STA258H5
Statistics with Applied Probability

Al Nosedal and Omid Jazi

Winter 2023

SAMPLING DISTRIBUTIONS RELATED TO A NORMAL POPULATION

Theorem 7.1

Let $Y_1, Y_2, ..., Y_n$ be a random sample of size n from a Normal distribution with mean μ and variance σ^2. Then

Sample mean

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

Sample mean

is Normally distributed with mean $\mu_{\bar{Y}} = \mu$ and variance $\sigma^2_{\bar{Y}} = \frac{\sigma^2}{n}$.

(or $Z = \frac{\bar{Y}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$)

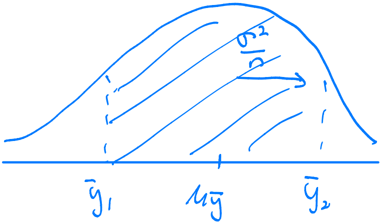
Sampling Distrib

Standard Normal

$\bar{y} \sim N\left(\mu_{\bar{y}} = \mu,\ \sigma_{\bar{y}}^2 = \frac{\sigma^2}{n}\right)$

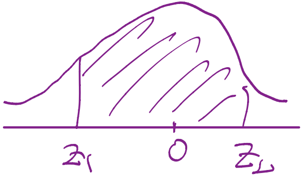
$Z \sim N(0, 1)$

$$z = \frac{\bar{y} - \mu_{\bar{y}}}{\sigma_{\bar{y}}}$$



$$\longrightarrow$$

$$\boxed{z = \frac{\bar{y} - \mu}{\sigma/\sqrt{n}}}$$



Example

NOT an individual value
but an average

σ

μ

Marks on a standardized test are Normally distributed with mean 75, standard deviation 15. What is the probability the class average, for a class of 30, is greater than 76?

$\bar{y}$

n

$\mu_y = 75$, $\sigma_y = 15$, $n = 30$

$$z = \boxed{\frac{\bar{y} - \mu_y}{\sigma_y / \sqrt{n}}}$$

$$P(\bar{y} > 76) = P\left(\boxed{\frac{\bar{y} - \mu_{\bar{y}}}{\sigma_y / \sqrt{n}}} > \frac{76 - \mu_{\bar{y}}}{\sigma_y / \sqrt{n}}\right)$$

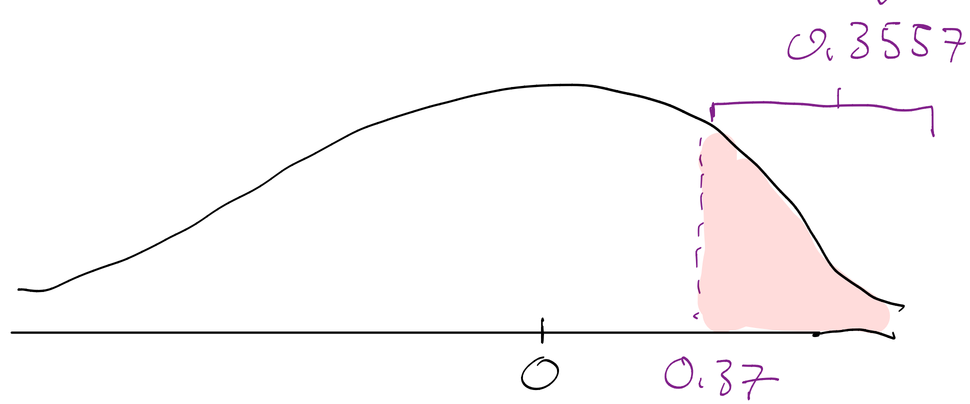
$$= P\left(z > \frac{76 - \mu_{\bar{y}}}{\sigma_y / \sqrt{n}}\right)$$

$$= P\left(z > \frac{76 - 75}{15 / \sqrt{30}}\right)$$

$$= P(z > \underbrace{0.37}_{z\text{-value}}) \longrightarrow \text{use table}$$

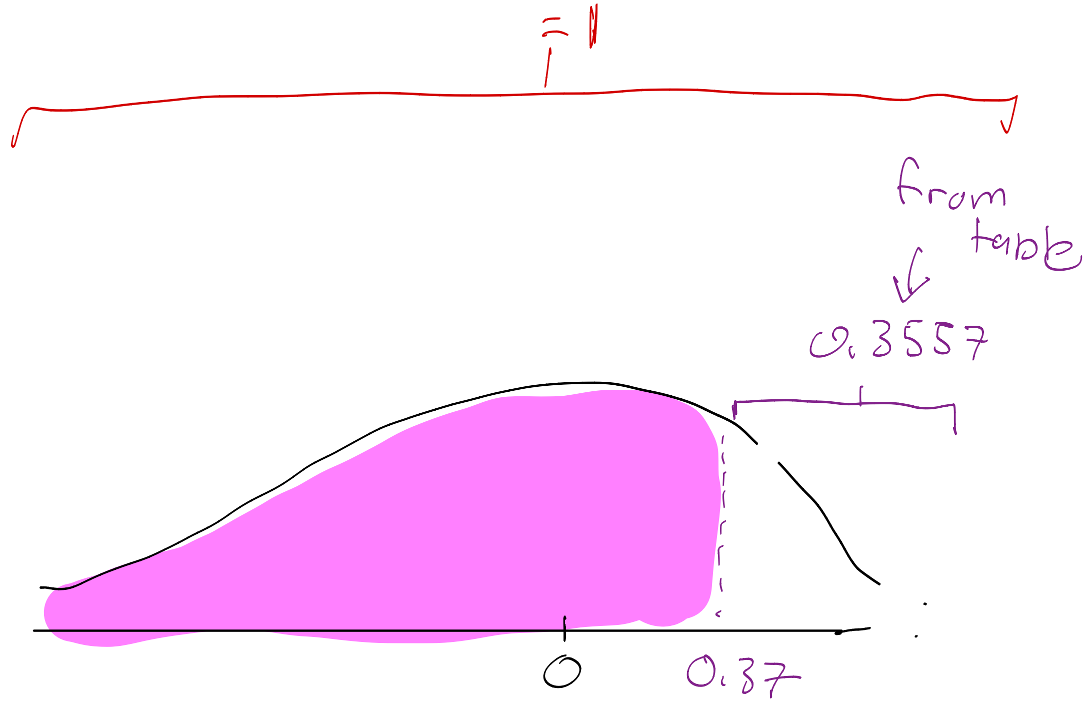
$$= 0.3557$$

from table



$P(z < 0.37)$

$= 1 - 0.3557$



Gamma Function (Not a distrib)

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t}\, dt \qquad x > 0$$

hard to integrate (software, numerical solver)

Properties

- $\Gamma(x) = x\Gamma(x-1)$
- $n \in \mathbb{Z}^+ : \quad \Gamma(n) = (n-1)!$
- $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

Gamma Distribution

$$f(x; \underset{\text{shape}}{\alpha}, \underset{\text{scale}}{\beta}) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}} \qquad x, \alpha, \beta > 0$$

χ^2

Chi-Square Distribution at n degrees of freedom

Gamma evaluated at $\alpha = \frac{n}{2}$, $\beta = 2$

Moment Generating Functions (MGF)

$$M_X(t) = E(e^{tx})$$

$e^x = \exp(x)$

$$= \begin{cases} \sum e^{tx} \cdot p(x) & \text{(discrete)} \\ \int_{-\infty}^{\infty} e^{tx} \cdot f(x)\, dx & \text{(cts)} \end{cases}$$

