

Statistics

| When Technology Computes, You Interpret

Prof. Dr. Mostafa Elhosseini

<https://youtube.com/drmelhosseini>

| Today's Focus

- Why Use Technology in Statistics?
- Why do you still have to learn about statistical methods?
- Understanding Data Files & Structure
- Variables in the Dataset
- Databases in Statistical Analysis
- Web Apps in Statistics
- Business Case: Digital Media Store & Online Advertising

◆ Why Use Technology in Statistics?

- Automates calculations and graphing
- Enables handling of large datasets
- Speeds up analysis and visualization
- Makes statistics accessible to non-experts



Common Tools Used

Tool	Use Case
TI Calculators (TI-83/84/89/Inspire)	Quick descriptive stats & graphs
Excel/Google Sheets	Organizing and analyzing datasets
StatCrunch / SPSS / MINITAB	Survey analysis, modeling
R / Python	Advanced statistical modeling & automation



Why Do You Still Have to Learn Statistical Methods?



Computers perform the calculations — but YOU must guide the analysis.



Technology $\neq$ Thinking

- Software runs the analysis, but
- ◆ Is it the right method?
- ◆ Are the assumptions met?
- ◆ Do you understand the results?



Why Do You Still Have to Learn Statistical Methods?



Your Role as the Analyst

- Decide which statistical method to use
- Select the correct software options
- Ensure the analysis is valid and appropriate
- Interpret the output correctly
- Draw meaningful conclusions



Conclusion: You — not the computer — are responsible for valid statistical thinking.



Understanding Data Files & Structure

Data is typically organized in a spreadsheet or CSV file

- Comma separated value (CSV) file
- Each row = One observation (◆ also called: instance, subject, or case)
 - e.g., a single person's responses or measurements
- Each column = One variable (📊 a characteristic being measured)
 - e.g., age, hours of TV watched, gender



Types of Data

-  Numerical (Quantitative):
 - Example: Hours of TV watching, Age, Salary
-  Categorical (Qualitative):
 - Example: Yes/No, Gender, Car type



Example Dataset

MINITAB Student - Untitled - [Worksheet 1.MTP ***]

File Edit Data Calc Stat Graph Editor Tools Window Help

	C1	C2-T	C3-T	C4-T	C5	C6	C7-T	C8-T	C9	C10	C11
	Student	Gender	Race	Age	GPA	TV	Veg	PolParty	Married?		
1	1	m	w	32	3.5	3	no	rep	1		
2	2	f	w	23	3.5	15	yes	dem	0		
3	3	f	w	27	3.0	0	yes	dem	0		
4	4	f	h	35	3.2	5	no	ind	1		
5	5	m	w	23	3.5	6	no	ind	0		
6	6	m	w	39	3.5	4	yes	dem	1		
7	7	m	b	24	3.7	5	no	ind	0		
8	8	f	h	31	3.0	5	no	ind	1		
9											
10											

Each column represents a given characteristic.

Each row gives data for a particular student.

Variables in the Dataset

Variable	Type	Examples / Notes
Gender	Categorical	f = female, m = male
Racial–Ethnic Group	Categorical	b = Black, h = Hispanic, w = White
Age (in years)	Numerical	... ,30 ,25 ,18
College GPA (0 to 4 scale)	Numerical	4.0 ,2.8 ,3.2
TV Hours (Avg. per week)	Numerical	5 ,22.5 ,10
Vegetarian	Categorical (Yes/No)	yes, no
Political Party	Categorical	dem, rep, ind
Marital Status	Categorical (Binary)	1 = married, 0 = unmarried



Databases in Statistical Analysis



What is a Database?

- An existing archive or collection of data files, often organized and stored for public or institutional use.



Why Use Databases?

- Saves time by using pre-collected, structured data
- Enables large-scale analysis
- Useful for benchmarking, comparative studies, and replication



Databases in Statistical Analysis

Caution: Not All Data Are Trustworthy

■ Before using data from a database:

- Verify the source's credibility
- Check data collection methods
- Ensure transparency and documentation

■ Examples of Trusted Databases:

- U.S. Census Bureau – Population and demographics
- Google Scholar – Academic data & research sources
- World Bank Open Data – Global development statistics
- Statista, WHO, OECD, UN Data

Web Apps in Statistics

What is a Web App?

- An interactive application that runs directly in a web browser — no installation needed.

Why Use Web Apps in Statistics?

- Useful for visualizing statistical concepts
- Helps students interact with data dynamically
- Simplifies simulation and experimentation
- Makes abstract ideas more tangible and intuitive.

Web Apps in Statistics

Simulation in Action

- Web apps can simulate real-world processes such as:
- Random sampling
- Confidence intervals
- Hypothesis testing
- Probability distributions

 This allows students to “see” what might happen if you drew many samples or ran repeated experiments.

Web Apps in Statistics

 Examples of Web Apps:

- CODAP – Exploratory data analysis for learners
- StatCrunch – Online statistical analysis
- Desmos – Graphing and distribution tools
- PhET Simulations – Interactive science and math tools



Business Case: Digital Media Store & Online Advertising



The Problem

- You're the manager of a digital media store (DVDs, video games).
- Sales have been declining due to competition from online retailers like Amazon.

Business Case: Digital Media Store & Online Advertising

Your Solution

- You launch a new website to reach more customers and boost sales.

The Strategy

- You sign an advertising agreement with Facebook:
- Your ad is shown to 1,000 Facebook users
- Anyone who clicks is directed to your website
- On your homepage: 10 featured items are displayed for purchase



Evaluating Your Facebook Ad Campaign



Profitability Goal:

- Your ad campaign is profitable if: Average spending per user $\geq$ \$0.75



Early Customer Behavior:

Customer	Click Ad?	Purchase Made?	Spending (\$)
1st	✗ No	✗ No	0
2nd	✓ Yes	✗ No	0
3rd	✓ Yes	✓ Yes	20\$ = 6\$×2 + 8\$×1

Person	item1	item2	item3	item4	item5	item6	item7	item8	item9	item10
1	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0
3	0	0	8	0	12	0	0	0	0	0



Evaluating Your Facebook Ad Campaign



Next Steps:

- You collect data from all 1000 users
- Each row in the data file shows spending on the 10 featured items
- Use statistical methods to:
 - Calculate average spending
 - Measure variability
 - Determine if your profitability goal was met



Statistical Insight Needed:

Are the ad costs justified?

Is your sample average significantly above or below the threshold?

? What If You Scale Up to 100,000 Users?

- Assume average sales per person = \$0.90 (from a sample of 1,000 users)

? What If You Scale Up to 100,000 Users?

- ✓ If future users behave similarly to the sample:
 - It's likely the average sales will remain at or above \$0.75
 - Your campaign would remain profitable

 But remember this is an estimate, not a guarantee