11/6/21

7. Sampling Distributions

Population: Collection of elements for investigating

sample: Part of population / subset of population which is collected to draw an inference about population.

Parameter: Statistical measure which is based on population $\mu, \sigma^2, \ldots$
↳ population

statistics: statistical measure based on sample.

sample mean $\bar{x}$, Sample variance s^2 etc

14/6/21

Simple Random Sampling: Each sample has equal chance of getting selecting

→ Sampling with Replacement
→ Sampling without Replacement

Statistic:

Def: Let $\{X_1, X_2, \ldots X_n\}$ be the sample of size n, function of these R.V's $f(x_1, x_2 \ldots x_n)$ is called statistic

It is a function of only known parameter.

Statistics: Let $X_1, X_2 \cdots X_n$ be the n no. of R.V's

X_1 ↓ 1st sample, X_2 ↓ 2nd sample, X_n ↓ nth sample

Sample Mean $\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$

Sample Variance $s^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{x})^2}{n-1}$

Ex: Population follows $X \sim b(1, P)$

$P(X=1) = P, \quad P(X=0) = 1-P$

0, 1, 1, 0, 1
$x_1 \; x_2 \; x_3 \; x_4 \; x_5$

Sample Mean $\bar{x} = \frac{3}{5} = 0.6$

Sample variance $s^2 = \frac{(-0.6)^2 + (0.4)^2 + (0.4)^2 + (-0.6)^2 + (0.4)^2}{4}$

$= \frac{0.36 + 0.16 + 0.16 + 0.36 + 0.36}{4}$

$s^2 = 0.3$

$s = \sqrt{0.3}$

$s = 0.55$

Ex $X \sim N(\mu, \sigma^2)$ both are unknown

Let $X_1, X_2, \ldots X_n$ be a sample

then $\frac{\sum_i X_i}{\sigma^2}$ is it statistic

Ans: No, since σ^2 is unknow

$\rightarrow -0.864, 0.561, 2.355, 0.582, -0.774$

Sample Mean $\bar{X} = 0.372$

Sample Variance $s^2 = 1.648$

$\rightarrow E(\bar{X}) = \frac{E(X_1+X_2+\ldots+X_n)}{n}$ X_i is iid

$= \frac{E(X_1)+E(X_2)+\ldots+E(X_n)}{n}$

$= \frac{n\mu}{n}$

$= \mu$ (known or unknown)

$Var(\bar{X}) = Var\left(\frac{X_1+X_2+\ldots+X_n}{n}\right)$

$= \frac{\sigma^2}{n}$ (known or unknown)

μ & σ^2 are mean & variance of population

which are constants

Use of sampling:

1. To estimate the unknown parameters of population.
2. To test the validity of a statement

Sampling distribution for normal population

* chisquare distribution
* F-distribution
* T-distribution.

chisquare distribution: Special case of gamma distribution with $\alpha = \frac{n}{2}$, $\beta = 2$.

A R.V is said to have chisquare distribution with 'n' degrees of freedom where $n>0$ if its pdf

$$f(x) = \begin{cases} \frac{1}{\Gamma(n/2)\, 2^{n/2}} \cdot e^{-x/2} \cdot x^{n/2 - 1} & x>0 \\ 0 & \text{otherwise} \end{cases}$$

$X \sim \chi^2(n)$ with 'n' degrees of freedom.

$$E(X^k) = \frac{1}{\Gamma(n/2)} \int_0^\infty \frac{x^k \cdot e^{-x/2} \cdot x^{n/2 - 1}}{2^{n/2}}$$

$$= \frac{1}{\Gamma(n/2)} \int_0^\infty \frac{e^{-x/2} \cdot x^{\frac{n}{2}+k-1}}{2^{n/2}}$$

$$\Gamma(\alpha) = \int_0^{\infty} \frac{e^{-x/\beta} \cdot x^{\alpha-1}}{\beta^{\alpha}} dx$$

$$= \frac{2^k}{\Gamma(n/2)} \int_0^{\infty} \frac{e^{-x/2} \cdot x^{n/2+k-1}}{2^{n/2+k}} dx$$

