Chi-Square & $F$ Distributions

Carolyn J. Anderson

EdPsych 580

Fall 2005
Chi-Square & $\mathcal{F}$ Distributions

... and Inferences about Variances

- The Chi-square Distribution
  - Definition, properties, tables of, density calculator
  - Testing hypotheses about the variance of a single population
    (i.e., $H_0 : \sigma^2 = K$). Example.

- The $\mathcal{F}$ Distribution
  - Definition, important properties, tables of
  - Testing the equality of variances of two independent populations
    (i.e., $H_0 : \sigma_1^2 = \sigma_2^2$). Example.
Chi-Square & $F$ Distributions

...and Inferences about Variances

- Comments regarding testing the homogeneity of variance assumption of the two independent groups t–test (and ANOVA).
- Relationship among the Normal, $t$, $\chi^2$, and $F$ distributions.
Chi-Square & $F$ Distributions

- **Motivation.** The normal and $t$ distributions are useful for tests of population means, but often we may want to make inferences about population variances.

- **Examples:**
  - Does the variance equal a particular value?
  - Does the variance in one population equal the variance in another population?
  - Are individual differences greater in one population than another population?
  - Are the variances in $J$ populations all the same?
  - Is the assumption of homogeneous variances reasonable when doing a $t$-test (or ANOVA) of two...
Chi-Square & $\mathcal{F}$ Distributions

- To make statistical inferences about populations variance(s), we need
  - $\chi^2 \rightarrow$ The Chi-square distribution (Greek “chi”).
  - $\mathcal{F} \rightarrow$ Named after Sir Ronald Fisher who developed the main applications of $\mathcal{F}$.

- The $\chi^2$ and $\mathcal{F}$–distributions are used for many problems in addition to the ones listed above.

- They provide good approximations to a large class of sampling distributions that are not easily determined.
The Big Five Theoretical Distributions

- The Big Five are Normal, Student’s $t$, $\chi^2$, $F$, and the Binomial $(\pi, n)$.

- Plan:
  - Introduce $\chi^2$ and then the $F$ distributions.
  - Illustrate their uses for testing variances.
  - Summarize and describe the relationship among the Normal, Student’s $t$, $\chi^2$ and $F$. 
The Chi-Square Distributions

- Suppose we have a population with scores $Y$ that are normally distributed with mean $E(Y) = \mu$ and variance $\text{var}(Y) = \sigma^2$ (i.e., $Y \sim \mathcal{N}(\mu, \sigma^2)$).

- If we repeatedly take samples of size $n = 1$ and for each “sample” compute

  \[ z^2 = \frac{(Y - \mu)^2}{\sigma^2} \]

  squared standard score

- Define $\chi^2_1 = z^2$

- What would the sampling distribution of $\chi^2_1$ look like?
The Chi-Square Distribution, $\nu = 1$

Standard Normal Distribution

Chi-Square Distribution, $\nu = 1$

Value of $Z$

Value of $X^2$
The Chi-Square Distribution, $\nu = 1$

- $\chi_1^2$ are non-negative Real numbers
- Since 68% of values from $\mathcal{N}(0, 1)$ fall between $-1$ to 1, 68% of values from $\chi_1^2$ distribution must be between 0 and 1.
- The chi-square distribution with $\nu = 1$ is very skewed.
The Chi-Square Distribution, $\nu = 2$

- Repeatedly draw independent (random) samples of $n = 2$ from $\mathcal{N}(\mu, \sigma^2)$.
- Compute $Z_1^2 = (X_1 - \mu)^2 / \sigma^2$ and $Z_2^2 = (X_2 - \mu)^2 / \sigma^2$.
- Compute the sum: $\chi^2_2 = Z_1^2 + Z_2^2$. 

