

Chapter 1

Chapter 2

Chapter 3

Statistical Methods Related to Variances

4.1 Introduction

All the proposed methods are applicable under the following conditions:

- The samples are random, simple, and independent.
- The parent populations are normally distributed.

4.2 Estimation of Population Variance and Confidence Interval

1. Definition of the Variance Estimator: The population variance estimator, based on a sample of size n , is:

$$\hat{\sigma}^2 = \frac{nS^2}{n-1} = \frac{SCE}{n-1},$$

where:

- $S^2 = \frac{\sum_{i=1}^{n_1} (x_i - \bar{x})^2}{n}$ is the sample variance (mean of squared deviations).
- $SCE = \sum_{i=1}^{n_1} (x_i - \bar{x})^2$ is the sum of squared deviations from the mean.

This estimator corrects the bias due to using n instead of $n - 1$.

2. Confidence Interval for the Variance: The goal is to estimate the true variance σ^2 with a confidence level of $(1 - \alpha)$.

- **For $n \leq 30$ (small sample):** the variance follows a χ^2 distribution with $n - 1$ degrees of freedom. The confidence interval is given by $[S_{\text{inf}}, S_{\text{sup}}]$

$$S_{\text{inf}} = \sqrt{\frac{SCE}{\chi_{1-\alpha/2, n-1}^2}}, \quad S_{\text{sup}} = \sqrt{\frac{SCE}{\chi_{\alpha/2, n-1}^2}}.$$

- **For $n > 30$ (large sample):** we can use the normal approximation, as the distribution of S^2 tends toward normal. The confidence interval becomes:

$$S_{\text{inf}}^2 = S^2 \left(1 - z_{\alpha/2} \sqrt{\frac{2}{n}} \right), \quad S_{\text{sup}}^2 = S^2 \left(1 + z_{\alpha/2} \sqrt{\frac{2}{n}} \right),$$

where $z_{\alpha/2}$ is the standard normal quantile.

Example (Small Sample)

Consider the weight (in kg) of 6 mice:

$$x = [20.1, 20.5, 19.8, 20.2, 20.0, 20.3]$$

1. Compute the sample mean \bar{x} , the variance S^2 and the corrected variance estimator $\hat{\sigma}^2$.
2. Determine the 95% confidence interval for the population variance.

Step 1: Sample Mean

$$\bar{x} = \frac{\sum x_i}{n} = \frac{20.1 + 20.5 + 19.8 + 20.2 + 20.0 + 20.3}{6} = 20.15$$

Step 2: Deviations and SCE

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
20.1	-0.05	0.0025
20.5	0.35	0.1225
19.8	-0.35	0.1225
20.2	0.05	0.0025
20.0	-0.15	0.0225
20.3	0.15	0.0225
SCE		0.295

The variance is:

$$S^2 = \frac{SCE}{n} = \frac{0.295}{6} = 0.04917$$

Step 3: unbiased Variance Estimator

$$\hat{\sigma}^2 = \frac{SCE}{n-1} = \frac{0.295}{5} = 0.059$$

$$\hat{\sigma} = \sqrt{0.059} \approx 0.243$$

Step 4: 95% Confidence Interval for the Variance

For a small sample ($n \leq 30$), use the Chi-square distribution with $df = n - 1 = 5$.

Given:

$$\chi_{0.025,5}^2 = 0.831, \quad \chi_{0.975,5}^2 = 12.833$$

Confidence interval bounds:

$$\sigma_{\text{inf}}^2 = \frac{0.295}{12.833} = 0.02299, \quad \sigma_{\text{sup}}^2 = \frac{0.295}{0.831} = 0.3553$$

Standard deviation interval:

$$\sigma_{\text{inf}} = \sqrt{0.02299} = 0.152, \quad \sigma_{\text{sup}} = \sqrt{0.3553} = 0.596$$

Step 5: Interpretation

The unbiased estimate of the population variance is:

$$\hat{\sigma}^2 = 0.059 \quad (\hat{\sigma} = 0.243 \text{ kg})$$

The 95% confidence interval for the true variance is:

$$0.02299 \leq \sigma^2 \leq 0.3553$$

The corresponding confidence interval for the standard deviation is:

$$0.152 \leq \sigma \leq 0.596$$

The 95% confidence interval for the true standard deviation is $[S_{\text{inf}}, S_{\text{sup}}] = [0.152, 0.596]$ kg.

