

IEMS 304 Reference Guide

Marina Siqueira

June 10, 2022

Contents

1	An Introduction to R	2
2	Fundamentals of Random Variables & Distributions	3
2.1	Reference Readings	3
2.2	Review Videos	3
2.3	Definitions	3
2.4	Common Distributions of Random Variables	10
2.4.1	Discrete	10
2.4.2	Continuous	13
2.5	Summary Computations	14
2.6	Summary Properties	14
3	Linear Algebra	15
3.1	Reference Readings	15
3.2	Operations	15
4	Linear Algebra & Random Variables	21
4.1	Manipulations	22
5	Statistics Overview	23
5.1	Reference Readings	23
5.2	Reference Videos	23
5.3	Definitions	23
5.4	Notation	24
5.5	Descriptive Statistics Formulas	25
5.6	Properties of Summations	25

6	Statistical Tests	26
6.1	Reference Readings	26
6.2	Reference Videos	26
6.3	Reference Formulas	26
6.4	Sums of Squares	27
6.5	Overview of Single Sample Hypothesis Testing	28
6.6	Z Test on Population Mean	29
6.7	One-Sample t Test	30
6.8	Error Types	31

How to Use This Reference Guide

IEMS 304 extensively builds on knowledge from previous probability and statistics courses. The purpose of this guide is to help provide a review of the key terms and concepts which will support your understanding of the material in this class. The guide is designed for your quick review of the topics and visuals to refresh your memory on the examples and terminology, and give you valuable additional resources such as videos or textbooks if you feel you need more in-depth exposure to certain subjects.

We recommend you start by looking through the table of contents and checking if any of our headings sound unfamiliar or distant in memory. We'd recommend you look through those sections first; there is no need to read this guide in order or in its entirety. This guide was not intended for a thorough read from front to back, but you are certainly welcome to do so if you wish. We encourage you to adapt this guide to your specific background in statistics, probability and linear algebra and to refer to it throughout the course as needed.

Good luck with IEMS 304 and hope you enjoy the read!

Sources

Many examples are adapted from [Probability and Statistics For Engineering and the Sciences](#) by Devore and Peck. Images found on the web are referenced by superscripts in-text with the appropriate corresponding source in the end-notes.

Acknowledgments

This was developed under the guidance of Jill Wilson with significant input and course materials from Barry Nelson, Bruce Ankeman, Edward Malthouse, Daniel Apley and Zhaoran Wang. We are also grateful for the feedback from Charlotte Oxnam, Alp Guneyman and Ruramai Zimuto.

1 An Introduction to R

R is a language and environment for statistical which provides a wide variety of statistical (linear and nonlinear modelling, classical statistical tests, time-series analysis, classification, clustering, ...) and graphical techniques and is used throughout IEMS 304. If you have not used R before, follow this [link](#) to install R and this [link](#) to install RStudio.

Here are a few more resources to get acquainted with R. Note that you do not have to use all of them - browse our suggestions and choose the option that best fits your learning style.

1. [Swirl Tutorial](#): built-in mini-courses that can be studied in-action directly from RStudio. You get to choose which courses you would like to install, and we recommend “R Programming E”. Simply run the following commands from your RStudio Console:

```
install.packages( 'swirl ' )  
library( swirl )  
install_course( 'R Programming E' )  
swirl()
```

2. [Data Camp](#): free interactive online modules on R through a browser window.
3. [Python to R](#) posts written to help make the transition between Python code and R code. Might be helpful if you have worked with Python previously.
4. [R Cheat Sheets](#): Short, consolidated reference sheets on R packages. We've provided the “Base-R” sheet as it provides the basic commands for beginners.

2 Fundamentals of Random Variables & Distributions

This section provides a brief overview of the most common discrete and continuous distributions, and provides reference formulas to compute values from distributions. Note that we use capital letters (e.g., X) for a random variable and lower case (e.g., x) for a particular numerical value. For example, $P(X = x)$ is the appropriate notation for the probability that the random variable called X takes the value x .

2.1 Reference Readings

[Introduction to Statistics and Data Analysis](#) Ch 7.1-7.3, 7.6-7.7, 8.1-8.2

[Probability and Statistics For Engineering and the Sciences](#) Ch 2-5

[Applied Probability and Statistics For Engineers](#) Ch 2-5

2.2 Review Videos

- [Random Variables](#)
- [Joint PMFs](#)
- [Conditional PMFs](#)

2.3 Definitions

- **Random variable:** a function X which assigns a real number to each element in the sample space
- **Discrete random variables:** a random variable X is defined to be discrete if its probability space is either finite or countable, meaning it takes only a finite or countable number of values.
 - ✎ For example, a random variable that only takes values 0, 1 and 2 is discrete. As a more realistic example, the two outcomes heads and tails of a coin flip can be modeled by a discrete random variable.
- **Continuous random variables:** a random variable which can take a possibly infinite set of values in an interval, and no single value has nonzero probability. If X is a continuous random variable, then $P(X = c) = 0$ for any possible c value.

✎ For example, a random variable that takes any real number in the interval $(3, 5]$ is continuous. As another example, people's heights can be modeled as continuous random variables.

- **Probability mass function (pmf):** function $p_X(x)$ that describes the probability that a discrete random variable X takes each of its values from its set of possible (countable) values.

$$P(X = x) = p_X(x).$$

Note that summing the pmf for all possible values of X equals 1: $\sum_x p_X(x) = 1$.

✎ For example, if we let X be the value of a single die roll, then X can take any value in the set $\{1, 2, 3, 4, 5, 6\}$, each with uniform probability $P(X = 1) \dots = P(X = 6) = 1/6$. The pmf describes this mapping between all the possible values the random variable X can take and the probability that X is that value.