![Chi-Square Distribution Graph](image)
The Chi-Square Distribution, $\nu = 2$

- All value non-negative
- A little less skewed than $\chi^2_1$.
- The probability that $\chi^2_2$ falls in the range of 0 to 1 is smaller relative to that for $\chi^2_1$... 

\[
P(\chi^2_1 \leq 1) = .68
\]
\[
P(\chi^2_2 \leq 1) = .39
\]

- Note that mean $\approx \nu = 2$....
Chi-Square Distributions

- **Generalize:** For $n$ independent observations from a $\mathcal{N}(\mu, \sigma^2)$, the sum of the squared standard scores has a Chi-square distribution with $n$ degrees of freedom.

- Chi–squared distribution only depends on degrees of freedom, which in turn depends on sample size $n$.

- The standard scores are computed using population $\mu$ and $\sigma^2$; however, we usually don’t know what $\mu$ and $\sigma^2$ equal. When $\mu$ and $\sigma^2$ are estimated from the sampled data, the degrees of freedom are less than $n$. 
Chi-Square Dist: Varying $\nu$

Chi-Square Distributions

Value of $X^2$ vs. Density
Properties of Family of $\chi^2$ Distributions

• They are all positively skewed.

• As $\nu$ gets larger, the degree of skew decreases.

• As $\nu$ gets very large, $\chi^2_\nu$ approaches the normal distribution.

Why?
Properties of Family of $\chi^2$ Distributions

- $E(\chi^2_\nu) = \text{mean} = \nu = \text{degrees of freedom}$.
- $E[(\chi^2_\nu - E(\chi^2_\nu))^2] = \text{var}(\chi^2_\nu) = 2\nu$.
- Mode of $\chi^2_\nu$ is at value $\nu - 2$ (for $\nu \geq 2$).
- Median is approximately $\frac{(3\nu - 2)}{3}$ (for $\nu \geq 2$).
Properties of Family of $\chi^2$ Distributions

IF

• A random variable $\chi^2_{\nu_1}$ has a chi-squared distribution with $\nu_1$ degrees of freedom, and

• A second independent random variable $\chi^2_{\nu_2}$ has a chi-squared distribution with $\nu_2$ degrees of freedom,

THEN

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

their sum has a chi-squared distribution with $(\nu_1 + \nu_2)$ degrees of freedom.
Percentiles of $\chi^2$ Distributions

Note: $0.95\chi_1^2 = 3.84 = 1.96^2 = z_{.95}^2$

- Tables
- http://calculator.stat.ucla.edu/cdf/
- pvalue.f program or the executable version, pvalue.exe, on the course web-site.
- SAS: PROBCHI(x,df<,nc>) where
  - $x$ = number
  - $df$ = degrees of freedom
  - If $p=PROBCHI(x, df)$, then
    $p = Prob(\chi_{df}^2 \leq x)$
SAS Examples & Computations

p-values:

DATA probval;
   pz=PROBNORM(1.96);
   pzsq=PROBCHI(3.84,1);
output;
RUN;

Output:

   pz      pzsq
  0.97500  0.95000

What are these values?
SAS Examples & Computations

...To get density values...

Probability Density;

data chisq3;
   do x=0 to 10 by .005;
      pdfxsq=pdf('CHISQUARE',x,3);
      output;
   end;
run;
Inferences about a Population Variance

or the sampling distribution of the sample variance from a normal population.

• **Statistical Hypotheses:**

\[ H_0 : \sigma^2 = \sigma_o^2 \quad \text{versus} \quad H_a : \sigma^2 \neq \sigma_o^2 \]

• **Assumptions:** Observations are independently drawn (random) from a normal population; i.e.,

\[ Y_i \sim \mathcal{N}(\mu, \sigma^2) \quad \text{i.i.d} \]
Inferences about $\sigma^2$ (continued)

Test Statistic:

- We know
  \[
  \sum_{i=1}^{n} \frac{(Y_i - \mu)^2}{\sigma^2} = \sum_{i=1}^{n} z_i^2 \sim \chi_n^2
  \]
  if $z \sim \mathcal{N}(0, 1)$.

