Introductory Distribution Theory

James H. Steiger

Department of Psychology and Human Development Vanderbilt University

P311, 2012

A B A B A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 B
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A

Introductory Distribution Theory

Introduction

- 2 The Chi-Square Distribution
 - Some Basic Properties
 - Basic Chi-Square Distribution Calculations in R
 - Convergence to Normality
 - The Chi-Square Distribution and Statistical Testing
- (3) The F Distribution
 - \bullet Characterization of the F Distribution
 - \bullet The $F\text{-}\mathrm{Ratio}$ Test
- In Noncentral Chi-Square Distribution
 - Introduction
 - Calculations with the Noncentral Chi-Square Distribution
 - The Effect of Noncentrality

\bigcirc Noncentral F Distribution

- Introduction
- Asymptotic Behavior
- \bullet Calculations in the Noncentral F

・ロト ・ 同 ト ・ ヨ ト ・ ヨ ト

Introduction

- In this module, we review basic facts about the central and noncentral χ^2 and F distributions, and how they are relevant to statistical testing.
- Some of this material was covered in Psychology 310.

Some Basic Properties Basic Chi-Square Distribution Calculations in R Convergence to Normality The Chi-Square Distribution and Statistical Testing

イロト イポト イヨト イヨト

The Chi-Square Distribution Basic Characterization

- Suppose you have an observation x taken at random from a normal distribution with mean μ and variance σ^2 that you somehow knew.
- We can characterize this as a realization of a random variable X, where $X \sim N(\mu, \sigma)$.
- Now suppose we were to transform X to Z-score form, i.e., $Z = (X \mu)/\sigma$. Then we would have a random variable $Z \sim N(0, 1)$.
- Finally, suppose we were to square Z. This random variable Z^2 is said to have a *chi-square distribution with* one degree of freedom.
- We write $Z^2 \sim \chi_1^2$.

Some Basic Properties Basic Chi-Square Distribution Calculations in R Convergence to Normality The Chi-Square Distribution and Statistical Testing

イロト イヨト イヨト

The Chi-Square Distribution Some Properties

- A χ_1^2 random variable is essentially a folded-over and stretched out normal.
- Here's a picture of the density function of a standardized normal random variable and a χ_1^2 random variable overlaid on the same graph.

Some Basic Properties Basic Chi-Square Distribution Calculations in R Convergence to Normality The Chi-Square Distribution and Statistical Testing

・ロト ・ 同 ト ・ ヨ ト ・ ヨ ト

The Chi-Square Distribution Some Properties

- > curve(dchisq(x,1),0,9,col="red",lty=1,xlim=c(-9,9),ylim=c(0,1.5))
- > curve(dnorm(x),-9,9,col="blue",lty=2,add=T)
- > abline(v=0)



Some Basic Properties Basic Chi-Square Distribution Calculations in R Convergence to Normality The Chi-Square Distribution and Statistical Testing

The Chi-Square Distribution Some Properties

- With a little thought, you can see that because the graph is "folded over", the 95th percentile of the χ_1^2 distribution is the square of the 97.5th percentile of the standard normal distribution.
- The mean of a χ_1^2 variable is 1.
- The variance of a χ_1^2 variable is 2.
- The sum of ν independent χ_1^2 variables is said to have a chi-square distribution with ν degrees of freedom, i.e.,

$$\sum_{j=1}^{\nu} \chi_1^2 \sim \chi_{\nu}^2$$
 (1)

• The preceding results, along with well-known principles regarding the mean and variance of linear combinations of variables, implies that, for independent chi-squares having ν_1 and ν_2 degrees of freedom,

$$\chi^2_{\nu_1} + \chi^2_{\nu_2} \sim \chi^2_{\nu_1 + \nu_2}$$
 (2)

$$E(\chi_{\nu}^2) = \nu \tag{3}$$

$$Var(\chi^2_{\nu}) = 2\nu \tag{4}$$

Some Basic Properties Basic Chi-Square Distribution Calculations in R Convergence to Normality The Chi-Square Distribution and Statistical Testing

