2. Statistics for Business Analytics

Welcome to the world of business statistics, where data transforms into meaningful insights, guiding decision-making and uncovering hidden truths. In this comprehensive exploration, we embark on a journey to demystify essential statistical concepts and techniques that underpin rigorous business data analysis. From understanding the intricacies of distributions to applying hypothesis testing and constructing confidence intervals, each chapter unfolds a new facet of statistical literacy.

At the heart of statistical analysis lies the normal distribution, a bell-shaped curve that permeates countless phenomena in nature and human behavior. In this part, we delve into the essence of the normal distribution, unraveling its properties and significance in statistical inference. Through visualizations and real-world examples, we illuminate the ubiquity of this fundamental distribution and its role as a cornerstone of statistical theory.

Standard deviation serves as a compass in the statistical landscape, guiding us through the variability inherent in datasets. In this chapter, we dissect the concept of standard deviation, unveiling its importance in quantifying dispersion and assessing the spread of data points. Armed with a deeper understanding of standard deviations, you will navigate data with confidence, discerning patterns and outliers with precision.

Variables form the building blocks of statistical analysis, each possessing distinct characteristics and implications. This chapter elucidates the dichotomy between continuous and discrete variables, shedding light on their respective roles in data modeling and interpretation. By grasping the nuances of variable types, you will harness the full potential of statistical techniques tailored to diverse data structures.

Sampling distribution serves as the bedrock of statistical inference, bridging the gap between sample observations and population parameters. In this chapter, we unravel the concept of sampling distribution, elucidating its relevance in making probabilistic statements about population characteristics. Through concrete examples, you will develop an intuitive understanding of sampling distribution's role in robust statistical analysis.

The Central Limit Theorem is a key concept in statistics that helps us make sense of uncertainty. This chapter explains the Central Limit Theorem in simple terms, showing how it makes sample means more predictable and aids in hypothesis testing. By understanding this concept, you'll be able to draw meaningful conclusions from data.

Understanding hypothesis testing is essential for making data-driven decisions. It allows us to determine whether observed patterns in data are meaningful or simply due to chance. By applying hypothesis testing, we can evaluate assumptions, compare groups, and assess the statistical significance of results, making it a vital tool in scientific research, business analysis, and many other fields.

Z-scores and Z-tables serve as navigational aids in the sea of standard normal distribution, facilitating standardized comparisons and probability calculations. This chapter elucidates the intricacies of Z-scores, empowering you to interpret standardized scores and harness Z-tables for statistical analysis. With proficiency in Z-scores, you will navigate the vast expanse of normal distribution with confidence and precision.

In situations where sample sizes are small or population standard deviations are unknown, t-scores and t-tables emerge as indispensable tools for statistical analysis. This chapter unravels the mysteries of t-scores, guiding you through their calculation and interpretation using t-tables. Armed with this knowledge, you will navigate the nuances of t-distributions with finesse, ensuring robust inference in diverse statistical scenarios.

Normal and t-distributions stand as pillars of probability theory, each possessing unique characteristics and applications. In this chapter, we elucidate the distinctions between these distributions, empowering you to discern when to employ each in statistical analysis. Through practical examples and comparative analyses, you will develop a nuanced understanding of normal and t-distributions, enriching your statistical toolkit.

Confidence intervals provide a window into the uncertainty surrounding population parameters, empowering us to quantify the precision of our estimates. In this chapter, we explore the construction of confidence intervals for means and proportions, unraveling the methodology and interpretation behind these essential statistical tools. By mastering confidence intervals, you will convey the uncertainty inherent in your findings with transparency and rigor.

While p-values offer a gateway to statistical inference, their misinterpretation can lead to erroneous conclusions and misinformed decisions. This chapter examines the potential pitfalls of overreliance on p-values, highlighting the importance of context and effect size in statistical analysis. Through critical examination and practical insights, you will navigate the complexities of p-values with vigilance, ensuring the integrity of your statistical conclusions.

Within these pages lie the keys to unlocking the mysteries of statistical analysis, empowering you to navigate the complexities of data with confidence and precision. As we embark on this journey together, let curiosity be our compass and inquiry our guiding light, illuminating the path towards deeper understanding and actionable insights.

2.1 Normal distribution

At the heart of statistical analysis lies the normal distribution, a ubiquitous probability distribution that serves as a benchmark for many statistical techniques. We will delve into its characteristics, its symmetrical bell-shaped curve, and its significance in understanding the distribution of data.