- We don’t know $\mu$, so we use $\bar{Y}$ as an estimate of $\mu$
  \[
  \sum_{i=1}^{n} \frac{(Y_i - \bar{Y})^2}{\sigma^2} \sim \chi_{n-1}^2
  \]
  or
  \[
  \sum_{i=1}^{n} \frac{(Y_i - \bar{Y})^2}{\sigma^2} = \frac{(n - 1)s^2}{\sigma^2} \sim \chi_{n-1}^2
  \]
Test Statistic for $H_0 : \sigma^2 = \sigma_o^2$

- So
  \[ s^2 \sim \frac{\sigma^2}{(n-1)} \chi_{n-1}^2 \]

- This gives us our test statistic:
  \[ \chi^2 = \frac{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}{\sigma_o^2} \]

  where $H_0 : \sigma^2 = \sigma_o^2$.

- Sampling distribution of Test Statistic: If $H_o$ is true, which means that $\sigma^2 = \sigma_o^2$, then
  \[ \chi^2 = \frac{(n-1)s^2}{\sigma_o^2} = \sum_{i=1}^{n} \frac{(Y_i - \bar{Y})^2}{\sigma_o^2} \sim \chi_{n-1}^2 \]
Decision and Conclusion, \( H_0 : \sigma^2 = \sigma_0^2 \)

- **Decision**: Compare the obtained test statistic to the chi-squared distribution with \( \nu = n - 1 \) degrees of freedom.

  or find the \( p \)-value of the test statistic and compare to \( \alpha \).

- **Interpretation/Conclusion**: What does the decision mean in terms of what you’re investigating?
Example of $H_o : \sigma^2 = \sigma_o^2$

• High School and Beyond: Is the variance of math scores of students from private schools equal to 100?

• Statistical Hypotheses:

$H_o : \sigma^2 = 100$ versus $H_a : \sigma^2 \neq 100$

• Assumptions: Math scores are independent and normally distributed in the population of high school seniors who attend private schools and the observations are independent.
Example of $H_o: \sigma^2 = \sigma^2_o$ (continued)

- **Test Statistic:** $n = 94$, $s^2 = 67.16$, and set $\alpha = .10$.

  $$\chi^2 = \frac{(n - 1)s^2}{\sigma^2} = \frac{(94 - 1)(67.16)}{100} = 62.46$$

  with $\nu = (94 - 1) = 93$.

- **Sampling Distribution of the Test Statistic:** Chi-square with $\nu = 93$.

  Critical values: $$.05\chi^2_{93} = 71.76 \& .95\chi^2_{93} = 116.51.$$
Example of $H_0 : \sigma^2 = \sigma_o^2$ (continued)

- Critical values: $0.05 \chi^2_{93} = 71.76$ & $0.95 \chi^2_{93} = 116.51$.

- Decision: Since the obtained test statistic $\chi^2 = 71.76$ is less than $0.05 \chi^2_{93} = 116.51$, reject $H_o$ at $\alpha = 0.10$. 
Confidence Interval Estimate of $\sigma^2$

- Start with

$$\text{Prob} \left( \frac{(n-1)s^2}{\sigma^2} \leq \frac{(1-\alpha/2)\chi^2_{\nu}}{\alpha/2} \right) = 1 - \alpha$$

- After a little algebra...

$$\text{Prob} \left[ \left( \frac{1}{(1-\alpha/2)\chi^2_{\nu}} \right) \leq \frac{\sigma^2}{(n-1)s^2} \leq \left( \frac{1}{(\alpha/2)\chi^2_{\nu}} \right) \right] = 1 - \alpha$$

- and a little more

$$\text{Prob} \left[ \left( \frac{(n-1)s^2}{(1-\alpha/2)\chi^2_{\nu}} \right) \leq \sigma^2 \leq \left( \frac{(n-1)s^2}{(\alpha/2)\chi^2_{\nu}} \right) \right] = 1 - \alpha$$
90% Confidence Interval Estimate of $\sigma^2$

- $(1 - \alpha)\%$ Confidence interval,

$$\frac{(n - 1)s^2}{(1 - \alpha/2)\chi^2_{\nu}} \leq \sigma \leq \frac{(n - 1)s^2}{\alpha/2\chi^2_{93}}$$

- So,

$$\frac{(94 - 1)(67.16)}{116.51}, \quad \frac{(94 - 1)(67.16)}{71.76} \rightarrow (53.61, 87.04),$$

which does not include 100 (the null hypothesized value).