✎ We often denote pmfs as piece-wise due to the discrete values the random variables take:

$$p_X(x) = \begin{cases} 1/6 & \text{if } x = 1 \\ 1/6 & \text{if } x = 2 \\ 1/6 & \text{if } x = 3 \\ 1/6 & \text{if } x = 4 \\ 1/6 & \text{if } x = 5 \\ 1/6 & \text{if } x = 6 \\ 0 & \text{otherwise.} \end{cases}$$

✎ We can also visualize pmfs on a number line. The height depicts the probability the discrete random variable takes on the value shown on the x axis. In Figure 1, the pmf shows that the discrete random variable takes the value 1 with probability 0.2, 3 with probability 0.5 and 7 with probability 0.3:

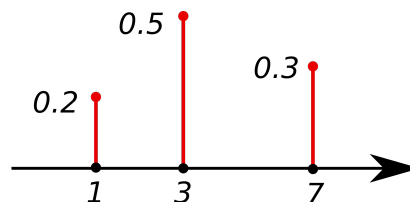


Figure 1: Illustration of a pmf.¹

- **Probability density function (pdf):** function $f(x)$ that describes the probability that a random variable takes on values between an interval. In other words, for $a \leq b$:

$$P(a \leq X \leq b) = \int_a^b f(x)dx.$$

✎ Example pdf:

$$f(x) = \begin{cases} \frac{1}{24}(x^2 + 1) & 1 \leq x \leq 4 \\ 0 & \text{otherwise.} \end{cases}$$

✎ We can also visualize pdfs. The area between two points on the x axis indicates the probability the random variable takes a value in the interval between the two points. For example, Figure 2 shows that the probability that a normal random variable takes a value between -1σ and 1σ is 68.27%:

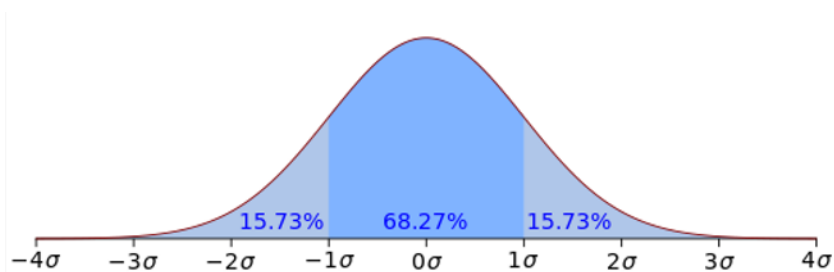


Figure 2: Illustration of the normal random variable pdf.²

- **Cumulative distribution function (cdf):** the probability that observed values from a random variable will be at most x : $F(x) = P(X \leq x)$

Discrete random variables:

$$F(x) = P(X \leq x) = \sum_{y:y \leq x} p(y).$$

Graphically, this is the area underneath the pmf for values less than or equal to x .

Continuous random variables:

$$F(a) = P(X \leq a) = \int_{-\infty}^a f(x)dx.$$

Graphically, this is the area underneath the pdf for values less than or equal to x . As the cdf is the area underneath the curve, then it is the integral of the pdf. This also

means that the cdf can be differentiated to compute the pdf:

$$\frac{d}{dx}F_X(x) = f_X(x).$$

See Figure 3 below for the visual connection between the pdf and the cdf.

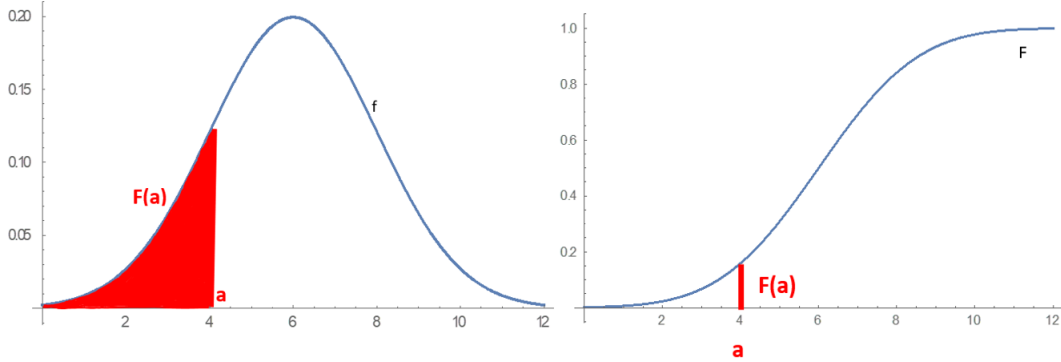


Figure 3: Illustration of the relationship between the pdf and the cdf for continuous random variables. The left graph plots the pdf, where the shaded area underneath the probability density function is the value of the cdf at that x-axis value, which is labeled as a . On the right, we see a plot of the cdf itself, where each value along the y axis is the area under the pdf curve for all values less than or equal to the x-axis value.³

📌 Important facts about cdfs:

- * $F(x)$ is a monotonically non-decreasing function: If $x_1 \leq x_2$, it follows that $F(x_1) \leq F(x_2)$
 - * $F(x) \rightarrow 0$ as $x \rightarrow -\infty$
 - * $F(x) \rightarrow 1$ as $x \rightarrow \infty$
 - * $F(x)$ is continuous from the right
 - * $0 \leq F(x) \leq 1$ for all $x \in \mathcal{R}$
- **Expected value:** an operator which sums all of the possible values in the sample space multiplied by the probability of each value's occurrence. This is a probability-centered term since we weight each value of an outcome by the probability it occurs, which requires knowing the underlying probability distribution. In reality, we expect the arithmetic mean to converge to the expected value over a large number of samples. Expected value can be computed for both discrete and continuous random variable:

$$E[X] = \begin{cases} \sum_x xp(x) & \text{where } X \text{ is discrete with pmf } p(x) \\ \int_{-\infty}^{\infty} xf(x)dx & \text{where } X \text{ is continuous with pdf } f(x). \end{cases}$$

✎ From Figure 1, we can compute the expected value of the random variable with the given distribution as:

$$E[X] = \sum_x xp(x) = 1 \cdot 0.2 + 3 \cdot 0.5 + 7 \cdot 0.3 = 3.8.$$

✎ From the example pdf shown above where:

$$f(x) = \begin{cases} \frac{1}{24}(x^2 + 1) & 1 \leq x \leq 4 \\ 0 & \text{otherwise.} \end{cases}$$

We compute the expected value as:

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx = \int_1^4 x \cdot \frac{1}{24}(x^2 + 1) dx = \frac{1}{24} \left(\int_1^4 x^3 dx + \int_1^4 x dx \right) = \frac{95}{32}.$$

Note that we can also take the expectation of functions of random variables:

$$E[h(X)] = \begin{cases} \sum_x h(x)p(x) & \text{where } X \text{ is discrete with pmf } p(x) \\ \int_{-\infty}^{\infty} h(x)f(x)dx & \text{where } X \text{ is continuous with pdf } f(x). \end{cases}$$