Can be used to characterize a family of distributions

Normal: $M_X(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)$

Gamma: $M_X(t) = (1 - \beta t)^{-\alpha}, \quad t < \frac{1}{\beta}$

Chi-Square (n df): $M_X(t) = (1 - 2t)^{-n/2}, \quad t < \frac{1}{2}$

Gaussian Integral (calculus)

$$\int_{-\infty}^{+\infty} e^{-x^2}\, dx = \sqrt{\pi}$$

$$\int_{-\infty}^{+\infty} e^{-kx^2} = \sqrt{\frac{\pi}{k}} \qquad k > 0$$

$$\int_{-\infty}^{+\infty} e^{kx^2} = \sqrt{\frac{\pi}{-k}} \qquad k < 0$$

χ_1^2 distribution

Chi-Square with 1 df

It can be shown that χ_1^2 distribution is the Gamma($\alpha = 1/2, \beta = 2$) distribution.

$$f(y) = \begin{cases} \frac{1}{\Gamma(1/2)2^{1/2}} y^{1/2-1} e^{-y/2} & y > 0 \\ 0 & y \leq 0 \end{cases}$$

$E(Y) = 1$ $\qquad$ $Var(Y) = 2.$

Example

$Z \sim N(0, 1^2)$

Let Z be a Normally distributed random variable with mean 0 and variance 1. Use the method of moment-generating functions to find the probability distribution of Z^2.

(a standard normal rv)$^2 \sim \chi^2_1$

$Z^2 \sim \chi^2_1$

proof:

$Z \sim N(\mu = 0, \sigma^2 = 1^2)$

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-z^2}{2}\right)$$

MGF of Z^2

$$M_{Z^2}(t) = E\left(e^{tZ^2}\right)$$

$$= \int_{z=-\infty}^{z=+\infty} e^{tz^2} f(z) \, dz$$

$$= \int_{z=-\infty}^{z=+\infty} \exp(tz^2) \cdot \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-z^2}{2}\right) dz$$

$e^a \cdot e^b = e^{a+b}$

$$= \frac{1}{\sqrt{2\pi}} \int_{z=-\infty}^{z=+\infty} \exp\left(tz^2 - \frac{z^2}{2}\right) dz$$

$\int_{-\infty}^{+\infty} e^{-x^2} = \sqrt{\pi}$

$$= \frac{1}{\sqrt{2\pi}} \int_{z=-\infty}^{z=+\infty} \exp\left(t - \frac{1}{2}\right) z^2 \, dz$$

$$= \frac{1}{\sqrt{2\pi}} \int_{z=-\infty}^{z=+\infty} \exp\left[-\left(\frac{1}{2} - t\right) z^2\right] dz$$

$u = kx$

$\frac{du}{dx} = k$

let $u = \left[\sqrt{\left(\frac{1}{2} - t\right)}\right] z$

$$du = \sqrt{\frac{1}{2} - t} \, dz \longrightarrow dz = \frac{1}{\sqrt{\frac{1}{2} - t}} du$$

$z = -\infty, \quad u = \sqrt{\left(\frac{1}{2} - t\right)} \cdot z = -\infty$

$z = +\infty \quad u = \quad = +\infty$

$$M_{Z^2}(t) = \frac{1}{\sqrt{2\pi}} \int_{u=-\infty}^{u=+\infty} \exp(-u^2) \frac{1}{\sqrt{\frac{1}{2}-t}} \, du$$

$$= \frac{1}{\sqrt{2\pi}} \cdot \frac{1}{\sqrt{\frac{1}{2}-t}} \underbrace{\int_{u=-\infty}^{u=+\infty} \exp(-u^2) \, du}_{\sqrt{\pi}}$$ form of Gaussian integral

$$= \frac{1}{\sqrt{2\pi\left(\frac{1}{2}-t\right)}} \cdot \sqrt{\pi}$$

$$= \frac{\cancel{\sqrt{\pi}}}{\cancel{\sqrt{\pi}}\sqrt{2\left(\frac{1}{2}-t\right)}} = \frac{1}{\sqrt{(1-2t)}} = \underbrace{(1-2t)^{-1/2}}_{\text{MGF of } \chi^2_1}$$

Exercise:

use $\int_{-\infty}^{+\infty} e^{-kx^2} = \sqrt{\frac{\pi}{k}}$ $\quad k>0 \quad$ instead of u-sub

Solution

$$M_{Z^2}(t) = E(e^{tZ^2}) = \int_{-\infty}^{\infty} e^{tz^2} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$$
$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z^2\left(\frac{1-2t}{2}\right)} dz$$

This integral can be evaluated using an "old trick" (we note that it looks like a Normally distributed random variable).

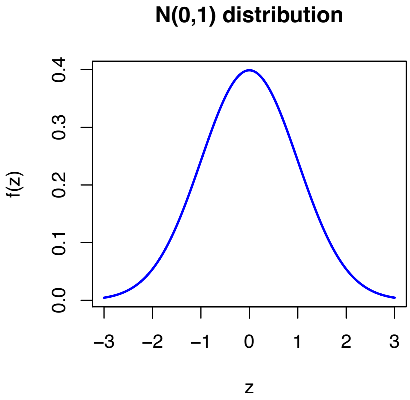
Solution

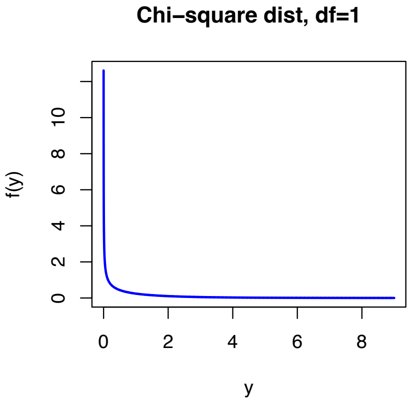
We realize that $e^{-z^2\left(\frac{(1-2t)}{2}\right)}$ is proportional to a Normal with $\mu = 0$ and $\sigma^2 = 1/(1-2t)$, then

$M_{Z^2}(t) = \frac{\sqrt{2\pi}\sqrt{1/(1-2t)}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sqrt{1/(1-2t)}} e^{-z^2\left(\frac{1-2t}{2}\right)} dz$

$M_{Z^2}(t) = \sqrt{\frac{1}{1-2t}} = (1-2t)^{-1/2}$ (Note. This is valid provided that $t < 1/2$).

$(1-2t)^{-1/2}$ is the moment-generating function for a gamma-distributed random variable with $\alpha = 1/2$ and $\beta = 2$. Hence, Z^2 has a χ^2 distribution with $\nu = 1$ degree of freedom.


N(0,1) distribution
f(z)
0.0
0.1
0.2
0.3
0.4
–3
–2
–1
0
1
2
3
z


Chi–square dist, df=1
f(y)
0
2
4
6
8
10
0
2
4
6
8
y

Example

above: $Z^2 \sim \chi^2_{(1)}$

Suppose that $\overset{Z_1}{\cancel{Y_1}}$ and $\overset{Z_2}{\cancel{Y_2}}$ are independent, standard Normal random variables. Find the probability distribution of $U = Y_1^2 + Y_2^2$.

