Warsaw Summer School 2023, OSU Study Abroad Program

Mean differences

ANOVA

Inferences about Two Means with Dependent Samples

1. What are dependent samples?

- Experiments (Before – After)

- Panels (Time 1, Time 2)

2. Pairing data

We use the difference (the subtraction) of the pairs in our data

For each pair, we subtract the values of a given variable for the same units of observation.

Mean & standard deviation of d

$$\bar{d} = \frac{\sum d_i}{n}$$

$$s_d = \sqrt{\frac{\sum (d - \bar{d})^2}{n-1}}$$

Testing

- **Just as we used the sample mean and the sample standard deviation in a one-sample t-test, <u>we will use the sample mean and sample standard deviation of the differences to test for matched pairs</u>.**

- **The assumption of normality must still be verified. The differences must be normally distributed or the sample size must be large enough ($n \geq 30$).**

Six steps of testing hypotheses

1) Assume random sample

2) State Ho and H1

3) Specify the sampling distribution & the test statistic

4) Choose alpha level & specify the rejection region in picture of the sampling distribution.

5) Compute test statistic from data.

6) Make decision & interpret results.

T-test

$$t = \frac{\bar{d} - \mu_d}{(s_d / \sqrt{n})}$$

Example

$$\bar{d} = -5.125$$

$$s_d = 6.081$$

H0:μd=0 (The mean of the differences is equal to zerono difference in xenophobia over time.)

H1:μd<0 The xenophobia is decreasing.)

This is a one-sided test (less than), so alpha is all in the left tail.

For 0.05 column with 7 df the ccritical value (tα) -1.895

t = - 2.38

Example – cont.

Conclusion: **The test statistic (-2.38) is less than the critical value (-1.895). It falls in the rejection zone.**

P-value approach. The p-value estimate = 0.025 is less than the level of significance 0.05). Reject the null hypothesis.

- **Confidence interval:**

$$\bar{d} \pm t_{\frac{\alpha}{2}} \left(\frac{s_d}{\sqrt{n}}\right)$$

- .

ANOVA

- **The analysis of variance, commonly referred to by the acronym ANOVA, was developed as a strategy for dealing with comparing the mean values in different samples.**

- **It is essentially an extension of the logic of t-tests to those situations where we wish to compare the means of three or more samples concurrently.**

Variablity

The analysis of variance focuses on variability.

It involves the calculation of several measures of variability, all of which come down to one or another version of the basic measure of variability already introduced in discussion of VARIANCE -the sum of squared deviates.

Sum of deviates or sum of squares

- **For any set of N values of X_i that derive from an equal-interval scale of measurement, a deviate is the difference between an individual value of X_i and the mean (M) of the set:**

  deviate = $X_i — M_X$

  a squared deviate is the square of that quantity:

  squared deviate = $(X_i — M_X)^2$

  and the sum of squared deviates is the sum of all the squared deviates in the set:

  $SS = \sum(X_i — M_X)^2$

ANOVA

Is there a significant difference between the mean values for some specific groups? How much of the variance is explained by belonging to these groups?

Note: "groups" can refer to the values of the nominal variable in which we consider the interval variable <u>distritution</u>.

Total sum of squared variation from the mean:

- $SS(total) = \Sigma\ [X - \tilde{X}(total)]^2$

ANOVA

The between group variation represents the squared deviations of every group mean from the total mean:

- **$SS(between) = \Sigma\ [\tilde{X}(group) - \tilde{X}(total)]^2$**

The within-group sum of squares is the sum of every raw score from its group mean:

