Statistics & Probability

Solutions of Seminar

Calculus

① $\int_1^4 f(x)\,dx$, where $f(x) = \frac{x^3}{5,000}$

$$\frac{x^{n+1}}{n+1}$$

$$\left(\frac{1}{5,000}\left[\frac{x^4}{\cancel{x}\,\cancel{5}\,4}\right]\right)\Bigg|_1^4$$

$$\left(\frac{1}{5,000}\right)\left[\frac{4^4}{\cancel{5}\,4} - \frac{1}{\cancel{5}\,4}\right]$$

anti der

$$f(x) = \frac{x^3}{5,000}$$

$$\frac{x^4}{4(5,000)}$$

$$\frac{x^4}{20,000}$$

② $$\int 10w^4 + 9w^3 + \sin(x)\, dx$$

$$= \frac{10w^5}{5} + \frac{9w^4}{4} - \cos(x)$$

$$= \underline{2w^5 + \frac{9w^4}{4} - \cos(x) + C}$$

③ $$\int (8x-12)(4x^2-12)^4\, dx$$

calculus II IBP

recommended graphing software wolfram.

④ derivative $e^{1-\cos(x)}$

e^{u}

$f(x) = 1-\cos(x)$

$f'(x) = +\sin(x)$

$f'(x) = e^{1-\cos(x)} \sin(x)$

(5) $t^2\ln(t^5)$

$f(x)g'(x) + g(x)f'(x)$

$\frac{t^2 \cdot 5\cancel{t^4}}{5t} + \ln(t^5)t^2$

$f'(t) = 5t + 2t\ln(t^5)$

$= 5t + 2t\ln(t^5)$

6

$f(x) = (6x^2+7)^4$

$f(u) = u^4$

$g(x) = 6x^2+7$

$g'(x) = 12x$

$f'(u) = 4u^3$

$f'(x) = 12x \cdot 4(6x^2+7)^3$

$f'(x) = 48x(6x^2+7)^3$

Set theory

① $A \subset C$ (T or T) → A is a subset of C.

a set that is contained in another set. $\{4\}$

② $A \cup B \cup C = \cancel{\{4\}}$

↓

$\{1, 2, 3, 4, 5, 6, 7, 8, 10\}$

③ $A \cap B \cap C = \{4\}$

④ $|A| = 4 \rightarrow$ 4 element in set

$|B| = 5 \rightarrow$ [5 elem, in set]

$|C| = 4 \rightarrow$ 4 element in set

5 power set of A

$\{\}$, $\{1\}$, $\{4\}$, $\{7\}$, $\{10\}$,

$\{1,4\}$, $\{1,7\}$, $\{1,10\}$

$\{4,7\}$, $\{4,10\}$, $\{7,10\}$,

$\{1,4,7\}$, $\{1,4,10\}$,

$\{1,7,10\}$, $\{1,7,10\}$,

$\{1,4,7,10\}$

→ recall power set is the set of A all subsets of A

$A = (0, 1, 2)$

$P(A) =$ $\{\}$, $\{0\}$, $\{1\}$, $\{2\}$,

$\{0,1\}$, $\{0,2\}$, $\{1,2\}$,

$\{0,1,2\}$

$$|A \times B| = 4 \cdot 4$$
$$= 16$$
$$|A| \cdot |B|$$

6 cartesian prod

$A \times \cancel{B}C$,	$A \times B$	$B \times C$
(1,2),(1,4)	your turn	your TURN
(1,6) (1,8)		
(4,2),(4,4)		
(4,6),(4,8)		
(7,2),(7,4)		
(7,6),(7,8),		
(10,2),(10,4)		
(10,6),(10,8)		
16		

① Find determinant

$$\begin{array}{ccc} + & - & + \end{array}$$
$$\begin{bmatrix} 1 & 2 & 5 \\ 1 & 2 & 3 \\ 1 & 2 & 1 \end{bmatrix}$$

$$1\begin{bmatrix} 2 & 3 \\ 2 & 1 \end{bmatrix} - 2\begin{bmatrix} 1 & 3 \\ 1 & 1 \end{bmatrix} + 5\begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix}$$

$$1\begin{bmatrix} 2-6 \end{bmatrix} - 2\begin{bmatrix} 1-3 \end{bmatrix} + 5\cancel{\begin{bmatrix} 2-2 \end{bmatrix}}$$

$$-4 + 4 = 0$$

② calculate product

inner

$$A = \begin{bmatrix} 3 \\ 0 \\ -2 \\ 1 \end{bmatrix} \qquad B = \begin{bmatrix} 8 \\ 2 \\ 3 \\ 1 \end{bmatrix}$$

$$24 + 0 - 6 + 1$$

$$= 19$$

③ $A = \begin{bmatrix} a & b \\ \overset{3}{c} & \overset{4}{d} \end{bmatrix}$ what is A^T

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

$$A^T = \begin{bmatrix} \overset{1}{A} & \overset{3}{C} \\ \overset{2}{B} & \overset{4}{D} \end{bmatrix}$$

What is Statistics and Why Is It Needed?

It is the discipline that concerns the collection, organization, displaying, analysis, interpretation and presentation of data.

Useful for ML, AI, and Data Science Because:

- We present the language of our results in statistical language
- The algorithms we use are based on statistical concepts

  (ex: give me a list of customers who are likely to accept the offer) (Likelihood)

Statistical Topics cover

- Basic Statistics in Preparation
- Statistics and Machine Learning
- Introduction to Statistics
- Descriptive Statistics
- Gaussian Distribution
- Correlation Between Variables
- Statistical Hypothesis Testing
- Estimation Statistics
- Nonparametric statistics

Basic Statistics Preparation

Central tendency - A number that describes the distributed “center” of data in a frequency table from which all the other data falls.

