MODULE 21

Financial Statistics

This module illuminates how statistics and probability are crucial tools in finance by teaching students how to analyze data from samples, make inferences, and apply these concepts in real-world decisions.

This knowledge is fundamental in navigating the uncertainties of financial markets and in making well-informed investment choices.

POWERED BY

Rapunzl

Rapunzl Investments LLC
http://www.rapunzlinvestments.com

Key Takeaways From This Module

Guiding Questions

- How can statistics help us understand a larger population by examining just a sample of it?
- What makes a sample representative of a population?
- What limitations and risks can sampling introduce?
- How do you use data from a random sample to make inferences about a population?
- What is the difference between correlation and causation?
- What are scatter plots and how do they help us visualize trends in data?

Enduring Understandings

- Statistics provide valuable insights into populations by analyzing representative samples, highlighting the importance of sample selection.
- Random sampling is crucial in statistics as it tends to produce the most representative and unbiased samples, leading to valid generalizations.
- Inferences about population characteristics can be drawn by examining and interpreting data from random samples.
- Correlation is not causation, so be sure to think about underlying data when you're evaluating relationships.

What Is Sampling?

Sampling is a process in statistics which involves selecting a smaller part of a larger population to estimate characteristics of the whole population.

The larger the sample size the greater the accuracy of any inferences we can draw, because our sample has more of the population, but it's also more expensive to sample more individuals.

Types of Sampling

Random Sampling

Explaining how each member of a population has an equal chance of being selected, ensuring unbiased financial data representation.

Stratified Sampling

Dividing a population into subgroups (strata) and randomly sampling from each, used for more precise financial analysis.

Cluster Sampling

Randomly selecting entire groups or clusters from a population, a cost-effective method in large-scale financial surveys.

Examples Of Real-World Sampling

Healthcare Companies
Doctors use random sampling during clinical trials, ensuring the drug's efficacy and safety across different demographics.

Environmental Agencies
Scientists use cluster sampling to monitor air quality in different regions, depending upon levels of industrial activity and population density.

Government Data
Governments use stratified sampling, categorizing the population by regions, industries, or demographics, to measure unemployment.

Election Polls
News teams use cluster sampling to predict election outcomes and public opinion trends.

Market Research
Firms use random sampling to gather data from a broad, unbiased cross-section of consumers.

Bias & Limitations Of Sampling

We need to be careful when we sample a population, because **a bad sample can be unrepresentative of a population, leading to biased results.**

Biased samples can wrongly influence investors, medical professionals, and election polls which is why it's really important to avoid any sources of bias that could jeopardize your study.

Non-Response Bias
Individuals who choose not to respond differ significantly from those who do, skewing the results.

Response Bias
Responses are different from true feelings due to question wording, survey methods, or personal embarrassment.

Selection Bias
Participants are not randomly selected, leading to a sample that is not representative of the larger population.

Important: There's a difference between bias and sampling error. With bias, there is a systemic issue with your study. With sampling error, you may need to increase the sample size to accommodate for potential outliers.

Imagine an investment analyst decides to conduct research for his job by asking a series of questions to strangers who pass them on the street in New York City.

What factors may introduce bias to this research?

The investment analyst now modifies their approach, and post a series of questions to a poll on social media.

What factors may still introduce bias to this research?

Sampling Best Practices

There are 5 main things to think about when sampling a population for a research project in order to avoid biases or other limitations.

1. Define a target population and a sampling frame which includes a list of individuals that the sample will be drawn from
2. Match the sampling frame to the target population as much as possible to reduce the risk of sampling bias.
3. Make the survey as short and accessible as possible for as many participants to avoid any potential bias.
4. Follow up on non-responders to avoid a no-response bias.
5. Don't be afraid to repeat your sample a second time to confirm any inferences you may make.

Drawing Inferences From Our Sample

Below are the performances of each of the FAANG stocks on a random day when the markets are open. Think through the following questions and prepare to discuss with the class.

+3.11%	+1.87%	+2.34%	+0.58%	+1.03%
META Meta Platforms Inc	**AAPL** Apple Inc	**AMZN** Amazon Inc	**GOOGL** Alphabet Inc	**NFLX** Netflix Inc

- Do you think that this is a random sample of the 8,000+ US Stocks that are publicly traded?
- Could you safely infer that healthcare companies' stocks also went up today? Why or why not?
- Since the stocks of the largest companies in the world are up today, can we assume that the S&P 500 is also up today? Explain your reasoning.

Causation & Correlation

Correlation is a statistical measurement indicating the extent to which two variables fluctuate together, but it does not imply cause and effect or that the two variables are related.

Causation is a relationship where one variable directly affects or causes a change in another, such that when one variable changes, the other also changes.

Correlations can either be positive or negative and vary in strength.

