Session 07-01 - Descriptive Statistics Essentials

Section 07: Probability & Statistics

Dr. Nikolai Heinrichs & Dr. Tobias Vlćek

Entry Quiz - 10 Minutes

Quick Review from Section 06

Test your understanding of Integration

1. Find $\int x \cdot e^x \, dx$ using integration by parts.
2. Evaluate $\int_0^1 (2x + 1) \, dx$
3. A company's marginal profit is $MP(x) = 60 - 2x$. Find the profit function if $P(0) = -100$.
4. Find the area between $y = x$ and $y = x^2$ from $x = 0$ to $x = 1$.

Welcome to Probability & Statistics!

New Section Overview

Section 07 covers essential exam topics:

- Session 07-01: Descriptive Statistics (today)
- Session 07-02: Basic Probability Concepts
- Session 07-03: Combinatorics & Counting
- Session 07-04: Conditional Probability
- Session 07-05: Bayes' Theorem
- Session 07-06: Contingency Tables
- Session 07-07: Binomial Distribution
- Session 07-08: Mock Exam 2

...

> ! Important
>
> Probability accounts for approximately 25% of the Feststellungsprüfung!

Learning Objectives

What You'll Master Today

- Calculate measures of central tendency: mean, median, mode
- Compute measures of spread: range, variance, standard deviation
- Interpret data distributions using histograms and box plots
- Work with frequency distributions and relative frequencies
- Apply statistical concepts to business scenarios

...

> **i** Note
>
> This is foundational material - brief coverage to prepare for probability!

Part A: Measures of Central Tendency

The Three Averages

How do we summarize a data set with a single number?

...

> **!** Three Measures of Center
>
> 1. Mean (Mittelwert): $\bar{x} = \frac{\sum x_i}{n}$
> 2. Median (Zentralwert): Middle value when data is sorted
> 3. Mode (Modalwert): Most frequently occurring value

Example: Sales Data

Monthly sales (in thousands €) for a store:

$$12, 15, 14, 18, 15, 22, 15, 16, 14, 19$$

...

Mean:

$$\bar{x} = \frac{12 + 15 + 14 + 18 + 15 + 22 + 15 + 16 + 14 + 19}{10} = \frac{160}{10} = 16$$

...

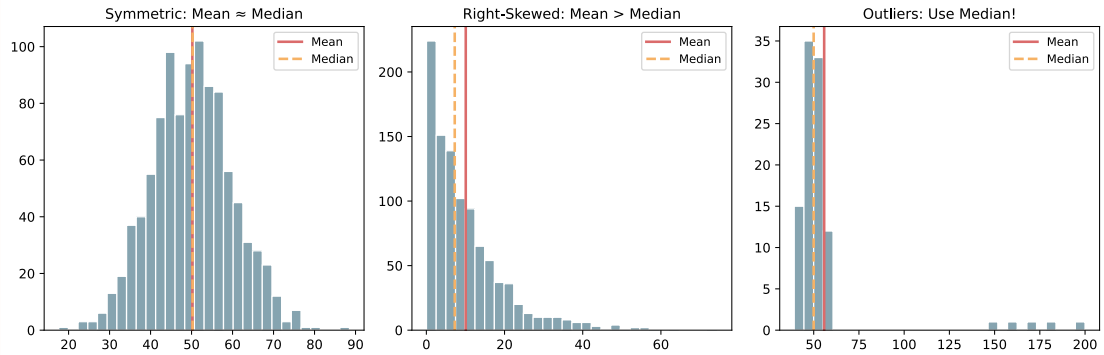
Median: Sort: $12, 14, 14, 15, 15, 15, 16, 18, 19, 22$

Middle values: $\frac{15+15}{2} = 15$

...

Mode: 15 (appears 3 times)

When to Use Each Measure



. . .

> **Tip**
>
> - Mean: Best for symmetric data without outliers
> - Median: Best for skewed data or data with outliers
> - Mode: Best for categorical data

Part B: Measures of Spread

How Spread Out Is the Data?