Mean = (a1+ … an) /(n) (average)

Median=Place the numbers in order and you find the middle number.

Mode= which number in the set is repeated more than once?

Factorial 5! = (1 x 2 x 3 x 4 x 5) = n(n-1)(n-2) ...1 (The product of an integer and all integers below it)

Range - For a set of data is simply the difference between the largest and smallest numbers in the set .

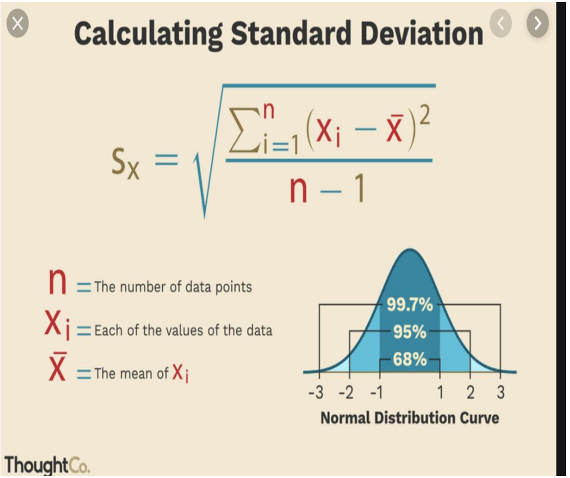
Standard deviation - it is a measure of how far data values of a population are from the mean or average value of the population. (small sd, tends to be values are close to mean value) (large sd, means values tend to be farther from the mean)

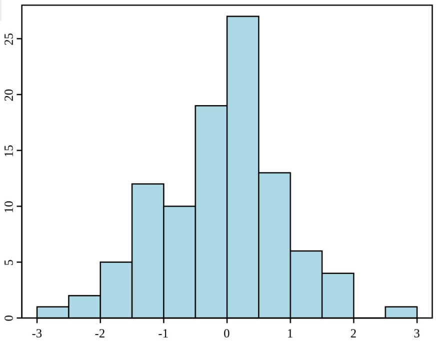
Variance- set of data which is the average of the squared differences from the mean of each value in the data. (It measures how far numbers are spread out from the average value)

Variance

$$s^2 = \frac{\Sigma (x - \bar{x})^2}{n - 1}$$ Sample Variance

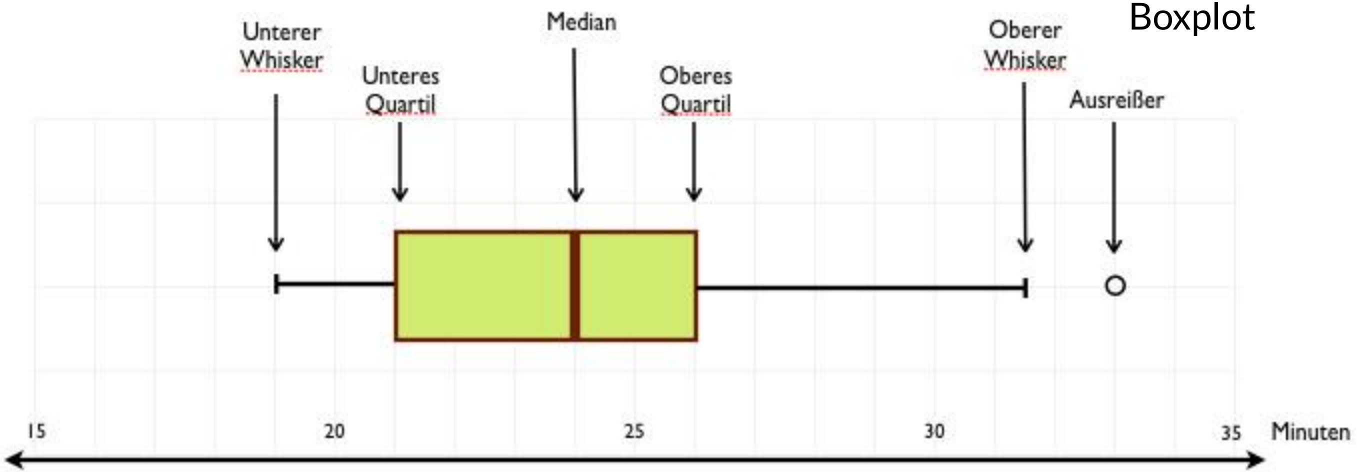
$$\sigma^2 = \frac{\Sigma (x - \mu)^2}{N}$$ Population Variance