- **Variance:** a measurement of the spread of data around the mean. It is computed using either of the two formulas below:

$$\text{Var}(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2.$$

✎ From Figure 1, we can compute the variance of the random variable as:

$$\text{Var}[X] = E[(X - E[X])^2] = 0.2 \cdot (1 - 3.8)^2 + 0.5 \cdot (3 - 3.8)^2 + 0.3 \cdot (7 - 3.8)^2 = 4.96.$$

✎ From the example pdf, we can compute the variance as:

$$\begin{aligned} \text{Var}(X) &= E[X^2] - (E[X])^2 \\ &= \int_{-\infty}^{\infty} x^2 f(x) dx - \left(\int_{-\infty}^{\infty} x f(x) dx \right)^2 \\ &= \int_1^4 x^2 \cdot \frac{1}{24}(x^2 + 1) dx - \left(\frac{95}{32} \right)^2 \\ &= \left(\frac{47}{5} \right) - \left(\frac{95}{32} \right)^2 \approx 0.5865. \end{aligned}$$

- **Standard Deviation:** a measure of average deviation from the mean, computed by taking the square root of the variance.
- **Covariance:** a measure of a linear relationship between two variables. It is computed through:

$$\text{Cov}_{XY} = \sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)].$$

A visual intuition of covariance is shown below in Figure 4:

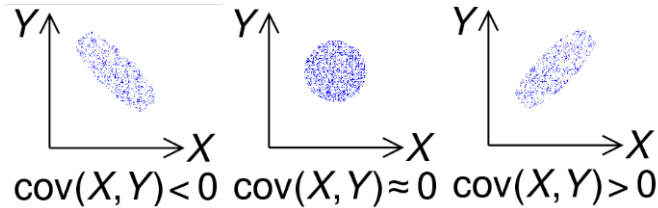


Figure 4: Visual representation of covariance between two random variables X and Y .⁴

- **Correlation:** a measure of the degree to which a pair of variables have a linear relationship. Correlation is essentially computed in the same way as covariance, but it is normalized for the standard deviations of the two random variables. Since we divide the covariance by the standard deviations, correlation values are between -1 and 1 , while covariances can be > 1 or < -1 . Correlation is computed through:

$$\text{Corr}_{XY} = \rho_{XY} = E[(X - \mu_X)(Y - \mu_Y)] / (\sigma_X \sigma_Y).$$

Note that, by our previous definitions,

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}.$$

- **Uncorrelated random variables:** random variables which have no **linear** relationship between them. Uncorrelated random variables have a correlation coefficient of zero, assuming none of the random variables have zero variance. For uncorrelated random variables, it follows that $E[XY] = E[X]E[Y]$.
- **Independent random variables:** random variables are independent if knowing information about one random variable does not affect the probability distribution for the other random variable. For independent random variables, it follows that $E[f(X)g(Y)] = E[f(X)]E[g(Y)]$ and $E[X|Y] = E[X]$.
- **Law of Total Probability:** Given A_1, A_2, \dots, A_n are disjoint events that partition the sample space and $P(A_i) > 0$ for all i , for any event B , the Law of Total Probability

states that:

$$\begin{aligned} P(B) &= P(A_1 \cap B) + \cdots + P(A_n \cap B) \\ &= P(A_1)P(B | A_1) + \cdots + P(A_n)P(B | A_n) \end{aligned}$$

- **Joint Distribution:** the joint distribution for two random variables describes the probability that two random variables take on values jointly

Discrete joint pmf for random variables X and Y :

$$P(X = x, Y = y) = p_{X,Y}(x, y)$$

Continuous joint pdf for random variables X and Y :

$$P(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f_{X,Y}(x, y) dy dx$$

- **Marginal Distribution:** the marginal distribution of a random variable is the distribution of the random variable alone, unconditional on other random variables.

Discrete marginal pmfs for random variables X and Y :

$$p_X(x) = \sum_y p_{X,Y}(x, y) \quad \text{and} \quad p_Y(y) = \sum_x p_{X,Y}(x, y)$$

Continuous marginal pdfs for random variables X and Y :

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy \quad \text{and} \quad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx$$

- **Conditional Distribution:** the distribution of a random variable when given another random variable

Discrete conditional distribution for random variables X and Y :

$$p_{X|Y}(x | y) = \frac{p_{X,Y}(x, y)}{p_Y(y)}, \quad p_Y(y) > 0$$

Continuous conditional distribution for random variables X and Y :

$$f_{X|Y}(x | y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}, \quad f_Y(y) > 0$$

- **Law of Total Probability for Distributions:**

$$p_X(x) = \sum_y p_{X,Y}(x, y) = \sum_y p_{X|Y}(x | y)p_Y(y)$$

$$f_X(x) = \int_y f_{X,Y}(x, y)dy = \int_y f_{X|Y}(x | y)f_Y(y)dy$$

- **Conditional Expectation:** taking the expectation of a jointly distributed random variable by using the conditional distribution.

Discrete conditional expectation

$$E[X | Y = y] = \sum_x Xp_{X|Y}(x | y)$$

Continuous conditional expectation

$$E[X | Y = y] = \int_x Xf_{X|Y}(x | y)dx$$

2.4 Common Distributions of Random Variables

2.4.1 Discrete

As a reminder, discrete random variables can take on a countable number of values. A summary table is provided on the next page.

- **Bernoulli:** a random variable that only takes the values 0 and 1, where the random variable is 1 with probability p and 0 with probability $(1 - p)$.

✎ A single coin flip where heads = 1 and tails = 0 is a Bernoulli random variable with $p = 0.5$.

- **Binomial:** a random variable that describes the outcome of repeating a success or failure Bernoulli variable a set number of times. Binomial random variables take two parameters: the n number of repetitions of the success or failure experiment and the p probability of success.

✎ The number of heads in a 5 coin flip experiment where “success” is defined as heads and “failure” is defined as tails follows a binomial distribution with $n = 5$ and $p = 0.5$.

- **Poisson:** a random variable that models count data in a fixed time or space. The Poisson only has a λ mean parameter.

- ✎ A random variable that describes the number of spelling mistakes in a page of writing can follow a Poisson distribution with mean $\lambda = 2.1$ mistakes per page.
- **Geometric:** a random variable that describes the number of failures until a success. A Geometric distribution has a single parameter p , which is the probability of success.
- ✎ A random variable that counts the number of failures until a player scores a free throw is described by a geometric random variable with a given p probability of success.