The normal distribution finds applications across various fields, including finance, psychology, engineering, and biology. From modelling stock prices to understanding human height distributions, the normal distribution serves as a versatile tool for analyzing and interpreting data.

Throughout this chapter, we will delve into the mathematical properties of the normal distribution, exploring how to calculate probabilities, percentiles, and z-scores. Moreover, we will discuss practical techniques for visualizing and interpreting normal distributions using histograms, density plots, and cumulative distribution functions.

By the end of this chapter, you will have a deep appreciation for the normal distribution and its significance in statistical analysis. Armed with this knowledge, you will be well-equipped to tackle more advanced statistical concepts and apply them to real-world datasets. Let's embark on this journey to unravel the mysteries of the normal distribution together.

A normal distribution, also known as a Gaussian distribution or bell curve, exhibits symmetrical data distribution without skewness. When graphed, the data forms a bell-shaped curve, with the majority of values congregating around the centre and decreasing as they move away from it.

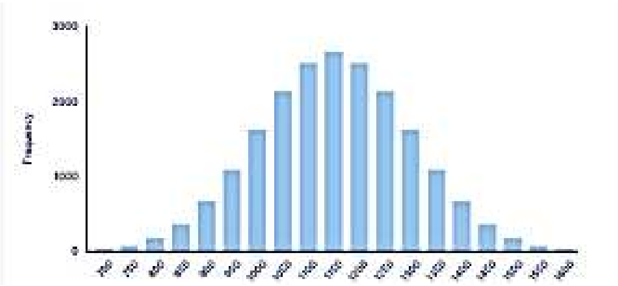

Figure 2.1 Example of a Gaussian distribution or bell curve.

Various variables in both natural and social sciences typically exhibit a normal distribution or an approximation thereof. Examples include height, birth weight, reading ability, job satisfaction, and SAT scores. Due to the prevalence of normally distributed variables, numerous statistical tests are tailored for such populations. Proficiency in comprehending the characteristics of normal distributions empowers individuals to employ inferential statistics for comparing groups and generating population estimates from samples.

Normal distributions have key characteristics that are easy to spot in graphs:

- The mean, median and mode are exactly the same.
- The distribution is symmetric about the mean—half the values fall below the mean and half above the mean.
- The distribution can be described by two values: the mean and the standard deviation.

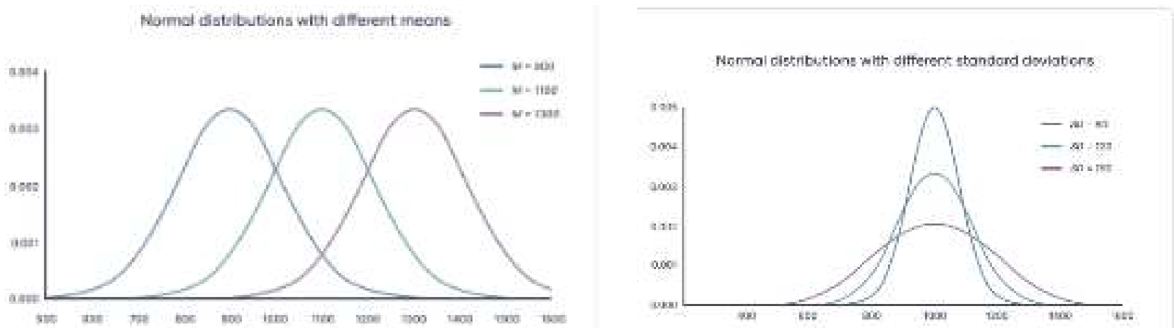

Figure 2.2 Normal distribution with different mean and with different stand deviations.

The mean serves as the location parameter, dictating the centre of the curve's peak. Adjusting the mean shifts the curve accordingly: increasing it shifts the curve to the right, while decreasing it shifts the curve to the left. Meanwhile, the standard deviation functions as the scale parameter, influencing the spread or width of the curve.

The standard deviation stretches or squeezes the curve. A small standard deviation results in a narrow curve, while a large standard deviation leads to a wide curve.

2.2 Empirical rule

The empirical rule, also known as the 68-95-99.7 rule, provides insight into the distribution of values within a normal distribution:

- Approximately 68% of values fall within 1 standard deviation from the mean.
- Roughly 95% of values lie within 2 standard deviations from the mean.
- About 99.7% of values are encompassed within 3 standard deviations from the mean.

