ANOVA (Analysis of Variance)

One way ANOVA

ANOVA: Introduction

- Many studies involve comparisons between more than two groups of subjects.
- If the outcome is numerical, ANOVA can be used to compare the means between groups.
- ANOVA is an abbreviation for the full name of the method: ANalysis Of Variance – Invented by R.A. Fisher in the 1920's

ANOVA: Introduction

- The response variable is the variable you're comparing
- The factor variable is the categorical variable being used to define the groups
  - we will assume k samples (groups)
- The one-way is because each value is classified in exactly one way
  - Examples include comparisons by gender, race, political party, color, etc.

ANOVA Assumptions

- The observations are from a random sample and they are independent from each other
- The observations are normally distributed within each group
- ANOVA is still appropriate if this assumption is not met when the sample size in each group is at least 30.
- It is not required to have equal sample sizes in all groups.
- The variances are approximately equal between groups
- If the ratio of the largest SD / smallest SD < 2, this assumption is considered to be met.

Why ANOVA instead of multiple t-tests

If you are comparing means between more than two groups, why not just do several two sample t-tests to compare the mean from one group with the mean from each of the other groups?

- Before ANOVA, this was the only option available to compare means between more than two groups.
- The problem with the multiple t-tests approach is that as the number of groups increases, the number of two sample t-tests also increases.
- As the number of tests increases the probability of making a Type I error also increases.

ANOVA: a single test for multiple comparisons

- The advantage of using ANOVA over multiple t-tests is that ANOVA will identify if any two of the group means are significantly different with a single test.
- If the significance level is set at 0.05, the probability of a Type I error for ANOVA = 0.05 regardless of the number of groups being compared.
- If the ANOVA F-test is significant, further comparisons can be done to determine which groups have significantly different means.

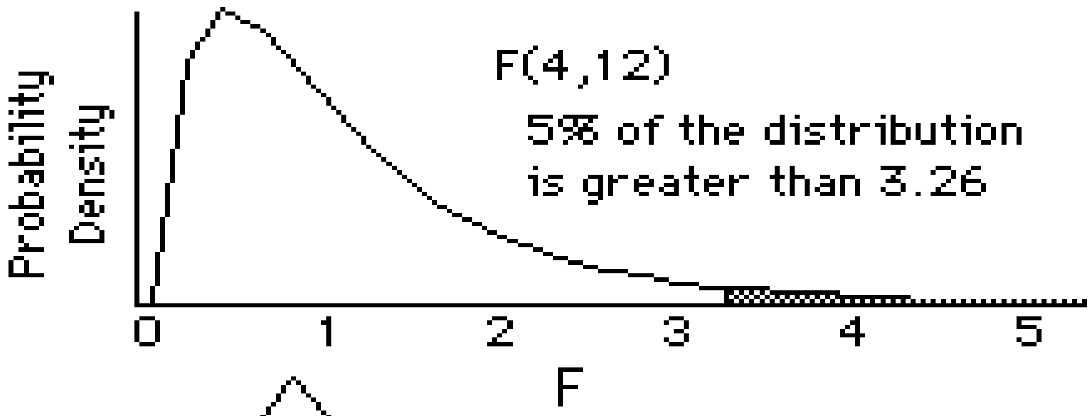
ANOVA Hypotheses

- The Null hypothesis for ANOVA is that the means for all groups are equal:

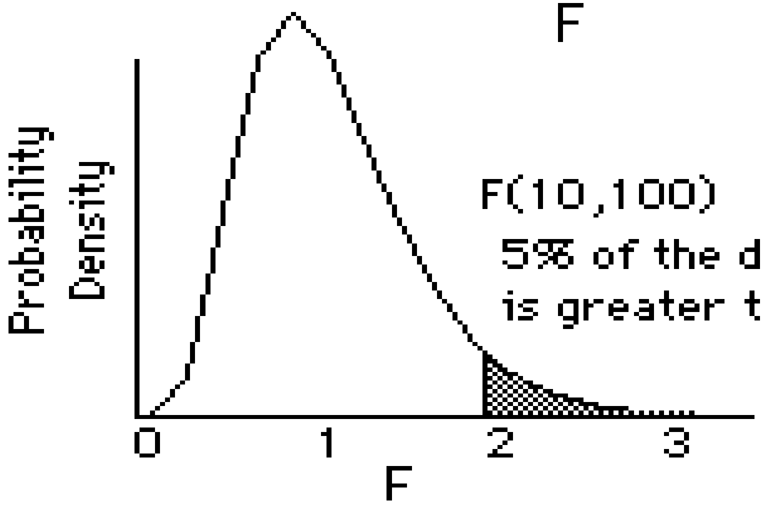
$$H_0 : \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_k$$

- The Alternative hypothesis for ANOVA is that at least two of the means are not equal.
- The test statistic for ANOVA is the ANOVA F-statistic.

