Statistics Foundations for A/B Testing

1. Population vs Sample

Population

The entire group being analyzed.

- Can be large (all people in the world) or small (all students in a classroom)
- Includes every member of the group you're interested in

Examples:

- All adults in the U.S. when studying average height
- All students in a school when researching test scores

Sample

A subset of a population selected for analysis.

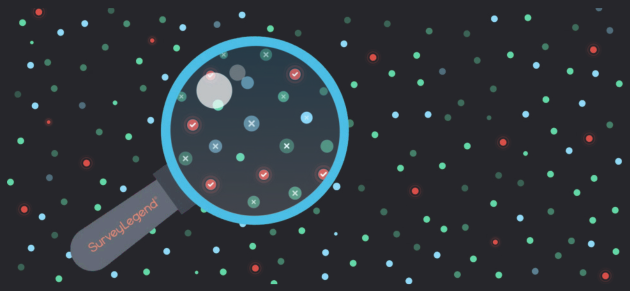
- Used when studying entire population is impractical
- Allows us to make inferences about the population

Example:

- Population: All students in a university
- Sample: 200 randomly selected students

Why Use Sampling?

- Cost and Time Efficiency
- Greater Feasibility
- Enables Statistical Inference



2. Types of Sampling Methods

Each sampling method has specific advantages and use cases in statistical analysis.

Simple Random Sampling

Every member of the population has an equal chance of being selected. Like drawing names from a hat.

Stratified Sampling

Population is divided into subgroups (strata) and samples are taken from each. Ensures representation of all subgroups.

Systematic Sampling

Selecting every nth member from an ordered population. Example: choosing every 10th person from a list.

Cluster Sampling

Population is divided into clusters, and some clusters are randomly selected for complete sampling.

Convenience Sampling

Selecting readily available subjects. Easiest but potentially biased method.

The choice of sampling method affects the reliability and validity of your statistical conclusions.

3. The Mean

Population Mean (μ)

Formula: $\mu = \Sigma X / N$

Imagine we're studying the ages of all 5 employees at a startup:

Ages: 25, 30, 35, 40, 50

- Sum all ages: 25 + 30 + 35 + 40 + 50 = 180
- Divide by number of employees (N=5): 180/5 = 36

The mean age of the entire population is 36 years.

Sample Mean ($\bar{X}$)

Formula: $\bar{X} = \Sigma X / n$

Let's find the mean of the following sample:

Values: 25, 30, 35

- Sum the values: 25 + 30 + 35 = 90
- Divide by observations (n=3): 90/3 = 30

The sample mean is 30.

Interpretation: The mean represents the central or average value in the data set.

4. Key Questions in Sampling

1. **Estimating Population Mean?**

   What is our best estimate of the population mean? The sample mean generated from our sample serves as the best estimate for the population mean.

2. **Confidence in Estimates?**

   How confident are we about that estimate? With small samples, we have less confidence that the sample mean accurately represents the true population mean.

3. **Sample Reliability?**

   How reliable is our sample? The size and selection method of our sample directly impacts how well it represents the broader population.

5. Standard Deviation & Variance (Population)

Key Concepts

- **Variance (σ^2):** Average squared deviations from mean
- **Standard Deviation (σ):** Square root of variance
- **Purpose:** Measures spread of data around mean

Formulas

Population Variance (σ^2):

$$\sigma^2 = \Sigma(X - \mu)^2 / N$$

where:

X = each value

μ = population mean

N = population size

Population Standard Deviation (σ):

$$\sigma = \sqrt{\Sigma(X - \mu)^2 / N}$$

Real World Example: Bank Wait Times

Consider wait times (in minutes) at three banks:

- **Bank 1:** 6, 6, 6, 6 (Mean = 6, SD = 0)
- **Bank 2:** 2, 4, 6, 10 (Mean = 6, SD = 3.4)
- **Bank 3:** 1, 3, 9, 11 (Mean = 6, SD = 4.7)

While all banks have the same mean wait time (6 minutes), their standard deviations reveal how consistent the service is. Bank 1 is most consistent, while Bank 3 has the most variable wait times.

6. Understanding Standard Deviation Formula

The standard deviation formula measures how spread out numbers are from their average (mean) value.

