PSY 201: Statistics in Psychology

Lecture 32
Analysis of Variance
Measure twice, cut once.

Greg Francis

Purdue University

Fall 2023

ANOVA VARIABLES

- independent variables: variable that forms groupings
- one-way ANOVA: one independent variable
- levels: number of groups, number of populations
- e.g. Method of teaching is an independent variable
- you may teach in 17 different ways (levels) and have 17 different sample groups with sample means

$$\overline{X}_1, \overline{X}_2, ...\overline{X}_{16}, \overline{X}_{17},$$

- so that for your hypothesis test you would want to test whether all the population means of the different levels are the same

ANOVA VARIABLES

- we need additional subscripts to keep track of variables

$$X_{ik}$$

- is the score for the ith subject in the kth level (group)

$$n_k$$

- is the number of scores in the kth level

$$\sum_i X_{ik}$$

- is the sum of scores in the kth level

$$\sum_k \sum_i^{n_k} X_{ik}$$

- is the sum of all scores

HYPOTHESES

- for one-way ANOVA the hypotheses are

$$H_0 : \mu_1 = \mu_2 = ... = \mu_K$$

$$H_a : \mu_i \neq \mu_k \text{ for some } i, \ k$$

- the null hypothesis is that all population means are the same
- the alternative hypothesis is that at least one mean is different from another

INTUITION

- the basic approach of ANOVA is to make two calculations of variance
  1. We can calculate variance of each group separately and combine them to estimate the variance of all scores. (within variance, s_W^2)
  2. We can also calculate the variance among all the group means, relative to a grand mean. (between variance, s_B^2)
- these estimates will be the same **if** H_0 is true!
- these estimates will be different **if** H_0 is not true!

INTUITION

- we compare the estimates using the F ratio

$$F = \frac{s_B^2}{s_W^2}$$

- f $F \approx 1$, do not reject H_0
- if $F > 1$, reject H_0
- how big depends on the sample sizes, significance, ...

SCORES

- what contributes to a particular score?
- assume a linear model

$$X_{ik} = \mu + \alpha_k + e_{ik}$$

  - X_{ik} is the *i*th score in the *k*th group
  - μ is the grand mean for the population, across all groups
  - $\alpha_k = \mu_k - \mu$ is the effect of belonging to group k
  - e_{ik} is random error associated with the score
- e_{ik} changes because of random sampling (normally distributed, mean of zero, σ^2)

SUM OF SQUARES

- we want to estimate σ^2 (variance of population if H_0 is true)
- need sum of squares

$$\Sigma_k \Sigma_i (X_{ik} - \overline{X})^2$$

- consider one score

$$(X_{ik} - \overline{X}) = (X_{ik} - \overline{X}_k) + (\overline{X}_k - \overline{X})$$

- so

$$(X_{ik} - \overline{X})^2 = [(X_{ik} - \overline{X}_k) + (\overline{X}_k - \overline{X})]^2$$

- or

$$(X_{ik} - \overline{X})^2 = (X_{ik} - \overline{X}_k)^2 + 2(\overline{X}_k - \overline{X})(X_{ik} - \overline{X}_k) + (\overline{X}_k - \overline{X})^2$$

SUM OF SQUARES

- if we sum across all subjects in category k

$$\sum_{i}^{n_k}(X_{ik}-\overline{X})^2 = \sum_{i}^{n_k}(X_{ik}-\overline{X}_k)^2+2(\overline{X}_k-\overline{X})\sum_{i}^{n_k}(X_{ik}-\overline{X}_k)+\sum_{i}^{n_k}(\overline{X}_k-\overline{X})^2$$

- since deviations from a mean equal zero, this reduces to

$$\sum_{i}^{n_k}(X_{ik}-\overline{X})^2 = \sum_{i}^{n_k}(X_{ik}-\overline{X}_k)^2 + \sum_{i}^{n_k}(\overline{X}_k-\overline{X})^2$$

- in addition,

$$\sum_{i}^{n_k}(\overline{X}_k-\overline{X})^2 = n_k(\overline{X}_k-\overline{X})^2$$

- so we get

$$\sum_{i}^{n_k}(X_{ik}-\overline{X})^2 = \sum_{i}^{n_k}(X_{ik}-\overline{X}_k)^2 + n_k(\overline{X}_k-\overline{X})^2$$

SUM OF SQUARES

- now, we sum across the k groups to get the total sum of squares

$$\sum_k \sum_i (X_{ik} - \overline{X})^2 = \sum_k \left(\sum_i^{n_k} (X_{ik} - \overline{X}_k)^2 + n_k(\overline{X}_k - \overline{X})^2 \right)$$

- which becomes

$$\sum_k \sum_i (X_{ik} - \overline{X})^2 = \sum_k \sum_i^{n_k} (X_{ik} - \overline{X}_k)^2 + \sum_k n_k(\overline{X}_k - \overline{X})^2$$

- or

$$SS_T = SS_W + SS_B$$

- where
    - SS_T is the total sum of squares.
    - SS_W is the within sum of squares. Deviation of scores from the group mean.
    - SS_B is the between sum of squares. Deviation of group means from the grand mean.