$$E(X^k) = \frac{2^k}{\Gamma(n/2)} \cdot \Gamma\left(\frac{n}{2}+k\right)$$

$$E(X) = \text{Mean} = \mu = \frac{2 \cdot \Gamma\left(\frac{n}{2}+1\right)}{\Gamma\left(\frac{n}{2}\right)}$$

$$= \frac{2 \cdot \frac{n}{2} \Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}$$

$$= n.$$

$$E(x^2) = \frac{2^2 \cdot \Gamma\left(\frac{n}{2}+2\right)}{\Gamma\left(\frac{n}{2}\right)}$$

$$= \frac{4 \cdot \left(\frac{n}{2}+1\right) \Gamma\left(\frac{n}{2}+1\right)}{\Gamma\left(\frac{n}{2}\right)}$$

$$= \frac{4 \cdot \left(\frac{n}{2}+1\right) \cdot \frac{n}{2} \Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}$$

$$= 2n\left(\frac{n}{2}+1\right)$$

$$E(X^2) = n^2+2n$$

$$Var = E(X^2) - (E(X))^2$$

$$= n^2 + 2n - n^2$$

$$Var = 2n.$$

Let $X_1, X_2 \dots X_n$ be iid R.V's

$$S_n = X_1 + X_2 + \dots + X_n$$

now $S_n \sim \chi^2(n) \Leftrightarrow X_1 \sim \chi^2(1)$

$\rightarrow X_i \sim N(0,1) \implies \sum_{i=1}^{n} X_i^2 \sim \chi^2(n)$

standard normal dist

$$Z_1, Z_2 \dots Z_n \sim N(0,1)$$

$$Z_1^2, Z_2^2 \dots Z_n^2 \sim \chi^2(n)$$

Properties:

1. sum of squares of normal distribution
2. Curve is non symmetrical skewed right

   as $n \to \infty$ curve looks like normal
3. Different curve for each n.
4. Always +ve
5. For $n \geq 30$, normal distribution can be used.
6. Mean is located to the right side of peek.

$$\text{skewness} \propto \frac{1}{\text{degrees of freedom}}$$

$$P(\chi^2(n) > \psi^2_{n,\alpha}) = \alpha.$$

Ex: $n = 25$

$$P(\psi^2(25) \leq 34.382) = ?$$

$$= 1 - P(\psi^2(25) > 34.382)$$

$$= 1 - 0.10$$

$$= 0.9$$

Ex: 2 samples

$$P(\psi^2(2) \geq 6) = 0.05$$

9 samples

$$P(\psi^2(9) \geq 6.3) = 0.70$$

15/6/21

Student's T-distribution:

A R.V is said to follow T-distribution if its pdf is given by

$$f_n(t) = \frac{\gamma\left(\frac{n+1}{2}\right)}{\gamma\left(\frac{n}{2}\right)\cdot\sqrt{n\pi}}\left(1+\frac{t^2}{n}\right)^{-\left(\frac{n+1}{2}\right)}$$

$-\infty < t < \infty$

$T \sim t(n)$

$n \rightarrow$ degrees of freedom.