Two datasets can have the same mean but different spreads:

. . .

Dataset A: $48, 49, 50, 51, 52$ (mean = 50)

Dataset B: $10, 30, 50, 70, 90$ (mean = 50)

. . .

> **Note**
>
> We need measures to quantify this difference!

Range

Simplest measure of spread:

$$\text{Range} = \text{Maximum} - \text{Minimum}$$

. . .

Dataset A: Range $= 52 - 48 = 4$

Dataset B: Range $= 90 - 10 = 80$

...

> ⚠ Warning
>
> Range only uses two values - sensitive to outliers!

Variance and Standard Deviation

> ! Variance (Varianz)
>
> Population variance:
>
> $$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$$
>
> Sample variance:
>
> $$s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}$$

...

> ! Standard Deviation (Standardabweichung)
>
> $$\sigma = \sqrt{\sigma^2} \quad \text{or} \quad s = \sqrt{s^2}$$

Calculation Example

Data: $4, 8, 6, 5, 3, 2, 8, 9, 2, 5$ (n = 10)

...

Step 1: Calculate mean

$$\bar{x} = \frac{4 + 8 + 6 + 5 + 3 + 2 + 8 + 9 + 2 + 5}{10} = \frac{52}{10} = 5.2$$

...

Step 2: Calculate deviations squared

$$(4 - 5.2)^2 + (8 - 5.2)^2 + \ldots = 1.44 + 7.84 + 0.64 + 0.04 + 4.84 + 10.24 + 7.84 + 14.44 + 10.24 + 0.04 = 57.6$$

...

Step 3: Variance and SD

$$s^2 = \frac{57.6}{9} = 6.4 \quad \Rightarrow \quad s = \sqrt{6.4} \approx 2.53$$

Part C: Frequency Distributions

Organizing Data

Raw data: Test scores of 30 students

$$65, 72, 78, 81, 65, 73, 85, 92, 78, 72, 65, 88, 91, 73, 78, 82, 76, 72, 85, 78, 65, 73, 82, 79, 88, 73, 78, 85, 92, 78$$

...

Question: How can we summarize this data effectively?

Frequency Table

Score Range	Frequency	Relative Frequency
60-69	4	4/30 = 13.3%
70-79	14	14/30 = 46.7%
80-89	9	9/30 = 30.0%
90-99	3	3/30 = 10.0%
Total	30	100%

...

> **Tip**
>
> Relative frequency = Frequency / Total = Probability interpretation!

Histogram Visualization

Break - 10 Minutes

Part D: Box Plots (Five-Number Summary)

The Five-Number Summary

> ! Five-Number Summary
>
> 1. Minimum (Min)
> 2. First Quartile (Q1) - 25th percentile
> 3. Median (Q2) - 50th percentile
> 4. Third Quartile (Q3) - 75th percentile
> 5. Maximum (Max)

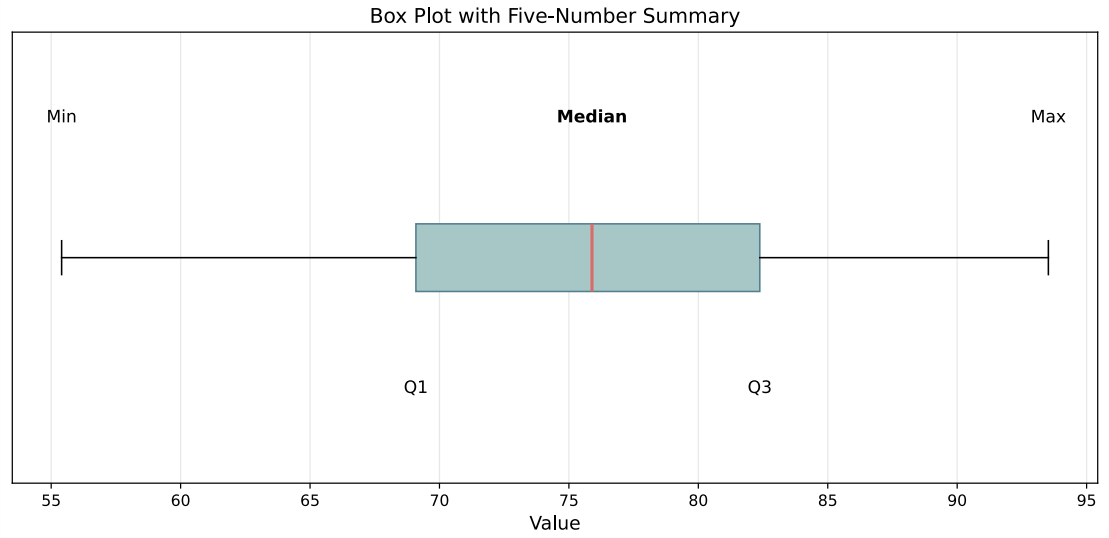
...

