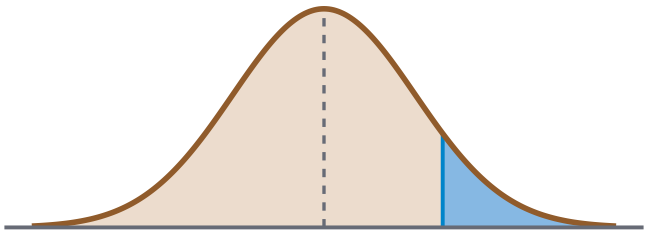
PROBABILITY AND STATISTICS

WORKBOOK

12 chapters · 170 problems · step-by-step solutions



Ertan Sinan Şahin

ertansinansahin.com

SAMPLE PAGES · Full workbook: ertansinansahin.com/product/olasilik-ve-istatistik-calisma-kitabi

How to use this book

- Each chapter opens with a reminder box summarizing the definitions and formulas the problems rely on.
- Difficulty: ★☆☆ basic ★★☆ intermediate ★★★ advanced. Each chapter has a few multiple-choice concept questions; its last problem is an exam-style mixed problem combining several topics.
- The dotted area below each problem is for your own work. Solve first; if stuck, go back to the reminder box and the lecture notes, and look at the solution last.
- All solutions are at the end of the book. The Solution → p. link next to a problem takes you to its solution, and ← Problem p. takes you back. Short answers are in the **ANSWER** box.
- **Tables:** Tables of the standard normal, t, chi-square and F distributions are at the end of the book; the solutions state the value read from the table and how it is rounded.
- **Midterm and final:** At most universities the midterm covers probability, random variables and distributions (Chapters 1–8), usually ending with the normal distribution. In courses with two midterms, the second covers distributions, sampling and estimation (Chapters 7–10). Confidence intervals, hypothesis testing and regression (Chapters 9–12) are mostly left to the final. Choose the chapters to study according to your own course syllabus.

Contents

1 Descriptive Statistics ..... 12 problems 4

2 Counting and Probability ..... 14 problems 11

3 Conditional Probability and Bayes' Rule ..... 16 problems 18

4 Random Variables ..... 14 problems 27

5 Expected Value and Variance ..... 14 problems 35

6 Joint Distributions ..... 14 problems 42

7 Discrete Distributions ..... 16 problems 50

8 Continuous Distributions ..... 16 problems 59

9 Sampling Distributions ..... 12 problems 67

10 Estimation and Confidence Intervals ..... 16 problems 74

11 Hypothesis Testing ..... 16 problems 83

12 Regression and Correlation ..... 10 problems 91

Solutions 98

Tables 157

SAMPLE PAGES · Full workbook: ertansinansahin.com/product/olasilik-ve-istatistik-calisma-kitabi

CHAPTER 1

DESCRIPTIVE STATISTICS

mean	$\bar{x} = \frac{1}{n}\sum x_i$	percentile position	$(n+1)p$
variance	$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$	interquartile range	$\mathrm{IQR} = Q_3 - Q_1$
shortcut	$s^2 = \frac{\sum x_i^2 - (\sum x_i)^2/n}{n-1}$	outlier fences	$Q_1 - 1.5\,\mathrm{IQR},\ Q_3 + 1.5\,\mathrm{IQR}$
coefficient of variation (%)	$\mathrm{CV} = 100 \cdot \frac{s}{\bar{x}}$	standardized value	$z = \frac{x - \bar{x}}{s}$
transformation $y = ax + b$	$\bar{y} = a\bar{x} + b,\ s_y = \lvert a \rvert s_x$	grouped mean	$\bar{x} \approx \frac{\sum f_k m_k}{n}$

The median is the middle value of the sorted data; the mode is the most frequent value. The sample variance has $n-1$ in the denominator. The position of a percentile in the ordered data is found from the position formula in the box; if the position is not an integer, linear interpolation between the two neighboring values is used. The first and third quartiles are the 25th and 75th percentiles; software may use other rules. Results are given to three significant figures.

Measures of central tendency

1.1 ★☆☆ Solution → p. 98

A data set has mean 40, median 36 and standard deviation 8. Every value is multiplied by 1.5 and then 4 is subtracted from each. Which of the following is true for the new data set?

A) The mean is 56 and the standard deviation is 12.

