The Normal Distribution

Understanding its definition, parameters and application

What is a normal distribution?

A normal distribution describes the probability distribution of a random variable, where values are standardized to the mean

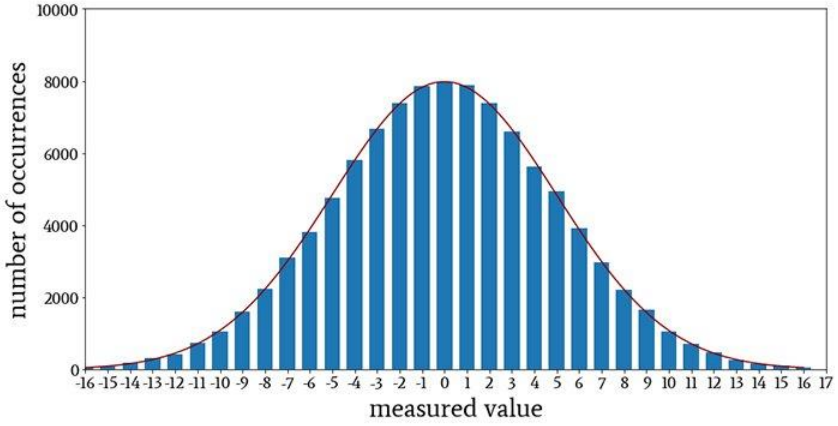
Random Variable: a variable whose value depends on a random process

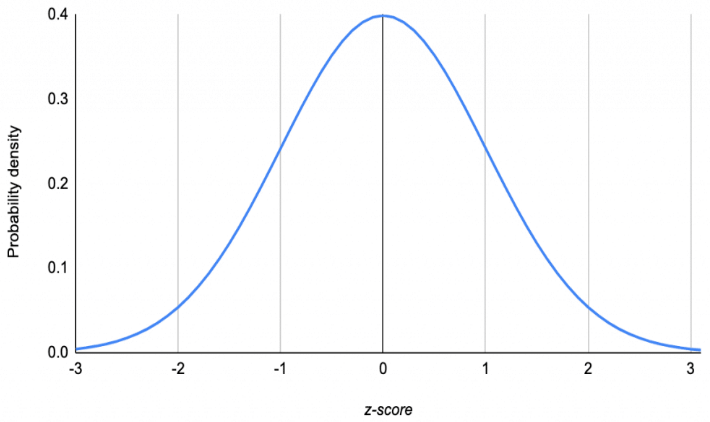
- Imagine that you are participating in an experiment where perform a coin toss. As the results of your toss is not impacted by the outcome of your previous toss, the result of the coin toss is a random variable

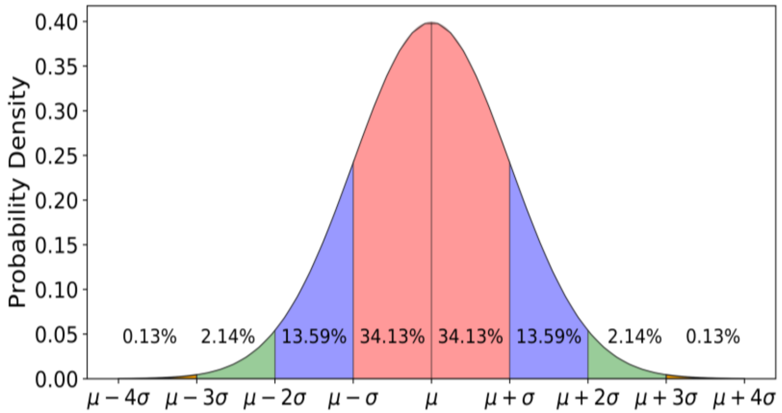
Probability distribution

- Where the x axis = values to be measured, y axis = probability of the measured values occurring

Shape of a normal distribution







Key characteristics of a normal distribution:

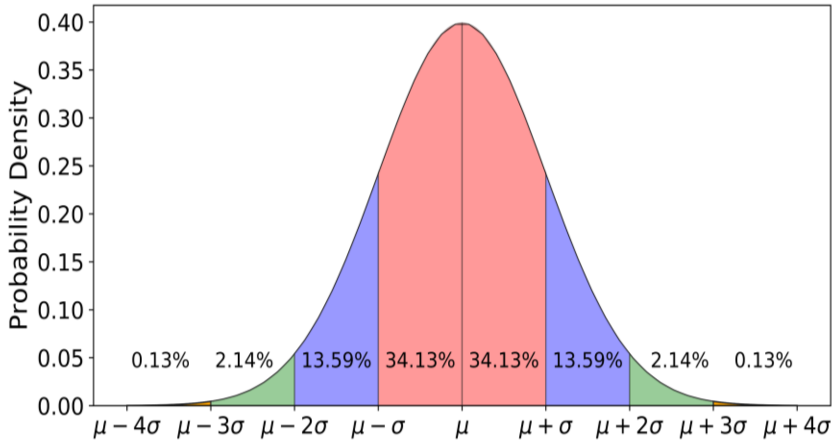
- Centre of the distribution will always equal the mean
- Symmetrical
- Approx. 68%, 95% and 99% of the values are within 1,2 and 3 standard deviation of the mean respectively