1 **Step 1: Calculate the Mean (μ)**

First find the average by adding all values and dividing by total count (N)

2 **Step 2: Find Deviations**

Subtract the mean from each value $(X - \mu)$ to get distances from average

3 **Step 3: Square the Deviations**

Square each difference $(X - \mu)^2$ to make all values positive

4 **Step 4: Find Average of Squares**

Add all squared deviations and divide by N to get variance (σ^2)

5 **Step 5: Take the Square Root**

Calculate $\sqrt{}(\sigma^2)$ to get final standard deviation (σ)

The complete formula: $\sigma = \sqrt{}[\Sigma(X - \mu)^2 / N]$

> 💡 A larger standard deviation means data points are more spread out, while a smaller value indicates they're clustered closer to the mean.

7. Standard Deviation (Sample)

Understanding the key differences between sample and population standard deviation is crucial for statistical analysis:

Population Standard Deviation (σ)

Used when you have data for the entire population

- Formula: $\sigma = \sqrt{[\Sigma(X - \mu)^2 / N]}$
- Divisor is N (total population)
- Uses population mean (μ)
- Gives exact measure of spread

Sample Standard Deviation (s)

Used when working with a sample of the population

- Formula: $s = \sqrt{[\Sigma(X - \bar{X})^2 / (n-1)]}$
- Divisor is (n-1) due to Bessel's correction
- Uses sample mean ($\bar{X}$)
- Estimates population spread

The key difference lies in using (n-1) instead of n in the sample calculation, known as Bessel's correction, which helps correct for the bias in estimation when working with samples.

8. Normal Distribution

The normal distribution is one of the most important probability distributions in statistics, appearing frequently in nature and statistical analyses.

Shape

Bell-shaped, symmetric curve with tails extending infinitely in both directions. The peak occurs at the mean.

Parameters

Defined by mean (μ) which determines center and standard deviation (σ) which determines spread. These two values completely characterize the distribution.

Empirical Rule

68% within $\pm 1\sigma$, 95% within $\pm 2\sigma$, 99.7% within $\pm 3\sigma$ of the mean. This "68-95-99.7 rule" is crucial for understanding data spread.

9. z-Scores

Definition

A z-score measures how many standard deviations an observation or data point is away from the mean of the distribution.

It helps standardize data to make meaningful comparisons across different datasets or measurements.

Formula

For population data: $z = (X - \mu) / \sigma$

Where X is the observed value, μ is the population mean, and σ is the population standard deviation.

For sample data: $z = (X - \bar{X}) / s$

Where $\bar{X}$ is the sample mean and s is the sample standard deviation.

Uses

- Standardizing scores across different scales
- Identifying outliers in datasets
- Comparing performance across different metrics

10. Example of z-Scores

LBJ's Height

Lyndon B. Johnson was 6'4" (76 inches)

Mean US president height: 71.5 inches
SD: 2.11 inches

z-score = (76-71.5)/2.11 = +2.13

LBJ was 2.13 standard deviations above presidential average height

Shaq's Height

Shaquille O'Neal is 7'2" (86 inches)

Mean Lakers height: 80 inches
SD: 3.3 inches

z-score = (86-80)/3.3 = +1.82

Shaq is 1.82 standard deviations above Lakers average height

Comparison

While Shaq (7'2") is absolutely taller than LBJ (6'4"), LBJ's z-score (+2.13) shows he was relatively taller compared to other presidents than Shaq (+1.82) is compared to Lakers players.

This demonstrates how z-scores help compare measurements across different reference groups, even when the absolute measurements and standard deviations differ.

11. Standard Normal Distribution

1

Basic Definition

A special case of normal distribution where mean (μ) = 0 and standard deviation (σ) = 1. This standardized form simplifies probability calculations, distribution comparisons, and finding percentiles.

2

Key Properties

Features a symmetric bell-shaped curve with precise data distribution: 68% within ±1 SD, 95% within ±2 SD, and 99.7% within ±3 SD from the mean.

3

Practical Application

Convert any normal distribution to standard form using z-scores. Example: For data points 40, 50, 60 (mean=50, SD=10), z-scores are -1, 0, +1 respectively.

The standard normal distribution is fundamental in statistics, particularly for hypothesis testing, confidence intervals, and probability calculations in A/B testing.

Original Value	Z-Score	Interpretation
40	-1	One SD below mean
50	0	At the mean
60	+1	One SD above mean

12. Summary So Far

Mean – Our Starting Point

The mean serves as our foundation, showing where data typically clusters. This central point becomes crucial for measuring variability.

Standard Deviation – Measuring Spread

Building on the mean, standard deviation tells us how data spreads out. This spread measurement becomes essential for understanding distributions.

Normal Distribution & Z-scores

When data follows a normal distribution, we can use the mean and standard deviation to calculate z-scores, transforming any normal distribution into the standard normal curve.