$Z_1^2 + Z_2^2 \sim \chi^2_{(2)}$

Solution

Self study

$M_U(t) = E[e^{Ut}] = E[e^{(Y_1^2+Y_2^2)t}]$
$= E[e^{Y_1^2 t}]E[e^{Y_2^2 t}]$ (Y_1 and Y_2 are independent)
$= M_{Y_1^2}(t)M_{Y_2^2}(t)$
$= [(1-2t)^{-1/2}][(1-2t)^{-1/2}] = (1-2t)^{-2/2}$.
Because moment-generating functions are unique, U has a χ^2 distribution with 2 degrees of freedom.

χ_n^2 distribution

$Z^2 \sim \chi^2_{(1)}$

$Z_1^2 + Z_2^2 \sim \chi^2_{(2)}$

$Z_1^2 + Z_2^2 + \ldots + Z_n^2 \sim \chi^2_{(n)}$

Definition. Let $Z_1, Z_2, ..., Z_n$ be iid standard Normal random variables.
Define $V = \sum_{i=1}^{n} Z_i^2$. V has a χ_n^2 distribution.
The χ_n^2 is the Gamma($\alpha = n/2, \beta = 2$) distribution.

Proof.
By uniqueness of moment generating functions.

From STA 256

If X_1 and X_2 are indep RV

$E(X_1 X_2) = E(X_1) \cdot E(X_2)$

Proof on slide 12 (Sum of n st. normals squared $\sim \chi^2_{(n)}$)

$$Z_1, Z_2, \ldots, Z_n \overset{iid}{\sim} N(0, 1^2)$$

want distrib of $V = \sum_{i=1}^{n} Z_i^2 = Z_1^2 + Z_2^2 + \ldots + Z_n^2$

MGF of V

$$M_V(t) = E[e^{tV}] = E[\exp(tV)]$$

$$= E\left[\exp\left(t \sum_{i=1}^{n} Z_i^2\right)\right]$$

$$= E\left[\exp\left(t\left(Z_1^2 + Z_2^2 + \ldots + Z_n^2\right)\right)\right]$$

$$= E\left[\exp\left(tZ_1^2 + tZ_2^2 + \ldots + tZ_n^2\right)\right] \qquad e^{a+b} = e^a \cdot e^b$$

indep.

$$= E\left[\exp(tZ_1^2) \cdot \exp(tZ_2^2) \cdot \ldots \exp(tZ_n^2)\right]$$

$$= \underbrace{E\left[\exp(tZ_1^2)\right]}_{M_{Z_1^2}(t)} \cdot \underbrace{E\left[\exp(tZ_2^2)\right]}_{M_{Z_2^2}(t)} \cdot \ldots \cdot \underbrace{E\left[\exp(tZ_n^2)\right]}_{M_{Z_n^2}(t)}$$

$$= \underbrace{(1-2t)^{-1/2} \cdot (1-2t)^{-1/2} \cdot \ldots \cdot (1-2t)^{-1/2}}_{n \text{ times}}$$

$$= \underbrace{(1-2t)^{-n/2}}_{\text{Form of MGF of } \chi^2_{(n)}} \quad \Bigg| \quad V \sim \chi^2_{(n)}$$

χ_n^2 distribution

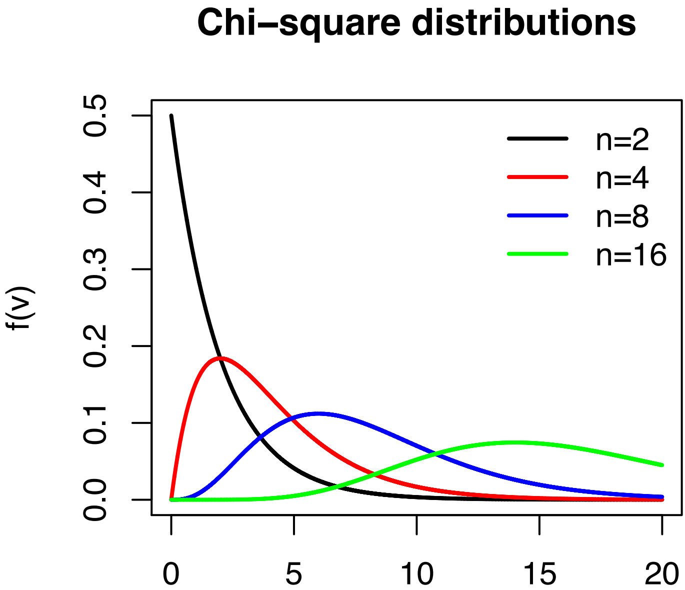
↳ Sum of n (st. normals)^2

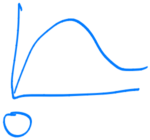
PDF

$$f(v) = \begin{cases} \frac{1}{\Gamma(n/2)2^{n/2}} v^{n/2-1} e^{-v/2} & v > 0 \\ 0 & v \leq 0 \end{cases}$$

$E(V) = n$ $\qquad$ $Var(V) = 2n.$

The subscript n is called the degrees of freedom; it is the number of "free" variables in the sum.


Chi–square distributions
f(v)
0.0
0.1
0.2
0.3
0.4
0.5
0
5
10
15
20
n=2
n=4
n=8
n=16


0

Theorem 7.2

Let $Y_1, Y_2, ..., Y_n$ be defined as in Theorem 7.1. Then $Z_i = \frac{Y_i - \mu}{\sigma}$ are independent, standard Normal random variables, $i = 1, 2, ..., n$, and

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

has a χ^2 distribution with n degrees of freedom (df).

It can be shown (see STA260) that

If $X_1, X_2, ..., X_n \sim$ iid $N(\mu, \sigma^2)$, then $V = \sum_{j=1}^{n} \left(\frac{X_j - \bar{X}}{\sigma}\right)^2 \sim X^2_{n-1}$

Theorem 7.3

$S^2 = \sum_{i=1}^{n} \frac{(y_i - \bar{y})^2}{n-1}$ sample var

Let $Y_1, Y_2, ..., Y_n$ be a random sample from a Normal distribution with mean μ and variance σ^2. Then

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

has a χ^2 distribution with $(n-1)$ df.
Also, $\bar{Y}$ and S^2 are independent random variables.

$$\boxed{\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}}$$

Since, if $Y_1, Y_2, ..., Y_n \sim$ iid $N(\mu, \sigma)$, then $V = \sum_{i=1}^{n} \left(\frac{Y_i - \bar{Y}}{\sigma}\right)^2 \sim \chi^2_{n-1}$
It follows that, if $Y_1, Y_2, ..., Y_n \sim$ iid $N(\mu, \sigma)$ and
$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \bar{Y})^2$, then $V = \sum_{i=1}^{n} \left(\frac{Y_i - \bar{Y}}{\sigma}\right)^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$.
$E(S^2) = \sigma^2$ and $Var(S^2) = \frac{2\sigma^4}{n-1}$.

Example

LD50

The Environmental Protection Agency is concerned with the problem of setting criteria for the amounts of certain toxic chemicals to be allowed in freshwater lakes and rivers. A common measure of toxicity for any pollutant is the concentration of the pollutant that will kill half of the test species in a given amount of time (usually 96 hours for fish species). This measure is called LC50 (lethal concentration killing 50% of test species). In many studies, the values contained in the natural logarithm of LC50 measurements are Normally distributed, and, hence, the analysis is based on ln(LC50) data.

