ACE 261
Fall 2002
Prof. Katchova

Lecture 13

Analysis of Variance and
Experimental Design

Outline

- ANOVA (Analysis of Variance)
- Experimental design
  - Completely randomized design
  - Randomized block design
  - Factorial experiments

2

Analysis of Variance

- Analysis of Variance (ANOVA) can be used to test for the equality of three or more population means.

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

  H_a: Not all population means are equal

- Rejecting H_0 means that at least two population means have different values. But we cannot conclude that all means are different.

3

ANOVA problem

- Question: Are average salaries the same or different in America, Europe, and Asia?

  H_0: $\mu_1 = \mu_2 = \mu_3$ (average salaries are equal)

  H_a: Not all salaries are equal

4

Definitions

- Response variable: *dependent variable*, the one that we would like to explain
  - salary
- Factor: *independent variable* that helps explain differences in the dependent variable
  - One factor: continent
  - Two factors: continent and education
- Treatments: specific levels of the factor
  - Three treatments: America, Europe, and Asia
  - Two treatments: Masters and Ph.D.

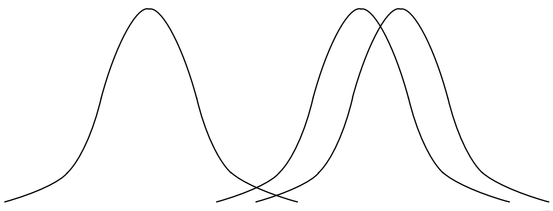
5

Salaries (in thousand dollars)

	America	Europe	Asia
	30	27	15
	34	24	20
	26	30	20
	30	27	25
Mean	30	27	20
Variance	10.67	6	16.67

6

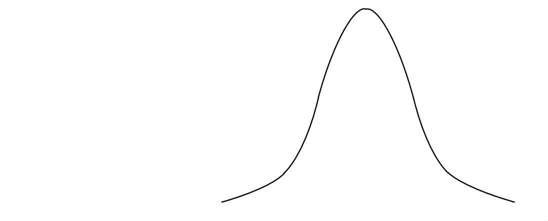
Sampling distributions of the means if the null hypothesis is false



- If the null hypothesis is false, the sample means will be very different. So the variance of the sample means will be large.

7

Sampling distributions of the means if the null hypothesis is true



- If the null hypothesis is true, the sample means will be close to each other. So we can combine the sample variances into one pooled variance, which will be the mean of the variances.

8

ANOVA calculations

- Overall mean = mean of means =(30+27+20)/3=25.67
- Between-treatments variance = variance of the means = s_{xbar}= = $[(30-25.67)^2 + (27-25.67)^2 + (20-25.67)^2]/(3-1) = 26.33$
- MSTR (Mean square due to treatments) = n*s_{xbar}= 4*26.33 =105.33
- MSE (Mean square due to error) = within-treatments variance = mean of the variances = (10.67+6+16.67)/3= 11.11

9

ANOVA calculations (continued)

- F=MSTR/MSE=105.33/11.11=9.48
- Degrees of freedom for the numerator = # treatments – 1 = k – 1 = 2
- Degrees of freedom for the denominator = # observations – # treatments = n_T – k = 12 – 3 = 9
- Critical value = $F_{0.05}$ = 4.26
- Reject H_0
- Conclusion: not all countries have the same average salary

10

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
America	4	120	30	10.66667
Europe	4	108	27	6
Asia	4	80	20	16.66667

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	210.6667	2	105.3333	9.48	0.006091	4.256492
Within Groups	100	9	11.11111			
Total	310.6667	11				

11

The ANOVA Table

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Squares	F
Treatment	SSTR	k - 1	MSTR	MSTR/MSE
Error	SSE	n_T - k	MSE	
Total	SST	n_T - 1		

SST=SSTR+SSE (total sum of squares = sum of squares due to treatments + sum of squares due to error)

F=MSTR/MSE = mean square due to treatment / mean square due to error

12

New ANOVA example: Rent for leased land in Illinois by type of lease

	Cash rent	Share rent	Cash/share rent
Mean	100	105	95
Variance	10	5	15
Number of farmers	5	5	5

13

New ANOVA example

- k = number of treatments =
- n = number of observations in a treatment =
- n_T = total number of observations =
- Overall mean =
- MSTR = n*(variance of means) =

- MSE = mean of variances =
- F=MSTR/MSE=
- $F_{critical}$ =

14

Notes

- ANOVA is about testing equalities of means. However, we use variances to tests these equalities of means.
- ANOVA with two treatments is the same as the test for differences in means in chapter 10.
- ANOVA assumes that each sample is normally distributed and has equal variances.
- F = MSTR/MSE = (#observations in a treatment)*(variance of means)/(mean of variances)

15

Multiple Comparison Procedures

- Suppose that analysis of variance has provided statistical evidence to reject the null hypothesis of equal population means.
- Fisher's least significance difference (LSD) procedure can be used to determine where the differences occur.

16

Experimental Design

- Statistical studies can be classified as being either experimental or observational.
- In an observational study, no attempt is made to control the factors.
  - Example: survey of food preferences
- In an experimental study, one or more factors are controlled so that data can be obtained about how the factors influence the variables of interest.
  - Example: conducting a food tasting experiment
- Cause-and-effect relationships are easier to establish in experimental studies than in observational studies.

17

Three types of experimental design

- Completely randomized designs
  - Assign 4 people to go to America, 4 to Europe, and 4 to Asia and observe their salaries.
  - Problem: people have different education
- Randomized block design
  - Assign the same 4 people to go first to America, then to Europe, and then to Asia and measure their salaries.
  - This design controls for different education because it's the same person (persons are the blocks)
- Factorial experiments
  - Two factors: measure not only the effect of different continents on salaries, but also the effect of education on salaries.

18

Completely randomized design (different people in different continents): the effect of continent on salary

America	Europe	Asia
30	27	15
34	24	20
26	30	20
30	27	25

19

ANOVA: single factor

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups (Treatments)	210.6667	2	105.3333	9.48	0.006091	4.256492
Within Groups (Error)	100	9	11.11111			
Total	310.6667	11				

SST=SSTR+SSE

H_0 is rejected, mean salaries are not equal for all continents

20

Randomized block design: same person in 3 continents

	America	Europe	Asia
Person 1	30	27	15
Person 2	34	24	20
Person 3	26	30	20
Person 4	30	27	25

21

ANOVA: two factor without replication

Source of Variation	SS	df	MS	F	P-value	F crit
Rows (Blocks)	17.33333	3	5.777778	0.419355	0.745885	4.757055
Columns (Treatments)	210.6667	2	105.3333	7.645161	0.022382	5.143249
Error	82.66667	6	13.77778			
Total	310.6667	11				

SST=SSBL+SSTR+SSE

Reject H_0, mean salaries are not equal for all continents

22

Two factor experiment: effect of continent and education on salary

	America	Europe	Asia
Ph.D.	30	27	15
	34	24	20
Masters	26	30	20
	30	27	25

23

ANOVA: two factor with replication

Source of Variation	SS	df	MS	F	P-value	F crit
Sample (factor A: continent)	5.333333	1	5.333333	0.64	0.45421	5.987374
Columns (Factor B: education)	210.6667	2	105.3333	12.64	0.007058	5.143249
Interaction (Factor A and B)	44.66667	2	22.33333	2.68	0.147339	5.143249
Within (Error)	50	6	8.333333			
Total	310.6667	11				

SST=SSA+SSB+SSAB+SSE

Accept H_0, mean salaries are equal for all continents

24