Lecture Notes for Chapter 7
Spring 2023

Probability Theory Facts

Let Z be $N(0,1)$. Then Z^2 has the (central) chi-squared distribution with 1 degree of freedom. This is denoted χ_1^2.

Let $Z_1, Z_2, \ldots, Z_n$ be $NID(0,1)$. Then $S_n = Z_1^2 + Z_2^2 + \cdots + Z_n^2 = \sum_{i=1}^{n} Z_i^2$ follows the (central) chi-square distribution with n degrees of freedom, denoted χ_n^2. The expected value of a χ_n^2 is *n*: $E(S_n) = E(Z_1^2) + E(Z_2^2) + \cdots + E(Z_n^2)$. Since $\text{var}(Z) = 1 = E(Z^2) - [E(Z)]^2 = 1$, then $E(Z^2) - [0]^2 = 1$. Using this in $E(S_n) = E(Z_1^2) + E(Z_2^2) + \cdots + E(Z_n^2) = n$. Further, the variance of a chi-square distribution with n degrees of freedom is *2n*: $\text{var}(S_n) = 2n$

Let Y be $N(\mu_Y, \sigma_Y^2)$. Then, $\frac{Y - \mu_Y}{\sigma_Y} = Z$ is $N(0,1)$. Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from Y, which is $N(\mu_Y, \sigma_Y^2)$. Then $\sum_{i=1}^{n} (\frac{Y_i - \mu_Y}{\sigma_Y})^2$ is χ_n^2. After factoring out σ_Y^2, $\frac{\sum_{i=1}^{n} (Y_i - \mu_Y)^2}{\sigma_Y^2}$ is also χ_n^2.

Since μ_Y is not known in applications, it must be estimated. An important property of a sample from a normal distribution is that $\frac{\sum_{i=1}^{n} (Y_i - \bar{Y}_n)^2}{\sigma_Y^2}$ is distributed as χ_{n-1}^2. That is, using the sample mean has reduced the degrees of freedom by one. From AMS 310, the unbiased estimator of the sample variance is $S^2 = \frac{\sum (Y_i - \bar{Y}_n)^2}{n-1}$. Since $(n-1)S^2 = (n-1)\frac{\sum (Y_i - \bar{Y}_n)^2}{n-1} = \sum_{i=1}^{n} (Y_i - \bar{Y}_n)^2$, $\frac{(n-1)S^2}{\sigma_Y^2}$ has a central chi-squared distribution with $n-1$ degrees of freedom when $Y_1, Y_2, \ldots, Y_n$ is a sample of size n from a $N(\mu_Y, \sigma_Y^2)$ distribution. This is our first important use of the chi-squared distribution. The tests in this chapter are usually one-sided.

Problem 1 from Chapter 7 Study Guide

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.

Solution: The test statistic is $TS = \frac{(n-1)S^2}{\sigma_Y^2}$, which has a central χ^2_{n-1}, where $n-1=8-1=7$. Since $H_0 : \sigma^2 = 400$, the null distribution of $TS = \frac{(n-1)S^2}{\sigma_Y^2} = \frac{(n-1)S^2}{400}$ is χ^2_7. The problem specifies a right-sided test with levels of significance 0.10, 0.05, and 0.01. From Table 7, the right-sided critical values are 12.02 (for the 0.10 level), 14.07 (for 0.05), and 18.48 (for 0.01). The value of the chi-squared test statistic is $ts == \frac{(8-1)\bullet 517.5}{400} = 9.056$. Since this is less than each of the critical values, we accept the null hypothesis at the 0.10, 0.05, and 0.01 levels. The value of the sample mean is not relevant. A common mistake is for a student to take the sample mean as a cue and answer with a one-sample t-test. This is not correct.

As usual, a confidence interval may be more informative than a statistical test. The next problem is an example of finding a confidence interval for a variance.

Problem 2 from Chapter 7 Study Guide

> 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.

Solution: Since the estimated variance has 6 degrees of freedom,

$\Pr\{0.6757 < \frac{\sum_{i=1}^{7}(Y_i - \overline{Y}_7)^2}{\sigma^2} < 18.55\} = \Pr\{0.6757 < \frac{6S^2}{\sigma^2} < 18.55\} = 0.99$. Then

$\Pr\{0.6757 < \frac{6S^2}{\sigma^2} < 18.55\} = \Pr\{\frac{1}{18.55} < \frac{\sigma^2}{6S^2} < \frac{1}{0.6757}\} = \Pr\{\frac{6S^2}{18.55} < \sigma^2 < \frac{6S^2}{0.6757}\}$.

The interval from $\frac{6S^2}{18.55}$ to $\frac{6S^2}{0.6757}$ is then the basis of a confidence interval for σ^2. In this problem, the left end of the confidence interval is $\frac{6s^2}{18.55} = 0.3235s^2 = 0.3235 \bullet 47.5 = 15.36$, and the right end is $\frac{6s^2}{0.6757} = 8.880s^2 = 8.880 \bullet 47.5 = 421.785$. The confidence interval extends from a factor of about 3 less than $s^2 = 47.5$ to a factor of about 9 greater than $s^2 = 47.5$. Again, the sample mean is not needed to answer the question.