Example

σ^2: pop var

Suppose that $n = 20$ observations are to be taken on ln(LC50) measurements and that $\sigma^2 = 1.4$. Let S^2 denote the sample variance of the 20 measurements.

a. Find a number b such that $P(S^2 \leq b) = 0.975$.

b. Find a number a such that $P(a \leq S^2) = 0.975$.

c. If a and b are as in parts a) and b), what is $P(a \leq S^2 \leq b)$?

Example (Slide 19-20)

Thm 7.3

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

$\underbrace{n=20, \quad \sigma^2 = 1.4}$. Find b. st

$$P(S^2 \le b) = 0.975$$

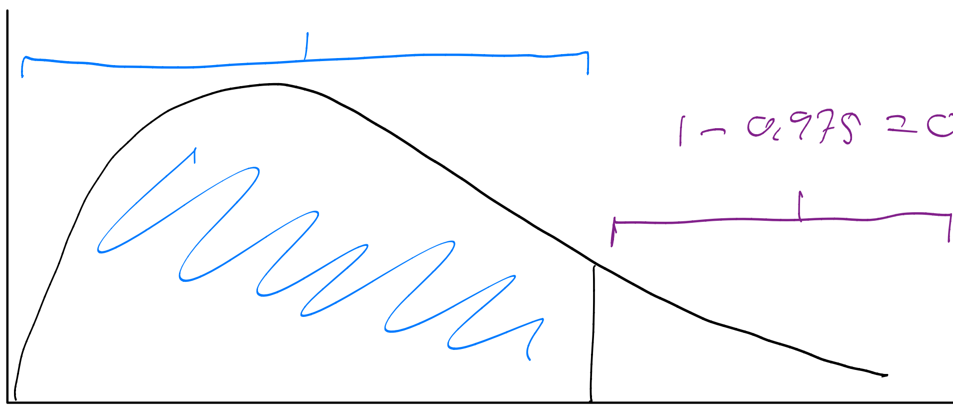
$$P\Big(\underbrace{\frac{(n-1)S^2}{\sigma^2}}_{\chi^2_{n-1} \,=\, 20-1 \,=\, 19} \le \frac{(n-1)b}{\sigma^2}\Big) = 0.975$$

$$P\Big(\chi^2_{19} \le \frac{(20-1)\,b}{1.4}\Big) = 0.975$$

$$P\Big(\chi^2_{19} \le \frac{95b}{7}\Big) = 0.975$$

χ^2 dist at 19 df

0.975



$1-0.975 = 0.025$ (use table)

$\frac{95b}{7} = 32.8523$ (table)

$$\frac{95b}{7} = 32.8523$$

$$b = 2.42$$

R

pchisq(value, df)

$\boxed{\text{Gives prob to the LEFT}}$

(b) Find (a) s.t. $P(a \le S^2) = 0.975$

Exercise. Ans $a = 0.656$

(c) $n=20, \quad \sigma^2 = 1.4 \qquad \underbrace{b = 2.42, \qquad a = 0.656}_{\text{part (a), (b)}}$

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

$$P(a \leq S^2 \leq b)$$

$$= P(0.656 \leq S^2 \leq 2.42)$$

$$= P\left(\frac{(n-1)\,0.656}{\sigma^2} \leq \underbrace{\frac{(n-1)S^2}{\sigma^2}}_{\chi^2_{n-1}} \leq \frac{(n-1)\,2.42}{\sigma^2}\right)$$

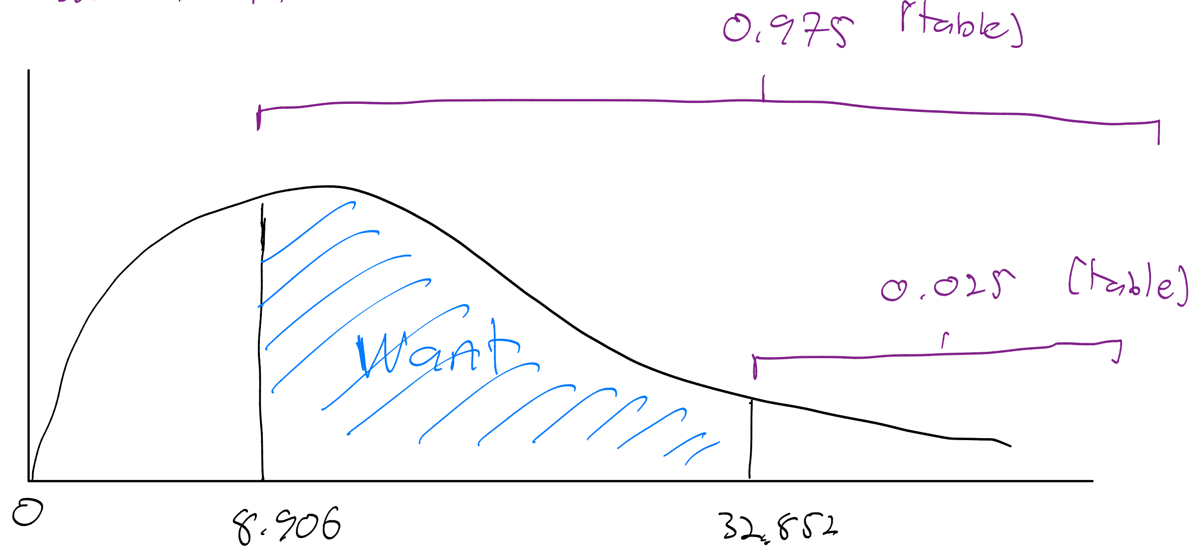
$$= P\left(\frac{(20-1)\,0.656}{1.4} \leq \chi^2_{19} \leq \frac{(20-1)\,2.42}{1.4}\right)$$

$$= P(8.906 \leq \chi^2_{19} \leq 32.852)$$

$$= 0.975 - 0.025$$

$$= 0.950$$

using χ^2 table at 19 df



Solution

These values can be found by using percentiles from the chi-square distribution.

With $\sigma^2 = 1.4$ and $n = 20$,

$\frac{n-1}{\sigma^2}S^2 = \frac{19}{1.4}S^2$ has a chi-square distribution with 19 degrees of freedom.

a. $P(S^2 \leq b) = P\left(\frac{n-1}{\sigma^2}S^2 \leq \frac{(n-1)b}{\sigma^2}\right) = P\left(\frac{19}{1.4}S^2 \leq \frac{19b}{1.4}\right) = 0.975$

$\frac{19b}{1.4}$ must be equal to the 97.5%-tile of a chi-square with 19 df, thus $\frac{19b}{1.4} = 32.8523$ (using Table 6). An so, $b = 2.42$

Solution

b. Similarly, $P(S^2 \geq a) = P\left(\frac{n-1}{\sigma^2}S^2 \geq \frac{(n-1)a}{\sigma^2}\right) = 0.975$. Thus, $\frac{19a}{1.4} = 8.90655$, the 2.5%-tile of this chi-square distribution, and so $a = 0.656$.

c. $P(a \leq S^2 \leq b) = P(0.656 \leq S^2 \leq 2.42) = 0.95$.

