

Class Note: Inferential Statistics for Insightful Decision-Making

Subject: Introduction to Inferential Statistics

Target Audience: Beginners with basic Excel knowledge

Objective: By the end of this class, students will understand **hypothesis testing**, **F-test**, **correlation**, and **covariance**, and be able to apply these tools in Excel to make data-driven business decisions.

Duration: 2 hours

Materials Needed: Laptops with Excel, sample dataset (provided), projector for demos

Introduction: Why Inferential Statistics Matters

What is Inferential Statistics?

Inferential statistics allows you to make educated guesses about a large group (a **population**) using a smaller sample of data. Imagine you're a business owner wanting to know if a new product will be a hit across the country. Testing it with every customer is impossible, so you survey a small group and use inferential statistics to predict its success.

- **Why it's useful:** Saves time, money, and effort by analyzing samples instead of entire populations.
- **How it helps decisions:** Provides evidence to support choices, like launching a product, optimizing marketing, or improving operations, by reducing guesswork.

Class Activity (5 mins): *Think-Pair-Share*

- **Think:** What's a business decision you'd love to make with data (e.g., "Should I open a new store?")?
 - **Pair:** Discuss with a partner how sampling a small group could help.
 - **Share:** One volunteer shares their idea with the class.
- Goal:* Connect inferential statistics to real-world problems.

Sample Dataset: A Business Scenario

To make learning hands-on, we'll use a dataset about a coffee shop chain testing two marketing campaigns: a **New Campaign** (e.g., social media ads) and a **Standard Campaign** (e.g., flyers). We'll analyze customer spending to decide which campaign drives more sales.

Dataset Description

- **Columns:**
 - **Customer ID:** Unique identifier.
 - **Campaign:** “New” or “Standard”.
 - **Spending:** Amount spent (\$) in a week.
 - **Visit Frequency:** Number of visits per week (to explore correlation/covariance).
- **Sample Size:** 10 customers per campaign (20 total).

Dataset

Customer ID	Campaign	Spending	Visit Frequency
1	New	50	4
2	New	55	5
3	New	52	4.5
4	New	60	6
5	New	48	4
6	New	58	5.5
7	New	45	3.5
8	New	53	5
9	New	49	4
10	New	62	6
11	Standard	40	3
12	Standard	42	3.5
13	Standard	45	4

Customer ID	Campaign	Spending	Visit Frequency
14	Standard	41	3
15	Standard	43	3.5
16	Standard	39	2.5
17	Standard	46	4
18	Standard	38	2
19	Standard	42	3
20	Standard	47	4

Setup in Excel:

1. Open Excel and create a new spreadsheet.
2. Enter the headers in row 1: “Customer ID” (A1), “Campaign” (B1), “Spending” (C1), “Visit Frequency” (D1).
3. Copy the data into rows 2–21.
4. Save as “Coffee_Shop_Data.xlsx”.

Interactive Twist: Students will act as “data analysts” for the coffee shop, deciding which campaign to scale up.

Concept 1: Hypothesis Testing

What It Is

Hypothesis testing is like a courtroom trial for data. You start with an assumption (the **null hypothesis**, H_0) and test if your data supports or rejects it. For our coffee shop, we’ll test if the **New Campaign** leads to higher spending than the **Standard Campaign**.

- **H_0 :** Average spending is the same for both campaigns.
- **H_1 :** New Campaign has higher average spending.

Why It Matters

It tells you if a difference (e.g., more spending) is real or just random chance, guiding decisions like which campaign to invest in.

Excel Steps

1. Organize Data:

- In column F, list “New Spending” (F1) and copy spending for “New” (50, 55, 52, 60, 48, 58, 45, 53, 49, 62) into F2:F11.
- In column G, list “Standard Spending” (G1) and copy spending for “Standard” (40, 42, 45, 41, 43, 39, 46, 38, 42, 47) into G2:G11.

2. Enable Data Analysis ToolPak:

- Go to **File > Options > Add-ins**, select “Analysis ToolPak”, click “Go”, and enable it.

3. Run t-test:

- Go to **Data > Data Analysis > t-Test: Two-Sample Assuming Equal Variances**.
- Set “Variable 1 Range” to F2:F11, “Variable 2 Range” to G2:G11.
- Set “Hypothesized Mean Difference” to 0, check “Labels”, and output to I1.
- Click “OK”.

4. Interpret:

- Check the “P(T<=t) one-tail” value. If $p < 0.05$, reject H_0 , meaning the New Campaign likely increases spending.
- **Expected result:** $p \approx 0.0001$, suggesting New Campaign is better.

Classwork (10 mins)

- **Task:** Run the t-test in Excel. Write down the p-value and decide: Should the coffee shop adopt the New Campaign? Why?
- **Bonus:** Change one “New” spending value (e.g., make 62 \rightarrow 40) and rerun the test. Does the decision change?

Goal: Practice hypothesis testing and see how data changes affect conclusions.

Business Question

A retail store tests two sales strategies: online ads vs. in-store discounts. How would you use hypothesis testing to decide which strategy drives more sales? What would a p-value of 0.03 tell you?

Concept 2: F-test

What It Is

The **F-test** compares the **variances** (how spread out data is) of two groups. We'll check if spending in the New vs. Standard Campaign has different variability. Lower variance means more consistent customer behavior, which is better for planning.

Why It Matters

Consistent spending (low variance) helps businesses predict revenue reliably, influencing budget decisions.

Excel Steps

1. **Use the same data:** F2:F11 (New Spending) and G2:G11 (Standard Spending).
2. **Run F-test:**
 - Go to **Data > Data Analysis > F-Test Two-Sample for Variances**.
 - Set "Variable 1 Range" to F2:F11, "Variable 2 Range" to G2:G11.
 - Check "Labels" and output to K1.
 - Click "OK".
3. **Interpret:**
 - If $p\text{-value} < 0.05$, variances are different.
 - **Expected result:** $p \approx 0.4$, suggesting similar variances.