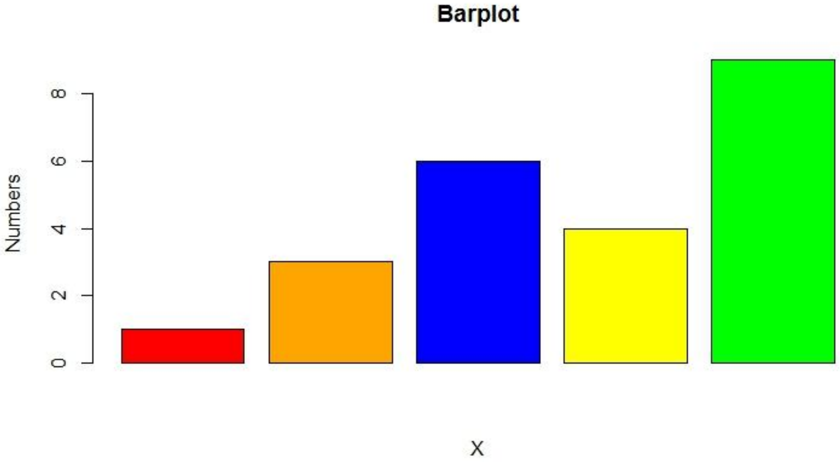




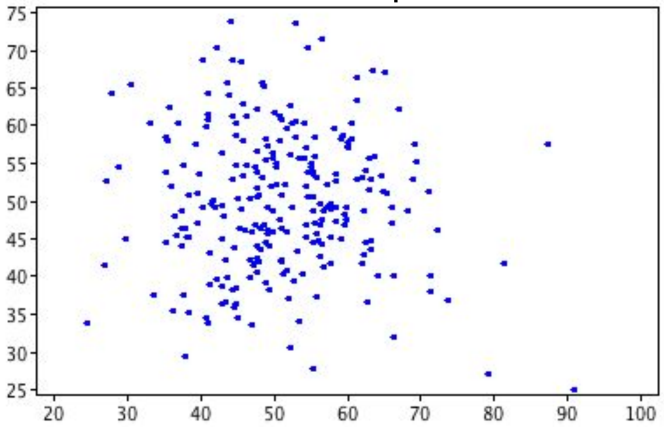
Histogram

Boxplot





Scatterplot



Statistics and Machine Learning

1. Statistics in Data Preparation
   - Outlier detection
   - Missing values
   - Variable encoding
   - Data scaling - Is a method used to normalize the range of independent variables or features of data.
2. Statistics in Model Evaluation
   - Data sampling - Is a statistical analysis technique used to select, manipulate, and analyze a representative subset of data points.
   - Data resampling - It is any of a variety of methods for doing one of the following (medians, variances,percentiles) by using subsets of available data or drawing randoming with replacement from data points. **(K fold cross validation ex:)**
   - Experimental Design - We can change deliberately change one or more process variables (or factors) in order to observe the effect the changes have on one or more response variables. (A response variable is a dependent variable)

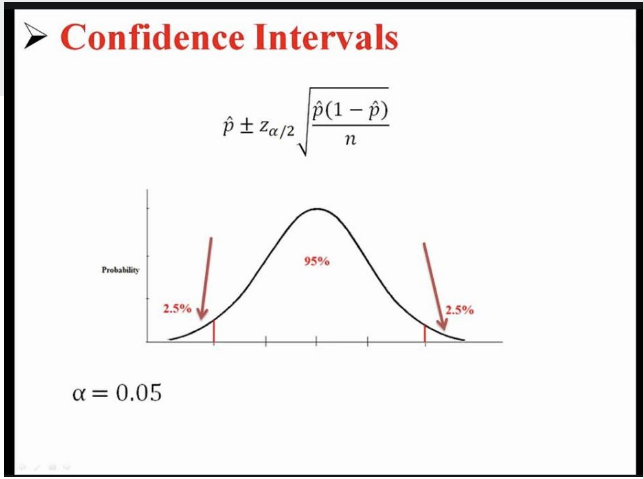
Statistics and Machine Learning

3. Statistics in Model Selection

- Checking for a significant differences between results
- Quantifying the size of difference between results

We can use a statistical hypothesis test:
It is a formal procedure used by statisticians to accept or reject a statistical hypothesis

Refer to https://stattrek.com/hypothesis-test/hypothesis-testing.aspx , for a deeper understanding



Statistics and Machine Learning

4. Statistics in Model Presentation

- Essential when presenting the skill of a final model to stakeholders
- Summarizing results of model on its expected skill (what is the accuracy?, how well does it perform?)
- Estimation can occur with confidence intervals, etc.

5. Statistics in Prediction

- They are required in making a prediction with a finalized model on a new data.

Introduction to Statistics

- Descriptive Statistics - You can use methods for summarizing raw observations into information that we can understand and share.
- Inferential Statistics - A method that aids in quantifying properties of the domain or population from a smaller set of observations called samples.

Gaussian distribution (also known as **normal distribution**) is a bell-shaped curve, and it is assumed that during any measurement values will follow a **normal distribution** with an equal number of measurements above and below the mean value.

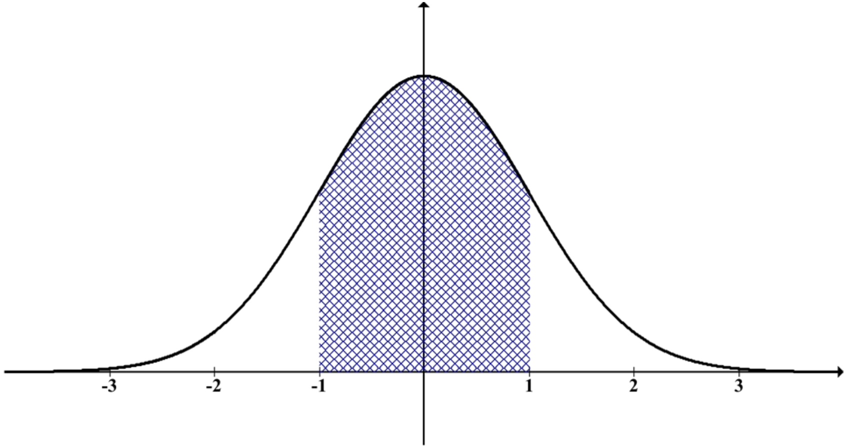
Gaussian Distribution & Descriptive Statistics

- When many observations fit a normal pattern or distribution it is called a normal distribution or formally the Gaussian distribution.
- Any data sample from the Gaussian distribution can be summarized with just 2 parameters
  1. Mean - The central tendency or the most likely value in the distribution
  2. Variance - The average difference that the observations have from the mean in the distribution. (How far out are the values spread?)