$$T = \frac{X}{\sqrt{\frac{Y}{n}}}$$

, $X \sim N(0,1)$ Mean $=0$, Variance $=1$

$Y \sim \psi^2(n)$ Mean $=n$, Variance $=2n$

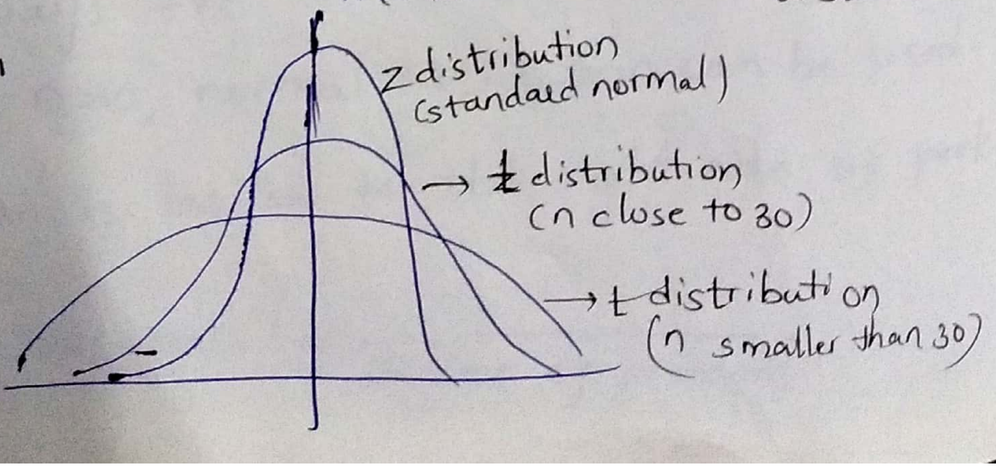
$n=1$

$$f_n(t) = \frac{\gamma\left(\frac{2}{2}\right)}{\gamma\left(\frac{1}{2}\right)\sqrt{\pi}}(1+t^2)^{(-1)}$$

$$= \frac{1}{\sqrt{\pi}\cdot\sqrt{\pi}\,(1+t^2)}$$

$$f_n(t) = \frac{1}{\pi(1+t^2)}$$ Cauchy's distribution

~~Prope~~



Properties;

① Symmetric about Mean. $\mu = 0$

② Mean = Median = Mode = 0

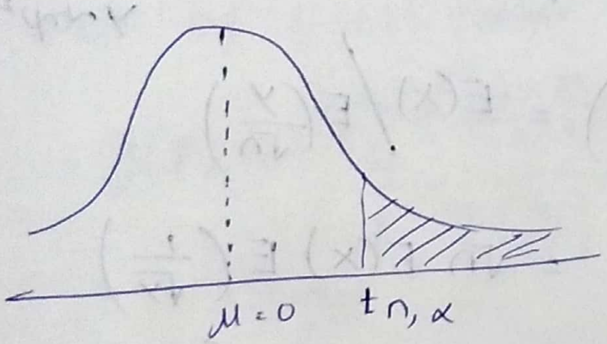
③ As $n \to \infty$ $f_n(t) \to 0$

④ As $n \to \infty$ it is close to normal distribution.

⑤ As $t \to \infty$ or $t \to -\infty$ $f_n(t) \to 0$ slower than tails of normal distribution.

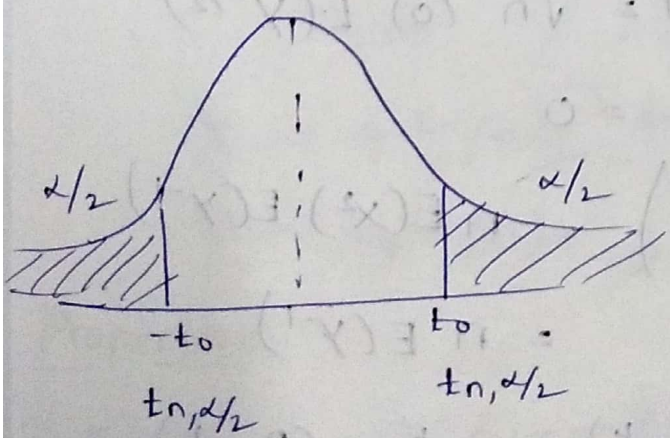
⑥ $P(|T| > t_0) \geq P(|Z| > t_0)$, $Z \sim N(0,1)$

Calculation of probabilities



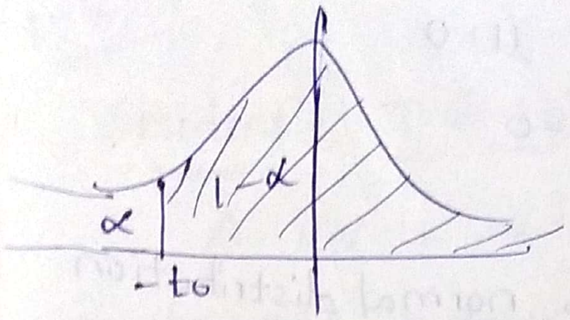
Table

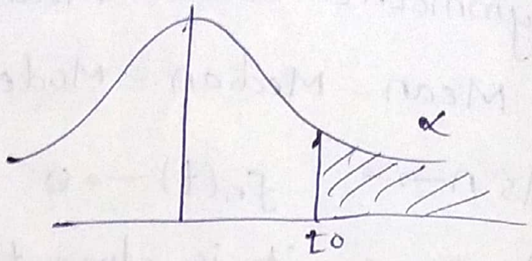
$T(t > t_{n,\alpha}) = \alpha$



$P\{|T| > t_0\} = \alpha$

$P\{-t_0 < T < t_0\}$





$$\boxed{t_{n,1-\alpha} = -t_{n,\alpha}}$$

$$t_{30,0.05} = 1.697$$

$$t_{30,0.95} = -1.697$$

Mean & Variance:

