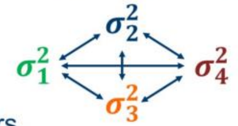
Comparing population variances

We want to know if the variances of more than two populations differ.

$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2$?

If it was just 2 populations, we could do a variance ratio F-test.



We *could* do a bunch of F-tests of all the pairs.

6 x 0.05 = 0.30

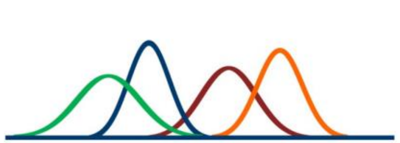
However: each test has a risk of type I error.

When we do many comparisons, the risk of type I error is too high.

We need a better test, where the overall risk of a type I error is 5%

Visualizing the test

$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2$?



Likely if variances equal
p > 0.05 for the stats test

Unlikely if variances equal
p < 0.05 for the stats test

The F_{MAX} distribution

When we choose two samples from populations with the same variances, and divide them by each other, the value follows the F distribution.

F test

$$F = \frac{s^2_{larger}}{s^2_{smaller}}$$

When we choose many samples from populations with the same variances, and divide the largest by the smallest, the value follows the F_{MAX} distribution.

F_{MAX} test

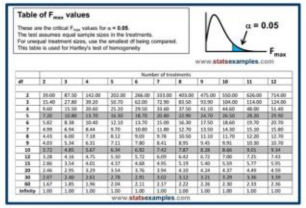
$$F_{MAX} = \frac{s^2_{largest}}{s^2_{smallest}}$$

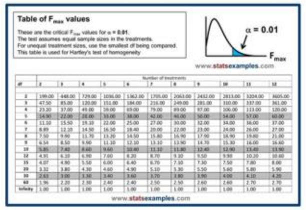
The F_{MAX} distribution

Tables show the critical values for specific alpha values.

There's a different distribution for each number of groups and df for each group.

Columns (groups) & rows (df)

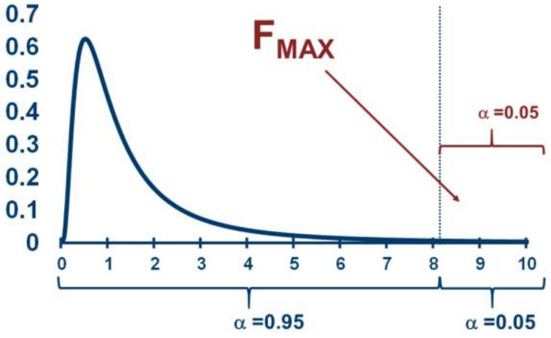




The F_{MAX} distribution

A different distribution for each number of groups and degrees of freedom for each group.

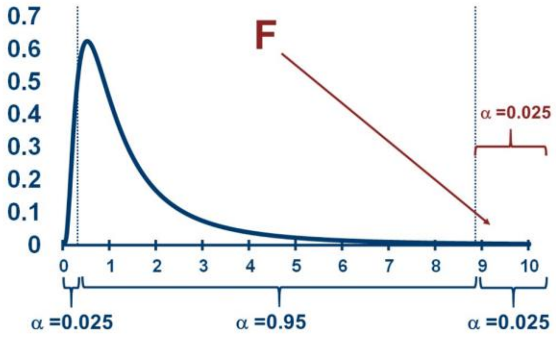
F_{MAX} distribution incorporates the many tails, use the α=0.05 area at top.



The F distribution

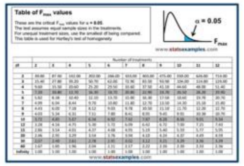
A different distribution for each pair of degrees of freedom values.

Two-tails meant using the α=0.025 area at top of the F distribution.

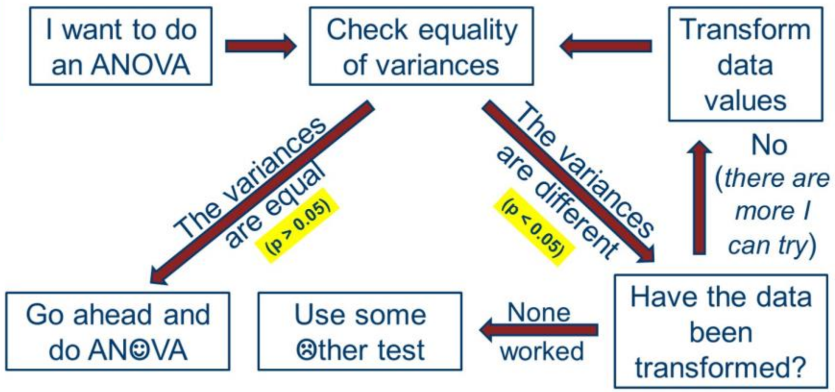


The variance ratio F test formal procedure

- Create null and alternative hypotheses:
  H_0: $\sigma_1^2 = \cdots = \sigma_{n-1}^2 = \sigma_n^2$
  H_A: $at\ least\ one\ is\ different$
- Calculate F_{MAX} value $\quad F_{MAX} = \frac{s^2_{largest}}{s^2_{smallest}}$
- Compare F_{MAX} to various F_{MAX} values (α=0.05 or α= 0.01)
- Determine probability, **p value**, of seeing F_{MAX} as large as we do.
- Decide to "reject H_0" or "fail to reject H_0" based on the p value.
  H_0: $\sigma_1^2 = \cdots = \sigma_{n-1}^2 = \sigma_n^2$ ← non-small p values.
  H_A: $at\ least\ one\ different$ ← small p values.



Columns: number of groups
Rows: df per group



Notes about the F_{MAX} test

- Also called Harley's F_{MAX} test for the homogeneity (or equality) of variances.
- F_{MAX} test requires normal population distributions (gives type I or II errors if not).
- F_{MAX} test requires equal sample sizes. If sizes are unequal, use Bartlett's, Levene's, or Browne-Forsythe test.
- F_{MAX} test is a pre-test for the ANOVA techniques which require equal variances in all compared groups.