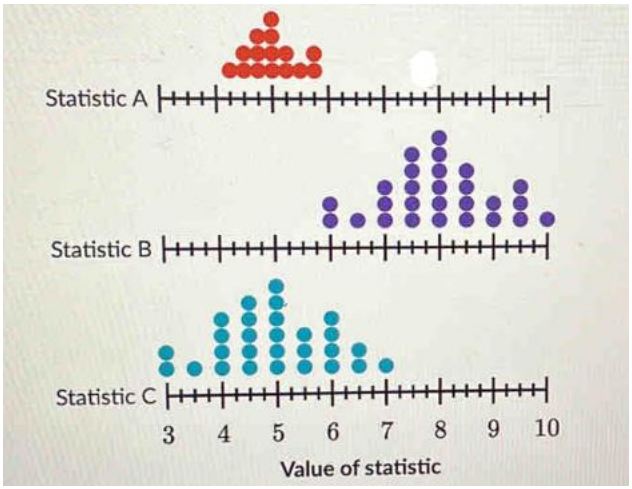
Chapter 12 Review Session 17: 4/16/25

Theorem

- The Logic of ANOVA
  - ANOVA is a test that compares individual samples to each other AND observes the variation within each individual sample.
    - When comparing individual samples to each other, we call that _____Mean Square Between_____ (MSB)
    - When observing the variation within each individual sample, we call that _____Mean Square Within_____ (MSW)



- Everything ANOVA (Table and Testing by Hand)
  - The _____response variable_____ is the interest that is measured in an experiment.
    - This is also known as the _____dependent variable_____.
  - The ____response____ is what is effected by the _____factor_____.
    - This is also known as the _____independent variable_____.
  - ____The treatments____ in an experiment are the levels in the ____category____.

	SS	Df	MS	TS	P-value
Between → Factor (Variable)	SSB	$Df_{numerator}$ or Df_B	MSB	Where between & within come together	
Within → Residual (Error)	SSW	$Df_{denominator}$ or Df_W	MSW		

  - Considerations

- There are T treatments which act as our levels
  - These treatments are under K category
- The overall sample size (n) for K are distributed for each T
- Each T also has a standard deviation & mean
  - To find the overall mean:

$$\bar{X} = \frac{\bar{x}_1 + \bar{x}_2 + \cdots + \bar{x}_I}{T}$$

- The hypotheses
  - Null: $\mu_1 = \mu_2 = \cdots = \mu_T$
  - Alternative: At least one differs from the others.

- Formulas
  - The Sum of Squares
    - Between (SSB)

$$SSB = \sum (n_i (\bar{x}_i - \bar{X})^2)$$

    - Within (SSW)

$$SSW = \sum ((n_i - 1) s_i^2)$$

  - The degrees of freedom
    - Between (Df_B)

$$Df_B = T - 1$$

    - Within (Df_w)

$$Df_w = n - T$$

  - The Mean Squares
    - Between (MSB)

$$MSB = \frac{SSB}{Df_B}$$

    - Within (MSW)

$$MSW = \frac{SSW}{Df_w}$$

- Test Statistic

$$F = \frac{MSB}{MSW}$$

- P-value

$p = Fcdf(TS, E99, Df_B, Df_W)$

- Total Sum or Squares (May or may not be asked to find this)

↳ $TSS = SSB + SSW$

- Total Degrees of Freedom (May or may not be asked to find this)

↳ $Total_{DF} = Df_B + Df_W$

- Decision & Interpretation
  - Rejecting ($\alpha > p$)

At α % level of significance, there is sufficient evidence to support that at least one of the treatment means differ.

  - Failing to Reject ($\alpha < p$)

At α % level of significance, there is insufficient evidence to support that at least one of the treatment means differ.

- ANOVA F-Test on the Calculator
  - You will need the data/list for each treatment
    - Input into the calculator

Stat --> 1: Edit --> Put in data for each treatment into individual lists

  - Then you go to the ANOVA function

Stat --> Tests --> A: ANOVA(--> List each treatment list with a comma between and close the parenthesis

    - Output Terms to know
      - F is the TS
      - P is the p-value
      - Factor stats (df, SS, and MS) is the between values
      - Error stats (df, SS, and MS) is the within values

- Ignore Sxp!!!

- Post-Hoc Analysis
  - Only used when you Reject Null in the ANOVA test!
  - This analysis looks at determining which treatment is different from the others, if possible for the particular scenario.
  - This class focuses on the Tukey HSD method and the Bonferroni method.
    - We use Tukey HSD for equal sample sizes between treatments.
    - We use Bonferroni for equal or unequal sample sizes between treatments.
  - Both methods are pairwise comparisons.
    - I.e. If there are 3 treatments Post-Hoc looks at:

(μ_1, μ_2) (μ_1, μ_3) (μ_2, μ_3)

  - This means the hypotheses look like:
    - Null: $\mu_1 = \mu_2$, $\mu_1 = \mu_3$, $\mu_2 = \mu_3$
    - Alternative: At least one mean is different.
  - For Post-Hoc, you will be given an R table with each treatment comparison written out. You will not be expected to calculate for each comparison!
  - Decision & Interpretation (Overall)
    - Rejecting ($\alpha > p$)
      - Means that $\mu \neq \mu$
    - Failing to Reject ($\alpha < p$)
      - Means that $\mu = \mu$

For each pairwise comparison

    - Thus, we say:

At α % level of significance, we could say that treatment(s) (diff μ) are different from treatment(s) (other μ).

Practice

1. Ella, a beauty CEO, is performing an experiment comparing the time it takes (in minutes) for different face mask formulas to dry to determine which face mask is best for quick use. Use the table below to answer the questions, keeping in mind that we are testing against a 5% significance level.