4.3 Variance Conformity Test

1. Tested Hypothesis: We want to check whether the observed variance differs from a theoretical value σ_0^2 :

$$H_0 : \sigma^2 = \sigma_0^2 \quad H_1 : \sigma^2 \neq \sigma_0^2$$

2. Case $n \leq 30$: We use the χ^2 distribution for the test statistic:

$$\chi_{\text{obs}}^2 = \frac{nS^2}{\sigma_0^2} = \frac{SCE}{\sigma_0^2}.$$

Decision:

$$\text{Reject } H_0 \quad \text{if } \chi_{\text{obs}}^2 \geq \chi_{1-\alpha/2, n-1}^2 \quad \text{or} \quad \chi_{\text{obs}}^2 \leq \chi_{\alpha/2, n-1}^2.$$

Otherwise, do not reject H_0 .

3. Case $n > 30$: For a large sample, we can use the normal approximation:

$$Z = \frac{S^2 - \sigma_0^2}{\sigma_0^2 \sqrt{2/n}}.$$

Exercise

A laboratory measures the blood pressure (mmHg) of $n = 6$ patients:

$$x = [120, 125, 118, 122, 121, 124]$$

The claimed population variance is $\sigma_0^2 = 9$.

Question: Test at the 5% significance level whether the population variance differs from the claimed value.

Solution: Variance Conformity Test

Data:

$$x = [120, 125, 118, 122, 121, 124], \quad n = 6, \quad \sigma_0^2 = 9$$

Step 1: Hypotheses

$$H_0 : \sigma^2 = 9, \quad H_1 : \sigma^2 \neq 9$$

Step 2: Sample Mean

$$\bar{x} = \frac{120 + 125 + 118 + 122 + 121 + 124}{6} = \frac{730}{6} \approx 121.67$$

Step 3: Corrected Sample Variance Compute deviations and squared deviations:

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
120	-1.67	2.7889
125	3.33	11.0889
118	-3.67	13.4889
122	0.33	0.1089
121	-0.67	0.4489
124	2.33	5.4289
SCE		33.35

Corrected sample variance:

$$\hat{\sigma}^2 = \frac{SCE}{n-1} = \frac{33.35}{5} \approx 6.67$$

Step 4: Test Statistic χ^2

$$\chi_{\text{obs}}^2 = \frac{nS^2}{\sigma_0^2} = \frac{SCE}{\sigma_0^2} = \frac{(n-1)\hat{\sigma}^2}{\sigma_0^2} = \frac{5 \cdot 6.67}{9} \approx 3.71$$

Step 5: Critical Values Degrees of freedom: $df = n - 1 = 5$

Two-sided test at $\alpha = 0.05$:

$$\chi_{0.025,5}^2 = 0.831, \quad \chi_{0.975,5}^2 = 12.833$$

Step 6: Decision

$$0.831 < 3.71 < 12.833 \Rightarrow \text{Do not reject } H_0$$

Step 7: 95% Confidence Interval for Variance

$$\sigma_{\text{inf}}^2 = \frac{(n-1)\hat{\sigma}^2}{\chi_{0.975,5}^2} = \frac{33.35}{12.833} \approx 2.60$$

$$\sigma_{\text{sup}}^2 = \frac{33.35}{0.831} \approx 40.14$$

95% CI for variance: [2.60, 40.14]

Conclusion At the 5% significance level, there is no evidence that the true population variance differs from the theoretical value of 9 mmHg². The confidence interval shows that the true variance could reasonably vary widely around the sample estimate.

3.1 Tests for comparing two variances and the confidence interval

Test of Equality of Two Variances

$$H_0 : \sigma_1^2 = \sigma_2^2$$

In practice, we compute the ratio of the two variances by placing the larger variance in the numerator and the smaller in the denominator:

$$F_{\text{obs}} = \frac{\hat{\sigma}_{\text{max}}^2}{\hat{\sigma}_{\text{min}}^2}$$

Case $n_1 \neq n_2$:

$$F_{\text{obs}} = \frac{\hat{\sigma}_{\text{max}}^2}{\hat{\sigma}_{\text{min}}^2}$$

Reject H_0 if:

$$F_{\text{obs}} \geq F_{1-\frac{\alpha}{2}}(n_1-1, n_2-1)$$

Case $n_1 = n_2$:

$$F_{\text{obs}} = \frac{SCE_{\text{max}}}{SCE_{\text{min}}}$$

Reject H_0 if:

$$F_{\text{obs}} \geq F_{1-\frac{\alpha}{2}}(n_1-1, n_2-1)$$

This test is called the **Fisher F-test**.

