T-Tests and ANOVA Formulas with Details

1 T-Tests

T-tests are statistical methods used to determine if there is a significant difference between means. This document covers the one-sample, independent (unpaired), and paired t-tests.

1.1 One-Sample T-Test

This test compares the mean of a single sample to a known population mean.

Formula:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \tag{1}$$

Details of the Formula:

- t: The t-statistic, indicating how many standard errors the sample mean is from the population mean. It follows a t-distribution with $n-1$ degrees of freedom.
- $\bar{x}$: The sample mean, calculated as the average of all observations, $\bar{x} = \frac{\sum x_i}{n}$.
- μ: The hypothesized population mean (a fixed value being tested against).
- s: The sample standard deviation, measuring data spread, calculated as $s = \sqrt{\frac{\sum(x_i-\bar{x})^2}{n-1}}$.
- n: The sample size (number of observations).
- The denominator $s/\sqrt{n}$ is the standard error of the mean, estimating the variability of the sample mean.

Assumptions: Data should be approximately normally distributed, observations are independent, and the variable is continuous. No significant outliers.

1.2 Independent Samples T-Test (Two-Sample, Unpaired)

This test compares the means of two independent groups. The formula assumes unequal variances (Welch's t-test).

Formula:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \tag{2}$$

Details of the Formula:

- t: The t-statistic, compared to a t-distribution. Degrees of freedom are approximated using the Welch-Satterthwaite equation, roughly $df \approx \frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1-1}+\frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2-1}}$.
- $\bar{x}_1$, $\bar{x}_2$: Means of the two groups, calculated as averages for each group.
- s_1, s_2: Standard deviations of the two groups.
- n_1, n_2: Sample sizes of the two groups.
- The denominator is the standard error of the difference in means.

For equal variances, use pooled variance: $s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$, and the formula becomes:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \tag{3}$$

with $df = n_1 + n_2 - 2$.

Assumptions: Normality of data in each group, independence of samples, homogeneity of variances (for pooled version), continuous variable, no outliers.

1.3 Paired Samples T-Test

This test compares means from the same group under different conditions or times.

Formula:

$$t = \frac{\bar{d}}{s_d/\sqrt{n}} \tag{4}$$

Details of the Formula:

- t: The t-statistic, with $df = n - 1$.
- $\bar{d}$: The mean of the differences between paired observations, where $d_i = x_{1i} - x_{2i}$, and $\bar{d} = \frac{\sum d_i}{n}$.
- s_d: The standard deviation of the differences, $s_d = \sqrt{\frac{\sum(d_i-\bar{d})^2}{n-1}}$.
- n: Number of pairs.
- The denominator is the standard error of the mean difference.

Assumptions: Differences are normally distributed, pairs are independent, continuous variable.

2 ANOVA (One-Way Analysis of Variance)

One-way ANOVA compares means across three or more independent groups to determine if at least one differs.

Formula for F-Statistic:

$$F = \frac{MS_B}{MS_W} \tag{5}$$

Details of the Formula:

- F: The F-statistic, following an F-distribution with degrees of freedom $df_B = k-1$ (between) and $df_W = n-k$ (within). Larger F indicates greater between-group variation relative to within-group.
- MS_B (Mean Square Between Groups): Average variation between group means, $MS_B = \frac{SS_B}{k-1}$.
- MS_W (Mean Square Within Groups): Average variation within groups, $MS_W = \frac{SS_W}{n-k}$.
- SS_B (Sum of Squares Between Groups): Measures variation due to group differences, $SS_B = \sum n_j(\bar{x}_j - \bar{x})^2$, where $\bar{x}_j$ is group mean, $\bar{x}$ is grand mean, n_j is group size.
- SS_W (Sum of Squares Within Groups): Measures variation within groups, $SS_W = \sum\sum(x_{ij} - \bar{x}_j)^2$.
- SS_T (Total Sum of Squares): $SS_T = SS_B + SS_W = \sum\sum(x_{ij} - \bar{x})^2$.
- k: Number of groups.
- n: Total sample size across all groups.

Assumptions: Normality of residuals, homogeneity of variances across groups (largest SD $\leq 2\times$ smallest), independence of observations, random samples. If significant, follow with post-hoc tests like Tukey's test.