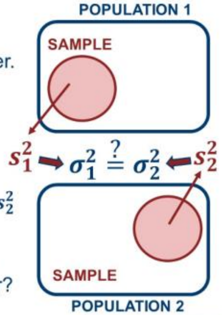
## Review: variance ratio F test

We want to know if population variances differ.
We can't measure the populations.
We take random samples.
We calculate sample variances.




$H_0$: $\sigma_1^2 = \sigma_2^2$ $H_A$: $\sigma_1^2 \neq \sigma_2^2$

Data: $s_1^2 \approx s_2^2$ Data: $s_1^2 \ll s_2^2$ or $s_1^2 \gg s_2^2$

What are the chances the pop. variances are the same (i.e., $H_0$), based on how much the sample variances differ from one another?

## Review: variance ratio F test

- Create null & alternative hypotheses:
  $H_0$: $\sigma_1^2 = \sigma_2^2$
  $H_A$: $\sigma_1^2 \neq \sigma_2^2$
  $$F_{calc} = \frac{s^2_{larger}}{s^2_{smaller}}$$
- Calculate $F_{calc}$
- Compare $F_{calc}$ to various $F_{crit}$ values (note: two-tailed test).
- Determine **p value**, of seeing $F_{calc}$ as large as we do.
- Decide to "reject $H_0$" or "fail to reject $H_0$" based on the p value.
  $H_0$: $\sigma_1^2 = \sigma_2^2$ consistent with non-small p values.
  $H_A$: $\sigma_1^2 \neq \sigma_2^2$ would give us small p values.

## Example #1 - meerkat lengths

Consider a pair of samples of meerkat lengths (cm):
Site 1: 32, 31, 30, 31, 29, 31, 28, 29, 31, 28
Site 2: 26, 23, 27, 29, 31, 35, 28, 33

Are the variances of the populations at these sites different or not?
With what degree of confidence do we make this conclusion?

First step, calculate sample variances and ratio:

$s_1^2 = 2.000$ $s_2^2 = 15.143$

$$F_{calc} = \frac{s_2^2}{s_1^2} = \frac{15.143}{2.000} = 7.571$$

$df_{num} = 8\text{-}1 = 7$
$df_{den} = 10\text{-}1 = 9$

## Example #1 - meerkat lengths $F_{calc} = 7.571$

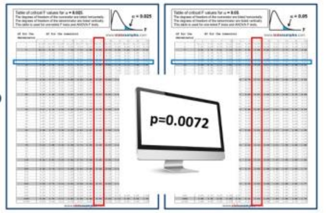
Are the variances of the populations at these sites different or not?
With what degree of confidence do we make this conclusion?

$df_{num} = 7$ $df_{den} = 9$

$F_{\alpha=0.05,7,9} = 4.20$
$F_{\alpha=0.1,7,9} = 3.29$ Remember to double $\alpha$ values

$F_{calc} = 7.571 > 4.20 = F_{\alpha=0.05,7,9}$

"The variance of lengths at site 2 is *significantly higher* than the variance of lengths at site 1 ( $p < 0.05$ )."




## Example #3 - bird parasites

Consider a pair of samples of birds with parasites:
Disturbed: 34, 29, 35, 37, 40
Preserved: 40, 38, 37, 37, 36, 38, 34, 36

Is the variance in the disturbed area larger than in the preserved area?
With what degree of confidence do we make this conclusion?

First step, calculate sample variances and ratio:

$s_D^2 = 16.500$ $s_P^2 = 3.143$

$$F_{calc} = \frac{s_D^2}{s_P^2} = \frac{16.500}{3.143} = 5.250$$

$df_{num} = 5\text{-}1 = 4$
$df_{den} = 8\text{-}1 = 7$

## One-tailed F test

- Different null & alternative hypotheses:
  $H_0$: $\sigma_1^2 \leq \sigma_2^2$
  $H_A$: $\sigma_1^2 > \sigma_2^2$
  $$F_{calc} = \frac{s_1^2}{s_2^2}$$
- Calculate $F_{calc}$
- Compare $F_{calc}$ to various $F_{crit}$ values (note: one-tailed test).
- Determine **p value**, of seeing $F_{calc}$ as extreme as we do.
- Decide to "reject $H_0$" or "fail to reject $H_0$" based on the p value.
  $H_0$: $\sigma_1^2 \leq \sigma_2^2$ consistent with non-small p values.
  $H_A$: $\sigma_1^2 > \sigma_2^2$ would give us small p values.