X & Y are independent $\quad X \sim N(0,1)$

$Y \sim \chi^2(n)$

$$T = \frac{X}{\sqrt{Y/n}}$$

$$E(T) = E\left(\frac{X}{\sqrt{Y/n}}\right) = E(X) \Big/ E\left(\frac{Y}{\sqrt{n}}\right)$$

$$= \sqrt{n}\, E(X)\, E\left(\frac{1}{\sqrt{Y}}\right)$$

$$= \sqrt{n}\, E(X)\, E\left(Y^{-1/2}\right)$$

$$E(T) = \sqrt{n}\, (0) \cdot E\left(Y^{-1/2}\right)$$

$$= 0$$

$$E(T^2) = E\left(\frac{X^2}{Y/n}\right) = n\, E(X^2)\, E(Y^{-1})$$

$$= n \cdot E(Y^{-1})$$

$$E(Y^{-1}) \qquad E(Y^k) = \frac{2^k\, \Gamma\left(\frac{n}{2}+k\right)}{\Gamma\left(\frac{n}{2}\right)}$$

$$E(Y^{-1}) = \frac{1}{2} \cdot \frac{\Gamma\left(\frac{n}{2}-1\right)}{\Gamma\left(\frac{n}{2}\right)}$$

$$= \frac{1}{2} \cdot \frac{\gamma\left(\frac{n}{2}-1\right)}{\left(\frac{n}{2}-1\right)\gamma\left(\frac{n}{2}-1\right)}$$

$\frac{n}{2}-1>0$

$\frac{n}{2}>1$

$n>2$

$$= \frac{1}{2\left(\frac{n}{2}-1\right)}$$

$$E(Y^{-1}) = \frac{1}{n-2} \quad \text{for } n>2$$

Mean $=0$, $E(T^2) = Var(T) = \frac{n}{n-2}$ for $n>2$

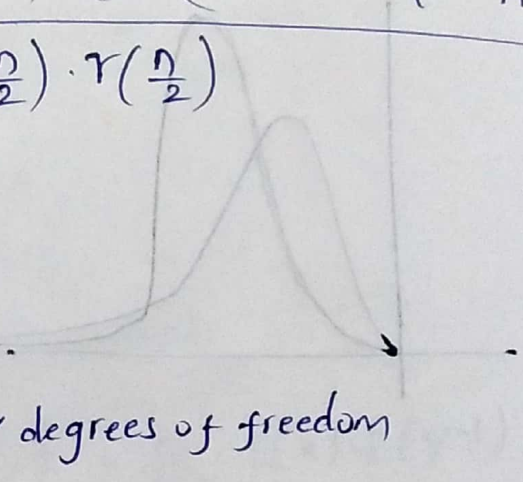
16/6/21

F-Distribution:

PDF of F-distribution is given by

$$g(f) = \begin{cases} \dfrac{\left(\gamma\left(\frac{m+n}{2}\right)\right)\cdot\left(\frac{m}{2}\right)\left(\frac{m}{n}f\right)^{\frac{m}{2}-1}\left(1+\frac{m}{n}f\right)^{-\left(\frac{m+n}{2}\right)}}{\gamma\left(\frac{m}{2}\right)\cdot\gamma\left(\frac{n}{2}\right)} & f>0 \\ 0 & f\leq 0 \end{cases}$$