For instance, consider a scenario where SAT scores from students in a new test preparation course are gathered, and the data conforms to a normal distribution with a mean score (M) of 1150 and a standard deviation (SD) of 150.

Applying the empirical rule yields the following insights:

- Around 68% of scores fall within the range of 1000 to 1300, corresponding to 1 standard deviation above and below the mean.
- Approximately 95% of scores are within the range of 850 to 1450, representing 2 standard deviations above and below the mean.
- Nearly all scores, around 99.7%, lie within the range of 700 to 1600, encompassing 3 standard deviations above and below the mean.

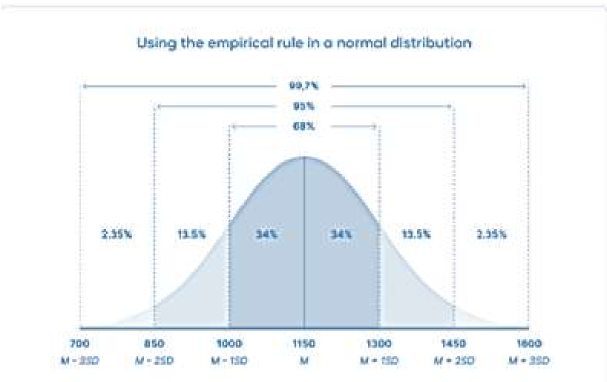

Figure 2.3 Empirical rule in normal distribution.

The empirical rule offers a rapid method to assess data, enabling detection of outliers or exceptional values that deviate from its expected pattern. In cases where data from small samples diverge significantly from this pattern, alternative distributions such as the t-distribution might be more suitable. Identifying the distribution of the variable allows for the application of relevant statistical tests.

2.3 Formula of the normal curve

To construct a normal curve based on a given mean and standard deviation, one can employ a probability density function, thereby accurately representing the distribution of the data.

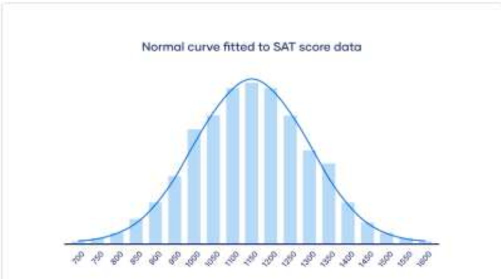

Figure 2.4 Normal curve fitted to SAT score data.

Within a probability density function, the area beneath the curve represents probability. Given that the normal distribution serves as a probability distribution, the cumulative area under the curve invariably sums up to 1 or 100%. Although the formula for the normal probability density function may appear intricate, utilizing it merely necessitates knowledge of the population mean and standard deviation. By substituting these parameters into the formula, one can determine the probability density associated with any given value of x.

- *f(x)* = probability
- *x* = value of the variable
- μ = mean
- σ = standard deviation
- σ^2 = variance

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Example:

Using the probability density function, you want to know the probability that SAT scores in your sample exceed 1380.

On your graph of the probability density function, the probability is the shaded area under the curve that lies to the right of where your SAT scores equal 1380.

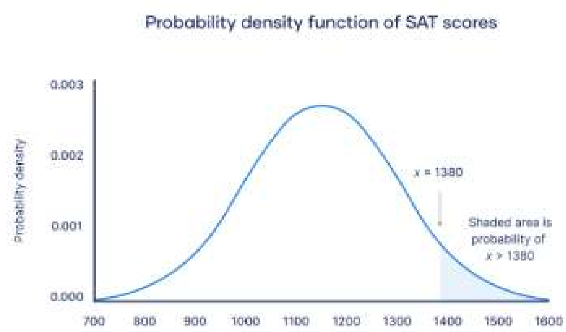

Figure 2.5 Probability density function of SAT scores graph.

You can find the probability value of this score using the standard normal distribution.

2.4 Standard normal distribution

The standard normal distribution, known as the **z-distribution**, is distinct for having a mean of 0 and a standard deviation of 1. Any normal distribution can be seen as a transformation of the standard normal distribution, undergoing adjustments in scale, position, or both.

In the context of the z-distribution, individual observations, which are typically denoted as x in normal distributions, are referred to as z-scores. These z-scores represent the number of standard deviations that each value deviates from the mean. Consequently, converting values from any normal distribution into z-scores facilitates comparison and analysis within the framework of the standard normal distribution.

