Key Concepts and Terminology

ANOVA: Analysis of variance is a parametric test that measures the difference in the means between two or more levels of a factor.

Factor: Thought of as the variable of interest or the independent variable (x).

Levels: These are the categories within the factor or independent variable. Example is the factor Weight group can have several levels ranging from (<100 Ibs to >201 Ibs).

Skewness: The measure of asymmetry of a distribution (right or left skewed).

$$Skewnes = \sqrt{\frac{\sum_{i=0}^{n}(X_n - \bar{X})^3}{(N-1) * \sigma^3}}$$

Kurtosis: Measure of "tailedness" or outliers for a distribution. Leptokurtic (>0) vs platykurtic (<0). Practice these calculations in Excel.

$$kurtosis \ = \frac{\sum_{i=0}^{n}(Y_n - \bar{Y})^4}{(\sum_{i=0}^{n}(Y_n - \bar{Y})^4)^2}$$

Homogeneity of Variance (Homoscedasticity): Test of the similarity in variance between the levels being assessed. Ensure the differences in variance are not extreme so as to question the sampling technique or if the levels are derived from the same population; imagine levels being assessed coming from a different population (monkeys and humans), not an APPLE-TO-APPLE comparison.

Lavene's Test: Used to test homoscedasticity with the Null being the data is homoscedastic. Implications are p > 0.01 (or 0.05) mean accepting the NULL.

Key Concepts and Terminology

Inter-level degrees of freedom (DF): $N_{levels}(k=6) - 1 = df_{bl}=5$

Intra-level DF: $N_{total}(382) - k(6) = df_{wl} = 376$

Total DF: $SUM(df_{bl}, df_{wl}) = 381$

F-ratio: Ratio of inter-level difference to intra-level difference. The F-ratio also addresses the concern of homoscedasticity but only does so relative to the inter-level difference. THE F-RATIO CAN ALMOST BE THOUGHT OF AS THE REVERSE OF LAVENE'S TEST.

$$F - ratio = \frac{\frac{\sum_{i=0}^{g=5} (X_g - \bar{X}_n)^2}{k-1(df_b)}}{\frac{\Sigma_{i=0}^{m} \sum_{i=0}^{0} (X_o - \bar{X})^2}{N-k(df_w)}} = \frac{29610.31}{1392.393} = 21.27$$

Eta-Squared (ω^2): A measure of effect size in Anova analysis. Measures the magnitude of the difference between groups. Welch Omega is used when Lavene's test yields rejection of the NULL but strictly speaking no further analysis should be conducted once Lavene's test fails.

$$\omega^2 = \frac{SS_b - (df_b)MS_w}{SS_t + MS_w} = \frac{148051.6 - (5 * 1392.393)}{671591 + 1392.393} = 0.209$$

$$Welch: Adjusted\ \omega^2 = \frac{(df_b)(F-1)}{(df_b)(F-1) + N_{total}} = \frac{128.746}{212.746} = 0.603$$

Robust Tests of Equality of Means

Tingwei Adeck Blood Pressure Difference (After - Before)

	Statistic[a]	df1	df2	Sig.
Welch	64.810	2	50.363	<.001
Brown-Forsythe	85.373	2	51.337	<.001

a. Asymptotically F distributed.

Key Concepts and Terminology

Post-Hoc Analysis: Analysis after the fact (rejection or failing to reject the NULL) on a level vs level basis. Easier with fewer levels. Popular post-hoc tests are Tukey, LSD, and Scheffe'. LSD is the most powerful (True positive detector) vs Scheffe' being the least powerful. It is a smart idea to match levels of significance for the F-ratio to the choice of post-hoc analysis. Tukey test for 2 means but then a T-test should do the trick. USE A MORE POWERFUL TEST WHEN LEVELS OF SIGNIFICANCE ARE LOW. In the weight_group example Scheffe' is a good choice since the level of significance is strong.

Post-Hoc Analysis Number of comparisons: The number of comparisons possible based on the number of levels. To avoid doing math, a little python program (numb_comparisons.py) was written to get the number of comparisons, new type I error rate and Bonferroni alpha level.

New Type I error Rate: Alpha value based on the number of comparisons.

$$Type\ I\ error\ rate = 1 - (1 - 0.05)^{\#comp} = 1 - (0.95)^3 = 0.143$$

Bonferroni alpha: A correction performed to ensure that chances of a Type I error are reduced.

$$\alpha_{new} = \frac{\alpha_{\mathrm{original}}}{\mathrm{n}} = \frac{0.05}{3} = 0.01667$$

Standard Error of Mean: Used when comparing several sample means relative to the population mean. Used to obtain the confidence interval.

$$\mathrm{SEM} = \text{Z-score} * \frac{\sigma}{\sqrt{\mathrm{n}}}$$

Key Concepts and Terminology

Confidence Interval: An interval derived from the mean plus minus the SEM. The confidence interval is arbitrary and increases based on the multiple of the standard deviation (1σ, 2σ, 3σ).