Chapter 13: Comparing Three or More Means

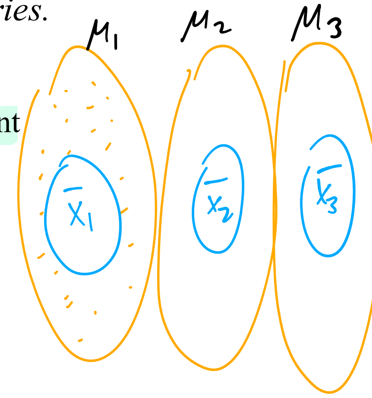
analysis of Variance

Section 13.1: Comparing Three or More Means: One-Way Analysis of Variance (ANOVA)

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

independent

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?

H_0: All of the μ's are equal H_A: At least one of the μ's is different

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.

$\bar{x}_1 \approx \bar{x}_2 \approx \bar{x}_3$

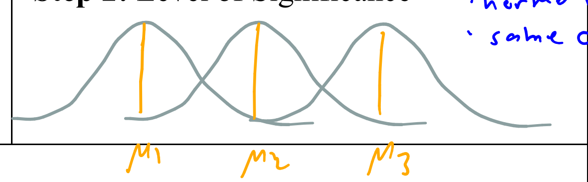
- If all of the $\bar{x}_k$'s are FAR APART, then REJECT H_0.

Steps for Hypothesis Test for ANOVA

Check Requirements:
- Each population must have a normal distribution.
- Each population must have the same variance, σ^2.
- Samples are independent, simple random samples

Step 2: Level of Significance

· normal
· same σ^2



Step 1: Hypotheses

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

H_A: At least one of the means is different from the others

ALWAYS RIGHT-TAILED TEST!

What to find:

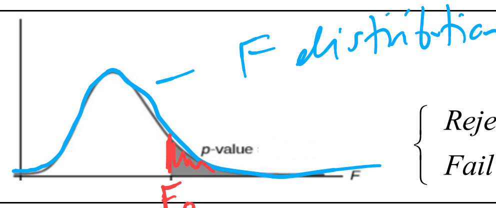
Source	df	Sum of Squares	Mean Square
Factor	$k-1$	$n\sum_{j=1}^{k}(\bar{x}_j - \bar{x})^2$	$MS(factor) = \dfrac{SS(factor)}{k-1}$
Error	$k(n-1)$	$\sum_{j=1}^{k}\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2$	$MS(error) = \dfrac{SS(error)}{k(n-1)}$

Step 3: Test Statistic

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

Step 4: Use P-Value Method

$Fcdf(F_0, UB, k-1, k(n-1))$



$$\begin{cases} \text{Reject } H_0 & \text{if } P\text{-value} \le \alpha \\ \text{Fail to Reject } H_0 & \text{if } P\text{-value} > \alpha \end{cases}$$

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

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 1: 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 $\alpha = 0.05$.

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	Variance *within* samples
$ns_{\bar{x}}^2 =$	$s_p^2 =$

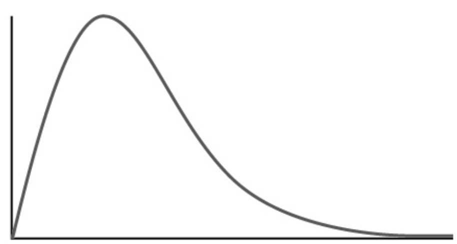
Check requirements

Null and Alternative Hypothesis

Test Statistic

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

P-value



Decision about Null Hypothesis

Conclusion

{H: ANOVA}

1-VARSTAT

Ex 2: 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.

Groups

1 μ_1	**State Farm**	75	83	102	90	84	77	91	78	$\bar{x}_1 = 85$
2 μ_2	**Geico**	81	92	78	72	94	85	93	77	$\bar{x}_2 = 84$
3 μ_3	**Farmers**	70	82	75	67	91	74	72	83	$\bar{x}_3 = 76.8$
4 μ_4	**AAA**	78	86	84	68	88	77	65	70	$\bar{x}_4 = 77$

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

Check requirements

- normal dist ✓
- equal variance ✓
- independent ✓
- SRS ✓

Null and Alternative Hypothesis

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

H_A: At least one of the means of insurance premiums is different than rest.

Test Statistic

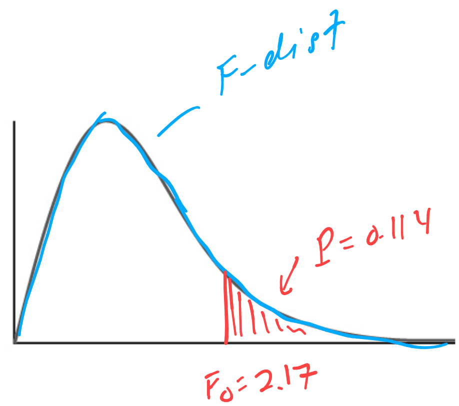
calc

$F_0 = \boxed{2.17}$

P-value

calc

$P\text{-Val} = \boxed{0.114}$



Decision about Null Hypothesis

LOS $\alpha = 0.05$

$P = 0.114$

$P > \alpha \rightarrow$ P high, Null will fly

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

Conclusion

"There is not enough statistical evidence to support the claim that at least one of the mean monthly insurance premiums is different than the others."

Remarks:

1) ProTip: Draw **box-plots** of the data! This helps you understand the
"between sample variability" vs "within sample variability"

2) *Which test do we use?*

- When we have only two groups and we are testing for means, we use the *T-Test*.
- When we have three or more groups and we are testing for means, we use *ANOVA*.

3) The frustrating thing about ANOVA is it says "at least one is different" but we don't know which!

4) For example, if we are testing: $H_0: \mu_1 = \mu_2 = \mu_3$ and we conclude that there is enough statistical evidence that H_A is true, then how do we know which is different? If we wanted to know, we'd have to do three T-Tests between two means:

$$\begin{cases} H_0: \mu_1 = \mu_2 \\ H_A: \mu_1 \neq \mu_2 \end{cases} \qquad \begin{cases} H_0: \mu_1 = \mu_3 \\ H_A: \mu_1 \neq \mu_3 \end{cases} \qquad \begin{cases} H_0: \mu_2 = \mu_3 \\ H_A: \mu_2 \neq \mu_3 \end{cases}$$