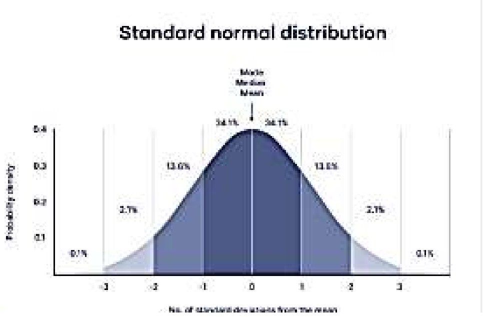

Figure 2.6 Standard normal distribution graph.

You only need to know the mean and standard deviation of your distribution to find the *z*-score of a value.

Z-score Formula Explanation

$$z = \frac{x - \mu}{\sigma}$$

- x = individual value
- μ = mean
- σ = standard deviation

We convert normal distributions into the standard normal distribution for several reasons:

- To find the probability of observations in a distribution falling above or below a given value.
- To find the probability that a sample mean significantly differs from a known population mean.
- To compare scores on different distributions with different means and standard deviations.

2.5 Finding probability using the *z*-distribution

Each z-score corresponds to a probability, often referred to as a p-value, indicating the likelihood of observing values below that specific z-score. By transforming an individual value into a z-score, one can determine the probability of all values up to that point occurring within a normal distribution.

For instance, consider a scenario where you wish to ascertain the probability of SAT scores in your sample surpassing 1380. Initially, you calculate the z-score using the mean and standard deviation of the distribution. With a mean of 1150 and a standard deviation of 150, the z-score reveals the number of standard deviations by which 1380 deviates from the mean.

Formula **Calculation**

$$z = \frac{x - \mu}{\sigma} = \frac{1380 - 1150}{150} = 1.53$$

For a *z*-score of 1.53, the *p*-value is 0.937. This is the probability of SAT scores being 1380 or less (93.7%), and it's the area under the curve left of the shaded area.

To find the shaded area, you take away 0.937 from 1, which is the total area under the curve.

Probability of x > 1380 = 1 – 0.937 = **0.063**

That means it is likely that only 6.3% of SAT scores in your sample exceed 1380.

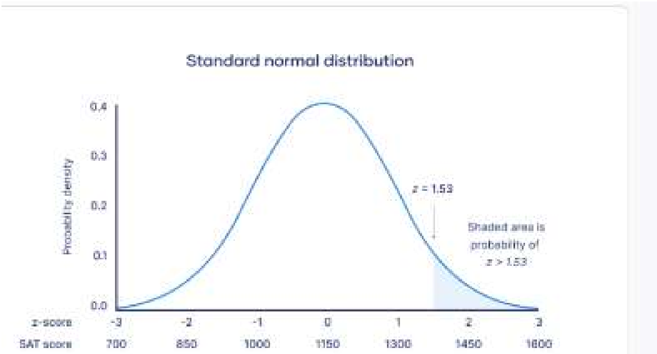

Figure 2.7 Standard normal distribution with SAT score indicated.

2.6 Sampling Distribution

Sampling distributions form the backbone of statistical inference, enabling us to draw conclusions about populations based on sample data. We will delve into the intricacies of sampling distributions, understanding how they reflect the variability of sample statistics and their pivotal role in hypothesis testing.

Sampling distribution refers to the distribution of a statistic, such as the sample mean or sample proportion, obtained from multiple samples of the same size drawn from a population. It provides insights into the behavior of sample statistics and their variability across different samples.

2.7 Central Limit Theorem and Sampling Distribution

The Central Limit Theorem (CLT) is a fundamental concept in statistics that underpins the behavior of sampling distributions. It states that the sampling distribution of the sample mean approaches a normal distribution as the sample size increases, regardless of the shape of the population distribution. This theorem enables us to make robust inferences about population parameters from sample data.

The central limit theorem serves as the cornerstone of understanding normal distributions in statistics. In research settings, obtaining an accurate estimation of a population mean often involves gathering data from numerous random samples within the population. These individual sample means collectively form what is known as a sampling distribution of the mean.

The central limit theorem delineates two key principles:

1. **Law of Large Numbers:** As the sample size or the number of samples increases, the sample mean tends to converge towards the population mean.
2. **Normality of Sampling Distribution:** Despite the original variable's distribution, when working with multiple large samples, the sampling distribution of the mean tends to approximate a normal distribution.

