Section 5.3: Tests about Several Variances

- We have seen tests designed to compare several populations in terms of their **means**.

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

Note that the null hypothesis

can be written as

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

- If we estimate $\mu_X$ and $\mu_Y$ (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 $\mu_X$ and $\mu_Y$ are estimated.

- This is the Talwar-Gentle test.

- Conover showed the power is improved by summing the **squared** 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

• 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.

**Example 1:** 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:

Current: 10.8, 11.1, 10.4, 10.1, 11.3

New: 10.8, 10.5, 11.0, 10.9, 10.8, 10.7, 10.8

Hypotheses:

• Talwar-Gentle test:
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 $\alpha = 0.05$.)

- 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 **highly sensitive** to the normality assumption: If the data distribution is actually heavy-tailed, the actual significance level may be __________ __________ than the nominal $\alpha$.

• Bartlett’s test is the parametric test comparing 3 or more variances – it is also **highly sensitive** to the normality assumption.

• Levene’s test is a parametric test that is somewhat less sensitive to the normality assumption.

Efficiency of the Conover Test

<table>
<thead>
<tr>
<th>Population</th>
<th>A.R.E.(Conover vs. F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td></td>
</tr>
<tr>
<td>Uniform (light tails)</td>
<td></td>
</tr>
<tr>
<td>Double exponential (heavy tails)</td>
<td></td>
</tr>
</tbody>
</table>

• 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.
Section 5.4: Measures of Rank Correlation

- Correlation is used in cases of paired data, to describe the association between the two random variables, say \( X \) and \( Y \).

For all measures of correlation:

- The correlation is always between -1 and 1.
- Positive correlation \( \Rightarrow \) The two variables are positively associated (large values of one variable correspond to large values of the other variable)
- Negative correlation \( \Rightarrow \) The two variables are negatively associated (large values of one variable correspond to small values of the other variable)
- Correlation near 0 \( \Rightarrow \) large values of one variable tend to appear randomly with either large or small values of the other variable.

How far the correlation is from 0 measures the strength of the relationship:

- nearly 1 \( \Rightarrow \) Strong positive association between the two variables
- nearly -1 \( \Rightarrow \) Strong negative association between the two variables
- near 0 \( \Rightarrow \) Weak association between the two variables

- When the correlation is zero, this sometimes (but not always) means that \( X \) and \( Y \) are independent.
• The **Pearson (product-moment) correlation coefficient** (denoted \( r \)) is a numerical measure of the **strength** and **direction** of the **linear** relationship between two variables.

Formula for \( r \) (the Pearson correlation coefficient between two paired data sets \( X_1, \ldots, X_n \) and \( Y_1, \ldots, Y_n \)):

This is the same as:

• If the bivariate distribution of \((X, Y)\) is unknown, then the Pearson correlation coefficient cannot be used for hypothesis tests and confidence intervals.

**Spearman Correlation Coefficient**

• An alternative measure of correlation simply ranks the two samples (**separately**, not combined) and calculates the Pearson measure on the ranks \( R(X_i) \) and \( R(Y_i) \) rather than on the actual data values.

• This produces the **Spearman Correlation Coefficient**,.
• Since the average of the $n$ ranks (1, 2, ..., $n$) in each sample is:

the formula for the Spearman Correlation Coefficient is

• We can use Spearman’s $\rho$ as a test statistic to test whether $X$ and $Y$ are independent.

Null Hypothesis:

3 Possible Alternatives

• The exact null distribution of $\rho$ is tabulated (for $n \leq 30$) in Table A10. Note $w_{1-p} =$
• For larger sample sizes (or with many ties), the approximate quantiles may be used:

where \( z_p \) is a standard normal quantile.

**Decision Rules**

<table>
<thead>
<tr>
<th>Two-tailed</th>
<th>Lower-tailed</th>
<th>Upper-tailed</th>
</tr>
</thead>
</table>

• Approximate P-values can be obtained from the normal distribution using one of equations (12)-(14) on pp. 317-318, or by interpolating within Table A10, but we will typically use software to get approximate P-values.

Example: The GMAT score and GPA for 12 MBA graduates are given on p. 316. Is there evidence of positive correlation between GMAT and GPA?

On computer: Use `cor.test` function in R with method=“spearman” (see code on course web page).
Kendall’s Tau

• Another measure of correlation, Kendall’s Tau, is based on the idea of concordant and discordant pairs.

• Consider two bivariate observations, say, \((X_i, Y_i)\) and \((X_j, Y_j)\).

• The two observations are concordant if both numbers in one observation are larger than the corresponding numbers in the other observation.

• The two observations are discordant if the numbers in observation \(i\) differ in opposite directions as the corresponding numbers in observation \(j\).

Examples:

If \(X_i < X_j\) and \(Y_i < Y_j\), then the \(i\)-th and \(j\)-th observations are:

If \(X_i < X_j\) and \(Y_i > Y_j\), then the \(i\)-th and \(j\)-th observations are:

If \(X_i > X_j\) and \(Y_i < Y_j\), then the \(i\)-th and \(j\)-th observations are:

If \(X_i > X_j\) and \(Y_i > Y_j\), then the \(i\)-th and \(j\)-th observations are:

Let \(N_c = \)

and \(N_d = \)
• There are possible pairs of bivariate observations.

• If there are no ties (no cases when $X_i = X_j$ or $Y_i = Y_j$), then

• A general definition of Kendall’s tau that allows for ties is

where we compute $N_c$ and $N_d$ by:

Examples on p. 316 data:
• We can use $T =$
as a test statistic to test for independence of $X$ and $Y$.

Null Hypothesis:

3 Possible Alternatives

• The exact null distribution of $T$ is tabulated (for $n \leq 60$) in Table A11. Note $w_{1-p} =$

• For larger sample sizes (or with many ties), the quantile for $T$ is approximately:

where $z_p$ is a standard normal quantile.
Decision Rules

| Two-tailed | Lower-tailed | Upper-tailed |

* Approximate P-values can be obtained from the normal distribution using one of equations (20)-(21) on p. 322, or by interpolating within Table A11, but we will typically use software to get approximate P-values.

Example: Recall the GMAT score and GPA for 12 MBA graduates on p. 316. Is there evidence of positive correlation between GMAT and GPA?

On computer: Use `cor.test` function in R with `method="kendall"` (see code on course web page).
Daniels Test for Trend

- The Daniels Test is a more powerful test for trend than the Cox-Stuart Test from Chapter 3.

- If we have a time-ordered sample $X_1, \ldots, X_n$, we create paired data: $(\text{Time}_1, X_1), \ldots, (\text{Time}_n, X_n)$.

- Then the test of independence based on Spearman’s rho or Kendall’s tau is performed, with

and the possible alternatives being:

Example on global temperature data again: Is there evidence of an increasing temperature trend?
Comparison to Competing Tests

• If the distribution of $X$ and $Y$ is ________________, a $t$-test based on Pearson’s correlation coefficient is used to test for independence.

• The A.R.E. of the tests based on Spearman’s and Kendall’s measures relative to that $t$-test are each _______ when the data are bivariate normal.

• However, the nonparametric tests can have better efficiency than the $t$-tests for many nonnormal distributions.

• These nonparametric tests only require the data to be ________________, rather than requiring normality.

• As measures of correlation, Spearman’s rho and Kendall’s tau are appropriate as long as the data are at least _______________ on the measurement scale.

• Kendall’s tau is often used as a measure of association when the data are binary and ordered (for example, Fail/Pass).

Example: 20 students each took both a Pass-Fail test in Math and a Pass-Fail test in History. Describe the association between the two tests.