When you work these problems, examine your answer and notice that the confidence interval for σ^2 is very wide. Specifically examine the ratio of the upper limit to the lower limit, here almost 28. It is remarkable that the t distribution stretches are as small as they are. One gets percentiles of the chi-squared distribution from Table 7. The Excel spreadsheet and all statistical packages have the percentiles available as well.

The F-distribution

Let X be $N(\mu_X, \sigma_X^2)$. Using the standard score transformation, $\frac{X - \mu_X}{\sigma_X} = Z$ is $N(0,1)$. Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from X, which is $N(\mu_X, \sigma_X^2)$. Then, $\frac{\sum_{i=1}^{n}(X_i - \mu_X)^2}{\sigma_X^2}$ is χ_n^2.

Let Y be $N(\mu_Y, \sigma_Y^2)$. Then, $\frac{Y - \mu_Y}{\sigma_Y} = Z$ is also $N(0,1)$. Let $Y_1, Y_2, \ldots, Y_m$ be a random sample of size m from Y, which is $N(\mu_Y, \sigma_Y^2)$. Then, $\frac{\sum_{i=1}^{m}(Y_i - \mu_Y)^2}{\sigma_Y^2}$ is χ_m^2. The definition of the central F distribution is that the random variable

$$F_{n,m} = \frac{\{[\sum_{i=1}^{n}(X_i - \mu_X)^2]/[n\sigma_X^2]\}}{\{[\sum_{i=1}^{m}(Y_i - \mu_Y)^2]/[m\sigma_Y^2]\}}$$ has a (central) F distribution with *n* numerator and *m* denominator degrees of freedom.

Application of the F distribution

The problem with this random variable is that the expected values are not known. As before, we use the sample averages as estimates of the expected values. The penalty for using sample data rather than expected values is a one degree reduction in both the numerator and denominator degrees of freedom. That is,

$$\frac{\{[\sum_{i=1}^{n}(X_i - \bar{X}_n)^2]/[(n-1)\sigma_X^2]\}}{\{[\sum_{i=1}^{m}(Y_i - \bar{Y}_m)^2]/[(m-1)\sigma_Y^2]\}} = \frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} = F_{n-1,m-1}$$ has a central F distribution with $n-1$ numerator and $m-1$ denominator degrees of freedom. Of course, there is still the issue of the unknown variances of *X* and *Y* that has to be dealt with.

One use of this random variable is to test the null hypothesis $H_0: \sigma_X^2 = \sigma_Y^2$. The most common alternative hypothesis is $H_1: \sigma_X^2 > \sigma_Y^2$. The test statistic for this hypothesis is $TS = \frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2}$. Under the null hypothesis $H_0: \sigma_X^2 = \sigma_Y^2$, $TS = \frac{S_X^2}{S_Y^2}$, and its null distribution is central $F_{n-1,m-1}$. Under the null hypothesis $E(S_X^2) = \sigma_X^2$ and $E(S_Y^2) = \sigma_Y^2$, so that $E_0(TS) \cong \frac{E(S_X^2)}{E(S_Y^2)} = 1$. Under the alternative hypothesis, $E_1(TS) \cong \frac{E(S_X^2)}{E(S_Y^2)} > 1$. That is, the test of $H_0: \sigma_X^2 = \sigma_Y^2$ against the alternative $H_1: \sigma_X^2 > \sigma_Y^2$ is a right-sided test. A value of *TS* near 1 (modulo statistical variation) supports the null hypothesis, and a value of *TS* much greater than 1 supports the alternative. The next problem illustrates the test.

Problem 3 from Chapter 7 Study Guide

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 $\mathrm{var}(Y)$. 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.

Solution: One has a choice of which sample variance to put in the numerator. When one puts the variance *hypothesized* to be larger in the numerator, then the test is right-sided. Here $ts = \frac{s_X^2}{s_Y^2} = \frac{917.6}{229.6} = 3.9965$, with 9 numerator and 8 denominator degrees of freedom. 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. The correct decision is to reject the null hypothesis at the 0.10 and 0.05 levels and accept it at the 0.01 level. As before, the sample means are not needed for the problem. Students who use the information given as the cue to their choice of statistical tests sometimes respond to a question like this with a two-sample t-test. This is incorrect.

Confidence interval for the ratio of variances $\frac{\sigma_X^2}{\sigma_Y^2}$

For this task, we use $TS = \frac{S_Y^2 / \sigma_Y^2}{S_X^2 / \sigma_X^2}$, which has an F-distribution with $m-1$ numerator and $n-1$denominator degrees of freedom. This choice may be counter-intuitive, but is necessary. The percentage points in Table 8 are based on right sided tail areas, so that $\Pr\{F_{1-\alpha/2,m-1,n-1} < \frac{S_Y^2 / \sigma_Y^2}{S_X^2 / \sigma_X^2} < F_{\alpha/2,m-1.n-1}\} = 1-\alpha$, and

$\Pr\{F_{1-\alpha/2,m-1,n-1} < \frac{\sigma_X^2}{\sigma_Y^2} \bullet \frac{S_Y^2}{S_X^2} < F_{\alpha/2,m-1.n-1}\} = 1-\alpha$.