13. Central Limit Theorem (CLT)

1

Definition

As sample size increases, the distribution of sample means approaches normal distribution, regardless of the original population's shape

2

Implications

Enables reliable statistical inference even with non-normal populations when sample size is sufficient ($n \geq 30$); crucial for A/B testing validity

3

Standard Error

Measures the variability of sample means: $SE = \sigma/\sqrt{n}$ when population SD (σ) is known, or $SE = s/\sqrt{n}$ when using sample SD (s)

Understanding CLT is crucial for A/B testing because it allows us to make valid statistical inferences even when dealing with non-normal data distributions. The larger the sample size, the more reliable our statistical conclusions become.

14. Understanding Z vs T Distributions

Z Distribution

- Standard normal distribution where $\mu=0$ and $\sigma=1$
- Used when population standard deviation (σ) is known
- Perfect symmetrical bell curve
- Commonly used for large samples ($n \geq 30$)

T Distribution

- Similar to z-distribution but with heavier tails
- Used when population standard deviation is unknown
- Shape depends on degrees of freedom ($df=n-1$)
- Essential for smaller samples ($n<30$)

Both distributions are crucial for A/B testing, with t-distribution being more commonly used in practice as we rarely know the population standard deviation. As sample size increases, the t-distribution approaches the z-distribution, becoming nearly identical when $n>30$.

Key Differences

T-distribution has heavier tails to account for additional uncertainty when estimating population parameters from small samples

Practical Application

In A/B testing, we typically use t-distribution for calculating confidence intervals and p-values, especially when dealing with smaller sample sizes

Selection Criteria

Choose z-distribution when σ is known and $n \geq 30$; otherwise, default to t-distribution for more conservative estimates

15. Hypothesis Testing

What is Hypothesis Testing?

A statistical method to test if observed effects are real or due to chance. Provides a framework for data-driven decisions in A/B testing to determine if differences between variants matter.

Key Components

H_0 (Null Hypothesis): Assumes no difference exists between variants.

H_1 (Alternative Hypothesis): Proposes a difference exists between variants.

We either reject H_0 in favor of H_1, or fail to reject H_0 based on evidence.

Critical Elements

- Significance Level (α): Risk threshold for false positives (typically 0.05)
- P-value: Probability of observing current results if H_0 is true
- Statistical Power: Ability to detect real effects ($1-\beta$)

16. Understanding P-Value

The p-value is a probability measure that helps us quantify the statistical significance of our results in A/B testing and hypothesis testing.

Definition

The p-value represents the probability of obtaining results at least as extreme as those observed, assuming the null hypothesis is true. A smaller p-value indicates stronger evidence against the null hypothesis.

- $P < 0.05$: Strong evidence against H_0
- $P \geq 0.05$: Insufficient evidence against H_0

Interpretation

Common misconceptions about p-value:

- It's NOT the probability that H_0 is true
- It's NOT the probability of making a mistake
- It's NOT the probability of replication

Practical Application

In A/B testing, the p-value helps determine if the observed difference between variants is statistically significant:

- $P = 0.01$ means 1% chance of seeing this difference by random chance
- Guides decision to implement changes or continue testing

17. Type I & Type II Errors in A/B Testing

Understanding these errors is crucial for making reliable decisions in A/B testing

Type I Error (False Positive)

Rejecting a true null hypothesis (α)

- Concluding variant B is better when it's not
- Risk: Implementing ineffective changes
- Controlled by significance level (typically 5%)

Type II Error (False Negative)

Failing to reject a false null hypothesis (β)

- Missing real improvements in variant B
- Risk: Overlooking valuable changes
- Reduced by increasing sample size

In A/B testing, balancing these errors is key: Type I errors lead to wasted resources on ineffective changes, while Type II errors mean missing valuable opportunities for improvement.

18. Statistical Power & Sample Size

Understanding Statistical Power

Statistical power is the probability of detecting a real effect when it exists. A power of 80% means we have an 80% chance of finding a true difference between variants.

- Higher power = Better chance of detecting real effects
- Lower power = Higher risk of missing important changes
- Industry standard is typically 80% power

Sample Size Impact

Larger sample sizes increase statistical power, improving our ability to detect meaningful differences between variants.

- Too small: Risk missing real effects (Type II error)
- Adequate: Reliable detection of meaningful differences
- Calculation depends on expected effect size

Running tests with insufficient sample sizes is a common pitfall that can lead to missed opportunities for optimization. Always calculate required sample size before starting a test.

Summary & Q&A