< ロト (同) (三) (三) (

Basic Calculations

- We perform basic calculations in R using the dchisq function to plot the density, pchisq to compute cumulative probability, and qchisq to compute percentage points.
- We've already seen an example of dchisq in our earlier chi-square distribution plot.
- Here we calculate the cumulative probability of a value of 3.7 in a χ^2_2 distribution.
 - > pchisq(3.7,2)
 - [1] 0.8427628
- Here we calculate the 95th percentile of a χ₅² variable.
 > gchisg(.95,5)
 - [1] 11.0705

Some Basic Properties Basic Chi-Square Distribution Calculations in R **Convergence to Normality** The Chi-Square Distribution and Statistical Testing

・ロト ・ 同 ト ・ ヨ ト ・ ヨ ト

Convergence to Normality

- Recall that the X_{ν}^2 variate is the sum of independent X_1^2 variates.
- Consequently, as degrees of freedom increase, the distribution of the χ^2_{ν} variate should tend toward normality, because of the "central limit" effect.
- Here is a picture of chi-square variates with 2,10,50, and 100 degrees of freedom.

Some Basic Properties Basic Chi-Square Distribution Calculations in R **Convergence to Normality** The Chi-Square Distribution and Statistical Testing

Convergence to Normality

- > par(mfrow=c(2,2))
- > curve(dchisq(x,2),0,qchisq(.995,2))
- > curve(dchisq(x,10),0,qchisq(.995,10))
- > curve(dchisq(x,50),0,qchisq(.995,50))
- > curve(dchisq(x,100),0,qchisq(.995,100))



・ロト ・ 一 ト ・ ヨ ト

→ Ξ →

Some Basic Properties Basic Chi-Square Distribution Calculations in R Convergence to Normality The Chi-Square Distribution and Statistical Testing

The Chi-Square Distribution and Statistical Testing

- We've sketched the basic properties of the χ^2 distribution, but how do we employ this distribution in statistical testing?
- A key result in statistical theory connects the χ^2 with the distribution of the sample variance s^2 .
- Suppose you have N independent, identically distributed (iid) observations from a normal distribution with mean μ and variance σ².
- Then

$$\frac{(N-1)s^2}{\sigma^2} \sim \chi^2_{N-1}$$
(5)

• This, in turn, implies that the sample variance has a distribution that has the same shape as a chi-square distribution with N-1 degrees of freedom, since we can rearrange the preceding equation as

$$s^{2} \sim \frac{\sigma^{2}}{N-1} \chi^{2}_{N-1} \tag{6}$$

Some Basic Properties Basic Chi-Square Distribution Calculations in R Convergence to Normality The Chi-Square Distribution and Statistical Testing

イロト イヨト イヨト

Test on a Single Variance

- The results on the preceding slide pave the way for a simple test of the hypothesis that $\sigma^2 = a$.
- If $\sigma^2 = a$, then

$$\frac{(N-1)s^2}{a} \sim \chi_{N-1}^2 \tag{7}$$

• So we have a simple method for testing whether $\sigma^2 = a$: Simply compare $(N-1)s^2/a$ with the upper and lower percentage points of the χ^2_{N-1} distribution.

Some Basic Properties Basic Chi-Square Distribution Calculations in R Convergence to Normality The Chi-Square Distribution and Statistical Testing

・ロト ・四ト ・ヨト

-

Test on a Single Variance An Example

Example (Test on a Single Variance)

Suppose you wish to test whether $\sigma^2 = 225$, and you observe N = 146 observations from the population which is assumed to be normally distributed. You observe a sample variance of 308.56. Perform the chi-square test with $\alpha = .05$.

Answer. The test statistic is

 $\frac{(146 - 1)308.56}{225} = 198.8498$

The area above 198.8498 in a χ^2_{145} distribution is

```
> 1-pchisq(198.9498,145)
```

[1] 0.001980618

Two get the two-sided $p\mbox{-value},$ we double this, obtaining a $p\mbox{-value}$ of

```
> 2*(1-pchisq(198.9498,145))
```

```
[1] 0.003961237
```

Since this is less than .05, the null hypothesis is rejected. We can also confirm that 198.9498 exceeds the critical value. Since the test is two-sided, the critical value is

```
> qchisq(.975,145)
[1] 180.2291
```

Some Basic Properties Basic Chi-Square Distribution Calculations in R Convergence to Normality The Chi-Square Distribution and Statistical Testing