WITHIN DEVIATIONS

$$SS_W = \sum_k \sum_i^{n_k} (X_{ik} - \overline{X}_k)^2$$

- what causes this to be greater than zero?
- since

$$X_{ik} = \mu + \alpha_k + e_{ik}$$

- $\mu + \alpha_k$ is fixed as i varies
- thus, deviations from $\overline{X}_k$ must be due to the e_{ik} term (random error)

ESTIMATE OF σ^2

- within each group, deviations from the mean are due to the error terms e_{ik}, so

$$s_k^2 = \frac{\sum_i (X_{ik} - \overline{X}_k)^2}{n_k - 1} \rightarrow \sigma^2$$

- to get a better estimate, pool across all groups (just like for two-sample t-test)

$$\frac{SS_W}{N - K} = MS_W \rightarrow \sigma^2$$

  - here MS_W stands for mean squares within
  - $N - K$ is the degrees of freedom

BETWEEN DEVIATIONS

$$SS_B = \sum_k n_k(\overline{X}_k - \overline{X})^2$$

- what causes this to be greater than zero?
- since

$$X_{ik} = \mu + \alpha_k + e_{ik}$$

- the mean of group k is

$$\overline{X}_k = \frac{\sum_i X_{ik}}{n_k} = \mu + \alpha_k + \frac{\sum_i e_{ik}}{n_k}$$

- as k changes, μ stays the same
- so any deviations from $\overline{X}$ are due to changes in α_k (changes between groups) or to changes in $\frac{\sum_i e_{ik}}{n_k}$ (random error)

ESTIMATE OF σ^2

- **if** H_0 is true, then all $\alpha_k = 0$ and any deviations must be due only to the random error terms ($\sum_i e_{ik}/n_k$)
- so we can again estimate σ^2 as

$$MS_B = \frac{SS_B}{K-1} = \frac{\sum_k n_k(\overline{X}_k - \overline{X})^2}{K-1} \rightarrow \sigma^2$$

  - here $K-1$ is degrees of freedom
- on the other hand, **if** H_0 is not true, then MS_B includes deviations due to α_k, so

$$MS_B > \sigma^2$$

F statistic

- so, we do not know what σ^2 is, but we have two estimates
  - MS_W: always estimates σ^2
  - MS_B: estimates σ^2 if H_0 is true. Larger than σ^2 if H_0 is false.
- compare the estimates by computing

$$F = \frac{MS_B}{MS_W}$$

- if H_0 is true, should get $F = 1$, if H_0 is not true, should get $F > 1$

F critical

- as always for inferential statistics, we need to know if F is significantly greater than 1.0
- depends on two degrees of freedom
- df numerator $= K - 1$
- df denominator $= N - K$
- look up p-value using the online F-distribution calculator

TESTING

- 4 STEPS
  1. State the hypothesis and set the criterion: $H_0 : \mu_1 = \mu_2 = ... = \mu_K$, $H_a : \mu_i \neq \mu_j$ for some i, j.
  2. Compute the test statistic $F = MS_B / MS_W$.
  3. Compute the p-value. Need to find the degrees of freedom.
  4. Make a decision.

EXAMPLE

- A college professor wants to determine the best way to present an important lecture topic to his class.
- He decides to do an experiment to evaluate three options. He solicits 27 volunteers from his class and randomly assigns 9 to each of three conditions.
- In condition 1, he lectures to the students.
- In condition 2, he lectures plus assigns supplementary reading.
- In condition 3, the students see a film on the topic plus receive the same supplementary reading as the students in condition 2.
- The students are subsequently tested on the material. The following scores (percentage correct) were obtained.

EXAMPLE

Lecture Condition 1	Lecture + Reading Condition 2	Film + Reading Condition 3
92	86	81
86	93	80
87	97	72
76	81	82
80	94	83
87	89	89
92	98	76
83	90	88
84	91	83

- No one does the calculations by hand. Always use a computer.