F table



Probability Density
F(4,12)
5% of the distribution
is greater than 3.26
0
1
2
3
4
5
F

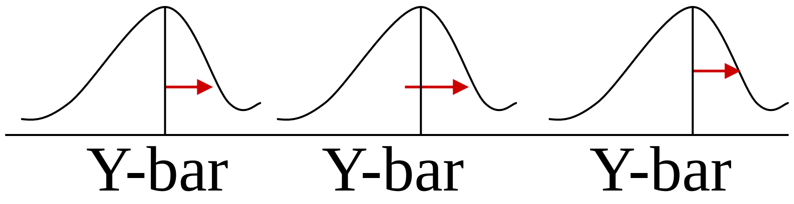


Probability Density
F(10,100)
5% of the distribution
is greater than 1.93
0
1
2
3
F

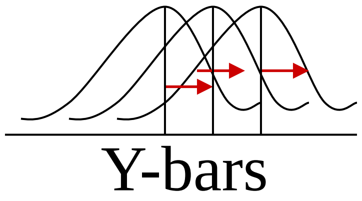
Analysis of Variance

- ANOVA is used to compare means between three or more groups, so why is it called Analysis of VARIANCE?
- The ANOVA F-test is a comparison of the average variability between groups to the average variability within groups.
- The variability within each group is a measure of the spread of the data within each of the groups.
- The variability between groups is a measure of the spread of the group means around the overall mean for all groups combined.
- F = average variability between groups average variability within groups

Analysis of Variance



Different groups, different means.



Similar groups, similar means.

ANOVA: F-statistic

- If variability between groups is large relative to the variability within groups, the F-statistic will be large.
- If variability between groups is similar or smaller than variability within groups, the F-statistic will be small.
- If the F-statistic is large enough, the null hypothesis that all means are equal is rejected.

Total variation in two parts

- The difference of each observation from the overall mean can be divided into two parts:
  - The difference between the observation and the group mean
  - The difference between the group mean and the overall mean
- ANOVA makes use of this partitioned variability.

SST = SSW + SSB

- This partitioned relationship is also true for the squared differences:

- The variability between each observation and the overall (or grand) mean is measured by the 'sum of squares total' (SST)

- The variability within groups is measured by the 'sum of squares within' (SSW). $$\sum_{i=1}^{k}(n_i - 1)s_i^2.$$

- The variability between groups is measured by the 'sum of squares between' (SSB). $$\sum_{i=1}^{k} n_i(\bar{x}_i - \bar{x})^2,$$

Mean Square Within and Mean Square Between

- The mean squares are measures of the average variability and are calculated from the sum of squares divided by the degrees of freedom.
- MSW = SSW/(N-j)
- MSW has N-j degrees of freedom where N= total number of observations and j= # groups
- MSB = SSB / (j-1)
- MSB = j-1 degrees of freedom where j= number of groups
- F-statistic = MSB / MSW

One-Way ANOVA Summary Table

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square (Variance)	F
Treatment	p - 1	SST	MST = SST/(p - 1)	$\frac{MST}{MSE}$
Error	n - p	SSE	MSE = SSE/(n - p)	
Total	n - 1	SS(Total) = SST+SSE		

ANOVA example:

mobility The hypothetical data below represent mobility scores (higher score indicates improved mobility) for 3 groups of patients: Control group did not receive any therapy Treatment group 1 received physical therapy, Treatment group 2 received counseling and physical therapy. Assume that the mobility scores are normally distributed.

Control	Treat #1	Treat #2
35	38	47
38	43	53
42	45	42
34	52	45
28	40	46
39	46	37

ANOVA

- State the Hypotheses

- Null Hypothesis: μ control = μ trt1 = μ trt2
- Alternative Hypothesis: at least two of the means (μ control , μ trt1 , μ trt2) are not equal
- ANOVA is always a two-sided test
- ANOVA will identify if at least two means are significantly different but will not identify which two (or more) means are different.

- Set the significance level – α = 0.05.

Check for Equality of Variance between groups

- Calculate the SD for each group
  - Control: 4.86
  - Trt 1: 4.94
  - Trt 2: 5.33
- Calculate the ratio:
  - largest SD / smallest SD = 5.33 / 4.86 = 1.1
  - Since the ratio < 2, assume equality of variance between groups.

