3.1 K-Sample Permutation Tests

- comparing the distributions of k different samples

$H_0: F_1(x) = F_2(x) = \cdots = F_k(x)$

$H_a: F_i(x) \le F_j(x)$ or $F_i(x) \ge F_j(x)$, for at least one pair of i and j.

$F(x-\mu_i)$

3.1.1 F Statistics

Sum of squares for treatments

$$SST = \sum_{i=1}^{k} n_i(\bar{X}_i - \bar{X})^2, \quad \bar{X} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i} X_{ij}}{N}$$

Mean Squares for treatment

$$MST = \frac{SST}{k-1}$$

Sum of Squares for error

$$SSE = \sum_{i=1}^{k}(n_i - 1)S_i^2$$

Mean Squares for error

$$MSE = \frac{SSE}{N-k}$$

F-Statistic

$$F = \frac{MST}{MSE}$$, F-distribution with $\overset{\text{numerator}}{k-1}$ and $\overset{\text{denominator}}{N-k}$ degrees of freedom

* we may calculate the p-value of the above F-statistic, but it is better to use a permutation F test when distributions are not normally distributed.

3.1.2 Steps in Carrying Out the Permutation F-test

1. Obtain the F-statistic for the original data. F_{obs}
2. Obtain all possible permutations of the N observations among k treatments in which there are n_i observations in treatment $i = 1, 2, 3, \dots, k$. $\dfrac{N!}{n_1!\, n_2! \cdots n_k!}$
3. compute the F statistic for each permutation.
4. Compute $\dfrac{\#\text{ of obs } F_i(x) \ge F_{obs}}{\text{all permutation}}$, because F distribution is always an upper-tail test.

* the number of permutation is large even for modest sample sizes, so we can run random sampling from permutation
$R = 1000$ is sufficient.

3.1.3 Alternative Forms of the Permutation F statistic

Let SS_{total} be the total sum of squares,

$$SS_{total} = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij}-\bar{X})^2 \quad , \text{ when } \quad SS_{total} = SST + SSE$$

SS_{total} has the same value for all permutations. Let $SS_{total} = C$,

$$F = \frac{SST/(k-1)}{(C-SST)/(N-k)} \quad \text{can be used for F statistic}$$

Bonus

$$SST = \sum_{i=1}^{k} n_i \bar{X}_i^2 - N\bar{X}^2$$

$$SSX = \sum_{i=1}^{k} n_i \bar{X}_i^2$$

3.2 Kruskal-Wallis Test

Kruskal-Wallis Statistic

$$KW = \underbrace{\frac{12}{N(N+1)}}_{\text{scaling factor}} \underbrace{\sum_{i=1}^{k} n_i\left(\bar{R}_i - \frac{N+1}{2}\right)^2}_{\text{treatment sum of squares}}$$

For small samples, most of the Kruskal-Wallis critical values will be somewhat smaller than the chi-square critical values. In this case, rejection of the null hypothesis using the chi-square critical value would assure that the exact level of significance is no greater than its p-value.

3.2.2 Adjustment for Ties

The process will be the same as any test adjustment for ties, except that,

$$KW_{ties} = \frac{1}{S_R^2}\sum_{i=1}^{k} n_i\left(\bar{R}_i - \frac{N+1}{2}\right)^2 \quad , \quad S_R^2 = \text{the sample variance of the combined adjusted ranks.}$$

An Alternative Formula for Tied Ranks

$$KW_{ties} = \frac{KW}{\left\{1 - \frac{\sum_{i=1}^{k}(t_i^3 - t_i)}{N^3 - N}\right\}} \quad , \quad KW = \frac{12}{N(N+1)}\sum_{i=1}^{k} n_i\left(\bar{R}_i - \frac{N+1}{2}\right)^2$$

t_i = # of tied obs in i^{th} group

The data have potential outliers
The population distributions have heavy tails
The population distributions are significantly skewed

} Kruskal-Wallis

The population distributions are normal
The population distributions are light tailed
The population distributions are symmetric

} F-test

3.2.3 An Intuitive Derivation of the Chi-Square Approximation for KW

- Under the assumption that observations come from a normal distributions come from a normal distribution with common variance σ^2, the quantity SST/σ^2 has a χ^2 distribution with $k-1$ degrees of freedom. For ranks, the mean rank is $(N+1)/2$, so the rank version of SST is given,

$$SST_R = \sum_{i=1}^{k} n_i \left(\bar{R}_i - \frac{N+1}{2}\right)^2$$

- It is reasonable to suppose that it should be possible to find a constant C so that $C(SST_R)$ has an approximate χ^2 distribution with $k-1$ d.f.