(1) HYPOTHESES

- for one-way ANOVA the hypotheses are

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

$$H_a : \mu_i \neq \mu_k \text{ for some } i,\ k$$

- Set $\alpha = 0.05$

(2) TEST STATISTIC

- Use the on-line calculator
- We have to format the data properly for the calculator
- One score to each line
- Indicate the level (no spaces) and then the score

Lecture	92
Lecture	86
...	
LectureReading	86
LectureReading	93
...	
FilmReading	81
FilmReading	80

- Order does not matter

(2) TEST STATISTIC

- Data could look like this when pasted into the calculator

```
Lecture	92
Lecture	86
Lecture	87
Lecture	76
Lecture	80
Lecture	87
Lecture	92
Lecture	83
Lecture	84
LectureReading	86
LectureReading	93
LectureReading	97
LectureReading	81
LectureReading	94
LectureReading	89
LectureReading	98
LectureReading	90
LectureReading	91
FilmReading	81
FilmReading	80
```

(2) TEST STATISTIC

- We read out the results of the analysis in the ANOVA summary table

Source	df	SS	MS	F	p-value
Between	2	408.0741	204.0370	7.2894	0.00336
Within	24	671.7778	27.9907		
Total	26	1079.8519			

- lots of information

(2) TEST STATISTIC

Source	df	SS	MS	F	p-value
Between	2	408.0741	204.0370	7.2894	0.00336
Within	24	671.7778	27.9907		
Total	26	1079.8519			

- We can double check things

$$F = \frac{MS_B}{MS_W} = \frac{204.0370}{27.9907} = 7.2894$$

$$MS_B = \frac{SS_B}{K-1} = \frac{408.0741}{3-1} = 204.0370$$

$$MS_W = \frac{SS_W}{N-K} = \frac{671.7778}{27-3} = 27.9907$$

(3) p VALUE

Source	df	SS	MS	F	p-value
Between	2	408.0741	204.0370	7.2894	0.00336
Within	24	671.7778	27.9907		
Total	26	1079.8519			

- between degrees of freedom (numerator)

$$df = K - 1 = 3 - 1 = 2$$

- within degrees of freedom (denominator)

$$df = N - K = 27 - 3 = 24$$

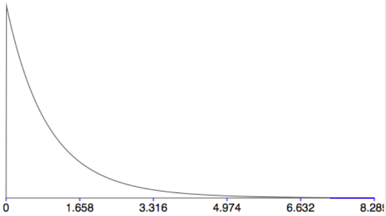
- Total degrees of freedom

$$df = N - 1 = 27 - 1 = 26$$

(3) p VALUE

Source	df	SS	MS	F	p-value
Between	2	408.0741	204.0370	7.2894	0.00336
Within	24	671.7778	27.9907		
Total	26	1079.8519			

- Check the p-value using the F distribution calculator



df numerator =	2
df denominator =	24
F =	7.2894
p =	0.00336

- Note, we just compute p from one tail, but this is equivalent to a two-tailed t-test.

(4) DECISION

- since

$$p = 0.00336 < .05 = \alpha$$

- we reject H_0. The methods of presentation are not equally effective.
- Note, does not tell us which pair of means are different!
- Look at means

Condition	Mean	Standard deviation	Sample size
Lecture	85.2222222222223	5.214829282387329	9
LectureReading	91	5.338539126015656	9
FilmReading	81.5555555555556	5.317685377847901	9

GENERALITY

- The great thing about ANOVA is that these basic steps stay the same even if you have many more means to be compared
- I happen to have data from 8 different classes that all completed an experiment where subjects responded as quickly as possible whether a set of letters formed a word or not
- The summary is the same format as above

GENERALITY

Source	df	SS	MS	F	p-value
Between	7	2324584.6485	332083.5212	6.6500	0.00000
Within	407	20324589.8142	49937.5671		
Total	414	22649174.4627			

Condition	Mean	Standard deviation	Sample size
Francis200F15	788.3333333333335	244.2585052255086	81
Francis200S16	756.0007352941174	204.17983832898088	68
Francis200F16	750.0464601769914	218.19667178177372	113
Francis200F17	756.6531914893621	214.33283856802967	94
FUSfall2018	766.1649999999998	172.00442964925605	30
Psy200Spring15	1167.3535714285715	360.9423454196428	14
FS16PSY200	776.26	224.8173218909571	10
PSY2008HKIED	849.6600000000002	191.92566073873397	5

- it would be the same format with 8000 classes!

CONCLUSIONS

- testing multiple means
- two estimates of population variance
- one estimate always estimates variance
- other estimate is true only if H_0 is true
- lets us test H_0

NEXT TIME

- interpreting ANOVA
- contrasts
- more multiple testing

Some thing versus which thing.