L1 → 1-Var Stat

Formula A		Formula B		Formula C	
2.35	$\bar{x}=2.44$	3.58	$\bar{x}=3.32$	1.25	$\bar{x}=1.99$
1.55	$s=0.56$	3.05	$s=0.43$	1.90	$s=0.69$
2.05	$n=8$	2.98	$n=6$	2	$n=7$
3		4		3.05	
2.88		2.89		1.5	
1.90		3.43		2.80	
3.05				1.45	
2.75					

a. Identify the Factor, Levels, and response variables.

Face Masks — Formulas A, B, & C — dry time

$\bar{x}_{overall} = \frac{2.44+3.32+1.99}{3} = 2.58$

b. State the ANOVA Hypotheses.

$H_0: \mu_A = \mu_B = \mu_C$ H_1: At least one μ differs.

c. Fill in the ANOVA table.

	SS	Df	MS	TS	p-value
Factor (Variable)	5.88	2	2.94	8.91	0.0020
Residual (Error)	5.98	18	0.33		

$SSB = (8(2.44-2.58)^2)+(6(3.32-2.58)^2)+(7(1.99-2.58)^2) = 5.88$

$SSW = ((8-1)0.56^2)+((6-1)0.43^2)+((7-1)0.69^2) = 5.98$

$Df_B = 3-1 = 2$ $Df_W = (8+6+7)-3 = 18$

$MSB = \frac{5.88}{2} = 2.941$ $MSW = \frac{5.98}{18} = 0.33$

$TS = \frac{2.94}{0.33} = 8.91$ $p = fcdf(8.91, E99, 2, 18) = 0.0020$

d. What is the decision?

$0.05 > 0.0020 \rightarrow$ Reject H_0

e. Choose the appropriate interpretation.

☒ At 5% level of significance, there is sufficient evidence to support that at least one of the treatment means differ. ← H_1

☐ At 5% level of significance, there is insufficient evidence to support that at least one of the treatment means differ.

f. Would we move onto Post-Hoc? If yes, continue.

Yes

g. Write the Post-Hoc Hypotheses.

H_0: $\mu_A = \mu_B$ H_1: $\mu_A \neq \mu_B$

$\mu_A = \mu_C$ $\mu_A \neq \mu_C$

$\mu_B = \mu_C$ $\mu_B \neq \mu_C$

h. Analyze the R output below.

Confidence Interval; what we need: α

H_0 μ pairs		diff	lwr.ci	upr.ci	pval	
A=B	:Formula B - Formula A	1.08	-1.455499	3.6154988	0.8149	> 0.05 Fail (A=B)
A=C	:Formula C - Formula A	-1.42	-3.955499	1.1154988	0.4616	> 0.05 Fail (A=C)
B=C	:Formula C - Formula B	-2.50	-4.814583	-0.1854169	0.0322	< 0.05 Reject (B≠C)

i. Interpret the scenario.

At 5% level of significance, we can say that while the means for mask A is not different from masks B and C, mask B is different from mask C.

2. Steven, a pharmacist, is trying to determine which ADHD medication is most effective out of the 4 most common types. To do so, the pharmacist asks 7 people for each medication how effective they would rate it. Use the statistics below to answer the questions, keeping in mind that we are testing against a 1% significance level.

Medication	Mean Score	Standard Deviation	Sample Size
#1	5.43	1.72	7
#2	4.14	1.35	7
#3	6.29	1.11	7
#4	6	1.41	7

a. Identify the Factor, Levels, and response variables.

ADHD Meds; Meds 1,2,3,&4; effectiveness

$\bar{x}_{overall} = \frac{5.43+4.14+6.29+6}{4} = 5.47$

b. State the ANOVA Hypotheses.

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$ H_1: At least one μ differs.

c. Fill in the ANOVA table.

	SS	Df	MS	TS	p-value
Factor (Variable)	19.07	3	6.36	0.93	0.4749

Residual (Error)	48.01	7	6.86	

$SSB = (7(5.43-5.47)^2)+(7(4.14-5.47)^2)+(7(6.29-5.47)^2)+(7(6-5.47)^2) = 19.07$

$SSW = ((7-1)1.72^2)+((7-1)1.35^2)+((7-1)1.11^2)+((7-1)1.41^2) = 48.01$

$Df_B = 4-1 = 3$

$MSB = \frac{19.07}{3} = 6.36$

$MSW = \frac{48.01}{7} = 6.86$

$Df_W = (7\times4)-4 = 7$

$TS = \frac{6.36}{6.86} = 0.93$

$p = fcdf(0.93, E99, 3, 7) = 0.4749$

d. What is the decision?

$0.01 < 0.4749 \rightarrow$ Fail to Reject

e. Choose the appropriate interpretation.

❑ At 1% level of significance, there is sufficient evidence to support that at least one of the treatment means differ.

☒ At 1% level of significance, there is insufficient evidence to support that at least one of the treatment means differ. ← H_1

f. Would we move onto Post-Hoc? If yes, continue.

No

g. Write the Post-Hoc Hypotheses.

h. Analyze the R output below.

	diff	lwr	upr	p adj
2-1	0.36250000	0.12528287	0.59971713	0.0010358
3-1	0.07833333	-0.15888380	0.31555047	0.8143113
4-1	0.22000000	-0.01721713	0.45721713	0.0778376
3-2	-0.28416667	-0.52138380	-0.04694953	0.0131752
4-2	-0.14250000	-0.37971713	0.09471713	0.3869986
4-3	0.14166667	-0.09555047	0.37888380	0.3921830

i. Interpret the scenario.