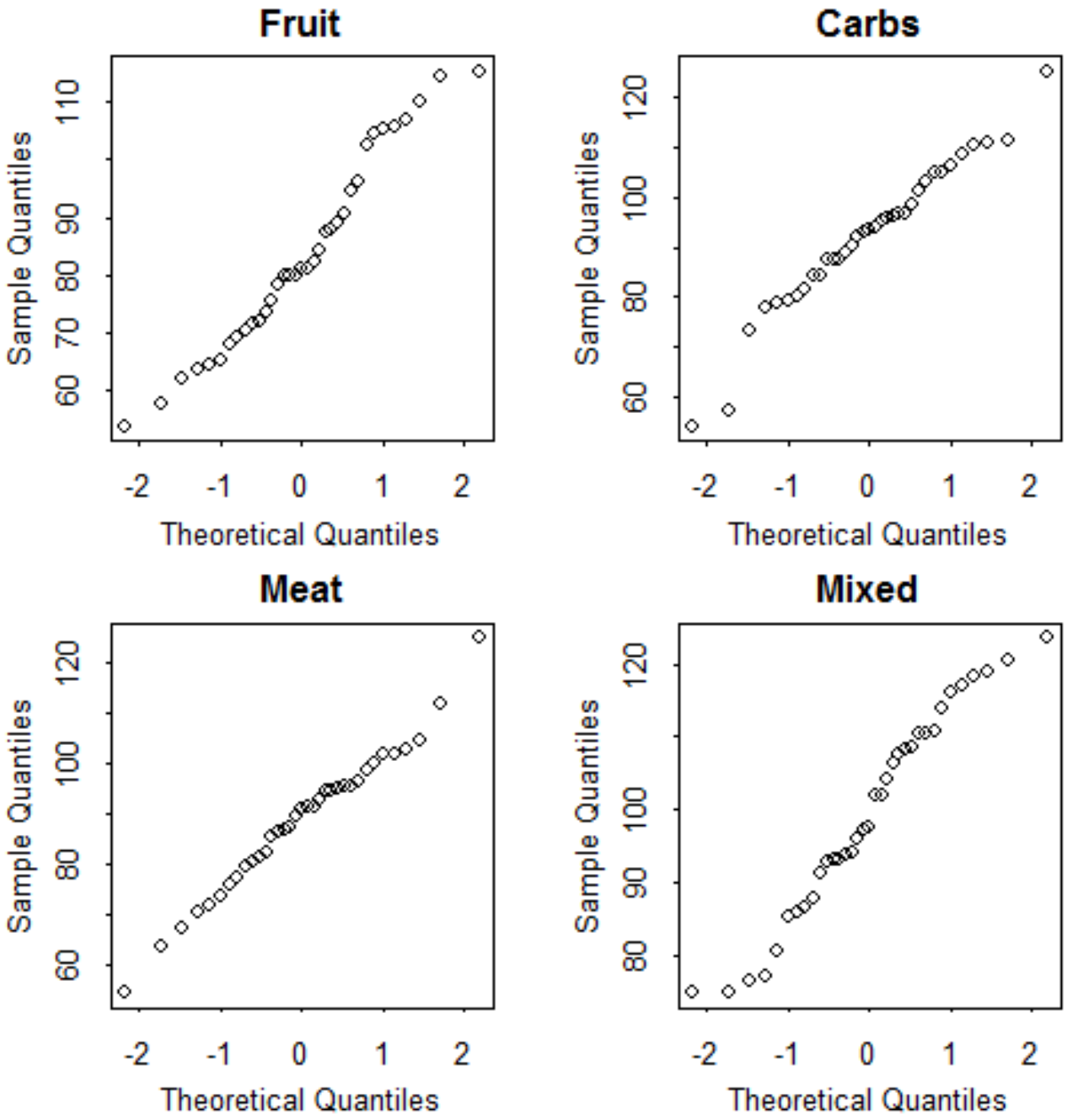

Figure 1: Normal quantile plots for four groups of rats.

In addition, the participants were asked questions about whether they identified (that is, related to) the character they played in the game. Their answers were turned into an "identification score" on a scale from 1 to 7.

The file `GameEmpathy.txt` (in the Data folder on Canvas) contains observations on 153 of the participants in the experiment. (Two observations are omitted because of missing data.) The variables:

- `sex`: Male or female.
- `game.type`: Neutral, Half-Life, or GTA.
- `identify`: A number on a scale from 1 to 7, with 1 meaning the least identification with the character they played, and 7 means the most identification.
- `empathy`: A number on a scale from 1 to 7, with 1 meaning the least empathy and 7 meaning the most empathy.

Read the data into R, e.g. by using `read.table()`.

(a) Do the different types of game lead to (population) differences in average empathy? Draw graphs, perform an ANOVA, and state your conclusion. Note: The samples aren't quite normal, but the samples are large enough that this shouldn't be a problem.

(b) Is there a relationship between identification and empathy for:

i. Students who played neutral games?

ii. Students who played *Half-Life*?

iii. Students who played *GTA*?

Draw graphs (or do calculations), and state your conclusions, remembering to adjust for multiple testing. Hint: If your data set is called `GameEmpathy`, you can pick out the data for individuals who played GTA with

```
GTA.players <- subset(GameEmpathy, game.type == "GTA")
```