Example

Let X equal the weight (in grams) of a nail of the type that is used for making decks. Assume that the distribution of X is $N(\mu = 8.78, \sigma^2 = 0.16)$. Let S^2 be the sample variance of the nine weights. Find constants a and b so that $P[a \leq S^2 \leq b] = 0.90$.

R: qchisq(prob, df)

based on area to left

Solution (using R)

```
b=qchisq(0.95,df=8)*0.16/8;
b

## [1] 0.3101463

a=qchisq(0.05,df=8)*0.16/8;
a

## [1] 0.05465274
```

HW?

Marks on a standardized test are Normally distributed with mean 75, standard deviation 15. What is the probability the class sample standard deviation, for a class of 31, is greater than 16?

S

Summary (a bit ahead)

qn's involving $S^2 \longrightarrow$ use χ^2 or F

" " mean $\longrightarrow$ use Z or t

Definition 7.2: Student's t distribution

Let

ratio of normal and the sqr root of a $\chi^2_{(r)}/r$

$Z \sim N(0,1)$

$$T = \frac{Z}{\sqrt{W/r}},$$

$W \sim \chi^2_{(r)}$

where Z is a random variable that is $N(0,1)$, W is a random variable that is $\chi^2(r)$, and Z and W are independent. Then T has a t distribution with r degrees of freedom, for which the pdf is

pdf

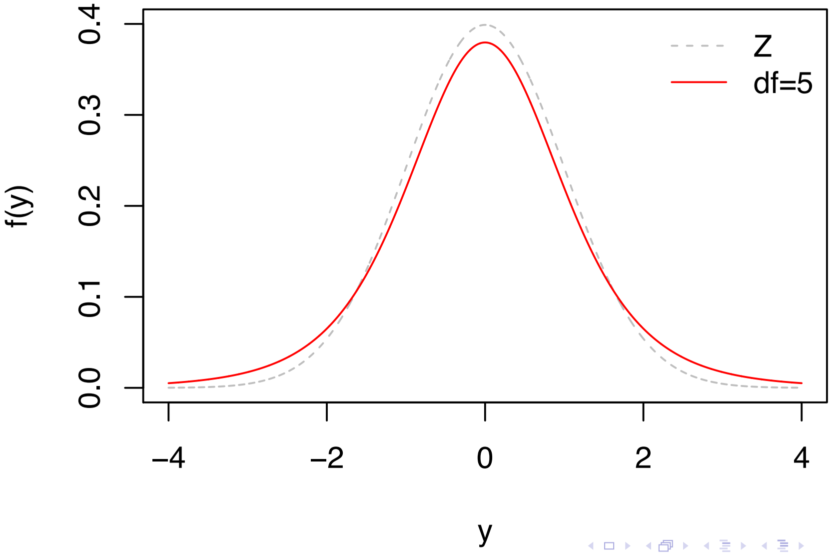
$$f(t) = \frac{\Gamma((r+1)/2)}{\sqrt{\pi r}\Gamma(r/2)} \frac{1}{(1+t^2/r)^{(r+1)/2}}, -\infty < t < \infty.$$

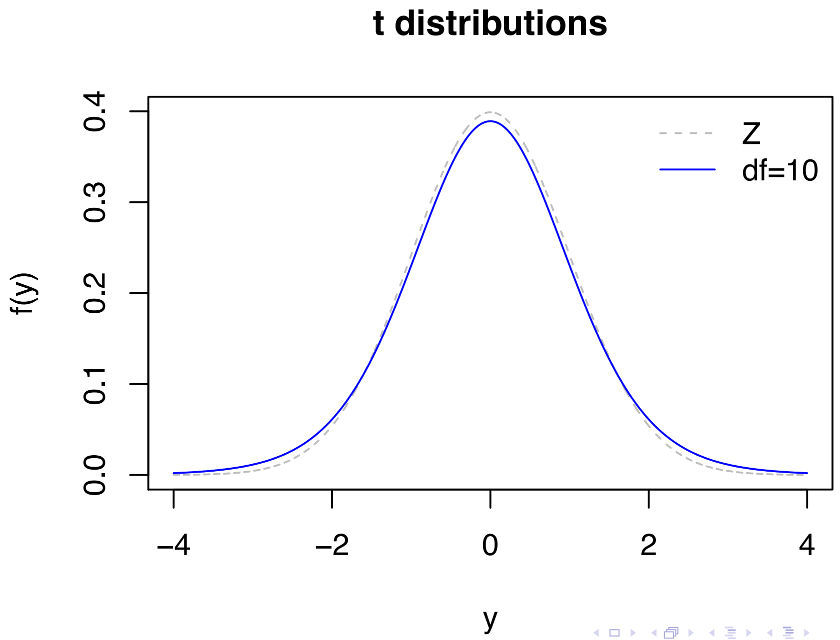
W.S. Gosset

resembles normal but with thicker tails

t distributions

(has df)

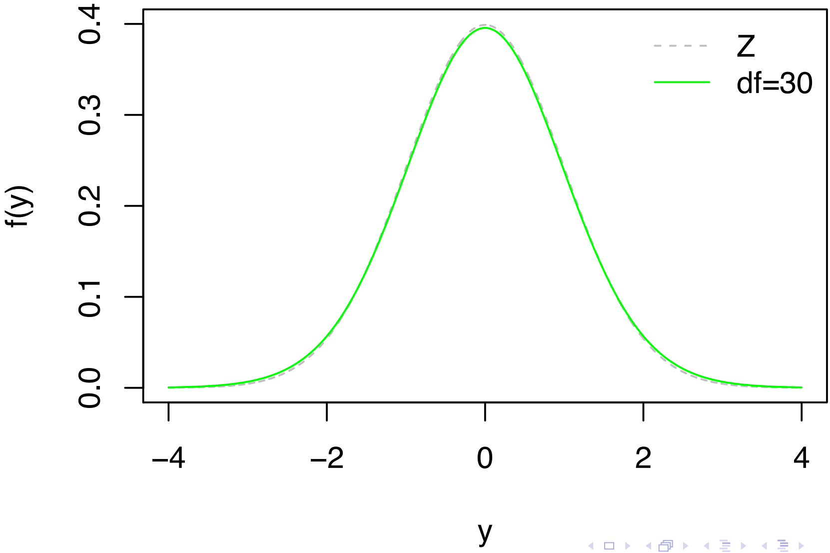



t distributions
Z
df=10
f(y)
0.0
0.1
0.2
0.3
0.4
y
−4
−2
0
2
4

as $df \longrightarrow \infty$

$t \longrightarrow Z$

t distributions



Properties of the t distribution, $T \sim t_\nu$

ν : degrees of freedom

$1 \leq \nu \leq 2 \rightarrow Var \rightarrow \infty$

$\nu < 1 \rightarrow$ var is undef.

- Support - the real numbers
- $E(T) = 0$
- $Var(T) = \frac{\nu}{\nu-2} > 1$ $Var(T) \rightarrow 1$ as $v \rightarrow \infty$

  $\nu > 2$
- The t_ν is shorter and fatter than the $N(0,1)$ distribution. The tails of the t_ν decrease to zero slower than those for the $N(0,1)$.
- $t_\nu \rightarrow Z$ (in distribution) as $\nu \rightarrow \infty$

For problems involving the mean

σ known $\longrightarrow$ Z

σ not known $\longrightarrow$ t

σ: pop st. dev

S: sample " "

Question: Why is the t_ν wider and more tail heavy than the $N(0,1)$.