3.2.4 Tests on General Scores

Skipped

$E[C(SST_R)] = k-1$, using this property,

$$E\left\{\left(\bar{R}_i - \frac{N+1}{2}\right)^2\right\} = Var(\bar{R}_i) = \frac{N-n_i}{N-1} \cdot \frac{\sigma_R^2}{n_i} \Rightarrow E(SST_R) = \sum_{i=1}^{k} \frac{N-n_i}{N-1} \cdot \sigma_R^2 = (k-1)\frac{N\sigma_R^2}{N-1}$$

Since $E[C(SST_R)] = k-1$, $C = \frac{N-1}{N\sigma_R^2} = \frac{1}{S_R^2}$, provided $S_R^2 = \frac{N(N+1)}{12}$

∴ Multiplying C by SST_R gives the Kruskal-Wallis statistic

3.3 Multiple Comparisons $\binom{k}{2}$ pairwise tests

Previous tests only verify if there are differences among treatments; they cannot verify which treatment/s differ.

Multiple comparisons using pairwise tests allow us to verify which treatment differs from the others, if any.

FWER ↓

Family-Wise Error Rate:
- the probability of making one or more false rejections (FWER > α)
FWER = 1 − P(no false rejection for k treatments)
= 1 − P(no rejection of H_0)$^{k(k-1)/2}$
= 1 − $(1-\alpha)^{k(k-1)/2}$

3.3.1 Three Rank-Based Procedures for Controlling Experiment-Wise Error Rate Assuming No Ties in the Data

1. Bonferroni Adjustment

- uses α' rather than α. $\alpha' = \dfrac{\alpha}{\frac{k(k-1)}{2}}$

- have an experiment-wise error rate no greater than α.

- If the data are normally distributed, we may use the t-test; otherwise, use the Wilcoxon rank-sum test or any other nonparametric approach.

2. Fisher's Protected Least Significant Difference (LSD)

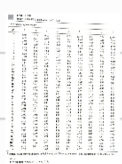
- applies to observations that have been selected from normal distributions, but it is often applied even when the assumption of normality is violated.

1) run F-test for equality of means as in the one-way analysis of variance.
2) if the F-statistic is significant at a desired level α, run all pairwise t-tests at level α.

$$|\bar{x}_i - \bar{x}_j| \geq \underbrace{t(\alpha/2, df)}_{\text{upper-tail with } N-k \text{ df}} \underbrace{\sqrt{MSE\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}}_{\text{Least Significant Difference}}$$

* - a large sample rank-based analogue of the LSD is to test first for equality of distributions using the "Kruskal-Wallis" test.
- if the Kruskal-Wallis test is significant at level α, then run the test below,

$$|\bar{R}_i - \bar{R}_j| \geq \underbrace{z(\alpha/2)}_{\text{upper-tail}} \sqrt{\underbrace{\frac{N(N+1)}{12}}_{\text{sample variance of combined ranks}}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$



3. Tukey's HSD Procedure (Honest Significant Difference)

- measuring the largest difference between sample means.

Appendix A8

$$Q = \max_{ij}\left\{\frac{\sqrt{n}\,|\bar{x}_i - \bar{x}_j|}{\sqrt{MSE}}\right\} = \frac{\sqrt{n}\,(\max_i\{\bar{x}_i\} - \min_i\{\bar{x}_i\})}{\sqrt{MSE}}$$

* critical value table in Appendix A8

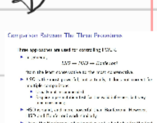
$$|\bar{x}_i - \bar{x}_j| \geq \overbrace{q(\alpha, k, \underbrace{df}_{\substack{\text{df of MSE}\\ = k(n-1)}})}^{\text{upper-tail}} \underbrace{\sqrt{\frac{MSE}{n}}}_{\text{Honest Significant Difference}} \quad \xLeftrightarrow{\text{alternative}} \quad |\bar{R}_i - \bar{R}_j| \geq q(\alpha, k, \infty)\sqrt{\frac{N(N+1)}{12n}}$$

Tukey-Kramer Procedure : to declare treatments i and j to be statistically significantly different.

$$|\bar{x}_i - \bar{x}_j| \geq q(\alpha, k, \underbrace{df}_{df = N-k})\sqrt{\frac{MSE}{2}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)} \quad \xLeftrightarrow{\text{alternative}} \quad |\bar{R}_i - \bar{R}_j| \geq q(\alpha, k, \infty)\sqrt{\frac{N(N+1)}{24}\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$$

comparing 3

↓



3.3.2 Multiple Comparisons for General Scores (Including Ties)
3.3.3 Multiple Comparison Permutation Tests

Bonferroni Adjustment for Permutation tests
- performing two-sample permutation tests on each of the pairs of treatments at the adjusted level of significance $\alpha' = 2\alpha / K(K-1)$.
- has an experiment-wise error rate of no more than α.