Classwork (10 mins)

- **Task:** Run the F-test. Is the variance in spending significantly different? If the New Campaign had lower variance, would you recommend it even if average spending was similar?
- **Bonus:** Calculate variances manually using `=VAR.S(F2:F11)` and `=VAR.S(G2:G11)`. Compare with F-test results.

Goal: Understand variance and its role in decision-making.

Business Question

A factory tests two machines for producing parts. Machine A's parts vary widely in size, while Machine B's are more consistent. How would an F-test help decide which machine to use? Why might consistency matter more than average output?

Concept 3: Correlation

What It Is

Correlation measures how strongly two variables move together, using the correlation coefficient (**r**, from -1 to 1). We'll check if **Visit Frequency** and **Spending** are correlated across all 20 customers.

- **$r \approx 1$** : More visits, more spending.
- **$r \approx 0$** : No relationship.
- **$r \approx -1$** : More visits, less spending.

Why It Matters

Identifying relationships (e.g., visits driving spending) helps businesses focus on key drivers of revenue.

Excel Steps

1. **Prepare Data**: Use columns C (Spending) and D (Visit Frequency) for all 20 customers.
2. **Calculate**:
 - In cell I13, enter `=CORREL(C2:C21,D2:D21)` .
 - Press Enter to get **r**.
3. **Interpret**:
 - **Expected result**: $r \approx 0.92$, a strong positive correlation.

Classwork (10 mins)

- **Task**: Calculate the correlation. Create a scatter plot (**Insert > Scatter**) of Spending vs. Visit Frequency. Does the plot confirm the correlation?
- **Bonus**: If $r = 0.92$, what business action would you suggest (e.g., loyalty programs)?
Goal: Visualize and interpret relationships.

Business Question

A gym finds a correlation of 0.8 between workout frequency and membership renewals. How could this guide their strategy? What if the correlation was -0.2?

Concept 4: Covariance

What It Is

Covariance shows whether two variables move together (positive) or oppositely (negative). We'll calculate covariance between **Spending** and **Visit Frequency**. Unlike correlation, it's not scaled, so we focus on the sign.

Why It Matters

Covariance confirms relationships, helping predict how changes in one variable (e.g., visits) affect another (e.g., spending).

Excel Steps

1. Run Covariance:

- Go to **Data > Data Analysis > Covariance**.
- Set "Input Range" to C2:D21 (include both columns).
- Check "Labels in first row" and output to M1.
- Click "OK".

2. Interpret:

- Find the covariance between Spending and Visit Frequency in the output matrix.
- **Expected result:** Positive value (e.g., 10), meaning more visits align with higher spending.

Classwork (10 mins)

- **Task:** Run the covariance analysis. Is it positive or negative? How does this support the correlation result?
- **Bonus:** If covariance is positive, what marketing strategy might you propose?
Goal: Connect covariance to business actions.

Business Question

A car dealership notices covariance between test drives and sales. How could this insight shape their sales process? What if covariance was negative?

Interactive Class Wrap-Up (15 mins)

Group Discussion: Applying Insights

- **Activity:** Divide into small groups. Each group gets a business scenario (e.g., “A hotel tests two pricing strategies”). Discuss:
 - Which inferential tool would you use?
 - What would the results tell you?
 - What decision would you make?
- **Share:** One group presents their scenario and solution.
Goal: Apply concepts creatively to new contexts.

Real-World Business Scenarios

1. **Retail:** A store tests two checkout systems. Use hypothesis testing to compare average wait times. How would you decide which system to keep?
2. **Manufacturing:** Two suppliers provide parts with different defect rates. Use an F-test to compare consistency. Why might consistency matter?
3. **Marketing:** A company tracks ad clicks and sales. Use correlation to see if clicks drive sales. What if $r = 0.1$?
4. **Finance:** A bank analyzes loan applications and approval rates. Use covariance to check if application volume affects approvals. How would this guide staffing?

Key Takeaways

- **Hypothesis Testing:** Confirms if differences are real, guiding choices like campaign selection.
- **F-test:** Checks consistency, ensuring reliable outcomes.
- **Correlation/Covariance:** Reveals relationships, helping prioritize key drivers (e.g., visits).
- **Excel:** Makes these analyses accessible with ToolPak and formulas.
- **Decision-Making:** These tools provide evidence to reduce risks and optimize strategies.

Final Classwork (Homework)

- **Task:** Collect a small dataset (e.g., 10 friends' study hours and grades). Run all four analyses in Excel. Write a 1-page report: What did you learn, and what decision would you make?
- **Submit:** Via email or class platform by next week.
Goal: Build confidence in applying inferential statistics independently.

Tips for Mastery

1. **Practice:** Tweak the dataset (e.g., add outliers) and rerun analyses to see how results change.
2. **Visualize:** Use Excel charts (scatter plots, bar charts) to make findings clearer.
3. **Think Critically:** Correlation doesn't mean causation, and low p-values don't guarantee success. Consider other factors.
4. **Explore:** Try Excel's `T.TEST` function or learn basic R/Python for advanced stats.

Additional Resources

- **Excel Tutorials:** YouTube channels like "ExcellsFun" for ToolPak demos.
- **Stats Basics:** Khan Academy's Statistics and Probability course.
- **Business Applications:** Read "Freakonomics" for fun examples of data-driven decisions.

Instructor's Note: Encourage questions and real-time Excel troubleshooting during class. Use the projector to demo each Excel step. Celebrate student insights during discussions to boost engagement.