Summary of Discrete Distributions

Distribution	Parameters	Uses	Mean and Variance	PDF and CDF
Bernoulli	p : success probability	Random variable that only takes the values 0 and 1.	$E[X] = p$ $\text{Var}[X] = p(1 - p)$	$P(x) = \begin{cases} p & \text{if } x = 1 \\ 1 - p & \text{if } x = 0 \end{cases}$
Binomial	n : number of trials p : success probability	Random variable that describes the outcome of repeating a success or failure Bernoulli variable a set number of times.	$E[X] = np$ $\text{Var}[X] = np(1 - p)$	$P(x) = \binom{n}{x} p^x q^{n-x} = \frac{n!}{(n-x)!x!} p^x q^{n-x}$ $F(x) = \sum_{i=0}^k \binom{n}{i} p^i (1 - p)^{n-i}$
Poisson	λ : mean	Random variable that models count data in a fixed time or space.	$E[X] = \lambda$ $\text{Var}[X] = \lambda$	$P(x) = \frac{e^{-\lambda} \lambda^x}{x!}$ $F(x) = \sum_{i=0}^k \frac{e^{-\lambda} \lambda^i}{i!}$
Geometric	p : success probability	Random variable that describes the number of failures until a success.	$E[X] = 1/p$ $\text{Var}[X] = (1 - p)/p^2$	$P(x) = (1 - p)^{k-1} p$ $F(x) = 1 - (1 - p)^k$

2.4.2 Continuous

- **Normal:** a probability distribution that is symmetric about the mean and has a bell-shaped pdf. The normal distribution is described by two parameters: a mean μ and a variance σ^2 .
- **Uniform:** a distribution in which every interval of the same length is equally likely. The probability that a random variable takes a value inside an interval is proportional to the length of the interval.
- **Exponential:** Distribution that describes the time between events in a Poisson process. Takes in a single parameter μ which is the expected time until the next event.

Summary of Continuous Distributions

Distribution	Parameters	Uses	Mean and Variance	PDF and CDF
Normal	μ : mean σ^2 : variance	Random variable that follows a distribution that is symmetric about the mean and has a bell-shaped pdf.	$E[X] = \mu$ $\text{Var}[X] = \sigma^2$	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$
Uniform	a, b : bounds	Random variable that has probability it takes a value inside an interval is proportional to interval length.	$E[X] = (a + b)/2$ $\text{Var}[X] = (b - a)^2/12$	$f(x) = \frac{1}{b-a} \quad a \leq x \leq b$ $F(x) = \frac{x-a}{b-a} \quad a \leq x \leq b$
Exponential	λ : average rate	Random variable that describes the time between events in a Poisson process.	$E[X] = 1/\lambda$ $\text{Var}[X] = 1/\lambda^2$	$f(x) = \lambda e^{-\lambda x}$ $F(x) = 1 - e^{-\lambda x} \quad a \leq x \leq b$

2.5 Summary Computations

	Discrete Random Variables	Continuous Random Variables
Expected Value	$\mu = \sum_{i=1}^n xf(x)$	$\mu = \int_{-\infty}^{\infty} xf(x)dx$
Variance	$\sigma^2 = \sum_{i=1}^n (x - \mu)^2 f(x)$	$\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx$
Standard deviation	$\sigma = \sqrt{\sigma^2}$	$\sigma = \sqrt{\sigma^2}$

2.6 Summary Properties

Given a and b are constants and X and Y are random variables, the following statements hold:

$$E[a] = a$$

$$E[aX] = aE[X]$$

$$E[aX + bY] = aE[X] + bE[Y]$$

$$E[aX + Y] = aE[X] + E[Y]$$

$$\text{Var}(a) = 0$$

$$\text{Var}(aX) = a^2 \text{Var}(X)$$

$$\text{SD}(aX) = |a|\text{SD}(X)$$

$$\text{Cov}(X, Y) = E[XY] - E[X]E[Y]$$

$$\text{Cov}(X, a) = 0$$

$$\text{Var}(aX \pm bY) = a^2 \text{Var}(X) + b^2 \text{Var}(Y) \pm 2ab \text{Cov}(X, Y)$$

$$\text{Cov}(aX, bY) = ab \text{Cov}(X, Y)$$

$$E(X) = E[E(X | Y)]$$

$$\text{Var}(X) = E[\text{Var}(X | Y)] + \text{Var}[E(X | Y)]$$

If X and Y are uncorrelated, then:

$$E[XY] = E[X]E[Y]$$

$$\text{Cov}(X, Y) = 0$$

$$\text{Var}(aX + bY) = a^2 \text{Var}(X) + b^2 \text{Var}(Y)$$

3 Linear Algebra

In the previous review section, the formulas and definitions primarily focused on one-dimensional applications. Now, we turn our attention to the higher-dimensional space. In IEMS 304, it will be important to understand and work with variables and parameters that are scalars, vectors, and matrices.

3.1 Reference Readings

[Applied Linear Algebra](#) Ch 1.2, 1.5, 1.6, 2.3


[Engineering Analysis I Textbook](#) Ch 1.1, 1.7, 2.1-2.3, 2.9, 3.1, 3.2, 4.3 and 4.6

3.2 Operations

Given an $m \times n$ matrix A :


- **Transpose A^T** : matrix formed by switching the row and column indices of A such that the new matrix is flipped across the diagonal.

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{bmatrix}$$

 **Example:**

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}, \quad A^T = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix}$$

- **Identity matrix I** : matrix with 1's along the diagonals and 0's in all off-diagonals. Useful for manipulations and transformations with matrices, and is similar to multiplying by 1 in scalar algebra. Note that given a matrix A , $AI = IA = A$.

 **Example:**


$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- **Matrix-Matrix Multiplication:** given a $m \times n$ matrix A and $n \times p$ matrix B , then

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & b_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{np} \end{pmatrix}$$

$$\mathbf{AB} = \begin{pmatrix} a_{11}b_{11} + \cdots + a_{1n}b_{n1} & a_{11}b_{12} + \cdots + a_{1n}b_{n2} & \cdots & a_{11}b_{1p} + \cdots + a_{1n}b_{np} \\ a_{21}b_{11} + \cdots + a_{2n}b_{n1} & a_{21}b_{12} + \cdots + a_{2n}b_{n2} & \cdots & a_{21}b_{1p} + \cdots + a_{2n}b_{np} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}b_{11} + \cdots + a_{mn}b_{n1} & a_{m1}b_{12} + \cdots + a_{mn}b_{n2} & \cdots & a_{m1}b_{1p} + \cdots + a_{mn}b_{np} \end{pmatrix}$$

Note that the inner dimensions of \mathbf{A} and \mathbf{B} must be equal (e.g., the number of columns in \mathbf{A} must be equivalent to the number of columns in \mathbf{B}).

