Section 5.2: Analyzing Several Independent Samples

- **The M-W test is designed to compare two populations.**

- **Sometimes we have *k* independent samples from *k* populations.**

- **We wish to test whether all *k* populations are identical in distribution.**

Kruskal-Wallis Test

- **We assume the *k* random samples are all mutually independent and that the measurement scale is at least ordinal.**

- **The K-W test is again based on the ranks.**

- **Denote Sample 1 as** $X_{11}, X_{12}, \ldots, X_{1n_1}$

Sample 2 as $X_{21}, X_{22}, \ldots, X_{2n_2}$

$\vdots$

Sample *k* as $X_{k1}, X_{k2}, \ldots, X_{kn_k}$

- **We combine all *k* samples and rank the observations in the combined sample from 1 (smallest) to *N* (largest).**

- **Let** $R_i = \sum_{j=1}^{n_i} R(X_{ij}) \qquad i = 1, 2, \ldots, k$

be the sum of the ranks for the i-th sample.

Hypotheses: H_0: All k populations have identical distributions

H_1: The k populations do not have identical means

Test Statistic:

$$T = \left[\frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i}\right] - 3(N+1)$$

Null Distribution of *T*

Note: Under H_0, $E(R_i) = \frac{n_i(N+1)}{2}$

$$\text{var}(R_i) = \frac{n_i(N+1)(N-n_i)}{12}$$

By the CLT, for large samples

$$\frac{R_i - E(R_i)}{\sqrt{\text{var}(R_i)}} \mathrel{\dot\sim} N(0,1)$$

and $$\frac{[R_i - E(R_i)]^2}{\text{var}(R_i)} = \frac{\left(R_i - \frac{n_i(N+1)}{2}\right)^2}{n_i(N+1)(N-n_i)/12} \mathrel{\dot\sim} \chi^2_1$$

If the R_i's were independent, summing these χ^2_1 r.v.'s over all $i=1,\dots,k$ would produce a χ^2_k r.v.

- Since $\sum_{i=1}^{k} R_i = \frac{N(N+1)}{2}$, there is a dependence among the R_i's.
- Kruskal showed $$T = \sum_{i=1}^{k} \frac{\left[R_i - \frac{n_i(N+1)}{2}\right]^2}{n_i(N+1)(N-n_i)/12} \mathrel{\dot\sim} \chi^2_{k-1}$$
- This T is algebraically equivalent to our test statistic T.

• So the asymptotic null distribution of T is χ^2 with $(k - 1)$ degrees of freedom.

Decision Rule

• T is large when the R_i's are fairly different from each other.

• This is evidence in favor of H_1.

So: Reject H_0 if $T > \chi^2_{1-\alpha}$ with $(k-1)$ d.f. ← found in Table A2.

Example 1: In an experiment, 43 newborn chicks were each given one of 4 diets. Weight gain in the first 21 days was measured (in grams). Is there evidence (at $\alpha = 0.05$) that the four diets produce different mean weight gains? H_0: All 4 diets produce same weight gain distributions

At least 2 of

H_1: The 4 diets have different mean weight gains.

$R(X_{1j})$'s: 19, 10, ..., 4 → sum to R_1

$R(X_{2j})$'s: 41, 12, ..., 7.5 → sum to R_2

$R(X_{3j})$'s: 29, 36.5, ..., 39 → sum to R_3

$R(X_{4j})$'s: 20, 34, ..., 30 → sum to R_4

⇒ $T = 10.97$ (tedious calculation, so get it from R).

$\chi^2_{1-\alpha} = \chi^2_{.95}$ with $k-1=3$ d.f.

$= 7.815$ from Table A2.

Since $10.97 > 7.815$, reject H_0. Conclude the 4 diets do not yield ~~different~~ the same mean weight gain.

On computer: Use `kruskal.test` function in R (see example code on course web page.

P-value $\approx .0119$ from R.

• If H_0 is rejected, we use multiple comparisons to infer which population means seem to differ.

• Populations *i* and *j* are significantly different if:

$$\left|\frac{R_i}{n_i} - \frac{R_j}{n_j}\right| > \underset{df = N-k}{t_{1-\alpha/2}} \left(S^2 \frac{N-1-T}{N-k}\right)^{1/2} \left(\frac{1}{n_i} + \frac{1}{n_j}\right)^{1/2}$$

• This can be checked readily in R.

Example 1 again:

From R: Population 1 and Population 3 are significantly different in terms of mean weight gain.

Population 1 and Population 4 are significantly different.

– All other pairs of population means do not differ significantly.

(Experimentwise significance level = 0.05)

• The K-W test can be used with categorical data (e.g., data in contingency tables) as long as the variable observed on each individual is ordinal so that the categories can be ranked in order.

Example 2: The grade distributions for 3 instructors were compared to see whether students tended to get similar grade distributions across instructor. The data are given on page 293.

• If we score A, B, C, D, F numerically as 4, 3, 2, 1, 0, then we can perform the K-W test on the data:

H_0: The 3 instructors tend to have the same grade distributions

H_1: The grade distributions of the 3 instructors differ.

From R: $T = 0.3209$

$\chi^2_{.95}$ (2 d.f.) $= 5.991$ (Table A2)

Since $0.3209 < 5.991$, we fail to reject H_0. We cannot conclude the grade distributions of the 3 instructors differ. P-value $\approx .8517$ from R.

Comparison to Other Tests

• When all *k* populations are normal, the usual parametric procedure to compare the *k* population means is the (one-way) analysis of variance (ANOVA) F-test.

• The F-test is robust against the normality assumption in terms of the actual significance level.

• But the F-test can have very low **power when the data are nonnormal (especially when heavy-tailed).**

• The A.R.E. of the K-W test relative to the F-test and relative to the median test is very similar to the A.R.E. of the M-W test relative to its competitors.