CHOOSING THE RIGHT ELEMENTARY STATISTICAL TEST

- The first step in determining what statistical test to use is to determine the type of research question to be answered by the statistical analysis. In elementary courses, the two basic types of questions are:
  1. **the degree of relationship or dependence among variables** (H_0 = there is no relationship or dependence, and the statistical test answers the question as to whether any relationship or dependence found is sufficiently different from zero that it can be considered "statistically significant".)
  2. **the significance of group differences** (H_0 = there is no difference between groups. The statistical test answers the question as to whether an observed difference is probably due just to random factors, or is large enough to be considered "statistically significant" and due to the treatment factor.)
- The next step is to determine the nature(s) of the variables under discussion, and whether they meet the assumptions of a particular test (e.g. the data are normally distributed).
- Types/levels of data:
  - **nominal** = unordered categories (e.g. religion; country of birth; etc.)
  - **ordinal** = ordered categories (e.g. level of agreement on an opinion survey; proficiency level at a martial art as measured by the colour of one's belt)
    - ➤ For nominal and ordinal data, what is usually recorded is the **number** of occurrences of a particular result (e.g. *number* of Christians, *number* of Buddhists etc. but these numbers are *not* the values of the variable. In this case, variable = religion, values = Christian, Buddhist, …and the numbers are the number of occurrences of a particular value.)
  - **interval** = ordinal + distance between values is of constant size (e.g. temperature)
  - **ratio** = interval + (i) there is a meaningful zero and (ii) the ratio between two numbers is meaningful (e.g. weight, distance, number of children)
    - ➤ ratio and interval data can be either **discrete** (i.e. there are gaps between values, e.g. number of children) or **continuous** (i.e. there are no gaps between values (e.g. weight, height).

Type of Question	Level of Data / Assumptions	Examples	Statistical Test
Degree of relationship or dependence among variables			
Is there a **relationship** between one **trait** and another? (I.e. does the observed frequency of one trait **depend** on the frequency of another?)	Both nominal or one nominal and one ordinal	(i) Is eye colour independent of hair colour? (ii) Is opinion about a government policy independent of political affiliation?	Contingency table χ^2
Is there a relationship between two continuous variables?	interval, ratio	(i) Does GPA depend on IQ? (ii) Does weight depend on height?	t-test for correlations using Pearson's correlation coefficient
Is there a relationship between two ordinal variables?	ordinal	(i) How strong is the relationship between the starting grid position of a race driver and their finishing position?	t-test for correlations using Spearman's r_s

Are the observed frequencies the same as an expected set of frequencies?	All (though often nominal)	(i) Are the absentee rates the same for each day of the week? (ii) Are the number of workplace accidents the same for each hour of the day? (iii) Is the observed set of scores normally distributed?	Goodness-of-fit χ^2
Significance of group differences			
Is a population mean or proportion (as estimated by a sample statistic) the same or different from a specified value?	- For means, data is interval or ratio level, the population is normal with a known variance. - For proportions, the variable is a binomial nominal variable	(i) Is the mean weight of soup in a can the same as stated on the label? (ii) Are 90% of pizza orders delivered within 30 minutes? (Success = delivered within 30 minutes, failure = delivered over 30 minutes.)	z-test for large samples, t-test for small samples.
Is there a difference between the means or proportion of "successes" for two populations? (Comparative experiments.)	Populations approximately normal and with same variances; samples independent	(i) Is there a difference in the mean lifetimes of Battery A and Battery B? (ii) Is there a difference in the proportion of consumers who bought Product A before and after an advertising campaign? (iii) Did a group of trainees score significantly better on some test after training compared to before training? (Dependent samples.)	If samples are independent, two-sample t-test (small samples) or z-test (large samples), related sample t-test if samples are dependent.
As above	Data at least ordinal; can be used if t-test assumptions not met		Mann-Whitney
Are the means of three or more populations (as estimated by samples) the same?	Interval or ratio level data	H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$	ANOVA
As above	Ordinal data and/or populations not normal		Kruskal-Wallis
As above, but more than one independent variable.		(i) Sleep score versus amount of exercise and whether exercise was in the morning or the afternoon. (ii) Time of travel from A to B versus route taken and driver.	Factorial ANOVA (Looking for 'main' and 'interaction' effects)
Multiple comparison of means	Interval or ratio level data	e.g. H_0: $\mu_1 = \mu_2 = \mu_3 > \mu_4 = \mu_5$	Linear contrast