Relationship	Correlation?	Causal?
Interest Rate Increases and Savings Account Deposits	Yes - Positive	Yes
Stock Market Performance and Luxury Car Sales	Yes - Positive	No
Credit Score and Loan Interest Rates	Yes - Negative	Yes
Value of Bitcoin & Small Business Revenue	No	No

Correlation is NOT Causation!

What other variables do you think might influence ice cream sales and shark attacks?

You collect data on sunburns and ice cream consumption. You find that higher ice cream consumption is associated with a higher probability of sunburn.

Does that mean ice cream consumption causes sunburn?

Causation Or Correlation?

1. Ice Cream Sales & Drowning Incidents
2. Number of Firefighters & Fire Damage
3. Pirate Population & Global Warming
4. Organic Food Sales & Autism Rates
5. Social Media Usage & Political Polarization
6. Shoe Size & Reading Ability in Children
7. Cell Phone Sales & College Tuition Costs
8. Number of Churches & Crime Rate in Cities
9. Internet Explorer Usage and Murder Rates
10. Length of Fingernails and IQ Intelligence

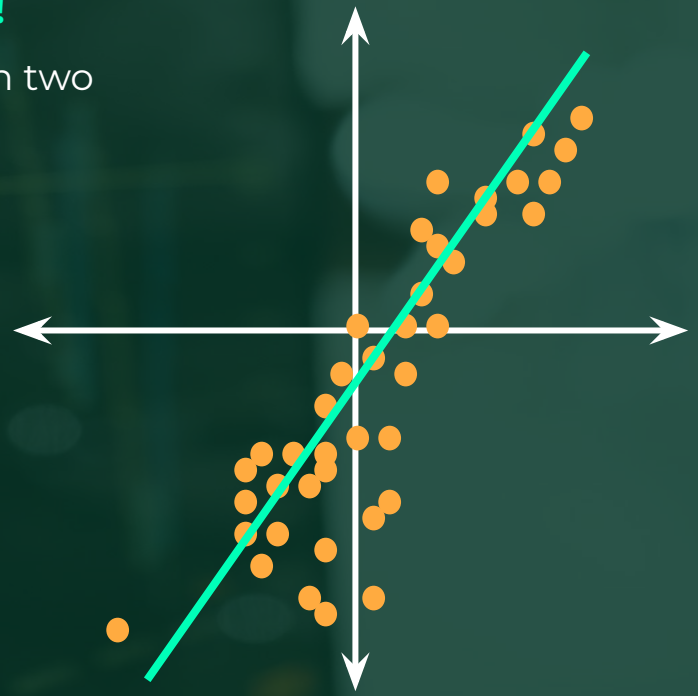
Finding Correlation With Scatter Plots

Let's introduce a new tool to our arsenal: scatter plots!

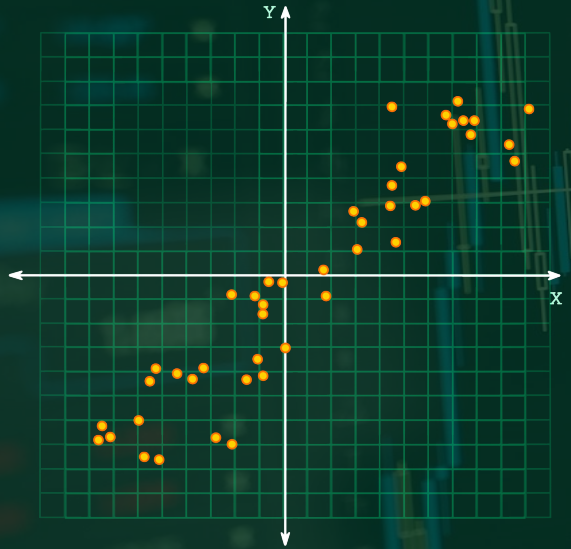
Scatter plots help us visually display the relationship between two variables to help determine if they share any relationship.

Steps To Make A Scatter Plot

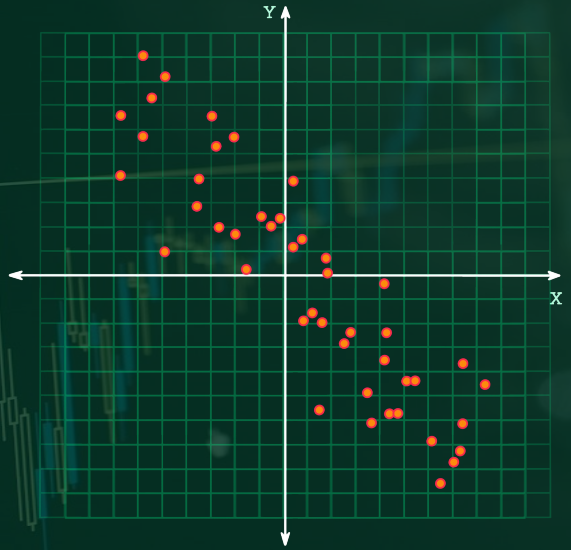
1. Draw Your Scatterplot...
2. **Plot Your Data:** You'll learn to plot individual data points along two axes
3. **Discover Relationships:** Observe how the arrangement of points reveals positive, negative, or no correlation between these variables.
4. **Draw a line of best fit!** This line of best fit can help approximate relationships and find trends.



