Chapter 12: Analysis of Variance

Section 12.1: One-Way ANOVA → similar to regression ↳ instead used for categorical data to predict a continuous one.

Stat 50

CONTINGENCY TABLES

Def ***One-way analysis of variance (ANOVA)*** is a method of testing the equality of three or more population means by analyzing sample variances.

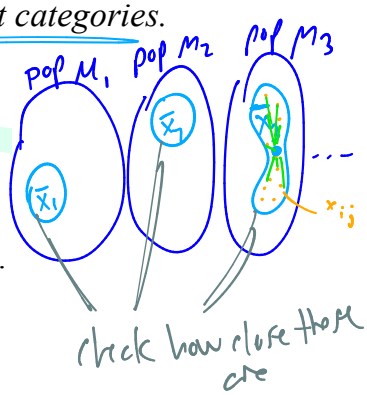
key

One-way ANOVA is used with data categorized with one factor, or treatment, so there is one characteristic used to separate the sample data into the different categories.

What are the hypotheses going to look like?

at least two are different

H_0: All of the $\mu's$ are equal H_1: Not all of the $\mu's$ are equal



Then find sample means for each population $(\bar{x}_1, \bar{x}_2, \ldots . \bar{x}_k)$

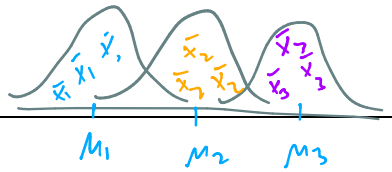
- If all of the $\bar{x}_k's$ are CLOSE together, then FAIL TO REJECT H_0.
  ↳ if H_0 is true, expect samples $\bar{x}_1 \approx \bar{x}_2 \approx \bar{x}_3 \approx \ldots$
- If all of the $\bar{x}_k's$ are FAR APART, then REJECT H_0.

Steps for Hypothesis Test for ANOVA

Check Requirements:

1. Each population must have a normal distribution.
2. Each population must have the same variance, σ^2.
3. Samples are independent, simple random samples

Step 2: Level of Significance α



Step 1: Hypotheses

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

$H_1: \ \textit{At least one of the means is different from the others}$

"Grand Mean" $\bar{x}_{GM} = \dfrac{\sum \bar{x}_i}{k}$

ALWAYS RIGHT-TAILED TEST!

What to find:

recall: $\sigma^2 = \dfrac{\sum (x_i - \bar{x})^2}{n}$

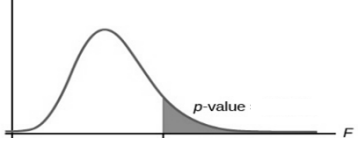
Source	df	Sum of Squares (SS)	Mean Square
Factor	$k-1$	$n\sum_{j=1}^{k}(\bar{x}_j - \bar{x})^2$ (GM)	$MS(factor) = \dfrac{SS(factor)}{k-1}$
Error	$k(n-1)$	$\sum_{j=1}^{k}\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2$ (S of SS)	$MS(error) = \dfrac{SS(error)}{k(n-1)}$

Step 3: Test Statistic

$$F_0 = \frac{MS(factor)}{MS(error)} = \frac{\text{variance between samples}}{\text{variance within samples}}$$

Step 4: Use P-Value Method

$Fcdf(F_0, UB, k-1, k(n-1))$ ↳ ignore



Reject H_0 ~ if *P-value* $\le \alpha$

Fail to Reject H_0 ~ if *P-value* $> \alpha$

Step 5: Write a CONCLUSION either rejecting or failing to reject H_0

Calculator STAT → Tests → ANOVA ANOVA(L1, L2, L3, ...)

Ex: Three groups of randomly selected students were given a different dosage of statistioprazole, a new drug that supposedly increases a student's ability to perform statistics. Each group was given a statistics pretest, then two groups were given a certain level of statistioprazole for two weeks while another group took a placebo. Then all groups took a post test (out of 10). The data is assumed to be from populations that are normally distributed with equal variances. Test the claim that the mean test scores are the same for each treatment with $\alpha = 0.05$.

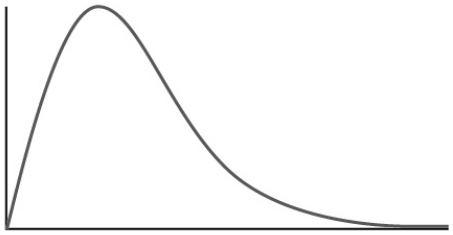
Null and Alternative Hypothesis

Placebo Group	**10 mg Group**	**20 mg Group**
6	9	8
8	6	10
5	6	9
3	8	9
$n_1 =$ $\bar{x}_1 =$ $s^2{}_1 =$	$n_2 =$ $\bar{x}_2 =$ $s^2{}_2 =$	$n_3 =$ $\bar{x}_3 =$ $s^2{}_3 =$
Variance *between* samples $ns_{\bar{x}}^2 =$		Variance *within* samples $s_p^2 =$

Test Statistic

$$F = \frac{ns_{\bar{x}}^2}{s_p^2}$$

P-value



Decision about Null Hypothesis

Conclusion

GRAPHING CALCULATOR (TI-83 OR 84)

Instructions:
(a) Input each sample into a separate list
(b) STAT $\Rightarrow$ TESTS $\Rightarrow$ H: ANOVA(
(c) Enter Lists separated by commas $\Rightarrow$ ENTER

Ex: A consumer agency randomly selected auto drivers who had similar driving records, cars, and insurance policies. The provided table gives the monthly insurance premiums (in dollars) by these drivers insured with one of four insurance companies.

n = 8

k = 4

L1
L2
L3
L4

State Farm	75	83	102	90	84	77	91	78
Geico	81	92	78	72	94	85	93	77
Farmers	70	82	75	67	91	74	72	83
AAA	78	86	84	68	88	77	65	70

At the 0.05 significance level, test the claim that the mean monthly insurance premiums are equal, assuming that the populations are normally distributed with equal variances.

[Calc] Stats → Tests → ANOVA(L1,L2,L3,L4)

Null and Alternative Hypothesis

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$

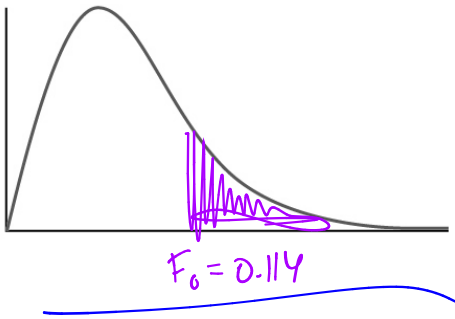
H_A: At least one monthly premium differs from the rest.

Test Statistic

$F_0 = 2.167\ldots$ $\boxed{F_0 = 2.17}$

P-value

$\boxed{p = 0.114}$



Decision about Null Hypothesis

$\alpha = 0.05$

$p = 0.114$

$p > \alpha \rightarrow$ p high, null fly

$\boxed{\text{Fail to Reject } H_0}$

Conclusion

"There is not enough statistical evidence to support the claim that the premiums (in $) differ between State Farm, Geico, Farmers, & AAA."