Parametric statistical tests conventionally assume that samples are derived from normally distributed populations. However, the central limit theorem obviates the necessity of this assumption for sufficiently large sample sizes. With large samples, parametric tests can be applied irrespective of the population's distribution, provided other pertinent assumptions are satisfied. A sample size of 30 or more is commonly deemed as sufficiently large.

Conversely, for small samples, ensuring the assumption of normality is crucial due to the uncertainty surrounding the sampling distribution of the mean. Accurate results necessitate confirmation that the population adheres to a normal distribution before employing parametric tests with small sample sizes.

Illustratively, the central limit theorem posits that by obtaining sufficiently large samples from a population, the means of these samples will exhibit a normal distribution, even if the underlying population distribution diverges from normality.

Example: Consider a population following a Poisson distribution (depicted in the left image). Upon drawing 10,000 samples from this population, each consisting of 50 observations, the distribution of sample means aligns closely with a normal distribution, in accordance with the central limit theorem (as illustrated in the right image).

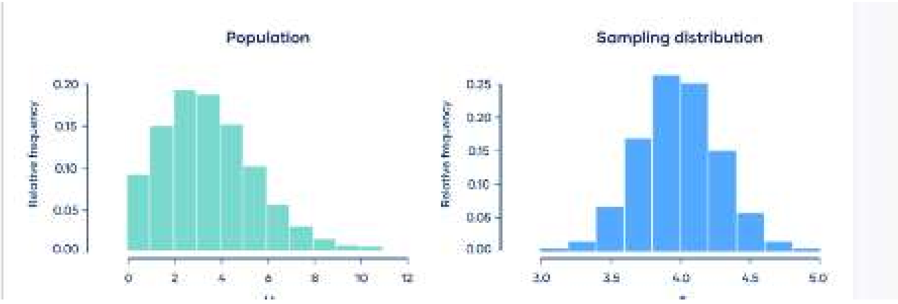

Figure 2.8 Example of population in Poisson distribution and normal distribution.

The central limit theorem hinges upon the notion of a sampling distribution, which represents the probability distribution of a statistic computed from numerous samples drawn from a population.

Conceptualizing an experiment can aid in grasping sampling distributions:

- Let's envision drawing a random sample from a population and computing a statistic, such as the mean.
- Subsequently, another random sample of identical size is drawn, and the mean is recalculated.
- This process is iterated numerous times, resulting in a plethora of means, each corresponding to a sample.

The aggregation of these sample means exemplifies a sampling distribution. According to the central limit theorem, the sampling distribution of the mean tends towards a normal distribution when the sample size is sufficiently large. Remarkably, irrespective of the population's distribution—be it normal, Poisson, binomial, or otherwise—the sampling distribution of the mean exhibit's normality.

Fortunately, one doesn't need to repeatedly sample a population to discern the shape of the sampling distribution. Instead, the parameters of the sampling distribution of the mean are contingent upon the parameters of the population itself.

- The mean of the sampling distribution is the mean of the population.

  $\mu_{\bar{x}} = \mu$

- The standard deviation of the sampling distribution is the standard deviation of the population divided by the square root of the sample size.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

We can describe the sampling distribution of the mean using this notation:

$$\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$$

Where:

- $\bar{X}$ is the sampling distribution of the sample means
- ~ means "follows the distribution"
- N is the normal distribution
- μ is the mean of the population
- σ is the standard deviation of the population
- n is the sample size

Sample size, denoted as n, represents the number of observations drawn from the population for each sample, maintaining uniformity across all samples. The sample size significantly influences the sampling distribution of the mean in two key aspects.

1. Sample Size and Normality:

- Larger sample sizes tend to yield sampling distributions that closely adhere to a normal distribution.
- Conversely, with small sample sizes, the sampling distribution of the mean may deviate from normality. This divergence arises because the central limit theorem's validity hinges on having a "sufficiently large" sample size.
- Conventionally, a sample size of 30 or more is considered "sufficiently large."
- When $n<30$, the central limit theorem doesn't apply, and the sampling distribution mirrors the population distribution. Hence, the sampling distribution is only normal if the population distribution is normal.
- Conversely, when $n \geq 30$, the central limit theorem holds true, and the sampling distribution approximates a normal distribution.