- $s^2 = 67.16$ isn’t in the center of the interval.
The $\mathcal{F}$ Distribution

• Comparing two variances: Are they equal?
• Start with two independent populations, each normal and equal variances....

\begin{align*}
Y_1 & \sim \mathcal{N}(\mu_1, \sigma^2) \quad \text{i.i.d.} \\
Y_2 & \sim \mathcal{N}(\mu_2, \sigma^2) \quad \text{i.i.d.}
\end{align*}

• Draw two independent random samples from each population,

\begin{align*}
n_1 & \quad \text{from population} \quad 1 \\
n_2 & \quad \text{from population} \quad 2
\end{align*}
The $\mathcal{F}$ Distribution (continued)

• Using data from each of the two samples, estimate $\sigma^2$.

$$s_1^2 \quad \text{and} \quad s_2^2$$

• Both $S_1^2$ and $S_2^2$ are random variables, and their ratio is a random variable,

$$F = \frac{\text{estimate of } \sigma^2}{\text{estimate of } \sigma^2} = \frac{s_1^2}{s_2^2} = \frac{\chi^2_{(n_1-1)}/(n_1 - 1)}{\chi^2_{(n_2-1)}/(n_2 - 1)} = \frac{\chi_{\nu_1}^2/\nu_1}{\chi_{\nu_2}^2/\nu_2}$$

• Random variable $F$ has an $\mathcal{F}$ distribution.
Testing for Equal Variances

- $\mathcal{F}$ gives us a way to test $H_0 : \sigma_1^2 = \sigma_2^2 (= \sigma^2)$.

- Test statistic:

$$F = \left( \frac{s_1^2}{s_2^2} \right) = \frac{1}{n_1-1} \sum_{i=1}^{n_1} (Y_{i1} - \bar{Y}_1)^2 \left( \frac{1}{\sigma^2} \right)$$

$$= \frac{1}{n_2-1} \sum_{i=1}^{n_2} (Y_{i2} - \bar{Y}_2)^2 \left( \frac{1}{\sigma^2} \right)$$

$$= \frac{\chi^2_{\nu_1} / \nu_1}{\chi^2_{\nu_2} / \nu_2}$$

- A random variable formed from the ratio of two independent chi-squared variables, each divided by its degrees of freedom, is an “$F$–ratio” and has an $\mathcal{F}$ distribution.
Conditions for an $F$ Distribution

• **IF**
  • Both parent populations are normal.
  • Both parent populations have the same variance.
  • The samples (and populations) are independent.

• **THEN** the theoretical distribution of $F$ is $F_{\nu_1, \nu_2}$ where
  • $\nu_1 = n_1 - 1 = \text{numerator degrees of freedom}$
  • $\nu_2 = n_2 - 1 = \text{denominator degrees of freedom}$
Eg of $\mathcal{F}$ Distributions: $\mathcal{F}_{2, \nu_2}$

F Distributions: $\nu_1 = 3$, $\nu_2 =$

- 10.5
- 10.10
- 10.20
- 10.100

![Graph of F Distributions](image-url)
Eg of $\mathcal{F}$ Distributions: $\mathcal{F}_{5,\nu_2}$
Eg of $\mathcal{F}$ Distributions: $\mathcal{F}_{50,\nu_2}$
Important Properties of $F$ Distributions

• The range of $F$–values is non-negative real numbers (i.e., $0$ to $+\infty$).

• They depend on 2 parameters: numerator degrees of freedom ($\nu_1$) and denominator degrees of freedom ($\nu_2$).

