

Exercise 3: Data Munging, Data Cleaning, Rankings & Scores

Submission Deadline: November 17 2025, 07:00 UTC

University of Oldenburg

Winter 2025/2026

Instructors: [Jannik Schröder](#), [Wolfram "Wolle" Wingerath](#)

Submitted by: <your names here>

Part 1: Data Munging & Data Cleaning / 40

1.) Find a source for energy prices over time (e.g. electric energy).

a) How would you assess the quality of the data set you found? Do you need to do any preparation before doing the analysis? (If so, what exactly did you do?)

Solution:

XXX

b) Analyze this data and make a projection about energy costs five years from now.

Solution:

XXX

c) What will energy prices be in 25 or 50 years?

Solution:

XXX

2.) What types of outliers might you expect to occur in the following data sets?

a) Student grades

Solution:

XXX

b) Salary data

Solution:

XXX

c) Lifespans in Wikipedia

Solution:

XXX

Part 2: Scores & Rankings

/ 60

3.) Let X represent a random variable drawn from the normal distribution defined by $\mu = 2$ and $\sigma = 3$. Suppose we observe $X = 5.08$.

Find the Z-score of x , and determine how many standard deviations away from the mean that x is.

Solution:

XXX

4.) What percentage of the standard normal distribution ($\mu = 0$, $\sigma = 1$) is found in each region?

a) $Z > 1.13$

Solution:

XXX

b) $Z < 0.18$

Solution:

XXX

c) $Z > 8$

Solution:

XXX

d) $|Z| < 0.5$

Solution:

XXX

5.) Amanda took the Graduate Record Examination (GRE), and scored 160 in verbal reasoning and 157 in quantitative reasoning. The mean score for verbal reasoning was 151 with a standard deviation of 7, compared with mean $\mu = 153$ and $\sigma = 7.67$ for quantitative reasoning. Assume that both distributions are normal.

a) What were Amanda's Z-scores on these exam sections? Mark these scores on a standard normal distribution curve.

Solution:

XXX

b) Which section did she do better on, relative to other students?

Solution:

XXX

c) Find her percentile scores for the two exams.

Solution:

XXX

6.) Identify three successful and well-used scoring functions in areas of personal interest to you. For each, explain what makes it a good scoring function and how

it is used to create rankings in that domain.

Solution:

XXX

Finally: Submission

Save your notebook and submit it (as both **notebook and PDF file**). And please don't forget to ...

- ... choose a **file name** according to convention (see Exercise Sheet 1) and to
- ... include the **execution output** in your submission!