 **Example:**

$$\begin{bmatrix} 3 & 4 \\ 2 & 1 \end{bmatrix} * \begin{bmatrix} 1 & 5 \\ 3 & 7 \end{bmatrix} = \begin{bmatrix} 3+12 & 15+28 \\ 2+3 & 10+7 \end{bmatrix} = \begin{bmatrix} 15 & 43 \\ 5 & 17 \end{bmatrix}$$

- **Inverse A^{-1} :** given a square matrix A , its inverse A^{-1} is the matrix such that $AA^{-1} = \mathbf{I}$. The example shown below is for a 2×2 matrix, but square matrices with any $n \times n$ dimensions may be invertible.

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

$$A^{-1} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$



$$A = \begin{bmatrix} 4 & 3 \\ 3 & 2 \end{bmatrix} \quad A^{-1} = \frac{1}{4*2 - 3*3} \begin{bmatrix} 2 & -3 \\ -3 & 4 \end{bmatrix} = \begin{bmatrix} -2 & 3 \\ 3 & -4 \end{bmatrix}$$

Verifying that AA^{-1} results in the identity matrix:

$$AA^{-1} = \begin{bmatrix} 4 & 3 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} -2 & 3 \\ 3 & -4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

- **Determinant:** scalar computed for a **square** matrix that describes many matrix

properties. Note that if the determinant of a matrix is 0, then the matrix has no inverse (e.g., A^{-1} does not exist) and the matrix is called **singular**.


For 2×2 matrices, the determinant can be computed by finding the difference between the product of the diagonals:

$$\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - cb$$

For 3×3 matrices, the determinant requires computing the determinant of three sub-matrices:

$$\det \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = a \cdot \det \begin{bmatrix} e & f \\ h & i \end{bmatrix} - b \cdot \det \begin{bmatrix} d & f \\ g & i \end{bmatrix} + c \cdot \det \begin{bmatrix} d & e \\ g & h \end{bmatrix}$$

This may be familiar to you from the cross product of i, j, k vectors from calculus.

 From the previous example where $A = \begin{bmatrix} 4 & 3 \\ 3 & 2 \end{bmatrix}$, the determinant was the difference between the product of the diagonals: $4 * 2 - 3 * 3 = -1$


However, we can also consider a matrix which has zero determinant:

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$$

$$\det(A) = 1 * 4 - 2 * 2 = 0$$

As the procedure to find A^{-1} requires finding $1/\det(A)$, we can reason that $\det(A) = 0$ would result in division by zero and therefore an inverse cannot be found.

- **Linear independence:** vectors v_1, v_2, \dots, v_n are linearly dependent if any of the vectors can be written as a linear combination of the other vectors. A square matrix that does not have all linearly independent rows and columns is singular.

 The following column vectors are linearly independent because no vector can be written as a linear combination of other vectors:

$$\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ -2 \end{bmatrix}, \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$$

In other words, there are no such constants c_1 and c_2 that would satisfy any:

$$c_1 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ 2 \\ -2 \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$$

$$c_1 \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \\ -2 \end{bmatrix}$$

$$c_1 \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} + c_2 \begin{bmatrix} 0 \\ 2 \\ -2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Note that, if there exists a c_1, c_2 pair such that one of the vectors can be written as a linear combination of the others, then we can find constants to write any vector as a linear combination of the other vectors.

However, if we include one more vector to the set, we see that the vectors are linearly dependent:

$$\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ -2 \end{bmatrix}, \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}, \begin{bmatrix} 4 \\ 2 \\ 3 \end{bmatrix}$$

We can find the constants 9, 5, 4 which express the fourth vector as a linear combination of the other three:

$$9 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} + 5 \begin{bmatrix} 0 \\ 2 \\ -2 \end{bmatrix} + 4 \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \\ 3 \end{bmatrix}$$

- **Column/Row Rank:** the number linearly independent column/row vectors in a matrix. If all of the columns in a matrix are linearly independent, we say that the matrix has **full column rank**.

✎ For example, the third column of matrix A below can be written as the sum of columns 1 and 2. Columns 1 and 2, or 2 and 3, or 1 and 3 considered separately are linearly independent, but no subset of three columns in matrix A are linearly independent. For this reason, $\text{rank}(\mathbf{A}) = 2$:

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & -3 & 1 \\ 4 & 1 & 7 \end{bmatrix}$$

✎ On the other hand, matrix B below has all three linearly independent columns. For this reason, matrix B has $\text{rank}(\mathbf{B}) = 3$ and is a full column rank matrix:

$$\mathbf{B} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & -3 & 1 \\ 4 & 1 & 1 \end{bmatrix}$$

- **Eigenvalues:** a λ is an eigenvalue of a matrix A if it satisfies the equation $\mathbf{A}x = \lambda x$ where x is the eigenvalue's corresponding eigenvector. To compute the eigenvalues of a matrix, take $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$ and solve for λ .

✎ Given $A = \begin{bmatrix} 1 & 4 \\ 3 & 2 \end{bmatrix}$, we solve for the eigenvalues λ :

$$\begin{aligned} \det(A - \lambda I) &= 0 \\ \det\left(\begin{bmatrix} 1 & 4 \\ 3 & 2 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) &= 0 \\ \det\left(\begin{bmatrix} 1 - \lambda & 4 \\ 3 & 2 - \lambda \end{bmatrix}\right) &= 0 \\ (1 - \lambda)(2 - \lambda) - 12 &= 0 \\ \lambda^2 - 3\lambda - 10 &= 0 \\ (\lambda - 5)(\lambda + 2) &= 0 \\ \lambda &= 5, -2 \end{aligned}$$

- **Eigenvectors:** a x is an eigenvector of a matrix A if it satisfies the equation $\mathbf{A}x = \lambda x$ where λ is the corresponding eigenvalue. We solve for the eigenvector x by setting $(\mathbf{A} - \lambda\mathbf{I})x = 0$.