Exercise

Two groups of students are tested for reaction time.

Group 1 ($n_1 = 10$) : 22, 25, 21, 23, 24, 26, 22, 25, 24, 23

Group 2 ($n_2 = 8$) : 30, 28, 29, 31, 32, 30, 27, 29

Test whether the variances are equal at significance level $\alpha = 0.05$.

Solution

1. Compute Sample Means

Group 1

Values X_i	Sum
22, 25, 21, 23, 24, 26, 22, 25, 24, 23	235

$$\bar{X}_1 = \frac{235}{10} = 23.5$$

Group 2

Values X_i	Sum
30, 28, 29, 31, 32, 30, 27, 29	236

$$\bar{X}_2 = \frac{236}{8} = 29.5$$

2. Variance Calculations

A) Group 1 Variance Table (Mean = 23.5)

X_i	$X_i - \bar{X}_1$	$(X_i - \bar{X}_1)^2$
22	-1.5	2.25
25	1.5	2.25
21	-2.5	6.25
23	-0.5	0.25
24	0.5	0.25
26	2.5	6.25
22	-1.5	2.25
25	1.5	2.25
24	0.5	0.25
23	-0.5	0.25
TOTAL	—	22.50

$$\hat{\sigma}_1^2 = \frac{22.5}{10 - 1} = \frac{22.5}{9} = 2.50$$

B) Group 2 Variance Table (Mean = 29.5)

X_i	$X_i - \bar{X}_2$	$(X_i - \bar{X}_2)^2$
30	0.5	0.25
28	-1.5	2.25
29	-0.5	0.25
31	1.5	2.25
32	2.5	6.25
30	0.5	0.25
27	-2.5	6.25
29	-0.5	0.25
TOTAL	—	18.00

$$\hat{\sigma}_2^2 = \frac{18}{8-1} = \frac{18}{7} = 2.5714$$

3. Compute the F Statistic

$$\hat{\sigma}_{\max}^2 = 2.5714, \quad \hat{\sigma}_{\min}^2 = 2.50$$

$$F_{\text{obs}} = \frac{2.5714}{2.5} = 1.0286$$

4. Critical Value

Degrees of freedom:

$$df_1 = n_{\max} - 1 = 7, \quad df_2 = n_{\min} - 1 = 9$$

Critical value (two-sided test):

$$F_{0.975}(7, 9) \approx 4.03$$

5. Decision

$$F_{\text{obs}} = 1.0286 < 4.03$$

Do NOT reject H_0

The population variances are statistically equal.

3.2 Two Tests for Equality of Multiple Variances

Two methods are commonly used:

- **Bartlett's Test:** applicable to samples of equal or unequal sizes.
- **Hartley's Test:** very fast, but only applicable for equal sample sizes.

a) Bartlett's Test

For p samples:

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_p^2$$

For each sample, compute:

$$\hat{\sigma}_i^2 = \frac{SCE_i}{n_i - 1}$$

The pooled variance:

$$\hat{\sigma}^2 = \frac{SCE_{\bullet}}{N - p}$$

Test statistic (case of unequal n_i):

$$\chi_{\text{obs}}^2 = \frac{2.3026(N - p) \log_{10} \hat{\sigma}^2 - \sum_{i=1}^p (n_i - 1) \log_{10} \hat{\sigma}_i^2}{1 + \frac{1}{3(p-1)} \left(\sum_{i=1}^p \frac{1}{n_i - 1} - \frac{1}{N - p} \right)}$$

The test statistic follows a chi-square distribution with $p - 1$ degrees of freedom:

$$\chi_{\text{obs}}^2 \sim \chi_{p-1}^2.$$

where $N = \sum_{i=1}^p n_i$.

Reject H_0 if:

$$\chi_{\text{obs}}^2 \geq \chi_{1-\alpha}^2(p - 1)$$

When: $n_1 = n_2 = \dots = n_p = n$,

$$H_0 : \hat{\sigma}_1^2 = \hat{\sigma}_2^2 = \dots = \hat{\sigma}_p^2$$

$$\chi_{\text{obs}}^2 = \frac{2.3026(n - 1) \left[p \log_{10} \left(\frac{SCE_{\bullet}}{p} \right) - \sum_{i=1}^p \log_{10} SCE_i \right]}{1 + \frac{p+1}{3p(n-1)}}$$

Example: Bartlett's Test with Unequal Sample Sizes

We have three independent samples drawn from normally distributed populations. We want to test the equality of the variances of the three populations at the significance level $\alpha = 0.05$.