Calculate Overall Mean and Group Means

- Overall mean = average of all 18 observations = 41.7
- Group means = average of the observations in each group
  - Control mean = 36
  - Treatment 1 mean = 44
  - Treatment 2 mean = 45

Calculate the Within Sum of Squares: SSW

- Square the difference between each observation and it's group mean and sum the 18 terms.
- The SSW for the control group (mean = 36)
- (35 - 36)2 + (38 – 36)2 + (42 - 36)2 + (34 – 36)2 + (28 – 36)2 + (39 – 36)2 = 118
- The SSW for the treatment 1 group (mean = 44)
- (38 - 44)2 + (43 – 44)2 + (45 - 44)2 + (52 – 44)2 + (40 – 44)2 + (46 – 44)2 = 122
- The SSW for the treatment 2 group (mean = 45)
- (47 - 45)2 + (53 – 45)2 + (42 - 45)2 + (45 – 45)2 + (46 – 45)2 + (37 – 45)2 = 142
- SSW = sum of SSW for each group = 118 + 122 + 142 = 382

Calculate the Between Sum of Squares: SSB

- The overall mean = 41.7
- The three group means are:
- Control: mean = 36
- Treatment 1: mean = 44
- Treatment 2: mean = 45
- For each group square the difference between the group mean and the overall mean and multiply by the group sample size, then sum these 3 terms for the SSB:
- SSB = 6*(36 - 41.7)2 + 6*(44 – 41.7)2 + 6*(45 - 41.7)2 = 292

Calculate MSB and MSW

- The Mean Square Between (MSB) is the average variability between groups
- MSB = SSB /(j-1) – SSB = 292, j = 3
- MSB = 292 / (3-1) = 146
- The Mean Square Within (MSW) is the average variability within groups:
- MSW = SSW /(N-j)
- SSW = 382, N = 18, j= 3
- MSW = 382 / (18 – 3) = 25.47

Calculate the ANOVA F-statistic

- The ANOVA F-statistic = MSB/MSW
- The ANOVA F-statistic will be large when there is more variability between the groups than within the groups.
- If the variability between groups and within groups is approximately equal the ANOVA F-statistic will be small (close to 1.0)
- The Null hypothesis of equal means between groups is rejected if the F-statistic is large enough.
- ANOVA F-statistic for example = 146/25.47 = 5.73

Sampling Distribution of ANOVA F-statistic

- The sampling distribution of the ANOVA F-statistic is the F-distribution
- non-negative since all F-statistics are positive.
- indexed by two degrees of freedom
- Numerator df = number of groups minus 1 (j-1)
- Denominator df = total sample size minus number of groups (N-j)
- The shape of the F-distribution varies depending on the two degrees of freedom

Find the p-value of the ANOVA F-statistic

- The p-value of the ANOVA F-statistic is the right tail area greater than the F-statistic under the F-distribution with (num. df, den. df)
- The F-statistic for the example data = 5.73
- The df for the F-distribution are (3-1) = 2 for the numerator and (18-3) = 15 for the denominator.
- the p-value = 0.014

Decision about the Hypotheses

- Since the p-value of 0.014 < the significance level of 0.05, the null hypothesis of equality between all three group means is rejected.
- We can conclude that that AT LEAST two of the means are significantly different.
- How many of the means are significantly different? Which of the means are different?
- Post-hoc tests are done to answer these questions.

ANOVA: post-hoc comparisons

- A significant ANOVA F-test is evidence that not all means are equal but it does not identify which means are significantly different.
- Methods used to find group differences after the ANOVA null hypothesis has been rejected are called post-hoc tests.
- Post-hoc is Latin for 'after-this'
- Post-hoc comparisons should only be done when the ANOVA F-test is significant.

Adjustments for Multiple Comparisons

- When multiple comparisons are being done it is customary to adjust the significance level of each individual comparison so that the overall experiment significance level remains at 0.05
- For an ANOVA with 3 groups, there are 3 combinations of t-tests.
- A conservative adjustment (Bonferroni adjustment) is to divide 0.05/3 so that alpha for each test = 0.017. Each comparison will be significant if the p-value < 0.017

Post-hoc comparison: Control and Treatment 1

- Results of the two-sample t-test to compare mean mobility scores between the control group and treatment 1 group:
  - Control mean score = 36
  - Treatment 1 mean score = 44
  - P-value for two-sample t-test = 0.0179
- This is not significant at the adjusted α-level of 0.017 but would be considered marginally significant since it is close to 0.017.
- Conclusion: After adjusting for multiple comparisons, the control group mean mobility score is marginally significantly less than the mean mobility score for the group that received physical therapy (p = 0.0179).

Post-hoc comparison: Control and Treatment 2

- A two-sample t-test is done to compare mean mobility scores between the control group and treatment 2 group:
  - Control mean score = 36
  - Treatment 2 mean score = 45
  - P-value for two-sample t-test = 0.012
- This is a significant difference at the adjusted α-level of 0.017
- Conclusion: After adjusting for multiple comparisons, the control group mean mobility score is significantly less than the mean mobility score for the group that received physical therapy and counseling ($p = 0.012$).