$$F = \frac{X/m}{Y/m}$$

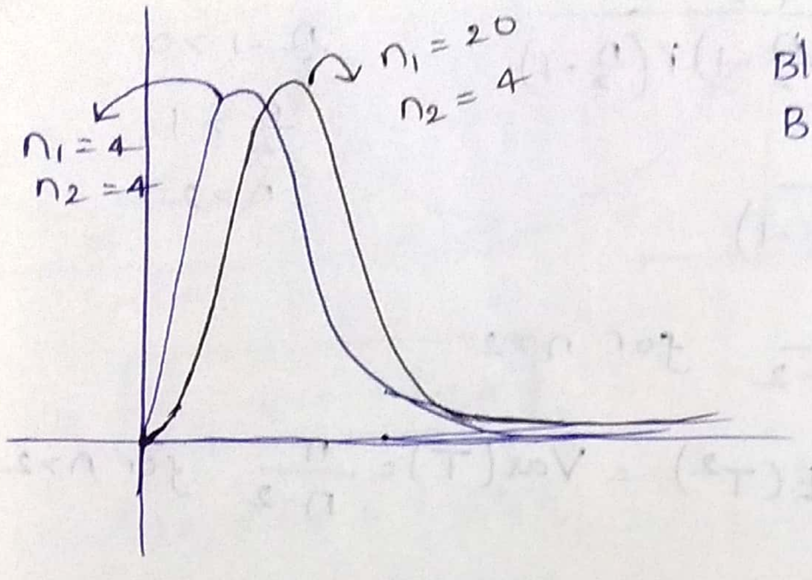


$X \sim \psi^2(m)$ degrees of freedom

$Y \sim \psi^2(n)$

Properties:

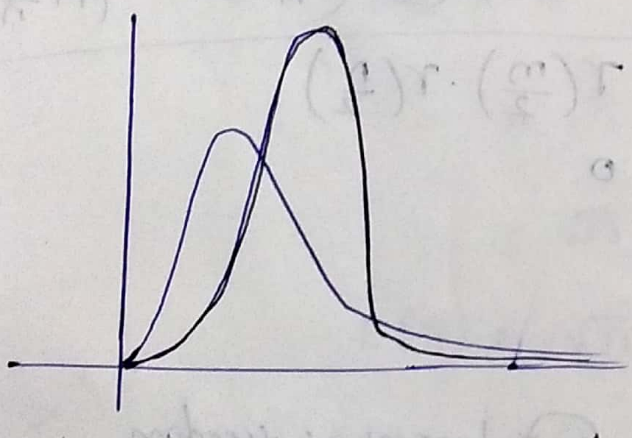
1. Curve is not symmetric
2. Different curves for different pairs of m & n.
3. $F \geq 0$
4. As degrees of freedom increases it approaches normal distribution.



Black $m = 20, n = 4$
Blue $m = 4, n = 4$

⑤ Mean decreases when n increases and remains same when m changes

⑥ As we increase both, height increases.



Black $m = 20, n = 20$
Blue $m = 4, n = 4$

① $X \sim F(m,n)$, $\frac{1}{X} \sim F(n,m)$

$$F = \frac{X/m}{Y/n} \quad , \quad \frac{1}{F} = \frac{Y/n}{X/m}$$

② $m = 1$, F dist. behaves like square of t-dist

$F = (t(n))^2$, $F(1,n)$ & $t^2(n)$ have same distribution.

$$T = \frac{X}{\sqrt{Y/n}} \quad , \quad T^2 = \frac{X^2/1}{Y/n} \Rightarrow F = \frac{Z/1}{Y/n}$$

$F \sim F(1, n)$ Random Variable.

$X \sim N(0,1)$ $\qquad$ $X^2 \sim \psi^2(1)$

$Y \sim \psi^2(n)$ $\qquad$ $Y \sim \psi^2(n)$.

$$\boxed{F_{m,n,1-\alpha} = \frac{1}{F_{n,m,\alpha}}}$$

Proof: $P(\overset{R.V}{F} > F_{m,n,\alpha}) = \alpha$

$X \sim F_{m,n}$

$\frac{1}{X} \sim F_{n,m}$

$$P\left(\frac{1}{F} < \frac{1}{F_{m,n,\alpha}}\right)$$

$$= 1 - P\left(\frac{1}{F} > \frac{1}{F_{m,n,\alpha}}\right)$$

$$= 1 - \alpha$$

Mean:

$$F = \frac{X/m}{Y/n} = \frac{n}{m} \cdot \frac{X}{Y}$$

X & Y are independent

$$E(F) = E\left(\frac{n}{m} \cdot \frac{X}{Y}\right) = \frac{n}{m} \cdot E(X) \cdot E(Y^{-1})$$

