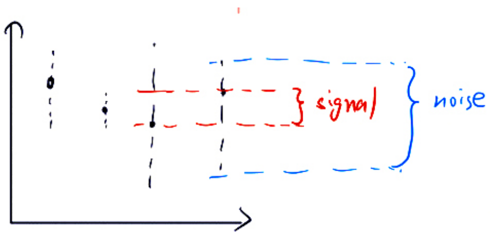
- **ANOVA**: Abbreviation of **Analysis of Variance**.
- **Methods for comparing means** of continuous responses between multiple groups.
- **F-ratio**: Signal/noise



- **Reasons why using "2+2+2=3" is undesirable**:
  - It's more work than we need to do. Three tests may not seem too bad, but to compare 10 groups we would have to do 45 different pairwise t-tests.
  - It can lead to lots of false positive results. Every test has the potential to incorrectly reject H0; i.e. falsely identify a difference between a pair of groups. If we do lots of tests then we risk generating lots of false positives.
- **ANOVA model**:

$$Y_{ij} = \mu_i + e_{ij}$$

  μ_i is the true mean response for the ith group at the population level. e_{ij} is the error term for the jth response in the ith group. The error terms are assumed to be independent and to follow a N(0, σ^2) with constant variance. The number of different groups is denoted K, and the number of responses in the ith group is denoted n_i.
- **Est. mean for ANOVA**: The "." = est. Value.

$$\hat{\mu}_i = \bar{y}_i.$$

- **Sample mean for the ith group:**

$$\bar{y}_i. = \frac{1}{n_i}\sum_{j=1}^{n} y_{ij}$$

- **Formula for residual sum of squares in ANOVA** (no need to memorise):

$$RSS = \sum_{i=1}^{K}\sum_{j=1}^{n_i}(y_{ij} - \hat{\mu}_i)^2 = \sum_{i=1}^{K}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i.)^2$$

- **Total sum of squares in ANOVA** (no need to memorise):

$$TSS = \sum_{i=1}^{K}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}..)^2$$

- $y..$ is the sample mean overall the data.

- **Formula for GSS in ANOVA** (no need to memorise):

$$GSS = TSS - RSS$$

$$GSS = \sum_{i=1}^{k} n_i(\bar{y}_i. - \bar{y}..)^2$$

GSS can be interpreted as a measure of the variation that is explained by differences between groups.

- **Setting up the hypotheses to test ANOVA:**

As usual, the null hypothesis will be the 'no difference' hypothesis:

$$H_0 : \mu_1 = \mu_2 = ... = \mu_K$$

The alternative is simply an expression that the null is incorrect:
$H_A : \mu_1, \mu_2, ..., \mu_K$ not all equal.

- **Equation for F statistic:**

$$F = \frac{GSS/(K-1)}{RSS/(n-K)}$$

- **GMS**: GSS/(K - 1) is the group mean square.

- **RMS**: RSS/(n - K) is the residual mean square.

- **What situations would let H_0 fail**: Large differences between group means. Relatively large value of GSS. A large value of F.

- **ANOVA table:**

Source	SS	DF	MS
Groups	GSS	$K-1$	$\frac{GSS}{K-1}$
Residuals	RSS	$n-K$	$\frac{RSS}{n-K}$
Total	TSS	$n-1$	

- **P-value is right censored**.

- **Blocking variable**: A second treatment variable that when included in ANOVA analysis will have the effect of reducing the SSE term (noise).