General normal distribution

Every normal distribution is a version of the standard normal distribution who domain has been stretched by a factor σ (the standard deviation) and then translated by μ (the mean value):

$$f(x \mid \mu, \sigma^2) = \frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right).$$



Correlation Between Variables

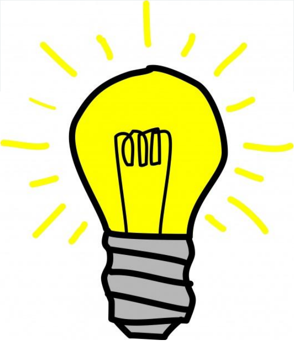
-Sometimes you will need quantify a relationship between two variables. (Correlation Coefficient)

Why is this important?

- It can be useful when modeling to better understand the relationships between variables.

A correlation can be positive, neural, negative

- Positive: Both variables, change in the same direction
- Neutral: No relationship in the change of the variables
- Negative: Variables change in the opposite directions



Performance can deteriorate if two or more variables are tightly related. (Multicollinearit y)

Pearson's Correlation Coefficient

$$r = \frac{\Sigma(X-\overline{X})(Y-\overline{Y})}{\sqrt{\Sigma(X-\overline{X})^2}\sqrt{(Y-\overline{Y})^2}}$$

Where, $\overline{X}$ = mean of X variable
$\overline{Y}$ = mean of Y variable

Properties of Coefficient of Correlation

- The value of the coefficient of correlation (r) always **lies between ±1**. Such as:
  r=+1, perfect positive correlation
  r=-1, perfect negative correlation
  r=0, no correlation

Statistical Hypothesis Tests

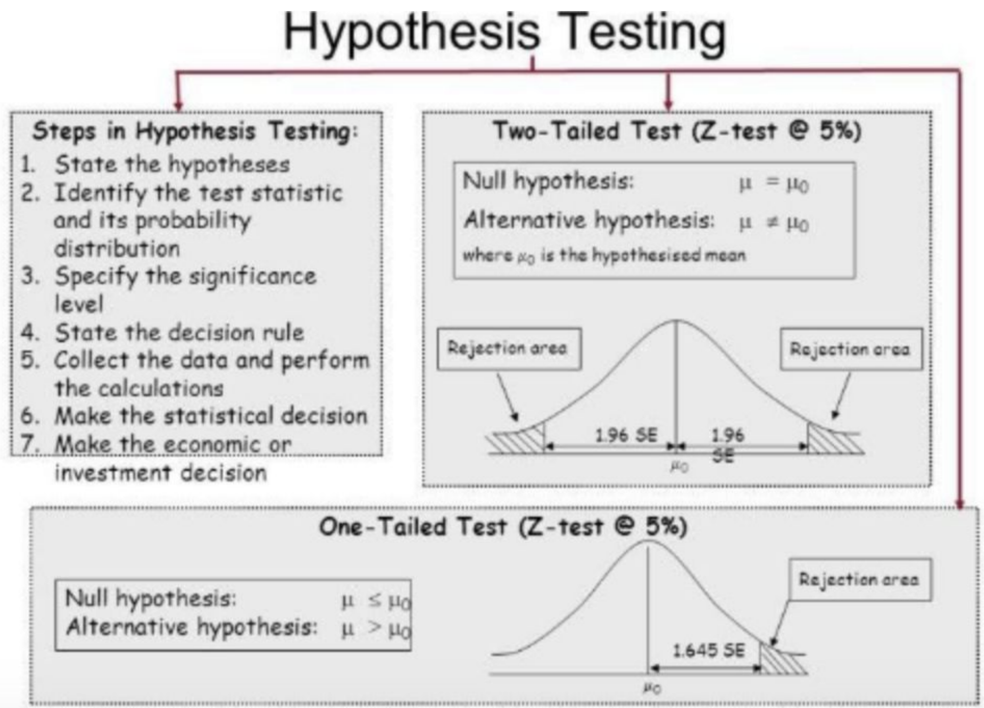
It is a hypothesis that is testable on the basis of observing a process that is modeled via a set of random variables.

Null Hypothesis- It is denoted by Ho, is usually the hypothesis that sample observations result purely on chance.

Alternative hypothesis: The alternative hypothesis is denoted by H1 or Ha is the hypothesis that sample observations are influenced by some non-random cause.

Type1 I Errors - when the research rejects the null hypothesis when it is true. Probability of committing a Type I error is called a significance level. This probability is also denoted by alpha

Type II When the research fails to reject the null hypothesis that is false. The probability of committing a Type II error is called Beta β. The probability committing a Type II error is called the Power of the test.

