Inference about Variances

Sampling Distribution of Variances:

You want to know the variance of a population. But you realize the population is too large to actually compute the true variance. So you decide to observe a sample from the population and use the sample variance as an estimate of the population variance. You need the sampling distribution of the sample variance in order to make inference about the population variance.

The population is normally distributed with mean $\mu$ and standard deviation $\sigma$. Let $y$ denote an observation from the population.

Draw sample of size $n \rightarrow y_1, y_2, \ldots y_n$.

Compute the sample variance $s^2 = \frac{\sum (y_i - \bar{y})^2}{n-1}$.

Then sampling distribution of $(n-1)s^2/\sigma^2$ is chi-square with n-1 degrees of freedom.
Inference about Variances

Use the table for the chi-square distribution or a computer program to get probabilities.

Examples: Suppose $X$ has chi-square distribution with df=4.

Then $P(X > 11.1433) = .025$.

Notation:

$X \sim \chi^2_k$

$X$ is distributed chi-square with $k$ degrees of freedom

$P\left(X > \chi^2_{a,k}\right) = a$

Examples:

$\chi^2_{.025,4} = 11.143$

$\chi^2_{.025,10} = 20.48$

$\chi^2_{.975,10} = 3.247$
Inference about Variances

95% Confidence Interval for a Variance:

\[ P\left( \chi^2_{0.025} < \frac{(n-1)s^2}{\sigma^2} < \chi^2_{0.975} \right) = 0.95 \]

\[ P\left( \frac{(n-1)s^2}{\chi^2_{0.975}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{0.025}} \right) = 0.95 \]

95% confidence interval for \( \sigma^2 \):

\[ \left( \frac{(n-1)s^2}{\chi^2_{0.025}}, \frac{(n-1)s^2}{\chi^2_{0.975}} \right) \]

Test of \( H_0: \sigma^2 = \sigma_0^2 \) versus \( H_a: \sigma^2 > \sigma_0^2 \)

Test statistic \( X = \frac{(n-1)s^2}{\sigma_0^2} \)

Large values are “significant”
Inference about Variances

Example: Construct 95% confidence interval for $\sigma^2$ for the hand-span for women

The summary statistics for the hand-span data are

\[
\begin{align*}
\bar{y}_m &= 8.99 & \bar{y}_f &= 8.27 \\
\bar{s}^2_m &= 0.470 & \bar{s}^2_f &= 0.682 \\
n_m &= 22 & n_f &= 13
\end{align*}
\]

\[
\left( \frac{(n-1)s^2}{\chi^2_{0.025}}, \frac{(n-1)s^2}{\chi^2_{0.975}} \right) = \left( \frac{(13-1).682}{\chi^2_{12,.025}}, \frac{(13-1).682}{\chi^2_{12,.975}} \right) \\
= \left( \frac{8.184}{23.34}, \frac{8.184}{4.404} \right) = (0.351, 1.858)
\]
Inference about Variances

Use the F distribution to compare two variances.

Reading probabilities from F table:

The F distribution has numerator and denominator degrees of freedom

\[ F \sim F_{df1, df2} \] means \( F \) has \( df1 \) numerator and \( df2 \) denominator degrees of freedom

\[ P(F > F_{a, df1, df2}) = a \]

Example:

\[ F_{0.05, 12, 21} = 2.25 \]

Property of F distribution:

\[ P(F < 1 / F_{a, df2, df1}) = a \]

Result: \[ F_{1-a, df1, df2} = 1 / F_{a, df2, df1} \]

Example:

\[ F_{0.95, 21, 12} = 1 / 2.25 = 0.44 \]
Inference about Variances

Test of hypothesis Test of $H_0: \sigma_1^2 = \sigma_2^2$ versus $H_a: \sigma_1^2 > \sigma_2^2$

Test statistic $F = s_1^2/s_2^2$

Large values are “significant”

Example: Test $H_0: \sigma_m^2 = \sigma_f^2$ versus $H_a: \sigma_m^2 > \sigma_f^2$

$F = .470/.682 = .689 < 1$ implies non-significant
Inference about Variances

Test of hypothesis Test of $H_0: \sigma_1^2 = \sigma_2^2$ versus $H_a: \sigma_1^2 \neq \sigma_2^2$

Test statistic $\quad F = \frac{\max(s_1^2, s_2^2)}{\min(s_1^2, s_2^2)}$

Large values are “significant”

Example: Test $H_0: \sigma_m^2 = \sigma_f^2$ versus $H_a: \sigma_m^2 \neq \sigma_f^2$

$$F = \frac{.682}{.470} = 1.45, \quad df1 = 12, \quad df2 = 21$$

$$F_{12,21,.25} = 1.38, \quad F_{12,21,.10} = 1.87 \quad \rightarrow .10 < p < .25$$