- **$SS(within) = \Sigma\ [X - \tilde{X}(group)]^2$**

Mean Squares

- **MSS(between) = SS(between) / df(between)**

  where df(between) = k – 1

- **MSS(within) = SS(within) / df(within)**

  where df(within) = N - k

F

F-statistic

$$F = \frac{MSS(between)}{MSS(within)}$$

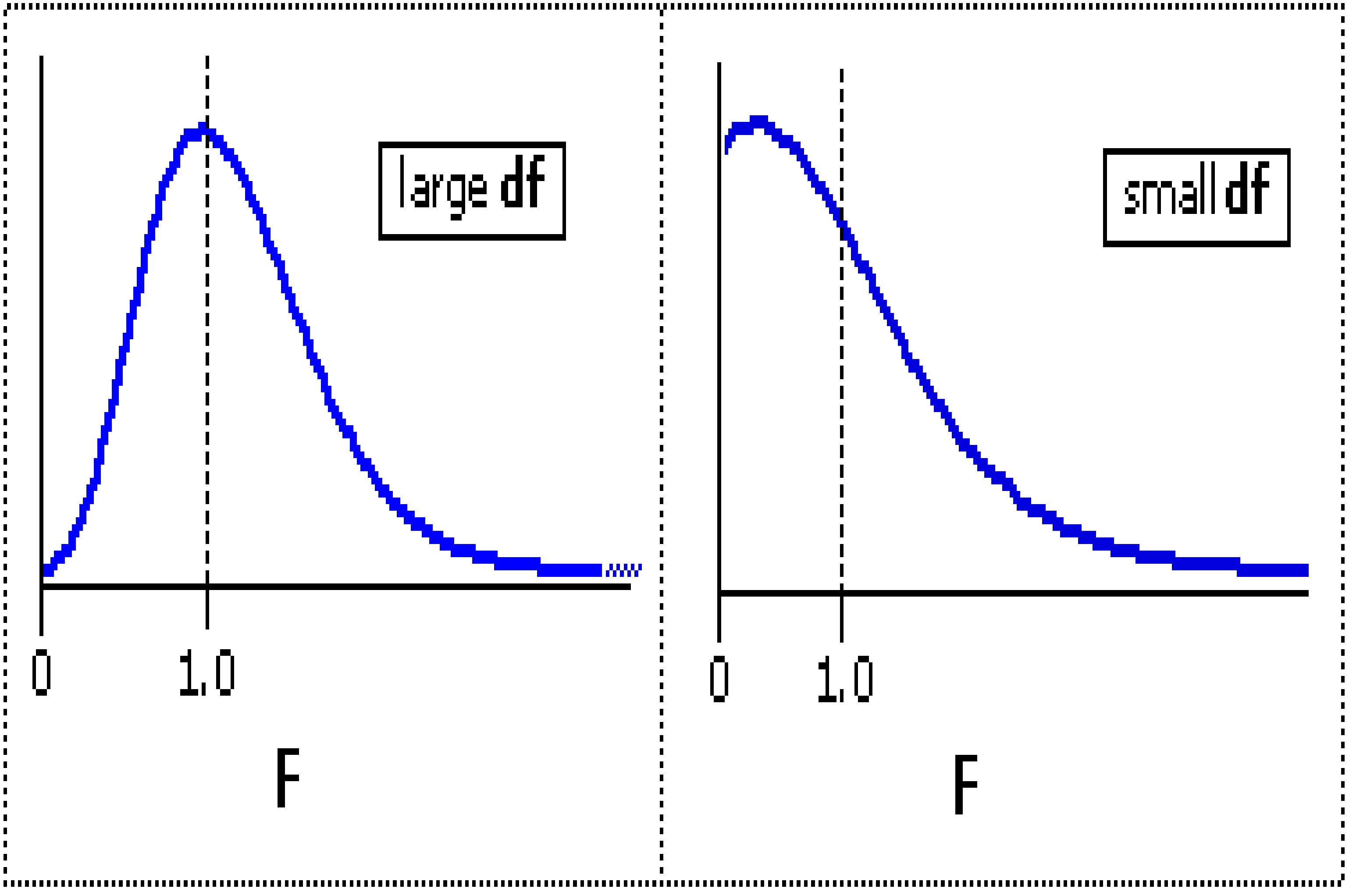
- **The larger the F-value, the greater the impact of a group on the dependent variable.**