・ロト ・聞 ト ・ ヨト ・ ヨトー

臣

Confidence Interval on a Single Variance

• The distributional result for a single variance implies that

$$\Pr\left(\chi^2_{N-1,\alpha/2} \le \frac{(N-1)s^2}{\sigma^2} \le \chi^2_{N-1,1-\alpha/2}\right) = 1 - \alpha \quad (8)$$

• If we take the reciprocal of all 3 sections of the inequality and reverse the direction of the signs, then multiply all 3 sections by $(N-1)s^2$, we obtain

$$\Pr\left(\frac{(N-1)s^2}{\chi^2_{N-1,1-\alpha/2}} \le \sigma^2 \le \frac{(N-1)s^2}{\chi^2_{N-1,\alpha/2}}\right) = 1 - \alpha \qquad (9)$$

• The outer two sections of this latter inequality are therefore the endpoints of a $1 - \alpha$ confidence interval for σ^2 .

Some Basic Properties Basic Chi-Square Distribution Calculations in R Convergence to Normality The Chi-Square Distribution and Statistical Testing

Confidence Interval on a Single Variance An Example

Example (Confidence Interval on a Single Variance)

Suppose you observe N = 146 observations from the population which is assumed to be normally distributed. You observe a sample variance of 308.56. What is a 95% confidence interval for σ^2

Answer. In R, we compute the lower and upper limits as

```
> s.squared <- 308.56
> N <- 146
> lower <- (N-1)* s.squared / qchisq(.975,N-1)
> upper <- (N-1)* s.squared / qchisq(.025,N-1)
> lower
[1] 248.2462
> upper
[1] 394.0022
```

The 95% confidence limits are thus 248.2462 and 394.0022.

Characterization of the F Distribution The F-Ratio Test

Characterization of the F Distribution

- The ratio of two independent chi-square variables $\chi^2_{\nu_1}$ and $\chi^2_{\nu_2}$, each divided by their respective degrees of freedom, is said to have an F distribution with ν_1 and ν_2 degrees of freedom.
- We will abbreviate this as F_{ν1,ν2}, and refer to ν1 and ν2 as the "numerator and denominator degrees of freedom," respectively.
- The expected value and variance of an F_{ν1,ν2} variate are, respectively,

$$E(F) = \frac{\nu_2}{\nu_2 - 2}$$
(10)

(assuming $\nu_2 > 2$ and

$$Var(F) = \frac{2\nu_2^2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 2)^2(\nu_2 - 4)}$$
(11)

assuming $\nu_2 > 4$.

 Since the F statistic is a ratio, and the positions of the items in numerator and denominator are, in a sense, arbitrary, it follows that the upper and lower quantiles of the F distribution satisfy the relationship

$$F_{\alpha_2,\nu_1,\nu_2} = \frac{1}{F_{1-\alpha_2,\nu_2,\nu_1}}$$
(12)

- That is, by reversing the degrees of freedom and taking the reciprocal, you may calculate lower percentage points of the F (which are generally not find in tables) from the upper percentage points (which are tabled).
- Modern software has rendered this point somewhat moot.

(日)

Characterization of the ${\cal F}$ Distribution The ${\it F}\mbox{-Ratio Test}$

The F-Ratio Test

- Suppose that you wish to compare the variances of two populations on the basis of two independent samples of sizes N₁ and N₂ taken at random from these populations.
- Under the assumption of a normal distribution, we can see that the ratio of the two sample variances will have an *F* distribution multiplied by the ratio of the two population variances.
- So, if the two population variances are equal, the ratio of the two sample variances will have an F distribution.

$$\frac{s_1^2}{s_2^2} \sim \frac{\frac{\sigma_1^2 \chi_{N_1-1}^2}{N_2-1}}{\frac{\sigma_2^2 \chi_{N_2-1}^2}{N_2-1}} \\ \sim \frac{\sigma_1^2 \chi_{N_2-1}^2/(N_1-1)}{\sigma_2^2 \chi_{N_2-1}^2/(N_2-1)} \\ \sim \frac{\sigma_1^2}{\sigma_2^2} F_{N_1-1,N_2-1}$$
(13)

Characterization of the F Distribution The F-Ratio Test

The F-Ratio Test

- The preceding result gives rise to an extremely simple test for comparing two variances.
- The null hypothesis is $H_0: \sigma_1^2 = \sigma_2^2$, and so the test as traditionally performed is two-sided.
- With modern software like R, one may simply compute the ratio s_1^2/s_2^2 , and compare the result to the $\alpha/2$ and $1 \alpha/2$ quantiles of the F distribution with $N_1 1$ and $N_2 1$ degrees of freedom.
- Alternatively, one may, after observing the data, simply put the larger of the two variances in the numerator (being careful to carry the numerator and denominator degrees of freedom into the proper positions).