✎ Following from the previous example, we set $(\mathbf{A} - \lambda\mathbf{I})x = 0$:

$$\lambda = 5$$
$$\begin{bmatrix} 1 - \lambda & 4 \\ 3 & 2 - \lambda \end{bmatrix} x = 0, \quad \begin{bmatrix} -4 & 4 \\ 3 & -3 \end{bmatrix} x = 0, \quad x = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\lambda = -2$$
$$\begin{bmatrix} 1 - \lambda & 4 \\ 3 & 2 - \lambda \end{bmatrix} x = 0, \quad \begin{bmatrix} 3 & 4 \\ 3 & 4 \end{bmatrix} x = 0, \quad x = \begin{bmatrix} -4 \\ 3 \end{bmatrix}$$

4 Linear Algebra & Random Variables

Let each X_i for $i = 1, 2, \dots, n$ be a Random Variable. We define the following:

- **Random vector:** vector of random variables. For example, each X_i can be a normally distributed random variable.

$$\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$$

- **Expectation of random vector:** vector of the expectations of each random variable X_i .

$$E[\mathbf{X}] = \begin{pmatrix} E[X_1] \\ \vdots \\ E[X_n] \end{pmatrix}$$

- **Variance Matrix:** n by n square matrix in which each i, j entry is $\text{cov}(X_i, X_j)$. I think it is more common to call this $\text{Var}(\mathbf{X})$ where \mathbf{X} is a vector; the covariance matrix is $\text{Cov}(\mathbf{X}, \mathbf{Y})$ where both are vectors. This way they are both consistent with the scalar case. Note at the bottom of p. 20 there is the same result twice stated as Var and Cov .

$$\begin{aligned} \text{Var}(\mathbf{X}) &\equiv [\text{cov}(X_i, X_j)] \\ &\equiv \begin{pmatrix} \text{var}(X_1) & \text{cov}(X_1, X_2) & \cdots & \text{cov}(X_1, X_n) \\ \text{cov}(X_2, X_1) & \text{var}(X_2) & \cdots & \text{cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(X_n, X_1) & \text{cov}(X_n, X_2) & \cdots & \text{var}(X_n) \end{pmatrix} \end{aligned}$$

Note three key properties of the covariance matrix:

- The matrix is **symmetric**. Each i, j entry is equal to each j, i entry.
- The diagonals are **variances**. Each i, i entry is the variance of X_i , and the square root of each diagonal is the standard deviation of X_i .
- The matrix is **positive semidefinite**. This means that all of its eigenvalues are nonnegative.

For example, the following illustrative covariance matrix allows us to identify the variance of x_1 and x_2 across the diagonals, and the covariance between x_1 and x_2 on the off-diagonals. We see that the variance of x_1 is 2.5, the variance of x_2 is 10 and the covariance of x_1 and x_2 is 3.0:

$$\text{Var}(\mathbf{X}) = \begin{pmatrix} 2.5 & 3.0 \\ 3.0 & 10.0 \end{pmatrix} = \begin{pmatrix} S_{x_1}^2 & S_{x_1x_2} \\ S_{x_1x_2} & S_{x_2}^2 \end{pmatrix}$$

- Correlation Matrix: n by n square matrix in which each i, j entry is $\text{corr}(X_i, X_j)$.

$$\begin{aligned} \text{corr}(\mathbf{X}) &\equiv [\text{corr}(X_i, X_j)] \\ &\equiv \begin{pmatrix} 1 & \text{corr}(X_1, X_2) & \cdots & \text{corr}(X_1, X_n) \\ \text{corr}(X_2, X_1) & 1 & \cdots & \text{corr}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{corr}(X_n, X_1) & \text{corr}(X_n, X_2) & \cdots & 1 \end{pmatrix} \end{aligned}$$

Note the correlation matrix is symmetric, positive semidefinite, and diagonals are 1. For example, the following illustrative correlation matrix allows us to identify that the correlation between x_1 and x_2 is 0.45 by looking at the off-diagonal elements.

$$\text{Corr}(\mathbf{X}) = \begin{pmatrix} 1 & 0.45 \\ 0.45 & 1 \end{pmatrix} = \begin{pmatrix} \rho_{x_1}^2 & \rho_{x_1x_2} \\ \rho_{x_1x_2} & \rho_{x_2}^2 \end{pmatrix}$$

4.1 Manipulations

Let \mathbf{X} be a $n \times 1$ column vector of random variables and A be a $n \times n$ matrix of constants. These properties and identities can be useful when deriving estimators.

- $E[A\mathbf{X}] = AE[\mathbf{X}]$
- $\text{cov}(A\mathbf{X}) = A \text{cov}(\mathbf{X})A^T$
- $\text{var}(A\mathbf{X}) = A \text{var}(\mathbf{X})A^T$
- $\mathbf{X}\mathbf{X}^{-1} = \mathbf{I}$ and $\mathbf{X}\mathbf{X}^{-1}\mathbf{Y} = \mathbf{Y}$

5 Statistics Overview

The section on the statistics glossary is a quick reference to terminology on descriptive statistics which you have likely encountered before. We make a clear distinction between the entire population of interest and the sample which we are able to gather data on, and summarize the notation used for key parameters and statistics as well as their formulas.

5.1 Reference Readings

[Probability and Statistics For Engineering and the Sciences](#) Ch 1.1, 1.3-1.4

[Applied Probability and Statistics For Engineers](#) Ch 6.1

5.2 Reference Videos

- [Basic Introduction](#)
- [Measures of Dispersion](#)

5.3 Definitions

Definitions based on [Probability and Statistics For Engineering and the Sciences](#).

- **Population:** collection of all objects that are of interest to a problem. Note that, in this context, a population does not necessarily have to be composed of people. For example, a population can be all computer chips produced by a factory.
- **Sample:** a subset of the whole population of interest. This is often a subset of the population which data is gathered on.
- **Variable:** a characteristic that may change between items in the population. For example, the thickness of a computer chip.
- **Inferential Statistics:** the branch of statistics that is concerned with generalizing information from a sample to the actual whole population.
- **Parameter:** true value that numerically characterizes a population value. This is often not possible to know exactly in practice.
- **Statistic:** numerical fact about a sample that approximates a parameter. For example, the average age computed from a sample to approximate the mean age of a population.

- **Mean:** measure of center computed as the arithmetic average of a set. We make the distinction between the sample mean, which is the arithmetic mean for a subset of the population, and the population mean, which is the true mean of all elements in the population.
- **Variance:** a measure of dispersion computed by finding each observation's squared deviation from the mean.
- **Standard deviation:** average deviation from the mean, which is the square root of the variance.