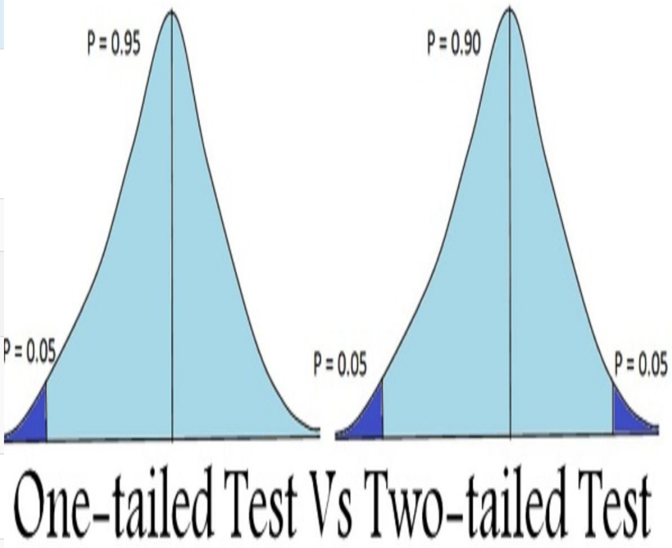


One tailed test & two tailed test

One tailed test - refers to a test of the null hypothesis in which the alternative hypothesis is articulated directionally. The critical region lies only one tail.

Two tailed test of the null hypothesis is where in the critical region is one both of the tails.

BASIS OF COMPARISON	ONE-TAILED TEST	TWO-TAILED TEST
Meaning	A statistical hypothesis test in which alternative hypothesis has only one end, is known as one tailed test.	A significance test in which alternative hypothesis has two ends, is called two-tailed test.
Hypothesis	Directional	Non-directional
Region of rejection	Either left or right	Both left and right
Determines	If there is a relationship between variables in single direction.	If there is a relationship between variables in either direction.
Result	Greater or less than certain value.	Greater or less than certain range of values.
Sign in alternative hypothesis	> or <	≠



Analysis Plan for Hypothesis Testing:

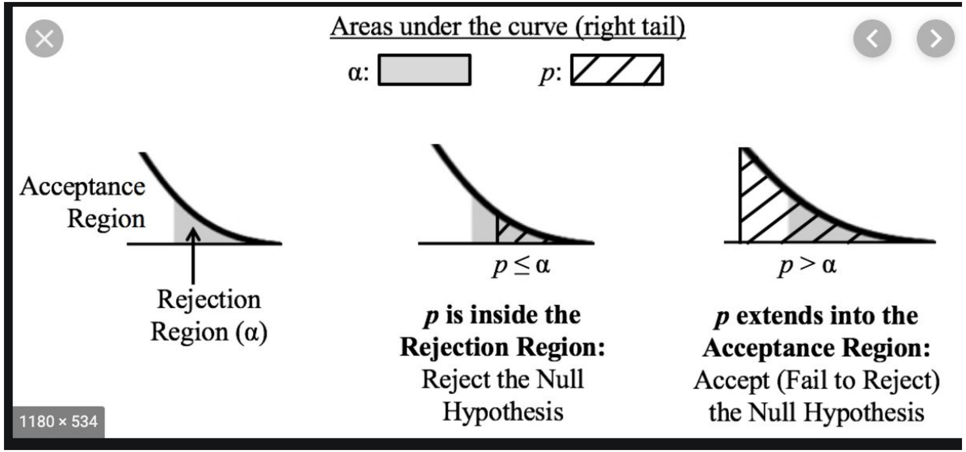
P-value - The strength of evidence in support of a null hypothesis is measured by the **P-value**

Region of Acceptance - The **region of acceptance** is a range of values. If the test statistic falls within the region of acceptance, the null hypothesis is not rejected. The region of acceptance is defined so that the chance of making a Type I error is equal to the significance level.

The set of values outside the region of acceptance is called the **region of rejection**. If the test statistic falls within the region of rejection, the null hypothesis is rejected. In such cases, we say that the hypothesis has been rejected at the α level of significance.

Includes

- Decision rules for rejecting the null hypothesis
- Decision rules in 2 ways (reference to P value or reference to region of acceptance)



Example:

Say if we want to determine whether a coin was fair and balanced.

A null hypothesis might be that half the filips result in heads and then tails.

Then the alternative hypothesis might be heads and tails are different

Ho: P=0.5

Ha: P !=0.5

Suppose we flipped the coin 50 times, resulting in 40 heads and 10 tails. Given this result, we would be inclined to reject the null hypothesis.   Based on evidence the coin was most likely not fair and balanced.

Estimation Statistics

It is a data analysis framework that uses a combination of effect sizes, confidence intervals, precision planning, and meta-analysis to plan experiments, analyze data and interpret results

- **Effect Size.** Methods for quantifying the size of an effect given a treatment or intervention.
- **Interval Estimation**. Methods for quantifying the amount of uncertainty in a value.
- **Meta-Analysis.** Methods for quantifying the findings across multiple similar studies.

Three main types of intervals: **Tolerance Interval**: The bou coverage of a proportion of a distribution with a specific lev confidence, **Confidence Interval**: The bounds on the estim a population parameter,**Prediction Interval**: The bounds or single observation.

Example of Calculating the Confidence Interval Formula for Proportions:

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$\hat{p}$ is a sample proportion
n is the sample size
z^* is the standardized normal distribution (obtained from commonly utilized tables)

Non Parametric Statistics

- Sometimes data will not come from a Gaussian Distribution , this is non parametric statistics where data is not required to fit a Normal Distribution.
- **Nonparametric statistics** uses data that is often ordinal, meaning it does not rely on numbers, but rather on a ranking or order of sorts

Before a Nonparametric statistical can be applied:

- Data must be converted into a rank format
- Statistical methods expect data in rank format are sometimes called rank statistics

Procedure:

1. Sort all data in the sample in descending order
2. Assign an integer from 1 to N for each unique value

An example of nonparametric test: the Mann-Whitney U test and it does not assume data comes from a normal

Refer to:
http://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_nonparametric/BS704_Nonparametric4.html

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

where R_1 = sum of the ranks for group 1 and R_2 = sum of the ranks for group 2.

$$U_1 = 5(5) + \frac{5(6)}{2} - 37 = 3$$

$$U_2 = 5(5) + \frac{5(6)}{2} - 18 = 22$$

What is Probability and why do we need it?

