AMS 315
Data Analysis
Chapter Seven Study Guide
Inferences about Population Variances
Spring 2023

Context

The variance of a random variable is a critical parameter in assessing risk and the extent to which the expected value of the random variable is known. Having a confidence interval for the population variance is a potentially useful tool. Since the two independent sample equal variance t-test assumes equal variances, it is logical to ask for a test of equality of variances. Standard statistical theory (specifically a procedure called the likelihood ratio test) suggests the ratio of sample variances is a good test.

In practice, the procedures given in this chapter are of limited value because the distributions of the test statistics are very sensitive to the assumption of normality. Specifically, the percentiles of the null distribution of the test statistics increase as the kurtosis of the sampled random variable increases. The procedures are of value for two reasons. First, the distribution of many random variables encountered in practice is normal or close to normal. Second, this material is an excellent introduction to using the central chi-squared distribution and the central F-distribution.

Chapter Seven

Single Population Variance (7.1-7.2)

The unbiased estimator of the sample variance $S^2 = \dfrac{\sum (Y_i - \bar{Y})^2}{n-1}$ is the core statistic. The key distributional fact is that $\dfrac{(n-1)S^2}{\sigma^2}$ has a central chi-squared distribution with $n-1$ degrees of freedom when sampling from a $N(\mu,\sigma^2)$ distribution. This is our first important use of the chi-squared distribution.

Note properties 1-3 of the chi-squared distribution listed in 7.2. The formula for the confidence interval for σ^2 in section 7.2 is the tool to answer homework and examination questions. When you work these problems, examine your answer and notice how wide the confidence interval for σ^2. Specifically examine the ratio of the upper limit to the lower limit. One gets percentiles of the chi-squared distribution from Table 7. The Excel spreadsheet and all statistical packages have the percentiles available as well.

Comparing Two Population Variances (7.3)

The assumptions of this test are that a random sample of size n_1 has been drawn from a random variable that is $N(\mu_1, \sigma_1^2)$ and that an independent random variable of size n_2 has been drawn from a random variable that is $N(\mu_2, \sigma_2^2)$. The test statistic of the null hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$ is the ratio of the sample variances $F = \frac{S_1^2}{S_2^2}$. The null distribution of this test statistic is a central F distribution with $n_1 - 1$ numerator and $n_2 - 1$ denominator degrees of freedom.

Note the properties of the F-distribution listed in section 7.3. The test of a two-sided alternative is complex. Example 7.5 shows you how to do this. Also pay attention to finding the confidence interval for the ratio of the population variances.

Comparing More than 2 Population Variances (7.4)

Hartley's $F_{\max} = \frac{S_{\max}^2}{S_{\min}^2}$ is used. This statistic does not have a central F-distribution under the null hypothesis that all variances are equal. You must use percentiles from Table 12. In practice, this test has a null distribution that is very dependent on the distribution of the sampled random variable. Your text recommends Levene's test. Statistical packages will calculate this test, and you should use it routinely. There is another more robust test of this null hypothesis that corrects for the estimated kurtosis of the sampled random variables. Some statistical packages report this test as well as or instead of Levene's test.

Summary (7.5)

You are responsible for key formulas 1-4 and 5a but not 5b.

Example Past Examination Questions

1. A research team took a sample of 8 observations from the random variable *Y*, which had a normal distribution $N(\mu, \sigma^2)$. They observed $\bar{y}_8 = 43.2$, where $\bar{y}_8$ is the average of the eight sampled observations and $s^2 = 517.5$ is the observed value of the unbiased estimate of σ^2, based on the sample values. Test the null hypothesis that $H_0 : \sigma^2 = 400$ against the alternative $H_1 : \sigma^2 > 400$ at the 0.10, 0.05, and 0.01 levels of significance.
   Answer: The value of the chi-squared test statistic is 9.056, with 7 degrees of freedom. The right-sided critical values are 12.02 (for the 0.10 level), 14.07 (for 0.05), and 18.48 (for 0.01). Accept the null hypothesis at the 0.10, 0.05, and 0.01 levels.

2. A research team took a sample of 7 observations from the random variable Y, which had a normal distribution $N(\mu, \sigma^2)$. They observed $\bar{y}_7 = 93.4$, where $\bar{y}_7$ was the average of the sampled observations and $s^2 = 47.5$ was the observed value of the unbiased estimate of σ^2, based on the sample values. Find the 99% confidence interval for σ^2.
Answer: The percentiles needed for 6 degrees of freedom are 0.6757 and 18.55. The 99% confidence interval for σ^2 based on the statistical estimate of 47.5 is between 15.36 (a factor of about 3 less than the estimate) and 421.785 (a factor of about 9 greater than the estimate).

3. A research team took a random sample of 9 observations from a normally distributed random variable Y and observed that $\bar{y}_9 = 91.2$ and $s_Y^2 = 229.6$, where $\bar{y}_9$ was the average of the nine observations sampled from Y and s_Y^2 was the unbiased estimate of $\text{var}(Y)$. A second research team took a random sample of 10 observations from a normally distributed random variable X and observed that $\bar{x}_{10} = 103.5$ and $s_X^2 = 917.6$, where $\bar{x}_{10}$ was the average of the ten observations sampled from X and s_X^2 was the unbiased estimate of $\text{var}(X)$. Test the null hypothesis $H_0 : \text{var}(X) = \text{var}(Y)$ against the alternative $H_1 : \text{var}(X) > \text{var}(Y)$ at the 0.10, 0.05, and 0.01 levels of significance.
Answer: The test statistic has the value 3.9965. The critical value for the 0.10 level is 2.56; for the 0.05 level 3.39; and for the 0.01 level 5.91. Reject the null hypothesis at the 0.10 and 0.05 levels; accept it at the 0.01 level.

4. A research team took a random sample of 9 observations from a normally distributed random variable Y and observed that $\bar{y}_9 = 91.2$ and $s_Y^2 = 529.6$, where $\bar{y}_9$ was the average of the nine observations sampled from Y and s_Y^2 was the unbiased estimate of $\text{var}(Y)$. A second research team took a random sample of 10 observations from a normally distributed random variable X and observed that $\bar{x}_{10} = 103.5$ and $s_X^2 = 894.3$, where $\bar{x}_{10}$ was the average of the ten observations sampled from X and s_X^2 was the unbiased estimate of $\text{var}(X)$. Find the 95% confidence interval for $\text{var}(X)/\text{var}(Y)$.
Answer: The 95% confidence interval for $\text{var}(X)/\text{var}(Y)$ is from 0.387 to 6.923. Since the confidence interval for $\text{var}(X)/\text{var}(Y)$ includes 1, we would accept the null hypothesis that the ratio of the variances was 1 at the 0.05 level of significance. Note that the confidence interval for the ratio of the variances has endpoints that are factors of roughly 4 times the ratio of the two variances.

End of Chapter 7 Study Guide