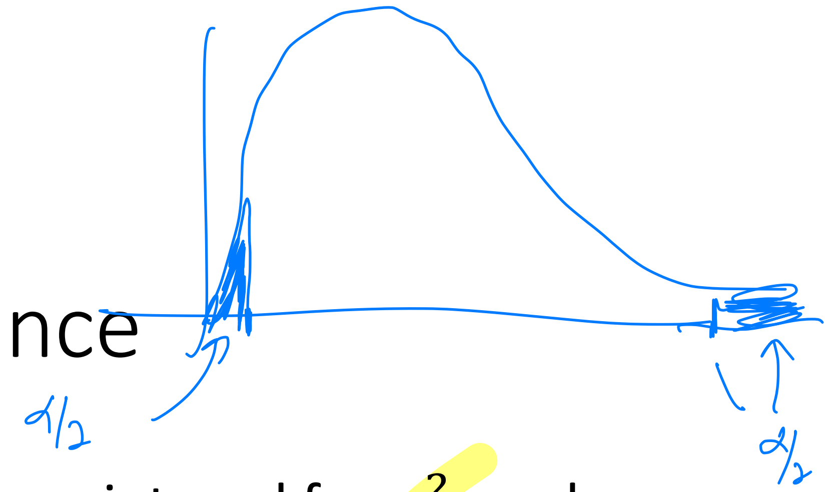
ESTIMATION

Confidence intervals for variances

Confidence intervals for variance



- If $Y_1, Y_2, \dots, Y_n \sim iid\ N(\mu, \sigma^2)$, then a confidence interval for σ^2 can be based on the pivot $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$.

prob. statement:

$$P\left(\chi^2_{n-1,\frac{\alpha}{2}} < \frac{(n-1)s^2}{\sigma^2} < \chi^2_{n-1,1-\alpha/2}\right) = 1-\alpha$$

Check: depends on s^2, $n-1$, σ^2 ✓

dist. known χ^2_{n-1} ✓

pivot!

$$P\left(\frac{\chi^2_{n-1,\frac{\alpha}{2}}}{(n-1)s^2} < \frac{1}{\sigma^2} < \frac{\chi^2_{n-1,1-\alpha/2}}{(n-1)s^2}\right) = 1-\alpha$$

$$P\left(\frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{n-1,\frac{\alpha}{2}}}\right) = 1-\alpha$$

$100(1-\alpha)\%$ C.I. for σ^2

$$\left(\frac{(n-1)s^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}, \frac{(n-1)s^2}{\chi^2_{n-1,\frac{\alpha}{2}}}\right)$$

Confidence intervals for variances: Example

$n = 12$

- Example 10: HDL cholesterol levels (mg/dL) were measured for each of a sample of 12 healthy male volunteers for a study on cardiovascular disease. For the sample, the mean level is 63.4 and the standard deviation is 19.4. Assuming that the data come from a normal distribution, calculate a 95% confidence interval for the true population variance.

$\bar{Y} = 63.4$

$s = 19.4$ $\alpha = 0.05$

$\chi^2_{11,\,0.975}$

$\chi^2_{11,\,0.025}$

qchisq (df = 11, 0.975) = 21.92

0.025 3.816

$$95\%\ \text{c.i.}: \left(\frac{11(19.4)^2}{21.92}, \frac{11(19.4)^2}{3.816} \right) \quad \boxed{(188.9,\ 1084.9)}$$

We know $U \sim \chi^2_{\nu_1}$, $W \sim \chi^2_{\nu_2}$ (ind)

$$\frac{U/\nu_1}{W/\nu_2} \sim F_{\nu_1, \nu_2}$$

Confidence intervals for the ratios of two variances

- Two independent samples: $Y_{1,1}, Y_{1,2}, \dots, Y_{1,n_1} \sim iid\ N(\mu_1, \sigma_1^2)$ and $Y_{2,1}, Y_{2,2}, \dots, Y_{2,n_2} \sim iid\ N(\mu_2, \sigma_2^2)$, all mutually independent.
- How to construct a confidence interval for σ_1^2/σ_2^2?
- Again, $\frac{(n_1-1)s_1^2}{\sigma_1^2} \sim \chi^2_{n_1-1}$ and $\frac{(n_2-1)s_2^2}{\sigma_2^2} \sim \chi^2_{n_2-1}$, and they are independent.