• The expected value (i.e, the mean) of a random variable with an $F$ distribution with $\nu_2 > 2$ is

$$E(F_{\nu_1, \nu_2}) = \mu_{F_{\nu_1, \nu_2}} = \frac{\nu_2}{(\nu_2 - 2)}.$$
Properties of \( F \) Distributions

- For any fixed \( \nu_1 \) and \( \nu_2 \), the \( F \) distribution is non-symmetric.
- The particular shape of the \( F \) distribution varies considerably with changes in \( \nu_1 \) and \( \nu_2 \).
- In most applications of the \( F \) distribution (at least in this class), \( \nu_1 < \nu_2 \), which means that \( F \) is positively skewed.
- When \( \nu_2 > 2 \), the \( F \) distribution is uni-modal.
Percentiles of the $\mathcal{F}$ Dist.

- http://calculators.stat.ucla.edu/cdf
- p-value program
- SAS probf

- Tables textbooks given the upper $25^{\text{th}}$, $10^{\text{th}}$, $5^{\text{th}}$, $2.5^{\text{th}}$, and $1^{\text{st}}$ percentiles. Usually, the
  - Columns correspond to $\nu_1$, numerator df.
  - Rows correspond to $\nu_2$, denominator df.
- Getting lower percentiles using tables requires taking reciprocals.
Selected $\mathcal{F}$ values from Table V

Note: all values are for upper $\alpha = .05$

<table>
<thead>
<tr>
<th>$\nu_1$</th>
<th>$\nu_2$</th>
<th>$F_{\nu_1,\nu_2}$</th>
<th>which is also...</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>161.00</td>
<td>$t_1^2$</td>
</tr>
<tr>
<td>1</td>
<td>20</td>
<td>4.35</td>
<td>$t_{20}^2$</td>
</tr>
<tr>
<td>1</td>
<td>1000</td>
<td>3.85</td>
<td>$t_{1000}^2$</td>
</tr>
<tr>
<td>1</td>
<td>$\infty$</td>
<td>3.84</td>
<td>$t_{\infty}^2 = z^2 = \chi_1^2$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\nu_1$</th>
<th>$\nu_2$</th>
<th>$F_{\nu_1,\nu_2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20</td>
<td>4.35</td>
</tr>
<tr>
<td>4</td>
<td>20</td>
<td>2.87</td>
</tr>
<tr>
<td>10</td>
<td>20</td>
<td>2.35</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
<td>2.12</td>
</tr>
<tr>
<td>1000</td>
<td>20</td>
<td>1.57</td>
</tr>
</tbody>
</table>
Test Equality of Two Variances

Are students from private high schools more homogeneous with respect to their math test scores than students from public high schools?

• Statistical Hypotheses:
  \[ H_0 : \sigma^2_{private} = \sigma^2_{public} \text{ or } \sigma^2_{public}/\sigma^2_{private} = 1 \]
  versus \[ H_a : \sigma^2_{private} < \sigma^2_{public} , (1\text{-tailed test}) \].

• Assumptions: Math scores of students from private schools and public schools are normally distributed and are independent both between and within in school type.
Test Equality of Two Variances

• Test Statistic:
\[ F = \frac{s_1^2}{s_2^2} = \frac{91.74}{67.16} = 1.366 \]

with \( \nu_1 = (n_1 - 1) = (506 - 1) = 505 \) and \( \nu_2 = (n_2 - 1) = (94 - 1) = 93 \).

Since the sample variance for public schools, \( s_1^2 = 91.74 \), is larger than the sample variance for private schools, \( s_2^2 = 67.16 \), put \( s_1^2 \) in the numerator.

• **Sampling Distribution** of Test Statistic is \( \mathcal{F} \) distribution with \( \nu_1 = 505 \) and \( \nu_2 = 93 \).
Test Equality of Two Variances

• Decision: Our observed test statistic, $F_{505,93} = 1.366$ has a $p$–value$= .032$. Since $p$–value $< \alpha = .05$, reject $H_o$.