## Example #2 - snake scales $F_{calc} = 2.320$

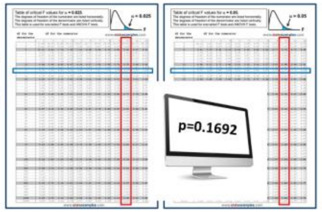
Are the variances of the populations at these sites different or not?
With what degree of confidence do we make this conclusion?

$df_{num} = 10$ $df_{den} = 12$

$F_{\alpha=0.05,10,12} = 3.37$
$F_{\alpha=0.1,10,12} = 2.75$ Remember to double $\alpha$ values

$F_{calc} = 2.320 < 2.75 = F_{\alpha=0.1,10,12}$

"The variances of the scale numbers in females and males are *not significantly different* ( $p > 0.1$ )."




## Example #2 - snake scales

Consider a pair of samples of snake scales:
Females: 21, 20, 17, 18, 20, 17, 23, 20, 14, 20, 19
Males: 16, 19, 17, 18, 16,14, 17, 15, 17, 19, 16, 18, 19

Are the scale number variances of the sexes different or not?
With what degree of confidence do we make this conclusion?

First step, calculate sample variances and ratio:

$s_F^2 = 5.800$ $s_M^2 = 2.500$

$$F_{calc} = \frac{s_F^2}{s_M^2} = \frac{5.800}{2.500} = 2.320$$

$df_{num} = 11\text{-}1 = 10$
$df_{den} = 13\text{-}1 = 12$

## Example #3 - bird parasites $F_{calc} = 5.250$

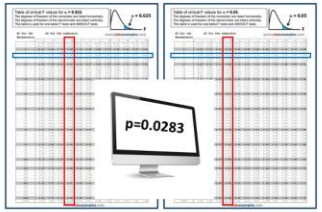
Are the variances of these populations different or not?
With what degree of confidence do we make this conclusion?

$df_{num} = 4$ $df_{den} = 7$

$F_{\alpha=0.025,4,7} = 5.52$
$F_{\alpha=0.05,4,7} = 4.12$ Do not double the $\alpha$ values

$F_{calc}$: $4.12 < 5.250 < 5.52$

"The variance of parasite number in the disturbed habitat is *significantly larger* than the variance in the preserved area (0.025<p<0.05 )."




## Example #4 - bacterial cultures

Samples of bacteria from doorknobs and keyboards:
Doorknobs: 15, 11, 13, 21, 12, 10, 7, 8, 11
Keyboards: 14, 13, 15, 17, 10, 15

Is the variance on doorknobs larger than keyboards?
With what degree of confidence do we make this conclusion?

First step, calculate sample variances and ratio:

$s_D^2 = 17.250$ $s_K^2 = 5.600$

$$F_{calc} = \frac{s_D^2}{s_K^2} = \frac{17.250}{5.600} = 3.080$$

$df_{num} = 9\text{-}1 = 8$
$df_{den} = 6\text{-}1 = 5$

## Example #4 - bacterial cultures $F_{calc} = 3.080$

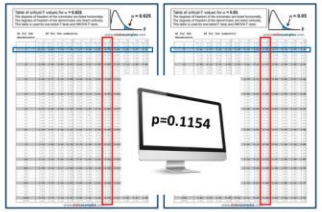
Is the variance on doorknobs larger than keyboards?
With what degree of confidence do we make this conclusion?

$df_{num} = 8$ $df_{den} = 5$

$F_{\alpha=0.025,4,7} = 6.76$
$F_{\alpha=0.05,4,7} = 4.82$ Do not double the $\alpha$ values

$F_{calc} = 3.080 < 4.82 = F_{\alpha=0.05,4,7}$

"The variance of the number of bacterial cultures on doorknobs is *not significantly larger* than the variance on keyboards ( $p > 0.05$ )."




## Caution about the F test

A strong assumption of the F test is normal population distributions.

If the populations are not normally distributed, the test can easily give type I or II errors.

It's OK as a first check, but if we really want to test for equality of variances, we should do a Levene's, Bartlett's, or Brown-Forsythe test (math is more complicated).

Note: F tests in ANOVA are OK due to Central Limit theorem.