Answer:

$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ contains variability from $\bar{X}$.

$\boxed{T = \frac{\bar{X}-\mu}{S/\sqrt{n}}} \sim t_{n-1}$ contains variability from $\bar{X}$ and from S.

Very important use of the t distribution

If $Y_1, Y_2, ..., Y_n \sim$ iid $N(\mu, \sigma^2)$, then

1. $\bar{Y} \sim N(\mu, \sigma^2/n)$ or $Z = \frac{\bar{Y}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$
2. $\frac{n-1}{\sigma^2} S^2 \sim \chi^2_{n-1}$
3. $\bar{X}$ and S^2 are independent (see STA260)

Thus $\frac{\frac{\bar{Y}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{n-1}{\sigma^2} S^2/(n-1)}} = \frac{\bar{Y}-\mu}{S/\sqrt{n}} \sim t_{n-1}$

Example

$\mu = 300$
(claim)

$\bar{x} = 290$

$n = 15$

The Acme Corporation manufactures light bulbs. The CEO claims that an average Acme light bulb lasts 300 days. A researcher randomly selects 15 bulbs for testing. The sampled bulbs last an average of 290 days, with a standard deviation of ~~50~~ 20 days. If the CEO's claim were true, what is the probability that 15 randomly selected bulbs would have an average life of no more than 290 days?
Note. Please, assume Normality.

$s = 20$

(σ NOT known $\to$ use t)

Example (slide 33) MODIFIED

$\mu = 300$ (claim) $n = 15$ $\bar{x} = 290$ $s = 20$

(σ NOT known)
↳ use t

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \sim t_{(n-1)}$$

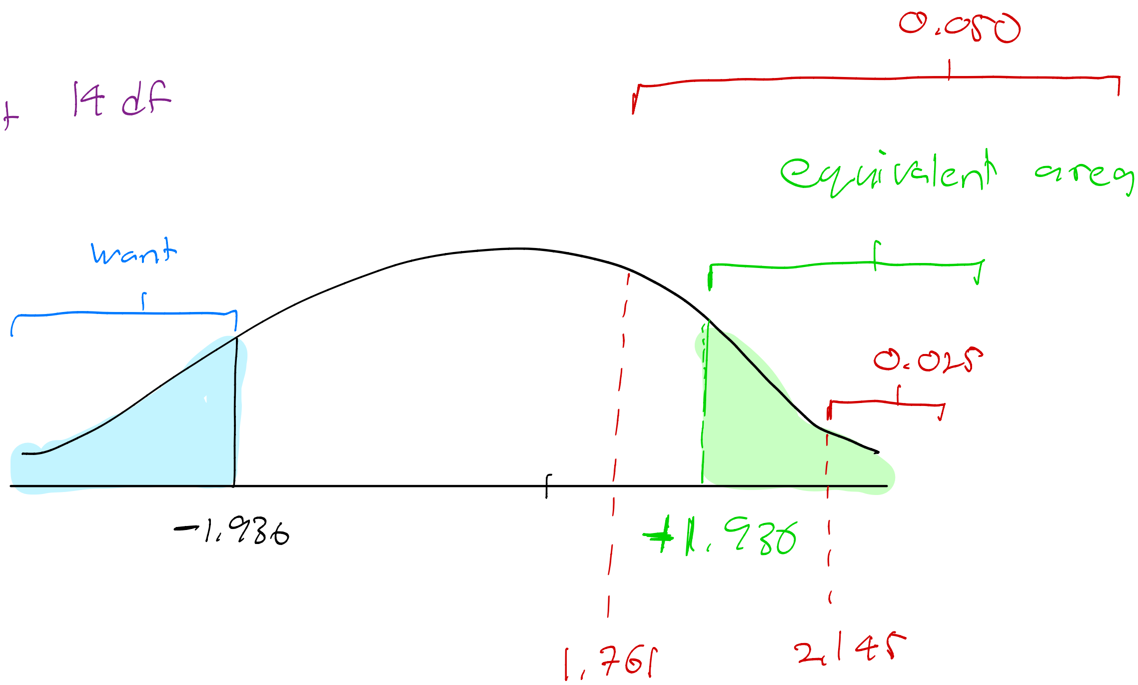
$$P(\bar{X} \leq 290)$$

$$= P\Big(\underbrace{\frac{\bar{X} - \mu}{s/\sqrt{n}}}_{t_{(n-1)}} \leq \frac{290 - \mu}{s/\sqrt{n}}\Big)$$

$$= P\Big(t_{14} \leq \frac{290 - 300}{20/\sqrt{15}}\Big)$$

$$= P(t_{14} \leq -1.936)$$

using t-dist at 14 df



$$0.025 \leq P(t_{14} \leq -1.936) \leq 0.05$$ (best with table)

table gives area to right

R: $\underbrace{\text{pt}(\text{value}, df)}_{\text{area to LEFT}}$

Solution

$X_1, X_2, ..., X_{15}$ iid random variables from a $N(\mu = 300, \sigma)$.

$$P(\bar{X} \leq 290) = P\left(\frac{\bar{X}-\mu}{s/\sqrt{n}} \leq \frac{290-300}{50/\sqrt{15}}\right)$$
$$= P(t_{14} \leq -0.7745) \text{ (since the } t \text{ distribution is symmetric)}$$
$$= P(t_{14} \geq 0.7745) \text{ (using our } t \text{ distribution table)}$$

The best that we can do using table is:

$0.20 < P(\bar{X} \leq 290) < 0.25$

```r
# pt = CDF of a £t£ distribution;

pt(-0.7745,14);

## [1] 0.2257589
```

Example

The Edison Electric Institute (EEI) has published figures on the annual number of kilowatt-hours expended by various home appliances. It is claimed that a vacuum cleaner expends an average of 46 kilowatt-hours per year. A random sample of 12 homes included in a planned study indicates that vacuum cleaners expend an average of 42 kilowatt-hours per year with a standard deviation of 11.9 kilowatt-hours.
If the EEI's claim were true, what is the probability that 12 randomly selected vacuum cleaners expend, on the average, less than 42 kilowatt-hours annually? Assume the population of kilowatt-hours to be Normal.

Solution

$X_1, X_2, ..., X_{12}$ iid random variables from a $N(\mu = 46, \sigma)$.

$$P(\bar{X} \leq 42) = P\left(\frac{\bar{X}-\mu}{s/\sqrt{n}} \leq \frac{42-46}{11.9/\sqrt{12}}\right)$$
$$= P(t_{11} \leq -1.1644) \text{ (since the } t \text{ distribution is symmetric)}$$
$$= P(t_{11} \geq 1.1644) \text{ (using our } t \text{ distribution table)}$$

The best that we can do using table from textbook is:

$P(\bar{X} \leq 42) > 0.10$

```r
# pt = CDF of a £t£ distribution;

pt(-1.1644 ,11);

## [1] 0.1344472
```

Attend tutorials → TA's will discuss the ans to a potential bonus

Quiz 2 end of this week

Test 2 next week

Past tests will be uploaded to quercus (no sol^{n}s)

↳ To give practice

past years were more theoretical

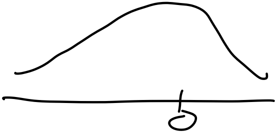
Review

T distrib

$Z \sim N(0,1)$

$$T = \frac{Z}{\sqrt{\frac{W_{(n)}}{n}}} \sim t_{(n)}$$

$\chi^2_{(n)}$



when σ is not known

$$T = t = \frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t_{(n-1)}$$

Definition 7.3

Let W_1 and W_2 be independent χ^2-distributed random variables with ν_1 and ν_2 df, respectively. Then

$$F = \frac{W_1/\nu_1}{W_2/\nu_2} \sim F_{\nu 1,\ \nu_2}$$

is said to have an F distribution with ν_1 numerator degrees of freedom and ν_2 denominator degrees of freedom.