• Or, we could compare the observed test statistic, $F_{505,93} = 1.366$, with the critical value of $F_{505,93}(\alpha = .05) = 1.320$. Since the observed value of the test statistic is larger than the critical value, reject $H_o$.

• Conclusion: The data support the conclusion that students from private schools are more homogeneous with respect to math test scores than students from public schools.
Example Continued

• Alternative question: “Are the individual differences of students in public high schools and private high schools the same with respect to their math test scores?”

• Statistical Hypotheses: The null is the same, but the alternative hypothesis would be

\[ H_a : \sigma^2_{public} \neq \sigma^2_{private} \] (a 2–tailed alternative)

• Given \( \alpha = .05 \), Retain the \( H_o \), because our obtained \( p \)–value (the probability of getting a test statistic as large or larger than what we got) is larger than \( \alpha/2 = .025 \).
Example Continued

- Given $\alpha = .05$, Retain the $H_0$, because our obtained $p$–value (the probability of getting a test statistic as large or larger than what we got) is larger than $\alpha/2 = .025$.

- Or the rejection region (critical value) would be any $F$–statistic greater than $F_{505,93}(\alpha = .025) = 1.393$.

- Point: This is a case where the choice between a 1 and 2 tailed test leads to different decisions regarding the null hypothesis.
Test for Homogeneity of Variances

\[ H_0 : \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_J^2 \]

- These include
  - Hartley’s \( F_{\text{max}} \) test
  - Bartlett’s test
  - One regarding variances of paired comparisons.

- You should know that they exist; we won’t go over them in this class. Such tests are not as important as they once (thought) they were.
Test for Homogeneity of Variances

• Old View: Testing the equality of variances should be a preliminary to doing independent \( t \)-tests (or ANOVA).

• Newer View:
  • Homogeneity of variance is required for small samples, which is when tests of homogeneous variances do not work well. With large samples, we don’t have to assume \( \sigma_1^2 = \sigma_2^2 \).
  • Test critically depends on population normality.
  • If \( n_1 = n_2 \), \( t \)-tests are robust.
Test for Homogeneity of Variances

- For small or moderate samples and there’s concern with possible heterogeneity → perform a Quasi-\(t\) test.
- In an experimental settings where you have control over the number of subjects and their assignment to groups/conditions/etc. → equal sample sizes.
- In non-experimental settings where you have similar numbers of participants per group, \(t\) test is pretty robust.
... and the central importance of the normal distribution.

- Normal, Student’s \( t_\nu \), \( \chi^2_\nu \), and \( F_{\nu_1, \nu_2} \) are all theoretical distributions.
- We don’t ever actually take vast (infinite) numbers of samples from populations.
- The distributions are derived based on mathematical logic statements of the form

\[
\text{IF} \quad \ldots \ldots \quad \text{Then} \quad \ldots \ldots
\]
Derivation of Distributions

• Example
  • IF we draw independent random samples of size (large) \( n \) from a population and compute the mean \( \bar{Y} \) and repeat this process many, many, many, many, many times,
  • THEN \( \bar{Y} \) is approximately normal.

• Assumptions are part of the “if” part, the conditions used to deduce sampling distribution of statistics.

• The \( t, \chi^2 \) and \( F \) distributions all depend on normal “parent” population.
Chi-Square Distribution

• \( \chi^2_\nu \) = sum of independent squared normal random variables with mean \( \mu = 0 \) and variance \( \sigma^2 = 1 \) (i.e., “standard normal” random variables).

\[
\chi^2_\nu = \sum_{i=1}^{n} z_i^2 \quad \text{where} \quad z_i \sim \mathcal{N}(0, 1)
\]

• Based on the Central Limit Theorem, the “limit” of the \( \chi^2_\nu \) distribution (i.e., \( n \to \infty \)) is normal.
The $\mathcal{F}$ Distribution

- $\mathcal{F}_{\nu_1,\nu_2} =$ ratio of two independent chi-squared random variables each divided by their respective degrees of freedom.

\[
\mathcal{F}_{\nu_1,\nu_2} = \frac{\chi^2_{\nu_1}/\nu_1}{\chi^2_{\nu_2}/\nu_2}
\]

- Since $\chi^2_{\nu}$’s depend on the normal distribution, the $\mathcal{F}$ distribution also depends on the normal distribution.