Probability is the likelihood or chance of an event occuring. The probability that you are certain that happen is 1 if it is impossible it is 0.

Life is full of uncertainties. We do not know the outcome of a particular situation til it happens.

In an uncertain world, it can help us know and understand chances of various events.

Ex: Will I get a promotion?

This is an example of an uncertain situation:

Common Terminology:

1. **Experiment -** uncertain situations, which could have multiple outcomes. (Does it rain on a daily basis?)
2. **Outcome -** is the result of a single trial. (If it rained today, then the outcome of the trial is that it rained)
3. **Event -** is one or more outcomes from an experiment. ("It rained" is an ex: of an event)
4. **Probability -** Measure of how likely an event is. (There is a 60% chance it will rain tomorrow)

Why is it useful for AI, Data Science and ML?

- Probability is the foundation and language needed for most statistics. (Basically without probability, we cannot do statistics, and without statistics we cannot express the basic languages of our algorithms, communicate our results, and understand the theory behind them)
- It helps us reason effectively, in situations where being certain is impossible.
- ML is the root in statistics, and probability.
- ML, we model our problems so that it's outputs are probabilities of things based on following intuitions.

Probability Topics Covered

- Probability axioms
- Discrete distributions
- Random distributions
- Continuous distributions
- Joint probability distributions
- Conditional/Marginal probability distributions
- Bayes rule
- Independence and conditional independence
- Distributions (we need to know)

Basic Probability Preparation

Continuous Random Variable - IS a R>V where the data can take infinitely many values. (Ex: a r.v measuring the time taken for something to be done is continuous since there are an infinite number of possible times taken) \

Random Variables- (variable who depends on a random event)

To calculate the likelihood of occurrence of an event, we need to put an framework to express the outcome in numbers. We can do this by mapping the outcome of an experiment to numbers.

Let's define X to be the outcome of a coin toss.

X = outcome of a coin toss

Possible Outcomes:

- 1 if heads
- 0 if tails

Sample Space: The set of all possible results or outcomes.

Probability Function- The function that helps in obtaining the probability of each and every outcome.

Probability Density Function - The PDF is the density of probability rather than the probability mass. (It is the derivative of the CDF)
Cumulative Distribution Function - It is the probability that function of X will take a value less than or equal to x. (It gives the area under the PDF from minus infinity to X.

Probability Formulas

Probability of an Event = (Number of Favorable outcomes)/(Total number of possible outcomes)

P(A)= n(E)/n(S) P(A) < 1

Here P(A) means finding the probability of an event A, n(E) means the number of favorable outcomes of an event and n(S) means the set of all possible outcomes of an event.

If the probability of an occurring event is P(A), then the probability of not occurring an event is P(A')=1-P(A)

Probability Formulas:

Event(A or B) P(A U B) = P(A) + P(B) - P(A ∩ B)

Event(A and B) P(A∩ B)= P(A)* P(B)

Event(A NOT B) P(A NOT B)== A-B

Event (B NOT A) P(B Not A)= B-A

*Probability of an Event Occurring is P(A)

Probability of an Event Not Occuring P(A')

Some **probability important formulas** based on them are as follows:

- P(A.A') = 0
- P(A.B) + P (A'.B') = 1
- P(A'B) = P(B) – P(A.B)
- P(A.B') = P(A) – P(A.B)
- P(A+B) = P(AB') + P(A'B) + P(A.B)

Mutually Exclusive - Means events cannot occur independently

Probability axioms

1. What is an axiom? - It is a statement or proposition which is regarded as being established, accepted, or true.

   Axiom 1: The probability of an event is a real number greater than or equal to 0. P(A) >= 0

   Axiom 2: The probability that at least one of all the possible outcomes of a process is 1. P(A) =1

   Axiom 3: The third axiom of probability deals with mutually exclusive events. If *A* and *B* are mutually exclusive, meaning that they have an empty intersection and we use U to denote the union, then *P*(*A* U *B*) = *P*(*A*) + *P*(*B*)

(1) $P(A) \geq 0$ for all $A \subset S$

(2) $P(S) = 1$

(3) If $A \cap B = \emptyset$,

then $P(A \cup B) = P(A) + P(B)$

Discrete Distributions

A discrete distribution is a statistical distribution that shows the probability of outcomes with finite values.

- The most common type include Binomial, Poisson, Bernoulli, & multinomial

Binomial - The **binomial distribution formula** is: $b(x; n, P) = {}_nC_x * P^x * (1 – P)^{n – x}$. Where: b =**binomial** probability. x = total number of “successes” (pass or fail, heads or tails etc.)

Poisson - Then, the **Poisson** probability is: $P(x; \mu) = (e^{-\mu}) (\mu^x) / x!$ where x is the actual number of successes that result from the experiment, and e is approximately equal to 2.71828.

Bernoulli - It is a random experiment that has only two outcomes (usually called a “Success” or a “Failure”). For example, the probability of getting a heads (a “success”) while flipping a coin is 0.5. The probability of “failure” is 1 – P (1 minus the probability of success, which also equals 0.5 for a coin toss). It is a special case of the binomial distribution for n = 1. In other words, it is a binomial distribution with a single trial (e.g. a single coin toss).