B) The difference between the mean and the median is still 4.

C) The median is 54.

D) The variance is 96.

E) The standard deviation is 8.

SAMPLE PAGES · Full workbook: ertansinansahin.com/product/olasilik-ve-istatistik-calisma-kitabi

1.2 ★☆☆ Solution → p. 98

The numbers of bugs closed by a software team in 9 consecutive sprints are:

$$14,\ 9,\ 17,\ 11,\ 9,\ 22,\ 13,\ 9,\ 16$$

a. Find the mean, the median and the mode.

b. Find the trimmed mean obtained by dropping the single smallest and the single largest value.

c. If 40 had been written by mistake instead of 22, which of the measures in parts a and b would change? Find the new values.

Measures of spread

1.3 ★☆☆ Solution → p. 98

The numbers of cars entering the car park of a shopping mall between 8 and 9 a.m. on 8 days are:

$$23,\ 31,\ 27,\ 19,\ 35,\ 28,\ 24,\ 30$$

a. Find the mean, the variance and the standard deviation using the sums $\sum x$ and $\sum x^2$.

b. Find the range.

c. Find the days on which the number of cars is more than one standard deviation from the mean.

SAMPLE PAGES · Full workbook: ertansinansahin.com/product/olasilik-ve-istatistik-calisma-kitabi

SOLUTIONS

1 DESCRIPTIVE STATISTICS

1.1

← Problem p. 4

The transformation is $y = 1.5x - 4$. Measures of location (mean, median) are affected by the whole transformation, while measures of spread are affected only by the multiplier:

$$\bar{y} = 1.5(40) - 4 = 56, \qquad s_y = 1.5(8) = 12.$$

A is correct.

- (B): The new median is $1.5(36) - 4 = 50$, so the difference is $56 - 50 = 6$. The multiplier also scales the difference by 1.5.
- (C): $54 = 1.5(36)$; the 4 was not subtracted. The correct value is 50.
- (D): The variance changes by the square of the multiplier: $1.5^2(64) = 144$. The value 96 comes from forgetting to square the multiplier.
- (E): Subtracting the same number from all values does not change the spread; the computation $1.5(8) - 4 = 8$ is wrong.

ANSWER A

1.2

← Problem p. 5

Sorted data: 9, 9, 9, 11, 13, 14, 16, 17, 22; total 120.

a. Mean, median ($n = 9$, 5th value) and mode:

$$\bar{x} = \frac{120}{9} = 13.3, \qquad \tilde{x} = 13.$$

9 occurs three times; the mode is 9.

b. Drop 9 and 22; the remaining 7 values add up to 89:

$$\bar{x}_{\text{tr}} = \frac{89}{7} = 12.7.$$

c. The total increases by 18; the mean changes:

$$\bar{x} = \frac{138}{9} = 15.3.$$

The median (13), the mode (9) and the trimmed mean (12.7) do not change: whatever the largest value is, it keeps its place in the ordering and is dropped by the trimming. The mean is sensitive to extreme values; the median and the trimmed mean are resistant.

ANSWER a. 13.3; 13; 9 b. 12.7 c. only the mean: 15.3

1.3

← Problem p. 5

a. Sums:

$$\sum x = 217, \qquad \sum x^2 = 6065.$$

$$\bar{x} = \frac{217}{8} = 27.125 \approx 27.1.$$

By the shortcut formula:

$$\begin{aligned} s^2 &= \frac{6065 - 217^2/8}{7} \\ &= \frac{6065 - 5886.125}{7} = \frac{178.875}{7} \\ &= 25.554 \approx 25.6, \end{aligned}$$

$$s = \sqrt{25.554} = 5.055 \approx 5.06.$$

b. Difference between the largest and the smallest value:

$$R = 35 - 19 = 16.$$

c. Values within at most one standard deviation of the mean lie in the interval:

$$\bar{x} \pm s = 27.125 \pm 5.055 = (22.07,\ 32.18).$$

The values 19 and 35 lie outside the interval; the distances of these two days from the mean (8.125 and 7.875) exceed one standard deviation.

ANSWER a. 27.1; $s^2 = 25.6$; $s = 5.06$ b. 16 c. 19 and 35

ertansinansahin.com

SAMPLE PAGES · Full workbook: ertansinansahin.com/product/olasilik-ve-istatistik-calisma-kitabi