イロト 不得下 イヨト イヨト 二日

Characterization of the ${\cal F}$ Distribution The ${\cal F}\mbox{-Ratio Test}$

The F-Ratio Test An Example

Example (An Example of the F-Ratio Test)

Suppose you observe sample variances of $s_1^2 = 134.69$ and $s_2^2 = 185.61$ in independent samples of sizes $N_1 = 101$ and $N_2 = 95$, respectively. Assuming that samples are independent and that population distributions are normal, test the null hypothesis that $\sigma_1^2 = \sigma_2^2$.

Answer. In this case, we have two options. We can compute

$$\frac{s_1^2}{s_2^2} = \frac{134.69}{185.61} \\ = 0.73$$

Since the value is less than 1, and assuming $\alpha = .05$, we compare the test statistic to the lower quantile of the F distribution, i.e., $F_{.025,100,94}$, which is 0.67. Since the the observed F statistic is not less than the critical value, we cannot reject "on the low side."

Somewhat surprisingly, although there is a substantial difference in the two variances, and the sample sizes are both close to 100, we cannot reject the null hypothesis.

・ロト ・聞ト ・ヨト ・ヨト

Characterization of the F Distribution The F-Ratio Test

The F-Ratio Test

• Here is the R code for performing the test.

```
> var.1 <- 134.69
```

- > var.2 <- 185.61
- > F <- var.1/var.2
- > F
- [1] 0.7256613
- > N.1 <- 101
- > N.2 <- 95
- > F.crit <- qf(.025,N.1-1,N.2-1)</pre>
- > F.crit

[1] 0.6707914

イロト イポト イヨト イヨト

Introduction

Calculations with the Noncentral Chi-Square Distribut The Effect of Noncentrality

The Noncentral χ^2 Distribution

- The χ^2 distribution we have studied so far is a special case of the *noncentral chi-square* distribution.
- This distribution has two parameters, the degrees of freedom and the noncentrality parameter.
- We symbolize a noncentral chi-square variate with the notation $\chi^2_{\nu,\lambda}$.

Introduction

Calculations with the Noncentral Chi-Square Distribut The Effect of Noncentrality

The Noncentral χ^2 Distribution

• The noncentral χ^2 distribution has a characterization that is similar to that of the central χ^2 distribution, i.e., if X_i are N independent observations taken from normal distributions with means μ_i and standard deviations σ_i . Then

$$\sum_{i=1}^{N} \left(\frac{X_i}{\sigma_i}\right)^2 \tag{14}$$

has a $\chi^2_{N,\lambda}$ distribution with

$$\lambda = \sum_{i=1}^{N} \left(\frac{\mu_i}{\sigma_i}\right)^2 \tag{15}$$

• When $\lambda = 0$, the noncentral chi-square distribution is equal to the standard ("central") chi-square distribution.

Introduction

Calculations with the Noncentral Chi-Square Distribut The Effect of Noncentrality

・ロト ・聞 ト ・ ヨト ・ ヨトー

æ

The Noncentral χ^2 Distribution

• The $\chi^2_{\nu,\lambda}$ distribution has mean and variance given by

$$E(\chi^2_{\nu,\lambda}) = \nu + \lambda \tag{16}$$

and

$$\operatorname{Var}(\chi^2_{\nu,\lambda}) = 2\nu + 4\lambda \tag{17}$$

Introduction Calculations with the Noncentral Chi-Square Distribut The Effect of Noncentrality

Calculations

- With R, the same functions that are used for the central χ^2 distribution are used for the noncentral counterpart, simply by adding the noncentrality parameter as additional input.
- For example, what is the 95th percentile of the noncentral χ^2 distribution with $\nu = 20$ and $\lambda = 5$?
 - > qchisq(.95,20,5)
 - [1] 38.92935

Introduction Calculations with the Noncentral Chi-Square Distribut **The Effect of Noncentrality**

・ロト ・ 同 ト ・ ヨ ト ・ ヨ ト

The Effect of Noncentrality

- The effect of the noncentrality parameter λ is to change the shape of the χ^2 variate's distribution and to move it to the right.
- In the plot below, we show four χ^2 variates, each with 15 degrees of freedom, and noncentrality parameters of 0,5,10,20.

