Section 5.3: Tests about Several Variances

- We have seen tests designed to compare several populations in terms of their <u>means</u>.

- Suppose we wish to compare two or more populations in terms of their variances.

Note that the null hypothesis

$$H_0: var(X) = var(Y)$$

can be written as

$$H_0: E[(X-\mu_X)^2] = E[(Y-\mu_Y)^2]$$

which is identical to the H_0 from the M-W test, with

$(X-\mu_X)^2$ playing the role of X

and $(Y-\mu_Y)^2$ playing the role of Y

- If we estimate μ_X and μ_Y (either with the group sample mean or sample median) then we could perform the M-W test on the values $(|X_1 - \mu_X|, \ldots, |X_n - \mu_X|)$ and $(|Y_1 - \mu_Y|, \ldots, |Y_n - \mu_Y|)$, where the μ_X and μ_Y are estimated.

- This is the Talwar-Gentle test.

- Conover showed the power is improved by summing the <u>squared</u> ranks of the first sample instead of the ranks. This is the test Conover presents in Section 5.3.

- **The Fligner-Killeen test is similar, but replaces the ranks R_i with the transformed ranks**

$$\left[\Phi^{-1}\left(\frac{n+m+1+R_i}{2(n+m+1)}\right)\right]^2$$

where $\Phi(\cdot)$ is the standard normal cdf.

- **In R, the `fligner.test` function performs this test (the function does not permit a one-tailed alternative).**

- **Any of these three tests (Talwar-Gentle, Conover, Fligner-Killeen) may be extended to three or more groups just as the M-W test is extended to the K-W test.**

<u>Example 1</u>: A cereal manufacturer is considering replacing its old packaging machine with a new one. The hope is to reduce the variability in the cereal amounts placed in the boxes. The data are:

X **Current: 10.8, 11.1, 10.4, 10.1, 11.3** $\Rightarrow \bar{X} = 10.74$

$|X-\hat{\mu}_X|$: .06, .36, .34, .64, .56

rank: 4 10 9 12 11 $\Rightarrow T = 46$

Y **New: 10.8, 10.5, 11.0, 10.9, 10.8, 10.7, 10.8** $\Rightarrow \bar{Y} = 10.79$

$|Y-\hat{\mu}_Y|$: .01, .29, .21, .11, .01, .09, .01

rank: 2 8 7 6 2 5 2

Hypotheses: $H_0: \operatorname{var}(X) \le \operatorname{var}(Y)$

$H_1: \operatorname{var}(X) > \operatorname{var}(Y)$

- **Talwar-Gentle test:** $n=5, \ m=7$

Table A7

Reject H_0 if $T > W_{.95} = 5(12+1) - W_{.05} = 65 - 22 = 43$

Since $T = 46 > 43$, reject H_0 and conclude the new machine has smaller variance. P-value $\approx .017$ from R.

Example 2: **Numerous specimens from four brands of golf ball were each hit by a machine in an experiment, and the distances (in yards) they traveled were recorded. Is there evidence that the four brands have different population variances? (Use $\alpha = 0.05$.)**

H_0: The four brands have equal variances $(\sigma_A^2 = \sigma_B^2 = \sigma_C^2 = \sigma_D^2)$

H_1: The variances of at least two brands differ.

From R: Talwar-Gentle p-value = 0.74

Fligner-Killeen p-value = 0.8979

We fail to reject Ho. Cannot conclude the four brands' variances differ.

- **The Fligner-Killeen test typically has more power than the Talwar-Gentle test.**

- **All three tests are robust against violations of the normality assumption.**

Comparison to Parametric Tests

- **If two populations are normal, an F-test can be used to compare their variances.**

• **This F-test is <u>highly sensitive</u> to the normality assumption: If the data distribution is actually heavy-tailed, the actual significance level may be** <u>two to three times greater</u> **than the nominal α.**

• **Bartlett's test is the parametric test comparing 3 or more variances – it is also <u>highly sensitive</u> to the normality assumption.**

• **Levene's test is a parametric test that is <u>somewhat</u> less sensitive to the normality assumption.**

Efficiency of the Conover Test

Population	A.R.E.(Conover vs. F)
Normal	0.76
Uniform (light tails)	1.00
Double exponential (heavy tails)	1.08

• **The efficiencies are the same in the case of 3 or more samples.**

• **Since the Fligner-Killeen test is usually somewhat more powerful than the Conover test, its A.R.E. should be similar (perhaps slightly better) than the A.R.E.'s given above.**