Recap of Past Weeks

- **t-test**: used for hypothesis testing when
  1. the population variance is unknown,
  2. the sampling distribution of the sample mean is at least approximately normally distributed.
     - sampled population normal $\implies$ sampling distribution normal for any sample size.
     - sampled population non-normal $\implies$ sampling distribution approximately normal for large sample sizes (CLT).
- Parametric procedures for the population mean require that the sampling distribution is (approx.) normally distributed.
- Test for normality can be done:
  i) Visually: histogram, Q-Q plots.
  ii) Numerically: mean $\lessgtr$ median; comparing skewness and kurtosis to normal distributions.
  iii) Statistically: Shapiro-Wilk test (under the null: data comes from a normal distribution).
- **Non-parametric tests**: sign test and Wilcoxon signed-rank test for the median.

THE UNIVERSITY OF MELBOURNE

Comparing Two Groups

- **Paired sample design**: linked samples but independent pairs.
  - If the population of differences is (at least approximately) normally distributed and its variance (σ_D) is unknown, we can use the matched-pairs t-test.
- **Independent sample design**: independent random samples from two populations.
  - If the sampled populations are (at least approximately) normally distributed and their variances (σ_1^2 and σ_2^2) are unknown, we can use the two-independent sample t-test.
  - If the variance of the two populations are equal, we estimate them from the pooled sample:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \rightarrow s_{\bar{X}_1 - \bar{X}_2}^2 = s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)$$

  - If the variance of the two populations are unequal, we estimate them separately:

$$s_{\bar{X}_1 - \bar{X}_2}^2 = \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}$$