Introduction Calculations with the Noncentral Chi-Square Distribut **The Effect of Noncentrality**

・ロト ・ 同 ト ・ ヨ ト ・ ヨ ト

The Effect of Noncentrality

- > curve(dchisq(x,15,0),0,qchisq(.999,15,20))
- > curve(dchisq(x,15,5),add=T,col=1)
- > curve(dchisq(x,15,10),add=T,col=2)
- > curve(dchisq(x,15,20),add=T,col=3)



Introduction Asymptotic Behavior Calculations in the Noncentral F

The Noncentral F distribution

- The ratio of a noncentral $\chi^2_{\nu_1,\lambda}$ to a central $\chi^2_{\nu_2}$ variate, each divided by their degrees of freedom, is said to have a *noncentral F distribution* with degrees of freedom ν_1, ν_2 , and noncentrality parameter λ .
- The noncentral F distribution is thus a three-parameter distribution family that includes the central F distribution as a special case when $\lambda = 0$. Thus,

$$\frac{\chi^2_{\nu_1,\lambda}/\nu_1}{\chi^2_{\nu_2}/\nu_2} \sim F_{\nu_1,\nu_2,\lambda} \tag{18}$$

Introduction Asymptotic Behavior Calculations in the Noncentral F

Mean and Variance of the Noncentral F

• The mean and variance of the noncentral F variate are given by

$$\mathbf{E}\left(F_{\nu_{1},\nu_{2},\lambda}\right) = \begin{cases} \frac{\nu_{2}(\nu_{1}+\lambda)}{\nu_{1}(\nu_{2}-2)} & \nu_{2} > 2\\ \text{Does not exist} & \nu_{2} \le 2 \end{cases}$$
(19)

and

$$\operatorname{Var}\left(F_{\nu_{1},\nu_{2},\lambda}\right) = \begin{cases} 2\frac{(\nu_{1}+\lambda)^{2}+(\nu_{1}+2\lambda)(\nu_{2}-2)}{(\nu_{2}-2)^{2}(\nu_{2}-4)} \left(\frac{\nu_{2}}{\nu_{1}}\right)^{2} & \nu_{2} > 4\\ \text{Does not exist} & \nu_{2} \leq 4\\ \end{cases}$$
(20)

イロト イポト イヨト イヨト

æ

Introduction Asymptotic Behavior Calculations in the Noncentral F

Asymptotic Behavior

- As $\nu_2 \to \infty$, the ratio χ^2_{ν}/ν converges to the constant 1. This is easy to see. Obviously, dividing a chi-square variate by its degrees of freedom parameter does not affect the shape of the distribution, since it multiplies it by a constant. We have already seen that, as $\nu \to \infty$, the shape of the χ^2_{ν} distribution converges to a normal distribution with mean ν and variance 2ν .
- Consequently, the ratio χ^2_{ν}/ν , which has a mean of 1 and a variance of $2/\nu$, has a limiting distribution that is N(1,0). That is, the distribution converges to the constant 1.
- An immediate consequence of the above is that, as $\nu_2 \to \infty$, the distribution of the variate $\nu_1 F_{\nu_1,\nu_2,\lambda}$ converges to a noncentral $\chi^2_{\nu_1,\lambda}$.

・ロト ・ 四ト ・ ヨト ・ ヨト ・ ヨ

Introduction Asymptotic Behavior Calculations in the Noncentral F

Calculations in the Noncentral F

- With R, the same functions that are used for the central F distribution are used for the noncentral counterpart, simply by adding the noncentrality parameter λ as additional input.
- For example, what is the 95th percentile of the noncentral F distribution with $\nu_1 = 20$, $\nu_2 = 10$ and $\lambda = 5$?
 - > qf(.95,20,10,5)

[1] 3.456638

イロト 不得下 イヨト イヨト