Post-hoc comparison: Treatment 1 and Treatment 2

- A two-sample t-test is done to compare mean mobility scores between the two treatment groups.
  - Treatment 1 mean score = 44
  - Treatment 2 mean score = 45
  - P-value for two-sample t-test = 0.743
- There is not a significant difference between the two treatment groups (at either the adjusted or un-adjusted level).
- Conclusion: There is no significant difference in mean mobility score between the two treatment groups (p = 0.743).

ANOVA summary

- ANOVA was done to evaluate differences in mean mobility score between three groups: a control group, a group that received physical therapy only and a group that received physical therapy and counseling. The significant ANOVA F- test result indicated that at least two of the mean mobility scores were significantly different. Post-hoc t-tests with adjusted α−level = 0.017 (Bonferroni adjustment for multiple comparisons) were done. Results of the post-hoc comparisons indicated a significant difference between the control group and the treatment group with both physical therapy and counseling ($p = 0.012$), a marginally significant difference between the control group and the treatment group with physical therapy only ($p = 0.0179$) and no significant difference between the two treatment groups.

Multiple Comparison Tests

- Bonferroni procedure
- Duncan Multiple range test
- Dunnett's multiple comparison test
- Newman-Keuls test
- Scheffe's test
- Tukey's test
- Holm t-test

Other ANOVA Procedures

- One-way ANOVA is Analysis of Variance for one factor
- More than one factor can be used for a two, three or four-way ANOVA
- A continuous variable can be added to the model
- this is Analysis of Covariance (ANCOVA)
- Repeated Measures ANOVA can handle replicated measurements on the same observation unit (subject)

Two way ANOVA

Factorial Design

- Experimental Units (Subjects) Are Assigned Randomly to Treatments
  - Subjects are Assumed Homogeneous
- Two or More **Factors** or Independent Variables
  - Each Has 2 or More Treatments (Levels)
- Analyzed by Two-Way ANOVA

The Two-way ANOVA

- We need to test for the *independent* and *combined* effects of multiple variables on performance. We do this with an ANOVA that asks:

  (i) how different from each other are the means for levels of Variable A?

  (ii) how different from each other are the means for levels of Variable B?

  (iii) how different from each other are the means for the treatment *combinations* produced by A and B together?

The Two-way ANOVA

- The first two of those questions are questions about *main effects* of the respective independent variables.

- The third question is about the *interaction* effect, the effect of the two variables considered simultaneously.

The Two-way ANOVA

- Main effect
  - A main effect is the effect on performance of one treatment variable considered in isolation (ignoring other variables in the study)

- Interaction
  - an interaction effect occurs when the effect of one variable is different across levels of one or more other variables

Two-way ANOVA – hypothesis test for A

H_0: No difference among means for levels of A

H_A: At least two A means differ significantly

Test statistic: $F = \dfrac{MS_A}{MS_E}$

Rej. region: $F_{obt} < F_{(2,\ 12,\ .05)} = 3.89$

Decision: Reject H_0 – variable A has an effect.

Two-way ANOVA – hypothesis test for B

H_0: No difference among means for levels of B

H_A: At least two B means differ significantly

Test statistic: $F = \dfrac{MS_B}{MS_E}$

Rej. region: $F_{obt} < F_{(1,\ 12,\ .05)} = 4.75$

Decision: Reject H_0 – variable B has an effect.

Two-way ANOVA – hypothesis for AB

H_0: A & B do not interact to affect mean response

H_A: A & B do interact to affect mean response

Test statistic: $F = \dfrac{MS_{AB}}{MS_E}$

Rej. region: $F_{obt} < F_{(2,\ 12,\ .05)} = 3.89$

Decision: Reject H_0 – A & B do interact…

Two-Way ANOVA without replication Summary Table

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F
A (Row)	a - 1	SS(A)	MS(A)	$\frac{MS(A)}{MSE}$
B (Column)	b - 1	SS(B)	MS(B)	$\frac{MS(B)}{MSE}$
Error	(a-1)(b-1)	SSE	MSE	
Total	n - 1	SS(Total)		

Two-Way ANOVA with replication Summary Table

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F
A (Row)	a - 1	SS(A)	MS(A)	$\frac{MS(A)}{MSE}$
B (Column)	b - 1	SS(B)	MS(B)	$\frac{MS(B)}{MSE}$
AB (Interaction)	(a-1)(b-1)	SS(AB)	MS(AB)	$\frac{MS(AB)}{MSE}$
Error	n - ab	SSE	MSE	
Total	n - 1	SS(Total)		