2. Sample Size and Standard Deviations:

- The sample size directly impacts the standard deviation of the sampling distribution, reflecting the variability or spread of the distribution.
- With smaller sample sizes, the standard deviation is typically higher, indicating greater variability among sample means due to their imprecise estimation of the population mean.
- Conversely, larger sample sizes correspond to lower standard deviations, indicating less variability among sample means owing to their more accurate estimation of the population mean.

Importance of the Central Limit Theorem:

Parametric tests such as t-tests, ANOVAs, and linear regression possess greater statistical power compared to most non-parametric tests. This enhanced statistical power stems from assumptions regarding the distribution of populations, which are grounded in the central limit theorem.

Continuous distribution

Let's consider the retirement age of individuals in the United States. The population consists of all retired Americans, and the distribution of this population could be represented as follows:

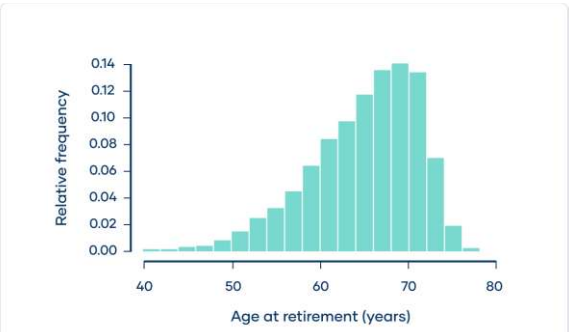

Figure 2.9 Continuous distribution graph.

The distribution of retirement ages skews leftward, with a majority retiring within approximately five years of the mean retirement age of 65 years. However, there exists an extended tail of individuals retiring much earlier, such as at 50 or even 40 years old. The population displays a standard deviation of 6 years.

Imagine conducting a small-scale sampling from this population. Five retirees are randomly selected, and their retirement ages are recorded. For instance: 68, 73, 70, 62, 63

The mean of this sample serves as an approximation of the population mean, albeit with limited precision due to the small sample size of 5. For example: Mean = (68 + 73 + 70 + 62 + 63) / 5 Mean = 67.2 years

Now, suppose this sampling process is repeated 10 times, with each sample comprising five retirees. The mean of each sample is computed, resulting in a distribution known as the sampling distribution of the mean. For instance: 60.8, 57.8, 62.2, 68.6, 67.4, 67.8, 68.3, 65.6, 66.5, 62.1

As this process is repeated numerous times, a histogram depicting the means of these samples will approximate a normal distribution.

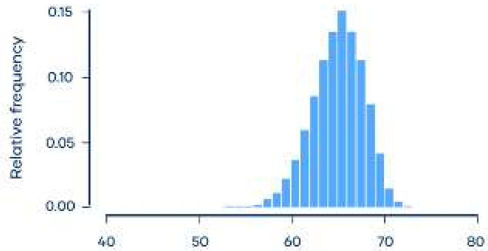

Figure 2.10 Normal distribution of means.

Despite the sampling distribution exhibiting a somewhat more normal shape compared to the population, it still retains a slight leftward skew. Additionally, it's evident that the variability in the sampling distribution is narrower than that of the population.

According to the central limit theorem, the sampling distribution of the mean tends to approximate a normal distribution as the sample size increases. However, the current sampling distribution of the mean deviates from normality due to its relatively small sample size.

2.8 Test statistics

A test statistic represents a numerical value derived from a statistical hypothesis test, indicating the degree of alignment between your observed data and the distribution expected under the null hypothesis of that test.

This statistic plays a crucial role in computing the p-value of your findings, facilitating the determination of whether to accept or reject your null hypothesis.

But what exactly constitutes a test statistic?

A test statistic articulates the similarity between the distribution of your data and the distribution anticipated under the null hypothesis of the statistical test employed. The distribution of data elucidates the frequency of each observation, characterized by its central tendency and the variability around it. Since different statistical tests anticipate different distribution types, selecting the appropriate test aligns with your hypothesis.

The test statistic condenses your observed data into a singular figure, leveraging measures such as central tendency, variability, sample size, and the number of predictor variables in your statistical model.

Typically, the test statistic emerges from discernible patterns in your data (e.g., correlations between variables or discrepancies among groups), divided by the data's variance (i.e., the standard deviation).

Consider this illustration:

You investigate the association between temperature and flowering dates in a specific type of apple tree. Analyzing a comprehensive dataset spanning 25 years, tracking temperature and flowering dates by randomly sampling 100 trees annually from an experimental field.