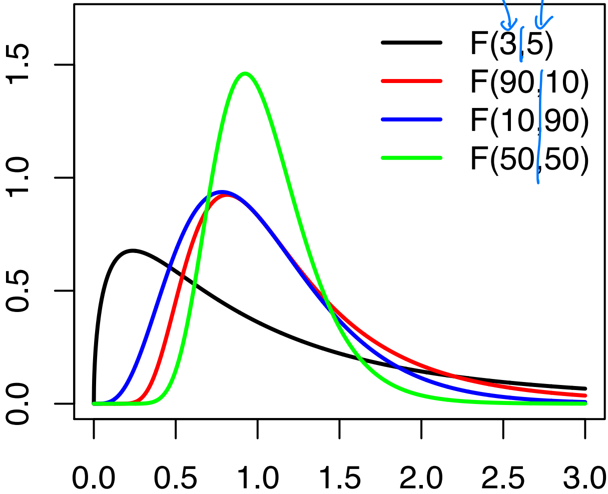
Ronald Fisher

pdf: complicated

F distributions

numerator df

denom df



Properties of F distribution, F_{ν_1,ν_2}

- Support - non-negative real numbers
- Relationship with its reciprocal: If $F \sim F_{\nu_1,\nu_2}$ then $W_1 = 1/F \sim F_{\nu_2,\nu_1}$
- Relationship with the t distribution: If $T \sim t_\nu$ then $W_2 = T^2 \sim F_{1,\nu}$
- Relationship with the exponential distribution: If U_1 and U_2 are independent exponential RVs with the same parameter, then $W_3 = U_1/U_2 \sim F_{2,2}$

Variances are SAME

size n

sample variance of X's

If

1. $X_1, X_2, ..., X_n \sim$ iid $N(\mu_X, \sigma^2)$, then $\frac{n-1}{\sigma^2} S_X^2 \sim \chi^2_{n-1}$
2. $Y_1, Y_2, ..., Y_m \sim$ iid $N(\mu_Y, \sigma^2)$, then $\frac{m-1}{\sigma^2} S_Y^2 \sim \chi^2_{m-1}$
3. Xs and Ys all independent

size m

sample variance of Y's

then

num df

$$F = \frac{\overbrace{\left(\frac{n-1}{\sigma^2} S_X^2\right) / (n-1)}^{\text{numerator}}}{\underbrace{\left(\frac{m-1}{\sigma^2} S_Y^2\right) / (m-1)}_{\text{denominator}}} = \frac{S_X^2}{S_Y^2} \sim F_{n-1, m-1}$$

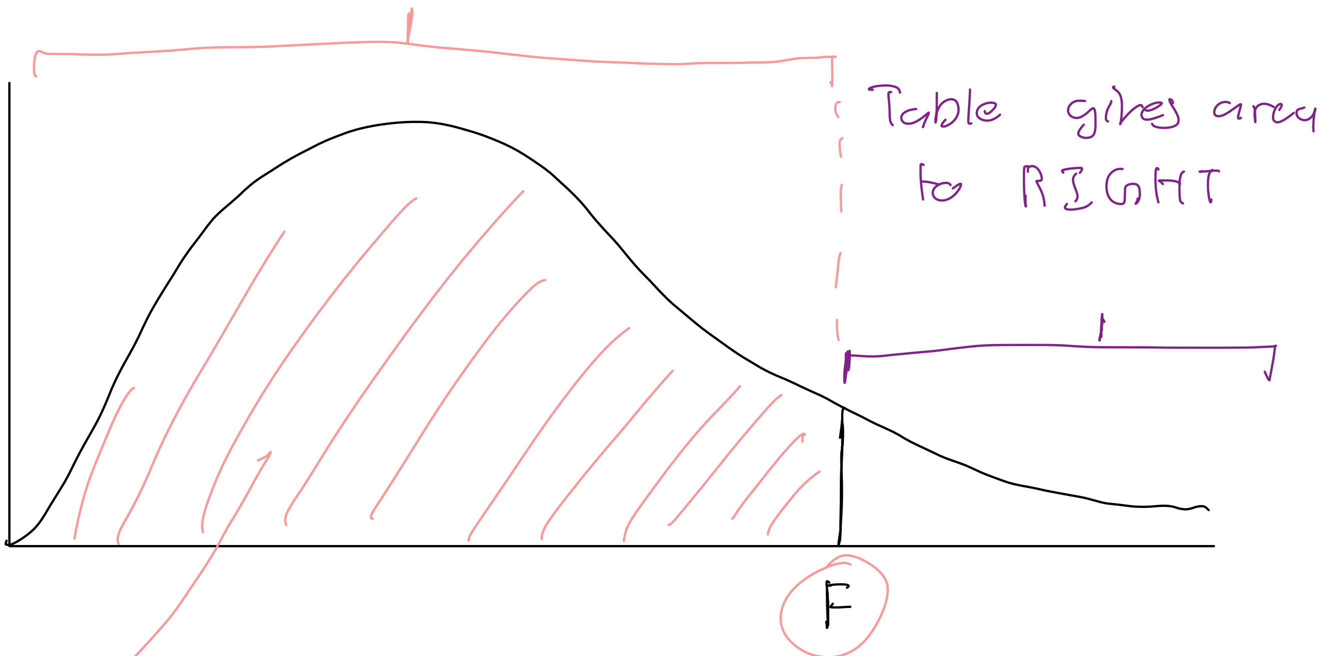
denom df

down ← denominator

Ratio of Sample Variance of X to the sample variance of Y

$$F = \frac{\left(\frac{n-1}{\sigma^2} \cdot S_x^2\right) / (n-1)}{\left(\frac{m-1}{\sigma^2} \cdot S_y^2\right) / (m-1)} \sim F_{n-1, m-1}$$

R gives area to right (by default)



pf(F, n-1, m-1)

F value

num df

denom df

$\chi^2_{(n)}$ used in qn's related to prob of a sample variance

$F_{(n,m)}$ " " " " " " "

ratios of 2 sample variances

Example

Marks on a standardized test are normally distributed with mean 75, standard deviation 15. The test is given to two classes, one class (A) has 31 students and the other (B) has 16 students.

- What is the probability the ratio of the two sample variances (A/B) is greater 2?
- What is the probability the ratio of the two sample variances (B/A) is less than 2?