Normal distribution formula:

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

$\mu =$ Mean
$\sigma =$ Standard Deviation
$\pi \approx 3.14159\cdots$
$e \approx 2.71828\cdots$

Normal distribution parameters



Population	Sample
Standard deviation $\sigma = \sqrt{\frac{\Sigma(X-\mu)^2}{n}}$	Standard deviation $s = \sqrt{\frac{\sum (x-\bar{x})^2}{n-1}}$
Mean $\mu = \frac{\sum_{i=1}^{N} x_i}{N}$	Mean $\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n}$

Z score – how far away a value is from its mean

Population

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

Sample

$$z_i = \frac{x_i - \bar{x}}{s}$$

We can also derive the probability of each Z score → p-value

Normal distribution applications – p value

P-value

- Normally used in hypothesis testing, it is the probability of obtaining a certain result, assuming that the null hypothesis is correct
  - Hypothesis testing is the process of validating a specific hypothesis where :
    - Null hypothesis → accepted fact
    - Hypothesis statement → what we think is true
  - Our hypothesis is validated if the p value of obtaining our results are lower than the p-value indicated by the significance level
    - Benchmark significance levels are usually p value = 0.05, however this may differ according to industry and project

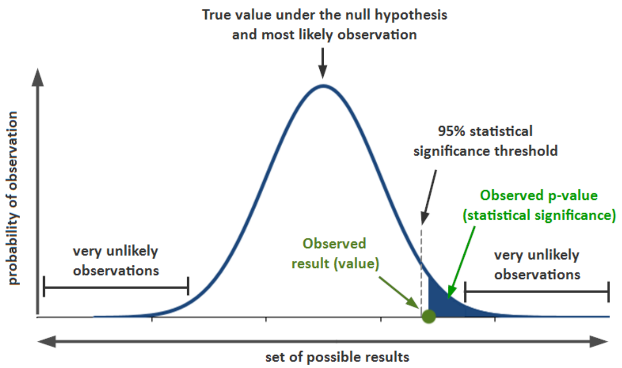
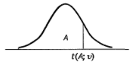


TABLE B.2 Percentiles of the t Distribution.

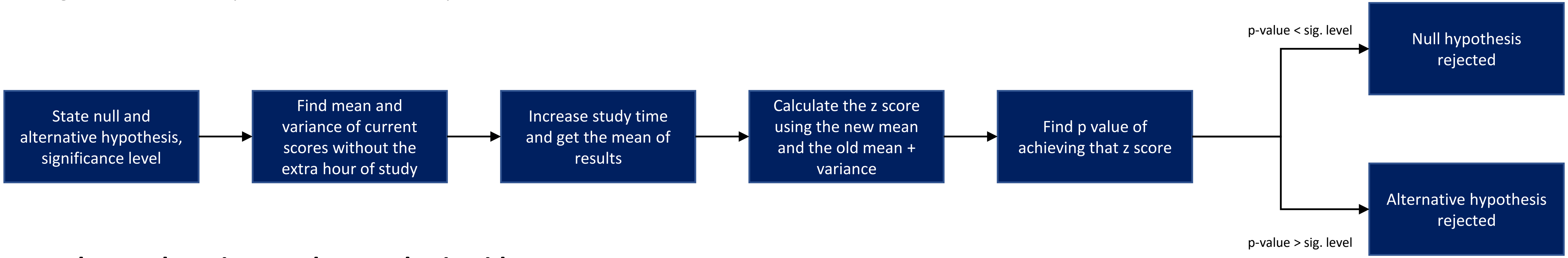
Entry is $t(A;\nu)$ where $P\{t(\nu) \le t(A;\nu)\} = A$