Sampling distribution of F

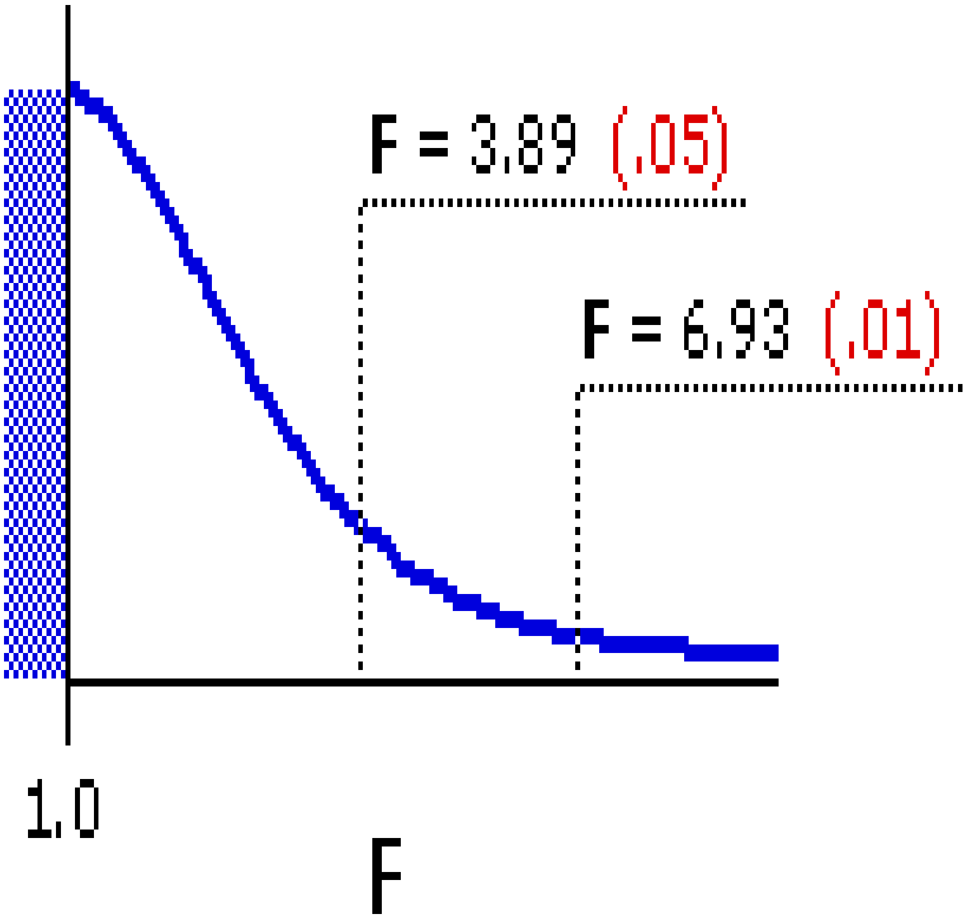
There is a separate sampling distribution of F for each possible pair of such numerator/denominator df values.

Thus, there is one sampling distribution for df=2,12, another for df=2,15, another for df=12,160, and so on.

The shapes of these various sampling distributions lie within the range of the two extreme forms


large df
0
1.0
F
small df
0
1.0
F

Sampling distribution of F for df = 2, 12



Eta coefficient

- $$F = \frac{MSS(between)}{MSS(within)}$$

- $$Eta^2 = \frac{SS(between)}{SS(total)}$$

Eta^2 = the proportion of variance in the dependent variable explained by the independent variable