Example (on slide 43)

Overall (population)

$\mu = 75 \quad \sigma = 15$

$\sigma^2 = 15^2$

From same population

Class A $\qquad$ Class B

$n_A = 31 \qquad n_B = 16$

want prob ratio of 2 sample variances → suggesting F

is greater than 2.

$$P\left(\frac{S_A^2}{S_B^2} > 2\right)$$

$$\frac{\left(\frac{n_A-1}{\sigma^2}\right) S_A^2 \Big/ (n_A-1)}{\left(\frac{n_B-1}{\sigma^2}\right) S_B^2 \Big/ (n_B-1)}$$

$$= P\left(\frac{\frac{1}{\sigma^2}\,\frac{n_A-1}{n_A-1}\cdot S_A^2}{\frac{1}{\sigma^2}\,\frac{n_B-1}{n_B-1}\cdot S_B^2} > 2\right)$$

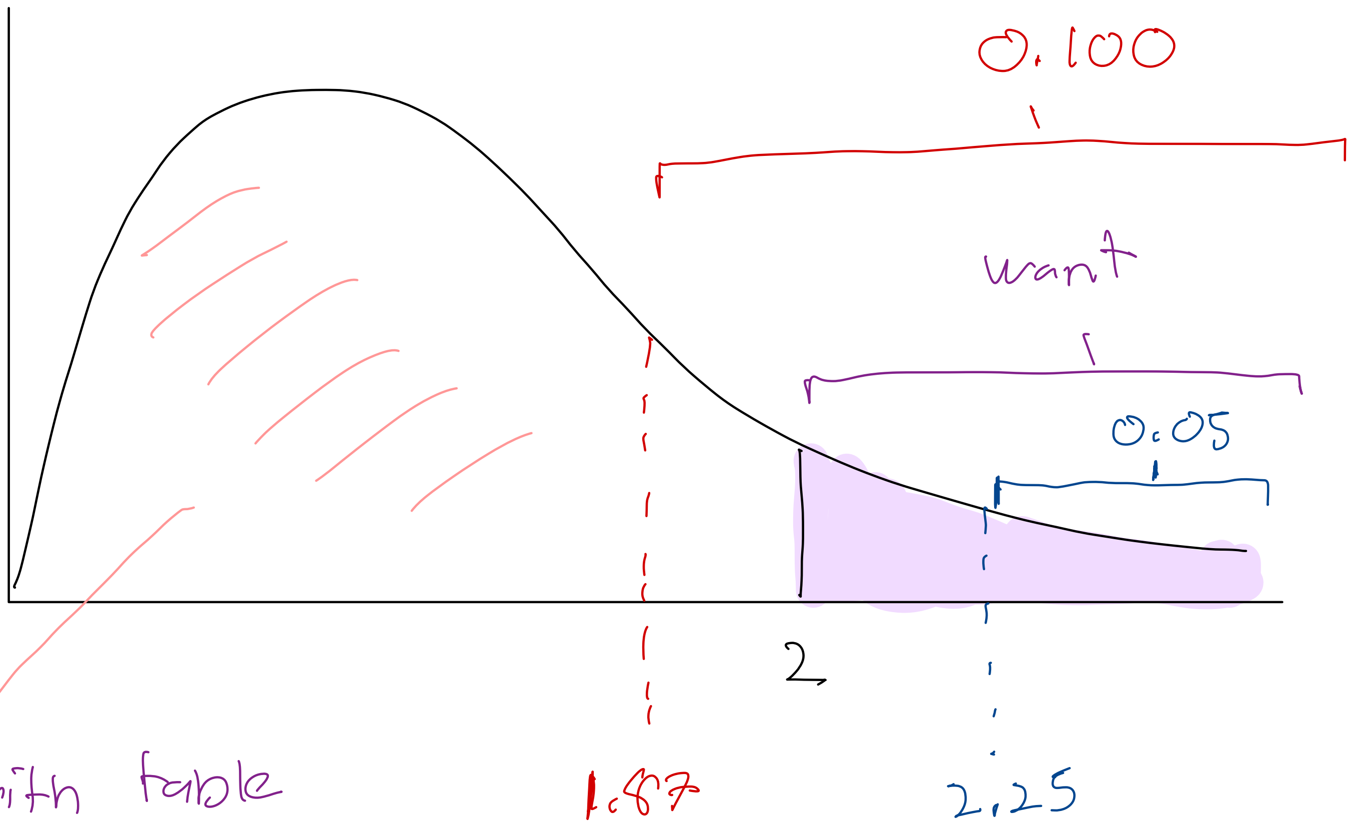
$$= P\left(\frac{\left(\frac{n_A-1}{\sigma^2}\right) S_A^2 \Big/ \overbrace{(n_A-1)}^{\text{num}}}{\left(\frac{n_B-1}{\sigma^2}\right) S_B^2 \Big/ \underbrace{(n_B-1)}_{\text{denom}}} > 2\right)$$

$\sim F_{n_A-1,\, n_B-1}$

$\sim F_{31-1,\, 16-1}$

$\sim F_{30,15}$

F distrib at num = 30 df and denom = 15 df



Best with table

$$0.05 < \underbrace{P\left(\frac{S_A^2}{S_B^2} > 2\right)}_{\text{prob we want}} < 0.10$$

using R

$$> \underbrace{1 - \text{pf}(2, 30, 15)}_{\text{area to left}}$$

> 1 - 0.921

> 0.0787

Solution (A/B)

$\mu = 75$, $\sigma = 15$, $n = 31$, $m = 16$.

$$
\begin{aligned}
P\left(\frac{S_X^2}{S_Y^2} > 2\right) &= P\left(\frac{\sigma^2}{\sigma^2}\frac{S_X^2}{S_Y^2} > 2\right) \\
&= P\left(\frac{S_X^2/\sigma^2}{S_Y^2/\sigma^2} > 2\right) \\
&= P\left(\frac{\frac{n-1}{\sigma^2}\frac{S_X^2}{n-1}}{\frac{m-1}{\sigma^2}\frac{S_Y^2}{m-1}} > 2\right) \\
&= P\left(\frac{\chi^2_{n-1}/(n-1)}{\chi^2_{m-1}/(m-1)} > 2\right) \\
&= P(F_{n-1,\ m-1} > 2) = P(F_{30,15} > 2)
\end{aligned}
$$

Using our Table, the best that we can do is:

$0.05 < P\left(\frac{S_X^2}{S_Y^2} > 2\right) < 0.10$.

```r
# pf = CDF of F ;

# area to the left of 2;

pf(2,30,15);

## [1] 0.9213152

# area to the right of 2;

1-pf(2,30,15);

## [1] 0.07868481
```

Solution (B/A)

$\mu = 75,\ \sigma = 15,\ n = 31,\ m = 16.$

$$
\begin{aligned}
P\left(\frac{S_Y^2}{S_X^2} > 2\right) &= P\left(\frac{\sigma^2}{\sigma^2}\frac{S_Y^2}{S_X^2} > 2\right) \\
&= P\left(\frac{S_Y^2/\sigma^2}{S_X^2/\sigma^2} > 2\right) \\
&= P\left(\frac{\frac{m-1}{\sigma^2}\frac{S_Y^2}{m-1}}{\frac{n-1}{\sigma^2}\frac{S_X^2}{n-1}} > 2\right) \\
&= P\left(\frac{\chi^2_{m-1}/(m-1)}{\chi^2_{n-1}/(n-1)} > 2\right) \\
&= P(F_{m-1,\ n-1} > 2) = P(F_{15,30} > 2)
\end{aligned}
$$

```r
# pf = CDF of F ;
# area to the left of 2;
pf(2,15,30);
## [1] 0.948209
# area to the right of 2;
1-pf(2,15,30);
## [1] 0.051791
```