The observed data are as follows:

Sample	Size n_i	Observations
1	10	12, 14, 11, 15, 13, 17, 16, 14, 12, 15
2	10	8, 9, 7, 10, 11, 9, 8, 10, 7, 9
3	10	20, 22, 19, 21, 23, 20, 24, 22, 21, 23

Solution: Bartlett's Test

We want to test:

$$H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 \quad \text{vs} \quad H_1 : \text{not all variances are equal}$$

at significance level $\alpha = 0.05$.

Step 1 — Sample means

$$\bar{x}_1 = \frac{12 + 14 + 11 + 15 + 13 + 17 + 16 + 14 + 12 + 15}{10} = 13.9$$

$$\bar{x}_2 = \frac{8 + 9 + 7 + 10 + 11 + 9 + 8 + 10 + 7 + 9}{10} = 8.8$$

$$\bar{x}_3 = \frac{20 + 22 + 19 + 21 + 23 + 20 + 24 + 22 + 21 + 23}{10} = 21.5$$

Step 2 — Sum of squares

$$SCE_1 = \sum (x_{1j} - \bar{x}_1)^2 = 32.9$$

$$SCE_2 = \sum (x_{2j} - \bar{x}_2)^2 = 15.6$$

$$SCE_3 = \sum (x_{3j} - \bar{x}_3)^2 = 22.5$$

—

Step 3 — Total sum of squares

$$SCE_{\bullet} = SCE_1 + SCE_2 + SCE_3 = 32.9 + 15.6 + 22.5 = 71.0$$

—

Step 4 — Bartlett's test statistic

For equal sample sizes, the formula is:

$$\chi_{\text{obs}}^2 = \frac{2.3026(n-1) \left[p \log_{10} \left(\frac{SCE_{\bullet}}{p} \right) - \sum_{i=1}^p \log_{10} SCE_i \right]}{1 + \frac{p+1}{3p(n-1)}}$$

- $n - 1 = 9$, $p = 3$ - $SCE_{\bullet}/p = 71/3 \approx 23.667$ - $\log_{10} SCE_1 \approx 1.517$, $\log_{10} SCE_2 \approx 1.193$, $\log_{10} SCE_3 \approx 1.352$

$$\sum \log_{10} SCE_i = 1.517 + 1.193 + 1.352 = 4.062$$

$$p \log_{10}(SCE_{\bullet}/p) = 3 \cdot 1.374 = 4.122$$

$$\text{Bracket term} = 4.122 - 4.062 = 0.06$$

—

Step 5 — Numerator and denominator

$$\text{Numerator} = 2.3026 \cdot 9 \cdot 0.06 \approx 1.244$$

$$\text{Denominator} = 1 + \frac{4}{3 \cdot 3 \cdot 9} = 1 + \frac{4}{81} \approx 1.049$$

$$\chi_{\text{obs}}^2 = \frac{1.244}{1.049} \approx 1.186$$

—

Step 6 — Critical value

Degrees of freedom: $p - 1 = 2$

$$\chi_{0.95}^2(2) = 5.991$$

Step 7 — Decision

$$\chi_{\text{obs}}^2 = 1.186 < 5.991 \Rightarrow \text{Fail to reject } H_0$$

—

Step 8 — Conclusion

There is no significant evidence that the population variances are different. Hence, the variances can be considered equal.

Remarks

- Bartlett's test is very sensitive to non-normality of the populations, regardless of sample sizes.
- The test is valid only if $n_i \geq 4$ and the number of samples is not too large.
- For two populations, Bartlett's test is equivalent to the F-test only if the two samples have the same sizes.

b) Hartley's Test

When all sample sizes are equal:

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_p^2$$

The test statistic is defined as:

$$F_{\text{obs}} = \frac{\hat{\sigma}_{\text{max}}^2}{\hat{\sigma}_{\text{min}}^2}$$

where $\hat{\sigma}_{\text{max}}^2$ and $\hat{\sigma}_{\text{min}}^2$ are the largest and smallest sample variances, respectively.

Decision rule: Reject H_0 if

$$F_{\text{obs}} \geq F_{1-\alpha}(p, n)$$

where p is the number of samples and n is the sample size per group.