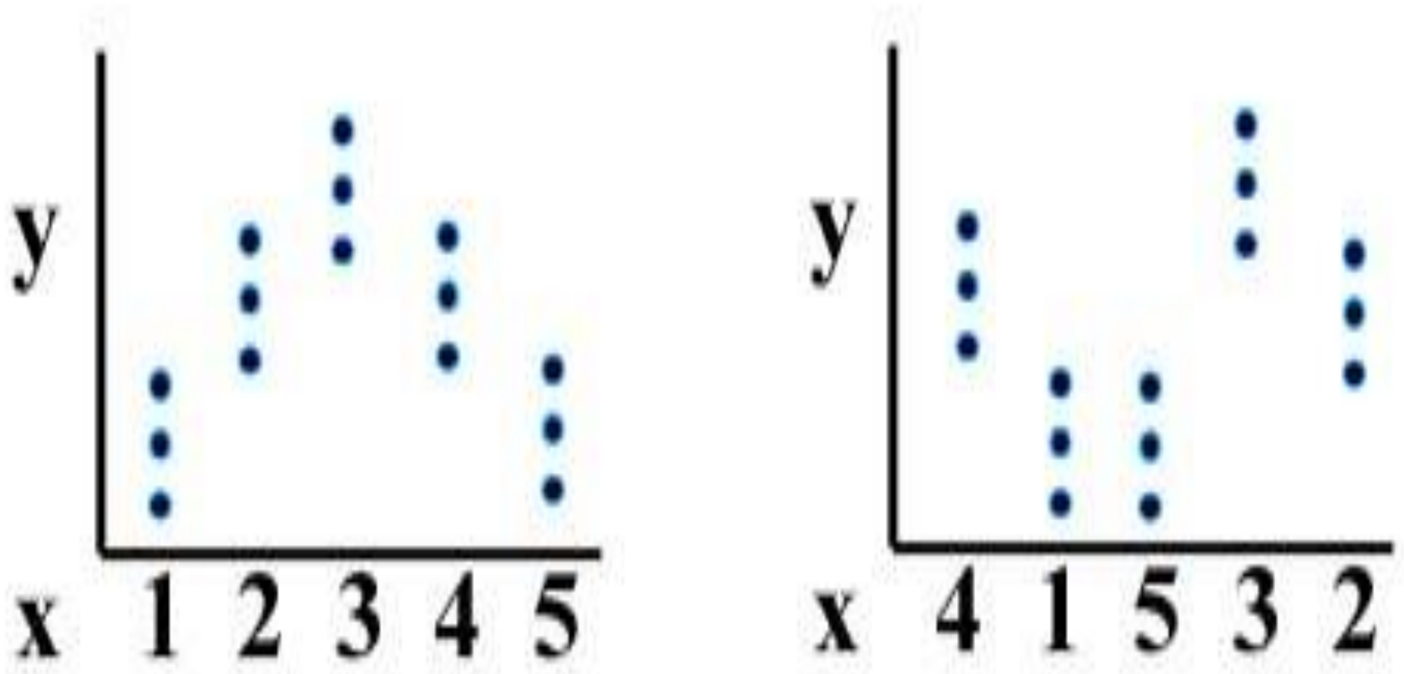
Eta

- **$Eta = \sqrt{Eta^2}$**

- **Eta is a coefficient of association, with values from 0 (no association) to 1 (perfect association)**

Eta

- **Eta is not sensitive to the order of the groups. An equally high eta would result from both patterns:**



Six steps of testing hypotheses

1) **Assume random sample**

2) **State Ho and H1**

3) **Specify the sampling distribution & the test statistic**

4) **Choose alpha level & specify the rejection region in picture of the sampling distribution.**

5) **Compute test statistic from data.**

6) **Make decision & interpret results.**

Step 1: Assumptions

Assumptions of ANOVA:

- **Independent cases**
- **Normality**
- **Equality (or "homogeneity") of variances, called homoscedacity — the variance of data in groups should be approximatly the same**

Step 2: Hypotheses

- **The null hypothesis is:**

Ho: $\mu 1 = \mu 2 = \mu 3$

This means that all the population means are the same.

The research hypothesis is: at least one of the population means is different.

Hr: $\mu 1 \neq \mu 2$ or $\mu 1 \neq \mu 3$ or $\mu 2 \neq \mu 3$

Note: research hypothesis is not Hr: $\mu 1 \neq \mu 2 \neq \mu 3$. This would be too restrictive. For large number of groups it is much easier to write out the alternate hypothesis in words, rather than symbols.

Step 3: Sampling distribution

F statistic with df = k – 1 (for between group MSS) and N - k (for within group MSS)

Step 4: Alpha level

- **Chosing $p < 0.05$ or $p < 0.01$**

Step 5: Calculating F statistics

Calculate Variance

- **a) Total variance**
- **b) Between-groups**
- **c) Within-groups**
-

Calculate Mean Squares

Calculate F-ratio

Step 6: Decision

- **Compare your computed value of F with the value from the table. If it is higher than in the table the Ho is rejected.**

F and t

The analysis of variance can also be applied to situations where you have only two groups—in particular, to situations of t-test.

The results for F-test are equivalent to what would be found in the corresponding t-test. The F-ratio obtained in such an analysis will be equal to the square of the corresponding t-ratio.

$$F = t^2$$

The only difference between t and F is that a t-test can be either directional or non-directional, whereas the ANOVA is intrinsically non-directional.