Describing Data:

- **Descriptive Statistics:** Describing categorical and quantitative variables and the relationships between them via 2 aspects:
    1. How to visualize data using graphs
    2. How to summarise data (key aspects) using numerical quantities (summary statistics)

Categorical Variables:

- **ONE CATEGORICAL VARIABLE:**
    - **Proportion:** Summary statistic helping describe categorical variables ○ Sum of proportions is 1

$$P/\hat{p} = \frac{no.\,in\,category}{total\,sample\,size}$$

    - **Relative Frequency:** Shows proportions ○ Enables making comparisons without referring to the sample size
    - **Bar/ Pie Charts:** Used to visually explain proportions

- **TWO CATEGORICAL VARIABLES:**
    - Relationship i.e. group of people 2 categories are male/ female
    - Useful measure of **ASSOCIATION**
    - **Two-Way Table:** Shows the relationship between 2 categorical variables
    - **Side by side chart:** Visual representation of this

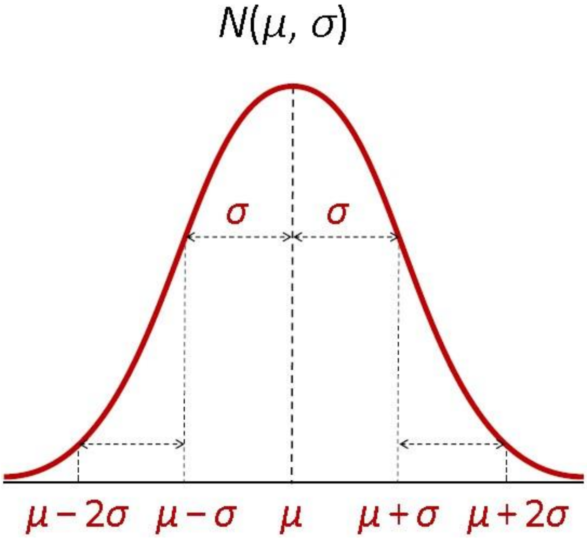
Testing in Normal Distributions:

Normal Distribution:

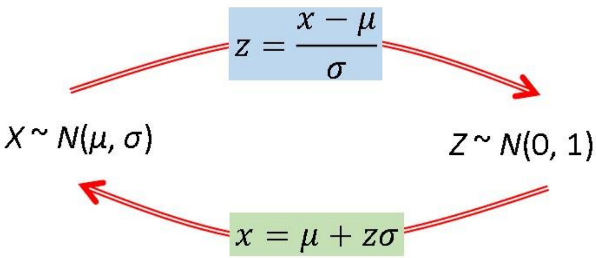
- Distributions of common statistics following a bell-shaped pattern ▪ Uses μ, σ → N (μ, σ)

- **Density Curve:** Reflects the location, spread and shape of the distribution
  - Total area under the curve is 1→ 100% of the distribution
  - Area <u>over</u> an interval is the proportion <u>within</u> the interval ▪ DOESN'T HAVE TO BE A NORMAL CURVE!

- **Normal Density Curve:** Follows a normal, bell-shaped distribution ▪ N (μ, σ)
  - Centered on μ
  - 95% of the data is μ ± 2σ



- **Standard Normal N(0,1):** z-distribution used as a basic way to compare different distributions
  - Convert normal distributions to z-scores
  - To convert from X ~ N (μ, σ) to Z ~ N(0,1) use **Z-SCORE FORMULA!!!**



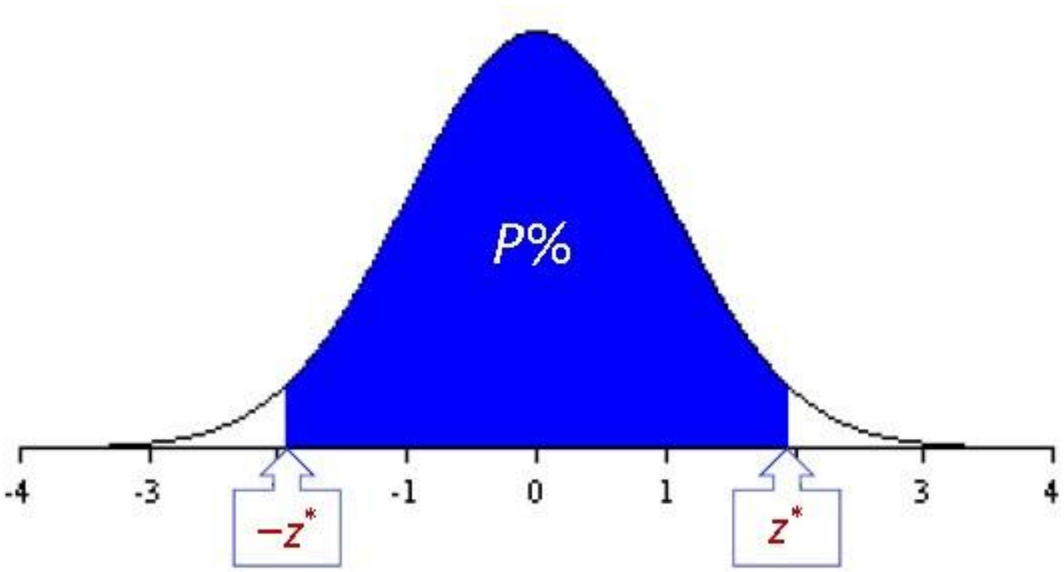
Confidence Intervals and P-Values Using Normal Distributions:

- For categorical data, the parameter = p for proportion
- For numerical data the parameter = μ for mean

- **Central Limit Theorem:** For a large enough sample size, the distribution of sample proportions and means will be approximately normally distributed and centered at the population parameter (p or μ)
    - $n \geq 30$ → for quantitative variables
    - $n \geq 10$ → for categorical variables
    - As n increases, the distribution gets more closely approximated to the normal distribution and σ decreases

- **Confidence Interval:**

$$\boldsymbol{Confidence\ Interval}: \boldsymbol{Sample\ Statistic} \pm \boldsymbol{Z}^{*} \times \boldsymbol{SE}$$



Z* area reflects the desired level of confidence

- **How to apply/ test for Confidence Intervals:**
    1. Confirm the sampling distribution can be approximated by normal distribution
        - Check CLT applies
    2. Find z* for the P% CI
    3. Obtain the P% CI using formula
- **P-Values in a normal distribution:** Standardise the value of the statistic from the original sample, using the Ho μ and the randomization SE

- **Hypotheses tests based on a normal distribution:** The normal distribution should be consistent with Ho so the μ is determined by Ho ▪ Compute standardized test statistic using:

$$Z = \frac{SS - Ho}{SE}$$

  - This formula calculates the number of SEs the statistic is from Ho, consequently allowing us to assess the extremity on a **COMMON SCALE**
    - ❖ If the statistic is normally distributed under Ho, the pvalue is the N (0,1) probability beyond z

In order to apply the CI and hypotheses testing for a normal distribution, then we need to **CALCULATE THE SE FOR DIFFERENT PARAMETERS!!!!!