- Null hypothesis (H_0): No correlation exists between temperature and flowering date.
- Alternate hypothesis (H_A or H_1): A correlation exists between temperature and flowering date.

To scrutinize this hypothesis, you undertake a regression test, yielding a t-value as the test statistic. This t-value juxtaposes the observed correlation between the variables against the null hypothesis of zero correlation.

2.9 Types of test statistics

Outlined below is a synopsis of prevalent test statistics, along with their corresponding hypotheses and the categories of statistical tests in which they are employed. While various statistical tests may employ distinct methodologies for computing these statistics, the fundamental hypotheses and interpretations of the test statistic remain consistent.

Test statistic	Null and alternative hypotheses	Statistical tests that use it
***t* value**	**Null:** The means of two groups are equal **Alternative:** The means of two groups are not equal	• *T* test • Regression tests
***z* value**	**Null:** The means of two groups are equal **Alternative:** The means of two groups are not equal	• *Z* test
***F* value**	**Null:** The variation among two or more groups is greater than or equal to the variation between the groups **Alternative:** The variation among two or more groups is smaller than the variation between the groups	• ANOVA • ANCOVA • MANOVA
X^2-value	**Null:** Two samples are independent **Alternative:** Two samples are not independent (i.e., they are correlated)	• Chi-squared test • Non-parametric correlation tests

In real-world scenarios, you'll typically compute your test statistic using a statistical software package such as R, SPSS, or Excel, which will also furnish the p-value associated with the test statistic. Nevertheless, formulas for manual computation of these statistics can be sourced online.

For instance, in testing your hypothesis concerning temperature and flowering dates, you conduct a regression analysis. The regression test yields: • a regression coefficient of 0.36 • a t-value comparing this coefficient to the anticipated range of regression coefficients under the null hypothesis of no relationship.

The resultant t-value from the regression test, 2.36, represents your test statistic.

2.10 Standard Error

The standard error of the mean (SE or SEM) serves as an indicator of the probable disparity between the population mean and a sample mean. It offers insight into the degree of variability one would anticipate in the sample mean if the study were replicated using fresh samples drawn from the same population.

While the standard error of the mean is the most frequently cited form of standard error, similar measures exist for other statistical parameters such as medians or proportions. Standard error functions as a prevalent gauge of sampling error, depicting the disparity between a population parameter and a sample statistic.

To mitigate standard error, increasing the sample size is recommended. Employing a large, randomized sample serves as the most effective strategy for minimizing sampling bias and enhancing the reliability of findings.

Standard error and standard deviation are both measures of variability:

- The **standard deviation** describes variability **within a single sample**.
- The **standard error** estimates the variability **across multiple samples** of a population.

The standard deviation serves as a descriptive statistic derived directly from sample data, whereas the standard error represents an inferential statistic, typically estimated unless the exact population parameter is known.

2.11 Standard error formula

The standard error of the mean is determined by employing the standard deviation alongside the sample size. Through the formula, it becomes apparent that the sample size and the standard error share an inverse relationship. In simpler terms, as the sample size increases, the standard error decreases. This phenomenon occurs because a larger sample tends to yield a sample statistic closer to the population parameter.

Various formulas are employed based on whether the population standard deviation is known. These formulas are applicable to samples comprising more than 20 elements ($n > 20$).

When population parameters are known

When the population standard deviation is known, you can use it in the below formula to calculate standard error precisely.

Formula	Explanation
$SE = \frac{\sigma}{\sqrt{n}}$	• *SE* is standard error • σ is population standard deviation • *n* is the number of elements in the sample

When population parameters are unknown

When the population standard deviation is unknown, you can use the below formula to only estimate standard error. This formula takes the sample standard deviation as a point estimate for the population standard deviation.

Formula	Explanation
$SE = \frac{s}{\sqrt{n}}$	• *SE* is standard error • *s* is sample standard deviation • *n* is the number of elements in the sample

Example: Using the standard error formula to estimate the standard error for math SAT scores. Follow next two steps.

First, find the square root of your sample size (*n*).

Formula	Calculation
$n = 200$	$\sqrt{n} = \sqrt{200} = 14.1$

Next, divide the sample standard deviation by the number you found in step one.

Formula	Calculation		
$SE = \frac{s}{\sqrt{n}}$	$s = 180$	$\sqrt{n} = 14.1$	$\frac{s}{\sqrt{n}} = \frac{180}{14.1} = 12.8$

The standard error of math SAT scores is 12.8.