Then, $\Pr\{(F_{1-\alpha/2,m-1,n-1})\frac{S_X^2}{S_Y^2} < \frac{\sigma_X^2}{\sigma_Y^2} < (F_{\alpha/2,m-1.n-1})\frac{S_X^2}{S_Y^2}\} = 1-\alpha$.

The values of $F_{\alpha/2,m-1,n-1}$are given in Table 8. These tables do not explicitly give $F_{1-\alpha/2,m-1,n-1}$. One needs to use a property of the F distribution to get this value. Since

$\Pr\{F_{1-\alpha/2,m-1,n-1} < \frac{S_Y^2 / \sigma_Y^2}{S_X^2 / \sigma_X^2}\} = 1 - \frac{\alpha}{2}$, $\Pr\{\frac{S_Y^2 / \sigma_Y^2}{S_X^2 / \sigma_X^2} < F_{1-\alpha/2,m-1,n-1}\} = \frac{\alpha}{2}$

$\Pr\{[1/[\frac{S_Y^2 / \sigma_Y^2}{S_X^2 / \sigma_X^2}] > [1/F_{1-\alpha/2,m-1,n-1}]\} = \frac{\alpha}{2}$. That is

$\Pr\{\frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} > \frac{1}{F_{1-\alpha/2,m-1,n-1}}\} = \frac{\alpha}{2}$. The distribution of $\frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2}$ is a central F with $n-1$ numerator degrees of freedom and $m-1$ denominator degrees of freedom. From the definition of the F percentage points in Table 8, $\Pr\{\frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} > F_{\alpha/2,n-1,m-1}\} = \frac{\alpha}{2}$.

Equating the two right hand sides of these inequalties shows that $F_{\alpha/2,n-1,m-1} = \frac{1}{F_{1-\alpha/2,m-1,n-1}}$, or equivalently, $F_{1-\alpha/2,m-1,n-1} = \frac{1}{F_{\alpha/2,n-1,m-1}}$. Then $\Pr\{(F_{1-\alpha/2,m-1,n-1})\frac{S_X^2}{S_Y^2} < \frac{\sigma_X^2}{\sigma_Y^2} < (F_{\alpha/2,m-1.n-1})\frac{S_X^2}{S_Y^2}\} = 1-\alpha$ reduces to $\Pr\{(\frac{1}{F_{\alpha/2,n-1,m-1}})\frac{S_X^2}{S_Y^2} < \frac{\sigma_X^2}{\sigma_Y^2} < (F_{\alpha/2,m-1.n-1})\frac{S_X^2}{S_Y^2}\} = 1-\alpha$. The interval that contains $\frac{\sigma_X^2}{\sigma_Y^2}$ with probability $1-\alpha$ is $(\frac{1}{F_{\alpha/2,n-1,m-1}})\frac{S_X^2}{S_Y^2}$ to $(F_{\alpha/2,m-1.n-1})\frac{S_X^2}{S_Y^2}$. We use the observed sample variances in the $1-\alpha$% confidence interval for $\frac{\sigma_X^2}{\sigma_Y^2}$: $(\frac{1}{F_{\alpha/2,n-1,m-1}})\frac{s_X^2}{s_Y^2}$ to $(F_{\alpha/2,m-1.n-1})\frac{s_X^2}{s_Y^2}$.

Problem 4 from Chapter 7 Study Guide

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 $\text{var}(X)/\text{var}(Y)$.

Solution: The sample variance $s_X^2 = 894.3$ is based on 9 degrees of freedom, and the sample variance $s_Y^2 = 529.6$ is based on 8 degrees of freedom. From Table 8, $F_{\alpha/2,m-1.n-1} = F_{0.025,8,9} = 4.10$, and $F_{\alpha/2,n-1.m-1} = F_{0.025,9,8} = 4.36$. The ratio $\frac{s_X^2}{s_Y^2} = \frac{894.3}{529.6} = 1.689$. The left endpoint is given by $\frac{1}{4.36}\frac{s_X^2}{s_Y^2} = 0.229 \bullet 1.689 = 0.387$

The right endpoint is given by $4.10\frac{s_X^2}{s_Y^2} = 4.10 \bullet 1.689 = 6.92$. The 95% confidence interval for $\text{var}(X)/\text{var}(Y)$ is from 0.387 to 6.92. 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 two-sided 0.05 level of significance. The confidence interval for the ratio of the variances has a right endpoint that is a factor of roughly 18 times the left endpoint. The sample averages do not enter into the solution of this problem. Some students respond incorrectly with a 95% confidence interval for the difference in means. This is not correct.

End of Chapter 7 Lecture Notes