Interquartile Range (IQR): $\text{IQR} = Q3 - Q1$

...

> Tip
>
> IQR contains the middle 50% of the data!

Box Plot Visualization



Detecting Outliers

Outliers are values that fall outside:

...

$$\text{Lower fence: } Q1 - 1.5 \times \text{IQR}$$

$$\text{Upper fence: } Q3 + 1.5 \times \text{IQR}$$

...

Example: If $Q1 = 65$, $Q3 = 85$, then IQR $= 20$

- Lower fence: $65 - 1.5(20) = 35$
- Upper fence: $85 + 1.5(20) = 115$

...

Any value below 35 or above 115 would be an outlier.

Part E: Business Applications

Quality Control Example

A factory measures the diameter of manufactured bolts (in mm):

$$10.2, 10.1, 10.0, 10.3, 9.9, 10.1, 10.0, 10.2, 10.1, 10.0$$

Target: 10.0 mm with tolerance ±0.3 mm

...

Calculate:

- Mean: $\bar{x} = 10.09$ mm
- Standard deviation: $s = 0.12$ mm

...

> **Note**
>
> If we assume normal distribution, approximately 99.7% of bolts will be within $\bar{x} \pm 3s = 10.09 \pm 0.36$ mm, which is within tolerance!

Sales Analysis Example

Weekly sales data for 8 weeks (in €1000):

$$45, 52, 48, 55, 62, 50, 48, 56$$

...

Measure	Value	Interpretation
Mean	€52,000	Average weekly sales
Median	€51,000	Typical week
Std Dev	€5,300	Sales variability
Range	€17,000	Max spread

Guided Practice - 15 Minutes

Practice Problems

Work in pairs

Problem 1: Customer wait times (minutes): $3, 5, 2, 8, 4, 6, 3, 7, 2, 10$

a) Calculate mean, median, and mode
b) Calculate variance and standard deviation
c) Is the mean or median a better measure of center? Why?

Problem 2: Create a frequency table for exam scores: $75, 82, 91, 78, 85, 68, 73, 88, 95, 79, 82, 76, 84, 90, 77$

Connection to Probability

From Statistics to Probability

Key connection:

. . .

$$\text{Relative Frequency} \approx \text{Probability}$$

. . .

Example: If 30% of customers wait more than 5 minutes, then the probability that a randomly selected customer waits more than 5 minutes is approximately 0.30.

. . .

> **! Important**
>
> This is the frequentist interpretation of probability - probability equals long-run relative frequency!

Wrap-Up & Key Takeaways

Today's Essential Concepts

- Mean, median, mode measure center differently
- Variance and standard deviation measure spread
- Box plots show distribution shape and outliers
- Relative frequency connects to probability
- Choose the right measure based on data characteristics

. . .

Coming Next

Session 07-02: Basic Probability Concepts - sample spaces, events, and probability rules!

Homework Assignment

Tasks 07-01

- Calculate descriptive statistics for business datasets
- Interpret measures in context
- Create and interpret frequency distributions
- Prepare for probability concepts

. . .

Note

This material is foundational - make sure you're comfortable before moving to probability!