5.4 Notation

In IEEMS 304, it will be important to identify whether we are referring to true parameters for the population or to sample statistics. This distinction is made through notation, and this will be helpful to understand proofs and derivations.

Note: this notation is common but not universally used. Verify the context of use.

Name	Parameter	Statistic
Mean	μ	\bar{x}
Total	N	n
Proportion (percent)	p	\hat{p}
Standard deviation	σ	s
Variance	σ^2	s^2
Covariance	σ_{ij}	s_{ij}
Correlation	ρ	r
Intercept	β_0	b_0
Slope	β_1	b_1

5.5 Descriptive Statistics Formulas

Name	Population	Sample
Mean	$\mu = \frac{\sum_{i=1}^N x_i}{N}$	$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$
Variance	$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N}$	$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$
Standard deviation	$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}}$	$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$
Covariance	$\sigma_{ij} = \frac{\sum (x_i - \bar{x})(y_j - \bar{y})}{n}$	$s_{xy} = \frac{\sum (x_i - \bar{x})(y_j - \bar{y})}{(n - 1)}$
Correlation	$\rho_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j}$	$r_{xy} = \frac{s_{xy}}{s_x s_y}$

5.6 Properties of Summations

$$E \left[\sum_{i=1}^n X_i \right] = \sum_{i=1}^n E(X_i)$$

$$\text{Cov} \left(\sum_{i=1}^n X_i, \sum_{j=1}^m Y_j \right) = \sum_{i=1}^n \sum_{j=1}^m \text{Cov}(X_i, Y_j)$$

If X_i are independent random variables:

$$\text{Var} \left(\sum_{i=1}^n X_i \right) = \sum_{i=1}^n \text{Var}(X_i)$$

6 Statistical Tests

To evaluate models in IEMS 304, it will be important to understand which statistical tests to apply where and how to execute them. Read this section for an overview of statistical tests and their formulas.

6.1 Reference Readings

[Introduction to Statistics and Data Analysis](#) Ch 8.3, 9.2-9.4, 10.2

[Probability and Statistics For Engineering and the Sciences](#) Ch 7-9

[Applied Probability and Statistics For Engineers](#) Ch 7-10

6.2 Reference Videos

- [Distribution of Sample Mean](#) (12 minutes)
- [Properties of Estimators](#) (6 minutes)
- [Normal Distribution and the z-table](#) (13 minutes)
- [Normal Distribution and the z-table Example \(Test Scores\)](#) (3 minutes)
- [A Confidence Interval for the Mean \(Potato Chip Example\)](#) (12 minutes)
- [Interpreting Confidence Intervals](#) (9 minutes)
- [A Confidence Interval for the Mean \(t-dist and Sub Example\)](#) (19 minutes)
- [Prediction Intervals](#) (16 minutes)
- [MSE and Bias](#) (30 minutes)
- [Paired t-test](#) (18 minutes)
- [Confidence Intervals](#) (20 minutes)

6.3 Reference Formulas

- Margin of error on the estimate of a sample mean at 5%

$$E = \pm 1.96s_{\bar{x}} = \pm 1.96 \frac{s_x}{\sqrt{n}}$$

- 95% Confidence interval on the estimate of a sample mean

$$\bar{x} \pm 1.96s_{\bar{x}} \quad \text{or} \quad \bar{x} \pm 1.96 \frac{s_x}{\sqrt{n}}$$

- 100(1 - α)% Confidence interval on the sample mean

$$\bar{x} \pm z_{\alpha/2}s_{\bar{x}} \quad \text{or} \quad \bar{x} \pm z_{\alpha/2} \frac{s_x}{\sqrt{n}}$$

Where $z_{\alpha/2}$ is the z score for $\alpha/2$.

- t-statistic for sample mean

$$t = \frac{\bar{x} - \mu_x}{s_x/\sqrt{n}} = \frac{\bar{x} - \mu_x}{s_{\bar{x}}}$$

The t-statistic follows a t distribution and has $n - 1$ degrees of freedom.

- Prediction interval for a new observation drawn from the same population

Lower and upper bounds for the interval, respectively

$$\bar{x} - t_{n-1, \alpha/2} s_x \sqrt{1 + 1/n} \quad \bar{x} + t_{n-1, \alpha/2} s_x \sqrt{1 + 1/n}$$

- Central limit theorem for sample means states that Z approximately follows a standard normal distribution for large n

$$Z = \frac{\bar{x} - \mu_x}{\sigma_x/\sqrt{n}} \sim N(0, 1)$$

- A chi-square random variable with $n - 1$ degrees of freedom (df) (written χ_{n-1}^2) is

$$\chi_{n-1}^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{\text{SST}}{\sigma^2}$$

6.4 Sums of Squares

We compute different types of error values in order to evaluate different models and to perform statistical tests. In the equations below, $x_{i,j}$ is the j th measurement taken from the i th experiment. Additionally, x_i is the sum of all $x_{i,j}$'s for a fixed i , and $x_{..}$ is the total sum of all $x_{i,j}$'s.

$$\begin{aligned}
\text{SST} &= \sum_{i=1}^I \sum_{j=1}^J (x_{ij} - \bar{x})^2 = \sum_{i=1}^I \sum_{j=1}^J x_{ij}^2 - \frac{1}{IJ} x^2 \\
\text{SSR} &= \sum_{i=1}^I \sum_{j=1}^J (\bar{x}_{i\cdot} - \bar{x})^2 = \frac{1}{J} \sum_{i=1}^I x_{i\cdot}^2 - \frac{1}{IJ} x^2 \\
\text{SSE} &= \sum_{i=1}^I \sum_{j=1}^J (x_{ij} - \bar{x}_{i\cdot})^2 \quad \text{where } x_{i\cdot} = \sum_{j=1}^J x_{ij} \quad x_{\cdot\cdot} = \sum_{i=1}^I \sum_{j=1}^J x_{ij} \\
\text{SST} &= \text{SSE} + \text{SSR}
\end{aligned}$$

6.5 Overview of Single Sample Hypothesis Testing

In order to make statistical statements about sample data and parameter estimates, we want to verify if our data support such claims. To do so, we develop an **alternative hypothesis**, H_a and test it against our **null hypothesis** H_0 , where the alternative and null hypothesis must cover all possibilities.