- The “limiting” distribution of $\mathcal{F}_{\nu_1,\nu_2}$ as $\nu_2 \to \infty$ is $\chi^2_{\nu_1}/\nu_1 \ldots \ldots$ because as $\nu_2 \to \infty$, $\chi^2_{\nu_2}/\nu_2 \to 1$. 
Students $t$ Distribution

Note that

$$ t_{\nu}^2 = \left( \frac{\bar{Y} - \mu}{s / \sqrt{n}} \right)^2 $$

$$ = \frac{(\bar{Y} - \mu)^2 n}{\sum_{i=1}^{n} (Y_i - \bar{Y})^2 / (n - 1)} $$

$$ = \frac{(\bar{Y} - \mu)^2 n}{\sum_{i=1}^{n} (Y_i - \bar{Y})^2 / (n - 1)} \left( \frac{1}{\sigma^2} \right) $$

$$ = \frac{(\bar{Y} - \mu)^2}{\sigma^2 / n} = \frac{z^2}{\chi^2 / \nu} $$
**Students $t$ Distribution** (continued)

- Student’s $t$ based on normal,

  $$t^2_\nu = \frac{z^2}{\chi^2_\nu / \nu} \quad \text{or} \quad t_\nu = \frac{z}{\sqrt{\chi^2_\nu / \nu}}$$

- A squared $t$ random variable equals the ratio of squared standard normal divided by chi-squared divided by its degrees of freedom. So...
Students $t$ Distribution (continued)

Since

$$t^2_{\nu} = \frac{z^2}{\chi^2_{\nu}/\nu} \quad \text{or} \quad t_{\nu} = \frac{z}{\sqrt{\chi^2_{\nu}/\nu}}$$

• As $\nu \to \infty$, $t_{\nu} \to \mathcal{N}(0, 1)$ because $\chi^2_{\nu}/\nu \to 1$.

• Since $z^2 = \chi^2_1$, $t^2 = \frac{z^2/1}{\chi^2_n/\nu} = \frac{\chi^2_1/1}{\chi^2_n/\nu} = \mathcal{F}_{1,\nu}$

• Why are the assumptions of normality, homogeneity of variance, and independence required for
  • $t$ test for mean(s)
  • Testing homogeneity of variance(s).
Summary of Relationships

Let \( z \sim \mathcal{N}(0, 1) \)

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Definition</th>
<th>Parent</th>
<th>Limiting</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \chi^2_{\nu} )</td>
<td>( \sum_{i=1}^{\nu} z_i^2 )</td>
<td>normal</td>
<td>As ( \nu \to \infty ), ( \chi^2_{\nu} \to ) normal</td>
</tr>
<tr>
<td>Independent ( z )'s</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( F_{\nu_1, \nu_2} )</td>
<td>( (\chi_{\nu_1}^2/\nu_1)/(\chi_{\nu_2}^2/\nu_2) )</td>
<td>chi-squared</td>
<td>As ( \nu_2 \to \infty ), ( F_{\nu_1, \nu_2} \to \chi^2_{\nu_1}/\nu_1 )</td>
</tr>
<tr>
<td>Independent ( \chi^2 )'s</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( t )</td>
<td>( z/\sqrt{\chi^2/\nu} )</td>
<td>normal</td>
<td>As ( \nu \to \infty ), ( t \to ) normal</td>
</tr>
</tbody>
</table>

Note: \( F_{1, \nu} = t^2_{\nu} \), also \( F_{1, \infty} = t^2_{\infty} = z^2 = \chi^2_1 \).