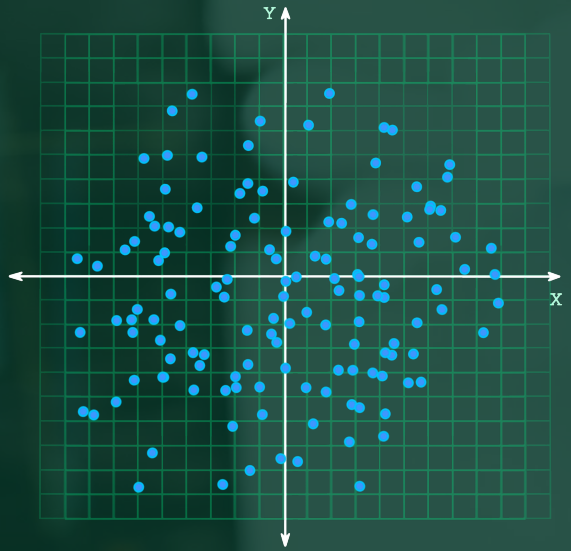
Analyzing Scatter Plot Correlations



Strong Positive Correlation



Strong Negative Correlation



No Correlation

What To Do With Outliers?

Outliers are specific data points that are significantly different from the majority of your dataset.

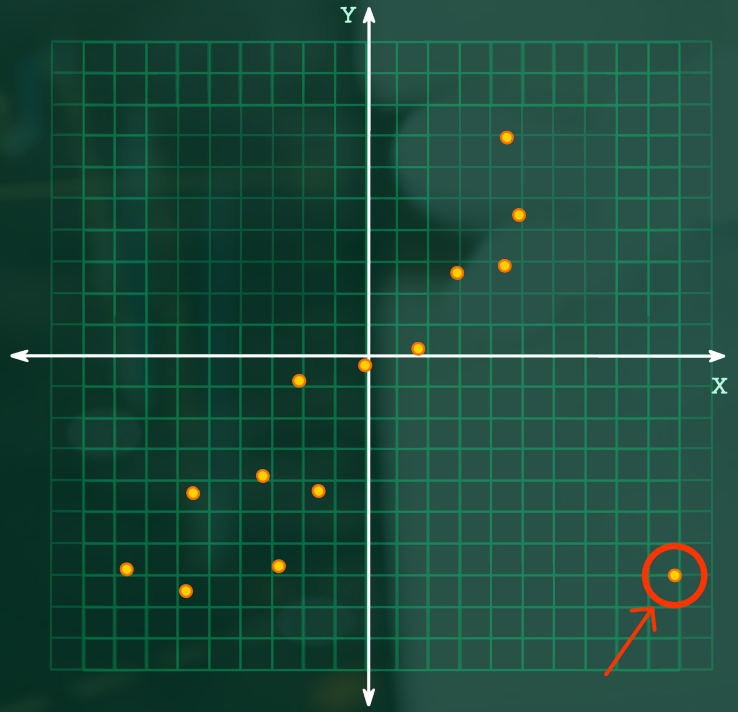
Excluding outliers can be necessary because they can distort the overall analysis and lead to misleading conclusions.

Sometimes outliers emerge as an accident or because of a flaw in an experiment's design.

Be cautious when excluding outliers; they sometimes represent important, real-world events that shouldn't be overlooked.

Can we exclude the circled data point? Why or Why Not?

Can you remove the outlier without knowing what is being charted?



Key Takeaways From This Module

CORE & FUNDAMENTALS

- Statistics allows us to understand broader populations by analyzing smaller, representative samples, but there are limits and risk of bias.
- Using data from random samples enables us to make informed inferences about general population trends.
- Differentiating between correlation and causation is essential to avoid misconceptions about relationships in data.

APPLIED KNOWLEDGE

- Skillful selection of samples is vital for reliable statistical analysis and drawing valid conclusions.
- Employing random sampling techniques helps minimize bias.
- Careful examination of correlation and causation is necessary to understand true relationships in data.
- Utilizing scatter plots can effectively reveal and visualize trends and patterns in complex data sets.

RELEVANCE FOR YOU

- Understanding samples helps evaluate the credibility of statistical claims in life, including understanding sampling risks and limitations such as bias.
- Knowing the difference between correlation and causation aids in understanding and questioning relationships presented in everyday life.
- Learning to interpret scatter plots enables you to visualize and grasp trends in real-world data, from economics to science.