$$\frac{\frac{\cancel{(n_2-1)} s_2^2}{\sigma_2^2} \Big/ \cancel{(n_2-1)}}{\frac{\cancel{(n_1-1)} s_1^2}{\sigma_1^2} \Big/ \cancel{(n_1-1)}} \sim F_{n_2-1, n_1-1}$$

$$\frac{s_2^2/\sigma_2^2}{s_1^2/\sigma_1^2} = \frac{s_2^2}{\sigma_2^2}\frac{\sigma_1^2}{s_1^2} = \boxed{\frac{s_2^2}{s_1^2}\frac{\sigma_1^2}{\sigma_2^2}} \sim F_{n_2-1, n_1-1}$$

P, Q, ?

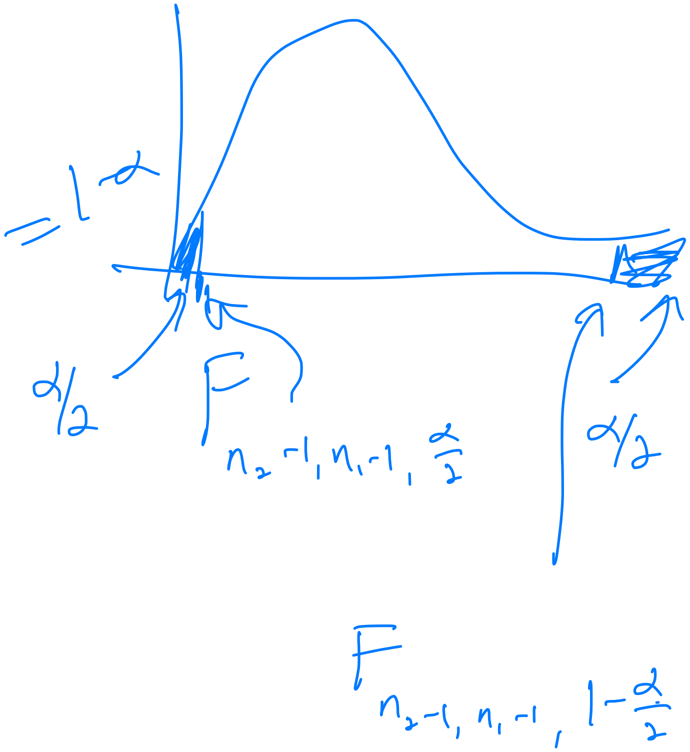
① depends on s_1^2, s_2^2, σ_1^2/σ_2^2 ✓

② dist. F_{n_2-1, n_1-1} ✓

YES

$$P\left(F_{n_2-1,\, n_1-1,\, \frac{\alpha}{2}} < \frac{S_2^2}{S_1^2}\,\frac{\sigma_1^2}{\sigma_2^2} < F_{n_2-1,\, n_1-1,\, 1-\frac{\alpha}{2}}\right) = 1-\alpha$$

$$P\left(\frac{S_1^2}{S_2^2}F_{n_2-1,\, n_1-1,\, \frac{\alpha}{2}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2}F_{n_2-1,\, n_1-1,\, 1-\frac{\alpha}{2}}\right) = 1-\alpha$$



$100(1-\alpha)\%$ c.i. for $\dfrac{\sigma_1^2}{\sigma_2^2}$

Confidence intervals for variances: Another example

- Example 11: Referring to Example 10, a separate sample was taken of 10 women, giving a mean level of 65.1 and a standard deviation of 23.4. Assuming all sample data are normally distributed, calculate a 95% confidence interval for the ratio of variances.

$n_2 = 10$ $s_2 = 23.4$ $n_1 = 12$ $s_1 = 19.4$

$F_{n_2-1,\, n_1-1,\, 0.975}$ 0.025

qf(df1 = 9, df2 = 11, 0.975)

3.588

0.025

0.2556

$$\left(\left(\frac{19.4}{23.4}\right)^2 (0.2556),\ \left(\frac{19.4}{23.4}\right)^2 3.588 \right)$$

$$\boxed{(0.1757,\ 2.466)}$$

Confidence intervals for variances: One final note

CLT

- As mentioned before, the “large sample” intervals for means hold reasonably well even if the data are not normally distributed.
- Confidence intervals for variances, however, depend more strongly on the normal distribution; these confidence interval formulas are not so “robust” to violations of the normality assumption.