$$= \frac{n}{m} \cdot m \cdot \frac{1}{n-2}$$

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

Variance:

$$E(F^2) = E\left(\left(\frac{n}{m} \cdot \frac{X}{Y}\right)^2\right)$$

$$= \frac{n^2}{m^2} \cdot E(X^2) \cdot E(Y^{-2})$$

$E(X^2) = m(m+2)$ $\qquad E(X^k) = \dfrac{2^k \Gamma\left(\frac{n}{2}+k\right)}{\Gamma\left(\frac{n}{2}\right)}$

$$E(Y^{-2}) = \frac{2^{-2}\,\Gamma\left(\frac{n}{2}-2\right)}{\Gamma\left(\frac{n}{2}\right)}$$

$$= \frac{1}{4} \cdot \boxed{\frac{\Gamma\left(\frac{n}{2}-2\right)}{\left(\frac{n}{2}-1\right)\left(\frac{n}{2}-2\right)\Gamma\left(\frac{n}{2}-2\right)}}$$

$$= \frac{1}{4} \cdot \frac{4}{(n-2)(n-4)}$$

$$E(Y^{-2}) = \frac{1}{(n-2)(n-4)}$$

$$Var(F) = E(F^2) - (E(F))^2$$

$$= \frac{n^2}{m^2} \cdot m(m+2) \cdot \frac{1}{(n-2)(n-4)} - \left(\frac{n}{n-2}\right)^2$$

$$= \frac{n^2}{n-2}\left[\frac{(m+2)}{m(n-4)} - \frac{1}{n-2}\right]$$

$$= \frac{n^2}{n-2} \cdot \left[\frac{(m+2)(n-2) - m(n-4)}{m(n-2)(n-4)}\right]$$

$$= \frac{n^2}{n-2}\left[\frac{mn - 2m + 2n - 4 - mn + 4m}{m(n-2)(n-4)}\right]$$

$$= \frac{n^2}{(n-2)} \cdot \frac{2(m+n-2)}{m \cdot (n-2)(n-4)}$$

$$Var(F) = \frac{n^2(2m+2n-4)}{m(n-2)^2(n-4)}$$

Distributions of $\bar{X}$ & s^2

Let $X_1, X_2, \ldots X_n$ be a sample from $N(\mu, \sigma^2)$

$$\bar{X} = \frac{1}{n} \sum X_i \quad , \quad s^2 = \frac{\sum (X_i - \bar{X})^2}{n-1}$$

① $X_1, X_2, \ldots X_n$ be iid $N(\mu, \sigma^2)$ R.V's

then $\bar{X}, X_1 - \bar{X}, X_2 - \bar{X}, \ldots X_n - \bar{X}$ are also independent

② $\bar{X}$ & s^2 are independent

③ $$\boxed{\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)}$$

$$s^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2$$

$$\frac{(n-1)s^2}{\sigma^2} = \frac{\sum (X_i - \bar{X})^2}{\sigma^2} \sim \chi^2(n-1)$$

$$\frac{X-\mu}{\sigma} \sim N$$

$$\left(\frac{X-\mu}{\sigma}\right)^2 \sim \chi^2$$

④ $$\boxed{\frac{\sqrt{n}(\bar{X} - \mu)}{s} \sim t(n-1)}$$

⑤ $X_1, X_2, \ldots X_m$ are iid $N(\mu_1, \sigma_1^2)$

$Y_1, Y_2, \ldots Y_n$ are iid $N(\mu_2, \sigma_2^2)$

Two samples were collected

$$\frac{(m-1)s_1^2}{\sigma_1^2} \sim \chi^2(m-1) \quad , \quad \frac{(n-1)s_2^2}{\sigma_2^2} \sim \chi^2(n-1)$$

$$F = \frac{\frac{(m-1)s_1^2}{\sigma_1^2} \Big/ (m-1)}{\frac{(n-1)s_2^2}{\sigma_2^2} \Big/ (n-1)} = \frac{\frac{s_1^2}{\sigma_1^2}}{\frac{s_2^2}{\sigma_2^2}} = \frac{\sigma_2^2}{\sigma_1^2} \cdot \frac{s_1^2}{s_2^2}$$

$$\sim F(m-1, n-1)$$