	A						
ν	.60	.70	.80	.85	.90	.95	.975
1	0.325	0.727	1.376	1.963	3.078	6.314	12.706
2	0.289	0.617	1.061	1.386	1.886	2.920	4.303
3	0.277	0.584	0.978	1.250	1.638	2.353	3.182
4	0.271	0.569	0.941	1.190	1.533	2.132	2.776
5	0.267	0.559	0.920	1.156	1.476	2.015	2.571
6	0.265	0.553	0.906	1.134	1.440	1.943	2.447
7	0.263	0.549	0.896	1.119	1.415	1.895	2.365
8	0.262	0.546	0.889	1.108	1.397	1.860	2.306
9	0.261	0.543	0.883	1.100	1.383	1.833	2.262
10	0.260	0.542	0.879	1.093	1.372	1.812	2.228
11	0.260	0.540	0.876	1.088	1.363	1.796	2.201
12	0.259	0.539	0.873	1.083	1.356	1.782	2.179
13	0.259	0.537	0.870	1.079	1.350	1.771	2.160
14	0.258	0.537	0.868	1.076	1.345	1.761	2.145
15	0.258	0.536	0.866	1.074	1.341	1.753	2.131
16	0.258	0.535	0.865	1.071	1.337	1.746	2.120
17	0.257	0.534	0.863	1.069	1.333	1.740	2.110
18	0.257	0.534	0.862	1.067	1.330	1.734	2.101
19	0.257	0.533	0.861	1.066	1.328	1.729	2.093
20	0.257	0.533	0.860	1.064	1.325	1.725	2.086
21	0.257	0.532	0.859	1.063	1.323	1.721	2.080
22	0.256	0.532	0.858	1.061	1.321	1.717	2.074
23	0.256	0.532	0.858	1.060	1.319	1.714	2.069

P-value calculation using example

- Let's assume that currently, a high school teacher is trying to find out if increasing the amount of study time for each student will lead to higher grades. To solve the problem, we take the steps below:



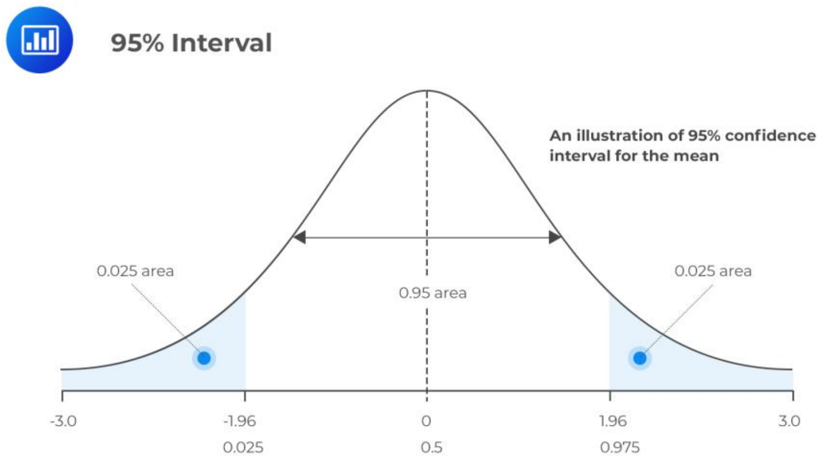
P-value explanations and examples in video

- https://www.youtube.com/watch?v=-FtlH4svqx4 – Khan Academy, Hypothesis testing and p values using an example, 11 mins

Normal distribution applications – confidence interval

Confidence intervals

- Used to estimate the range of values of which an unknown variable will take, based on the level of confidence
    - “I am 95% sure that the stock price of Tesla will not fall between $100 and $200
- Degree of confidence (e.g. 95%, 90%) plays a significant role as range will change depending on the value of the interval
    - You can also interpret degree of confidence as the probability
- To calculate our interval, we will need to translate the **degree of confidence into z scores and the translate them into actual values**
    - Remember as the normal distribution is a probability density function, area of the graph = probability



Area under the curve = probability

Confidence interval calculation using example

- Let’s assume that you are a working as an investment analyst, and your boss wants to know the confidence interval for the price of Share X, assuming that you are 95% confident. To calculate the confidence interval:

Find the mean and variance of the distribution → Find the area that are not included in the 95% → For that area/p value calculate the z score → From the z score, use the mean and variance to calculate the actual values → This will result in 2 values, which we use in our interval → “ With 95% confidence, Share X will not fall between $y and $z “

Confidence interval explanations and examples in video

- https://www.youtube.com/watch?v=SeQeYVJZ2gE&t=1s – Khan Academy, Confidence interval and margin of error, 20 mins
- https://www.youtube.com/watch?v=27iSnzss2wM&t=62s – JB Statistics, introduction to confidence intervals
- https://www.youtube.com/watch?v=TqOeMYtOc1w&t=252s – StatQuest, Confidence intervals