Binomial Formula and Binomial Probability

The **binomial probability** refers to the probability that a binomial experiment results in exactly x successes. For example, in the above table, we see that the binomial probability of getting exactly one head in two coin flips is 0.50.

Given x, n, and P, we can compute the binomial probability based on the binomial formula:

Binomial Formula. Suppose a binomial experiment consists of n trials and results in x successes. If the probability of success on an individual trial is P, then the binomial probability is:

$$b(x; n, P) = {}_nC_x * P^x * (1 - P)^{n - x}$$

or

$$b(x; n, P) = \{ n! / [x! (n - x)!] \} * P^x * (1 - P)^{n - x}$$

Example 1

Suppose a die is tossed 5 times. What is the probability of getting exactly 2 fours?

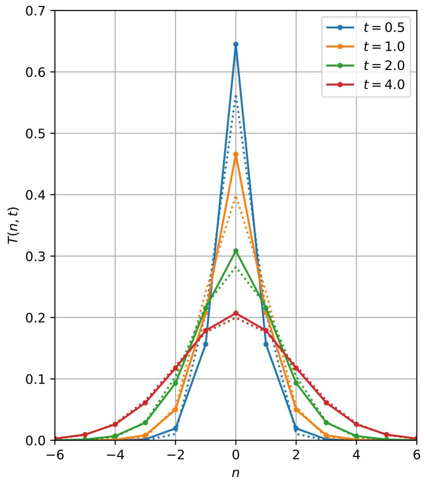
Solution: This is a binomial experiment in which the number of trials is equal to 5, the number of successes is equal to 2, and the probability of success on a single trial is 1/6 or about 0.167. Therefore, the binomial probability is:

$$b(2; 5, 0.167) = {}_5C_2 * (0.167)^2 * (0.833)^3$$
$$b(2; 5, 0.167) = 0.161$$

Continuous Distributions

What is a continuous distribution?

A continuous distribution describes the probabilities of the possible values of a continuous random variable. A continuous random variable is a random variable with a set of possible values (known as the range) that is infinite and uncountable.



Joint Probability Distributions

Example: The probability that a card is a five and black = p(five and black) = 2/52 = 1/26. (There are two black fives in a deck of 52 cards, the five of spades and the five of clubs).

Joint Probability and Joint Distributions: Definition, Examples

Probability > Joint Probability / Joint Distribution

What is Joint Probability?

Joint probability is the probability of two events happening together. The two events are usually designated *event A* and *event B*. In probability terminology, it can be written as:

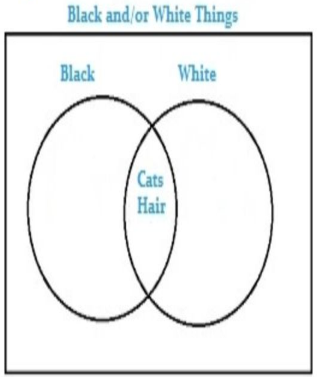
p(A and B)

or

p(A ∩ B)

Joint probability is also called the **intersection** of two (or more) events. The intersection can be represented by a Venn diagram:

Joint probability is also called the **intersection** of two (or more) events. The intersection can be represented by a Venn diagram:



Conditional and Marginal Probability Distributions

- A conditional distribution is a probability distribution for a **subpopulation**. In other words, it shows the probability that a randomly selected item in a **subpopulation** has a characteristic you're interested in.
- The marginal distribution of a subset of a collection of random variables is the probability distribution of the variables contained in the subset.
-

Conditional Distributions vs. Marginal Distributions.

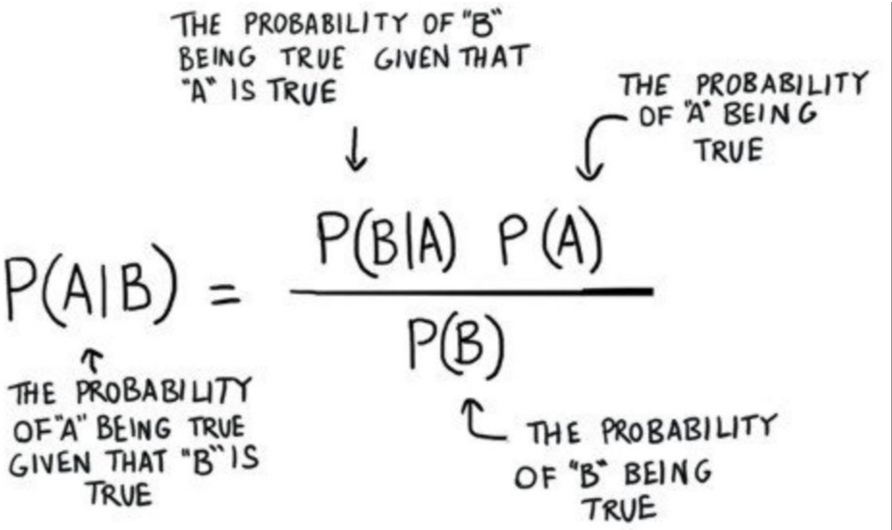
Marginal and conditional distributions can be found the same table. Marginal distributions are the totals for the probabilities. They are found in the margins (that's why they are called "marginal"). The following table shows probabilities for rolling two dice. The total probabilities in the margins are the marginal distributions.