You can present the standard error alongside the mean or incorporate it into a confidence interval to convey the uncertainty surrounding the mean.

For instance: Example: Presenting the mean and standard error The mean math SAT score for a random sample of test takers is 550 ± 12.8 (SE).

Reporting the standard error within a confidence interval is preferable as it eliminates the need for readers to perform additional calculations to derive a meaningful range.

A confidence interval denotes a span of values where an unknown population parameter is anticipated to lie most frequently if the study were to be replicated with new random samples.

At a 95% confidence level, it's expected that 95% of all sample means will fall within a confidence interval encompassing ± 1.96 standard errors of the sample mean. This interval serves as an estimate within which the true population parameter is believed to lie with 95% confidence.

For example: Example: Constructing a 95% confidence interval You construct a 95% confidence interval (CI) to estimate the population mean math SAT score. Given a normally distributed characteristic like SAT scores, roughly 95% of all sample means fall within approximately 4 standard errors of the sample mean.

Confidence interval formula

$CI = \bar{x} \pm (1.96 \times SE)$

$\bar{x}$ =sample mean = 550
SE = standard error = 12.8

Lower limit	**Upper limit**
$\bar{x} - (1.96 \times SE)$	$\bar{x} + (1.96 \times SE)$
550 − (1.96 × 12.8) = **525**	550 + (1.96 × 12.8) = **575**

With random sampling, a 95% CI [525 575] tells you that there is a 0.95 probability that the population mean math SAT score is between 525 and 575.

References Chapter 2

- *Introductory Statistics.* Bentham Science Publishers, Kahl, A. (Publish 2023). DOI:10.2174/97898151231351230101

- Introductory Statistics 2e, Openstax, Rice University, Houston, Texas 77005, Jun 23, senior contributing authors: Barbara Illowsky and Susan dean, De anza college, Publish Date: Dec 13, 2023, (https://openstax.org/details/books/introductory-statistics-2e);

- Introductory Statistics 4th Edition, Susan Dean and Barbara Illowsky, Adapted by Riyanti Boyd & Natalia Casper (Published 2013 by OpenStax College) July 2021, (http://dept.clcillinois.edu/mth/oer/IntroductoryStatistics.pdf);

- Journal of the Royal Statistical Society 2024**,** A reputable journal publishing cutting-edge research and articles on various aspects of statistics, including theoretical advancements and practical applications. Recent issues have featured studies on sampling and hypothesis testing.

- Introductory Statistics 7th Edition, Prem S. Mann, eastern Connecticut state university with the help of Christopher Jay Lacke, Rowan university, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030-5774, 2011

- Introduction to statistics, made easy second edition, Prof. Dr. Hamid Al-Oklah Dr. Said Titi Mr. Tareq Alodat, March 2014

- Statistics for Business and Economics, Thirteenth Edition, David R. Anderson, Dennis J. Sweeney, Thomas A. Williams, Jeffrey D. Camm, James J. Cochran, 2017, 2015 Cengage Learning®

- Statistics for Business, First edition, Derek L Waller, 2008 Copyright © 2008, Derek L Waller, Published by Elsevier Inc. All rights reserved

Additional links to literature and Youtube videos Chapter 2

- https://open.umn.edu/opentextbooks/textbooks/196
- https://www.scribbr.com/category/statistics/
- https://stats.libretexts.org/Bookshelves/Introductory_Statistics
- https://assets.openstax.org/oscms-prodcms/media/documents/IntroductoryStatistics-OP_i6tAI7e.pdf
- https://saylordotorg.github.io/text_introductory-statistics/
- https://drive.uqu.edu.sa/_/mskhayat/files/MySubjects/20178FS%20Elementary%20Statistics/Introductory%20Statistics%20(7th%20Ed).pdf
- https://dept.clcillinois.edu/mth/oer/IntroductoryStatistics.pdf
- https://www.geeksforgeeks.org/introduction-of-statistics-and-its-types/
- https://onlinestatbook.com/Online_Statistics_Education.pdf

- https://www.researchgate.net/profile/Tareq-Alodat-2/publication/340511098_INTRODUCTION_TO_STATISTICS_MADE_EASY/links/5e8de3dc4585150839c7b58a/INTRODUCTION-TO-STATISTICS-MADE-EASY.pdf
- https://byjus.com/maths/statistics/
- https://www.khanacademy.org/math/statistics-probability