- 📖 For example, if we estimate a parameter θ , then want to test whether the parameter is not equal to zero, we define the following hypotheses:

$$H_a : \theta \neq 0, H_0 : \theta = 0$$

- 📖 If we estimate a parameter β , then want to test whether the parameter is greater than 5, we define the following hypotheses:

$$H_a : \beta > 5, H_0 : \beta \leq 5$$

We use a **test statistic** to decide whether to **reject** H_0 or to **fail to reject** H_0 , and accept H_a . Using the test statistic, we compute **P-values**, the probability of obtaining the sampled data assuming that the null hypothesis is true. The lower the P-value, the less likely it is that the null hypothesis holds true for the generated data. We compare the P-value from the test-statistic to our threshold called the **significance level**, denoted by α :

- If $P \leq \alpha$ reject H_0 in favor of H_a
- If $P > \alpha$, fail to reject H_0

Common levels for α are often 5% and 1%.

At a high level, the steps for hypothesis testing given data is already available are:

1. Define null and alternative hypotheses

2. Compute test statistic p-value
3. Compare to threshold significance level
4. Reject or fail to reject null hypothesis


6.6 Z Test on Population Mean

We assume the population follows a normal distribution with known standard deviation. We test the relationship between the population mean μ and a null value of interest μ_0 .

Test statistic:

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

H_0	H_a	P-Value Determination
$\mu \leq \mu_0$	$\mu > \mu_0$	Area under the standard normal curve to the right of z
$\mu \geq \mu_0$	$\mu < \mu_0$	Area under the standard normal curve to the left of z
$\mu = \mu_0$	$\mu \neq \mu_0$	$2 \times$ Area under the standard normal curve to the right of $ z $

 **Example:** A computer manufacturer states the length of a piece of hardware is 130 cm. The manufacturer tests 9 pieces from the factory and gets an average length of 131.08 cm. Given that the lengths of the pieces follow a normal distribution with standard deviation of 1.5cm, is the manufacturer's statement valid at $\alpha = 1\%$?

1. Define null and alternative hypotheses:

$$H_0 : \mu = 130 \quad H_a : \mu \neq 130$$

2. Compute test statistic and p-value:

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{\bar{x} - 130}{1.5/\sqrt{n}}$$

Substitute $n = 9$ samples and the sample mean of 131.08:

$$z = \frac{131.08 - 130}{1.5/\sqrt{9}} = \frac{1.08}{.5} = 2.16$$

Compute P-value using test statistic:

$$\text{P-value} = 2[1 - \Phi(2.16)] = 2(.0154) = .0308$$

3. Compare to threshold significance level

$$\text{P-value} = 0.0308 \geq 0.01$$

4. Reject or fail to reject null hypothesis:

We fail to reject the null hypothesis.


6.7 One-Sample t Test

When we have a small n , we should use the t-test as the Central Limit Theorem (CLT) cannot be applied. To apply the one-sample t test, we assume that the data is a random sample from a normal distribution.

Test statistic:

$$T = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

H_0	H_a	P-Value Determination
$\mu \leq \mu_0$	$\mu > \mu_0$	Area under the t_{n-1} distribution, to the right of the computed t
$\mu \geq \mu_0$	$\mu < \mu_0$	Area under the t_{n-1} distribution, to the left of the computed t
$\mu = \mu_0$	$\mu \neq \mu_0$	$2 \times$ Area under the t_{n-1} distribution, to the right of the computed $ t $

 **Example:** Biologists studying evolution collect data on the lengths of birds' beaks to test whether the mean differs from the 300-year old mean of 45mm. The average beak length from their data composed of 20 sampled birds is 59.30 with sample standard deviation of 9.84. Is the true average today different from the true average from 300 years ago at the 5% level?

1. Define null and alternative hypotheses:

$$H_0 : \mu = 45 \quad H_a : \mu \neq 45$$

2. Compute test statistic and p-value:

$$T = \frac{\bar{X} - 45}{s/\sqrt{n}}$$

Substitute $n = 20$, the sample mean of 59.3 and the sample standard deviation of 9.84:

$$T = \frac{59.3 - 45}{9.84/\sqrt{20}} = 6.5$$

Find P-value using test statistic $t = 6.5$ and the degrees of freedom $df = n - 1 = 20 - 1 = 19$. t-tables can be found online and in Table A.8 in

[Probability and Statistics For Engineering and the Sciences](#)

$$\text{P-value for } t = 6.5 = 2 * 0$$

3. Compare to threshold significance level

$$\text{P-value} = 0 \leq 0.05$$

4. Reject or fail to reject null hypothesis:

We reject the null hypothesis in favor for the alternative hypothesis.

There are multiple other tests in the suggested readings, and we encourage you keep reference statistical test material within reach so that you may look up other tests when needed.

6.8 Error Types

Even though we develop hypothesis and perform statistical tests, our conclusions of which hypothesis to accept or reject may be incorrect. We classify these errors as:

- **Type I Error:** occurs when we reject the null hypothesis H_0 when H_0 is in fact true. This is a false positive.
- **Type II Error:** occurs when we fail to reject the null hypothesis H_0 when H_0 is in fact false. This is a false negative.

By using a test procedure where P-value $\leq \alpha$ results in the rejection of H_0 , then the probability of a type I error is α .

A confusion matrix, as shown below, helps us visually classify the relationship between rejecting and failing to reject H_0 with the actual result of whether or not H_0 is true. The confusion matrix visually demonstrates the correct values, which are the true positives and true negatives in which our predictions match the the actual values, and false positives (Type I) and false negatives (Type II) as described previously.

	Reject H_0	Fail to Reject H_0
H_0 Is True	Type I Error α (False Positive)	Correct $1 - \alpha$ (True Negative)
H_0 Is False	Correct $1 - \beta$ (True Positive)	Type II Error β (False Negative)

Notes

1. By Oleg Alexandrov - Own work, Public Domain, <https://commons.wikimedia.org/w/index.php?curid=2073424>
2. By Jhguch at en.wikipedia, CC BY-SA 2.5, <https://commons.wikimedia.org/w/index.php?curid=14524285>
3. By ShristiV - Own work, CC BY-SA 4.0, <https://commons.wikimedia.org/w/index.php?curid=84530726>
4. By Cmglee - Own work, CC BY-SA 4.0, <https://commons.wikimedia.org/w/index.php?curid=90452334>