$i \backslash j$	1	2	3	4	5	6	$p_X(i)$
1	1/36	1/36	1/36	1/36	1/36	1/36	1/6
2	1/36	1/36	1/36	1/36	1/36	1/36	1/6
3	1/36	1/36	1/36	1/36	1/36	1/36	1/6
4	1/36	1/36	1/36	1/36	1/36	1/36	1/6
5	1/36	1/36	1/36	1/36	1/36	1/36	1/6
6	1/36	1/36	1/36	1/36	1/36	1/36	1/6
$p_Y(j)$	1/6	1/6	1/6	1/6	1/6	1/6	

Marginal distributions (yellow borders). Image: Penn State.

Bayes Rule

describes the probability of an event, based on prior knowledge of conditions that might be related to the event. For example, if cancer is related to age, then, using Bayes' theorem, a person's age can be used to more accurately assess the probability that he has cancer than can be done without knowledge of the person's age.



Independence & Conditional Independence

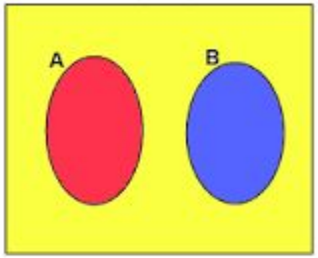
Conditional Independence

- Dependent events can become independent given certain other events.
- Example,
  - Size of shoe
  - Age
  - Size of vocabulary
- Two events A, B are conditionally independent given a third event C iff

  $P(A|B, C) = P(A|C)$

When two events are said to be **independent** of each other, what this means is that the **probability** that one event occurs in no way affects the **probability** of the other event occurring. An example of two **independent** events is as follows; say you rolled a die and flipped a coin.

S = Sample Space



Distributions we should know Uniform and Exponential

Exponential Distribution

Let's consider the call center example one more time. What about the interval of time between the calls ? Here, exponential distribution comes to our rescue. Exponential distribution models the interval of time between the calls.

Other examples are:

1. Length of time beteeen metro arrivals,
2. Length of time between arrivals at a gas station
3. The life of an Air Conditioner

Exponential distribution is widely used for survival analysis. From the expected life of a machine to the expected life of a human, exponential distribution successfully delivers the result.

A random variable X is said to have an **exponential distribution** with PDF:

$$f(x) = \{ \lambda e^{-\lambda x},\ x \geq 0$$

and parameter **λ>0** which is also called the rate.

Uniform Distribution

When you roll a fair die, the outcomes are 1 to 6. The probabilities of getting these outcomes are equally likely and that is the basis of a uniform distribution. Unlike Bernoulli Distribution, all the n number of possible outcomes of a uniform distribution are equally likely.

A variable X is said to be uniformly distributed if the density function is:

$$f(x) = \frac{1}{b-a} \qquad for\ -\infty < a \leq x \leq b < \infty$$

The graph of a uniform distribution curve looks like

So, why?

Why do we need to know distributions?

1. They allow data scientists to discover patterns otherwise completely random variables
2. Remember Data Science and Machine Learning is based up on certain assumptions about the probability distributions of your data.

Practice Problems

1. What is 1 method that can be used for descriptive statistics and explain it.
2. What is 1 method that can be used for inferential statistics and explain it.
3. Find the correlation coefficient

SUBJECT	AGE X	GLUCOSE LEVEL Y
1	43	99
2	21	65
3	25	79
4	42	75
5	57	87
6	59	81

4. A sample of size n = 100 produced the sample mean of $\bar{X}$ = 16. Assuming the population standard deviation σ = 3, compute a 95% confidence interval for the population mean μ.

5. Use Bayes Formula: You might be interested in finding out a patient's probability of having liver disease if they are an alcoholic. "Being an alcoholic" is the **test** (kind of like a litmus test) for liver disease.

- **A** could mean the event "Patient has liver disease." Past data tells you that 10% of patients entering your clinic have liver disease. P(A) = 0.10.
- **B** could mean the litmus test that "Patient is an alcoholic." Five percent of the clinic's patients are alcoholics. P(B) = 0.05.
- You might also know that among those patients diagnosed with liver disease, 7% are alcoholics. This is your **B|A:** the probability that a patient is alcoholic, given that they have liver disease, is 7%.

6. Calculate the conditional distribution of pet preference among women.

	Cats	Fish	Dogs	
Men	2	4	6	12
Women	5	3	2	10
	7	7	8	22

Upcoming Months:

Seminar 3 - Essentials of SQL / and Data Structures for Machine Learning
Seminar 4 - Learning Python
Seminar 5- Learning the DevOps Side for Machine Learning
Seminar 6: Machine Learning Terminology
Seminar 7: Building a Basic Classification M

.....

Work cited

1. https://machinelearningmastery.com/statistics-for-machine-learning-mini-course/
2. https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/gaussian-distribution
3. https://www.thoughtco.com/what-are-probability-axioms-3126567
4. https://www.statisticssolutions.com/correlation-pearson-kendall-spearman/
5. https://businessjargons.com/karl-pearsons-coefficient-of-correlation.html
6. https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/correlation-coefficient-formula/
7. https://www.datasciencecentral.com/profiles/blogs/importance-of-hypothesis-testing-in-quality-management
8. https://stattrek.com/hypothesis-test/hypothesis-testing.aspx
9. https://keydifferences.com/difference-between-one-tailed-and-two-tailed-test.html
10. https://www.statisticsfromatoz.com/blog/acceptance-and-rejection-regions
11. https://stattrek.com/probability-distributions/binomial.aspx
12. https://www.statisticshowto.datasciencecentral.com/joint-probability-distribution/