Permutation LSD
- If the F-test at level of α is significant, select a statistic T_{ij} for comparing the mean scores of 2 treatments.

1. obtain all possible samples without replacement of sizes n_i and n_j, or if this is not feasible, obtain an appropriately large randomly selected subset of such samples.
2. Obtain the permutation distribution of $|T_{ij}|$, and run,

$$|T_{ij}| \geq t^*(\alpha)$$

3. the p-value for comparing treatment i with j is the fraction of the permutation distribution greater than or equal to the observed value of $|T_{ij}|$.

* T_{ij} may be the difference of two means, the two sample t, standardized Wilcoxon rank-sum, or the like.

Permutation HSD
1. Permute the data as in the permutation F-test, and for each permutation, obtain

$$Q^* = \max_{ij} |T_{ij}|$$

2. obtain the upper-tail $q^*(\alpha)$.

$$|T_{ij}| \geq q^*(\alpha)$$

3. a p-value for comparing treatment i to j is the fraction of the permutation distribution of Q^* greater than or equal to the observed value of $|T_{ij}|$
4. For the special case of the two-sample t-statistics for T_{ij}, find the permutation distribution of

$$Q^* = \max_{ij} \left\{ \frac{|\bar{X}_i - \bar{X}_j|}{\sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}} \right., \quad T_{ij} = \left. \frac{\bar{X}_i - \bar{X}_j}{\sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}} \right\}$$

3.3.4 Variance of a Difference of Means When Sampling from a Finite Population

- Let $\bar{T}_1$ and $\bar{T}_2$ denote the sample means of the two samples n_1 and n_2. Then,

$Var(\bar{T}_1 - \bar{T}_2) = Var(\bar{T}_1) + Var(\bar{T}_2) - 2Cov(\bar{T}_1, \bar{T}_2)$, here

$Var(\bar{T}_i) = \left(\frac{N-n_i}{N-1}\right)\frac{\sigma^2}{n_i}$, $i = 1, 2$, and $cov(\bar{T}_1, \bar{T}_2) = -\frac{\sigma^2}{N-1}$, therefore

$$Var(\bar{T}_1 - \bar{T}_2) = \frac{N\sigma^2}{N-1}\left(\frac{1}{n_1} + \frac{1}{n_2}\right).$$

- Assuming approximate normality for the difference of sample means, we can assert the quantity,

$$Z = \frac{\bar{T}_1 - \bar{T}_2}{\sqrt{\frac{N\sigma^2}{N-1}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \Rightarrow \overbrace{Z_{\alpha/2}\sqrt{\frac{N\sigma^2}{N-1}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}^{LSD}.$$

Rank Version :

$$LSD_{ranks} = Z_{\alpha/2}\sqrt{\frac{N(N+1)}{12}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

* a similar approach applies to the Bonferroni and Tukey procedures.

3.4 Ordered Alternatives

- if treatments are not equal, it may be possible to anticipate the direction in which the treatments differ.
- We are interested in alternative hypotheses in which observations from treatment 1 tend to be smaller than observations from treatment 2, and so on.

$H_a: F_1(x) \geq F_2(x) \geq F_3(x) \geq \cdots \geq F_k(x)$

- If we have shift alternatives, like $F_i(x) = F(x - \mu_i)$, then

$H_a: \mu_1 \leq \mu_2 \leq \mu_3 \leq \cdots \leq \mu_k$

- A general form of a test statistic for testing the hypothesis is the sum of the pairwise statistics

$T = \sum_{i<j} T_{ij}$

3.4.1 Jonckheere - Terpstra Test

- a test statistic of the form T in which T_{ij}'s are one-sided Mann-Whitney statistics.

$$T = \sum_{i<j} T_{ij}$$

Obtaining a p-value for JT

1. Compute JT_{obs}, the observed value of JT from the original data.
2. obtain all possible samples without replacement of sizes n_i and n_j, or if this is not feasible, obtain an appropriately large randomly selected subset of such samples.

3. To obtain the upper-tail p-value, compute the fraction of the JT's in step 2 that are greater than or equal to JT_{obs}.

3.4.2 Large-Sample Approximation

When the data have no ties,

$$E(JT) = \sum_{i<j} \frac{n_i n_j}{2} = \frac{N^2 - \sum_{i=1}^{k} n_i^2}{4}$$

, and the variance of JT is,

$$Var(JT) = \frac{N^2(2N+3) - \sum_{i=1}^{k} n_i^2(2n_i+3)}{72}$$

, the standardized test statistic is

$$Z = \frac{JT - E(JT)}{\sqrt{Var(JT)}}$$

, or were we to compare ranks,

$$E(JT_{wilcoxon}) = \sum_{i<j} \frac{n_j(n_i+n_j+1)}{2}$$