STAT 500

ANOVA Table - Inference for Multiple Means

ANOVA: Cell Means Model

$$Y_{ij} = \mu_i + \epsilon_{ij}$$

The variation in Y_{ij} comes from two sources:

- **Population Means** - $\mu_1, \mu_2, \ldots, \mu_r$
- **Random errors** - $\epsilon_{ij} = Y_{ij} - \mu_i$
- Research Question - Do the populations or treatment groups have the same mean values for the variable?
  - Yes - variation in Y_{ij} comes from random errors
  - No - variation in Y_{ij} comes from random errors and population means

ANOVA: Analysis of Variance

- Random errors and population means are unknown.
- Study sample means and residuals.
  - Sample Means: $\bar{Y}_{1\cdot}, \bar{Y}_{2\cdot}, \ldots, \bar{Y}_{r\cdot}$
  - Residuals: $e_{ij} = Y_{ij} - \bar{Y}_{i\cdot}$ for all i, j

ANOVA: Analysis of Variance

- Research Question - Do the populations or treatment groups have the same mean values for the variable?
  - Yes
    - Sample Means will vary, but should be similar in value.
    - Most variation in Y_{ij} will be in residual e_{ij}.
  - No
    - Sample Means will vary, differences will reflect differences in population means.
    - Will still have variation in Y_{ij} from residual e_{ij}.

ANOVA: Analysis of Variance

- Calculate three variations based on observations $\boldsymbol{Y_{ij}}$
  - Variation due to group means
  - Variation due to residuals
  - Total Variation

ANOVA: Analysis of Variance

- Variation due to group means:

$$SS_{\text{among groups}} = \sum_{i=1}^{r} \sum_{j=1}^{n_i} (\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2 = \sum_{i=1}^{r} n_i (\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2$$

- Also called SS_{model}

- If the population means are the same (different), this value should be small (large).

ANOVA: Analysis of Variance

- Variation due to residuals:

$$
\begin{aligned}
SS_{\text{within groups}} &= \sum_{i=1}^{r} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i\cdot})^2 \\
&= \sum_{i=1}^{r} (n_i - 1) S_i^2 \\
&= \sum_{i=1}^{r} \sum_{j=1}^{n_i} e_{ij}^2
\end{aligned}
$$

- also called SS_{error} or $SS_{\text{residuals}}$

ANOVA: Analysis of Variance

- Total variation - (Corrected) total sum of squares:

$$SS_{\text{total}} = \sum_{i=1}^{r} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{\cdot\cdot})^2$$

ANOVA: Analysis of Variance

Key result:

$$SS_{\text{total}} = SS_{\text{model}} + SS_{\text{error}}$$

$$\begin{aligned}
\sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{..})^2 &= \sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{i.} + \bar{Y}_{i.} - \bar{Y}_{..})^2 \\
&= \sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{i.})^2 + \sum_{i=1}^{r}\sum_{j=1}^{n_i}(\bar{Y}_{i.} - \bar{Y}_{..})^2 \\
&\quad +2\sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{i.})(\bar{Y}_{i.} - \bar{Y}_{..}) \\
&= \sum_{i=1}^{r} n_i(\bar{Y}_{i.} - \bar{Y}_{..})^2 + \sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{i.})^2
\end{aligned}$$

ANOVA Table

source of variation	degrees of freedom	sums of squares	mean square	F
Model	$r-1$	SS_{model}	$MS_{\text{model}} = \frac{SS_{\text{model}}}{(r-1)}$	$\frac{MS_{\text{model}}}{MS_{\text{error}}}$
Error	$N-r$	SS_{error}	$MS_{\text{error}} = \frac{SS_{\text{error}}}{(N-r)}$	
Total	$N-1$	SS_{total}		

Note: $MS_{\text{error}} = S_p^2$

Model Assumptions

- Assumptions on random error terms
  - ϵ_{ij} are i.i.d. from a normal distribution with mean $\mathbf{0}$ and variance $\boldsymbol{\sigma}^2$.
  - $\boldsymbol{\epsilon}$ is multivariate normal with mean $\mathbf{0}$ and variance $\boldsymbol{\sigma}^2\boldsymbol{I}$.
- This implies that
  - $\boldsymbol{Y}_{ij}$ are i.i.d. from a normal distribution with mean $\boldsymbol{\mu}_i$ and variance $\boldsymbol{\sigma}^2$.
  - $\mathbf{Y}$ is multivariate normal with mean $\boldsymbol{X}\boldsymbol{\beta}$ and variance $\boldsymbol{\sigma}^2\boldsymbol{I}$.
- In addition, we assume samples are independent of each other.

Results: Mean Squares

- $E(MS_{\text{error}}) = E(S_p^2) = \sigma^2$

- $E(MS_{\text{model}}) = \sigma^2 + \frac{1}{r-1} \Sigma_{i=1}^{r} n_i(\mu_i - \bar{\mu})^2$

  where $\bar{\mu} = \frac{1}{N} \Sigma_i n_i \mu_i$

- MS_{error} and MS_{model} are independent

- $\frac{E\left(MS_{\text{model}}\right)}{E(MS_{\text{error}})} = \frac{\sigma^2 + \frac{1}{r-1} \Sigma_{i=1}^{r} n_i(\mu_i - \bar{\mu})^2}{\sigma^2}$

Hypothesis Test

- $H_o : \mu_1 = \mu_2 = \cdots = \mu_r$
- $H_a :$ at least one μ_i is different $i = 1, \ldots, r$
- Test Statistic:

$$F = \frac{MS_{\text{model}}}{MS_{\text{error}}}$$

- Large values of F provide evidence against the null hypothesis.

Definition of Central F Distribution

- Let W_1 has a χ^2 distribution with ν_1 degrees of freedom.
- Let W_2 has a χ^2 distribution with ν_2 degrees of freedom.
- Assume W_1 and W_2 are independent.

$$F = \frac{W_1/\nu_1}{W_2/\nu_2}$$

has a central F distribution with ν_1 numerator and ν_2 denominator degrees of freedom.

Distribution of Test Statistic

Under model assumptions and $H_o : \mu_1 = \mu_2 = \cdots = \mu_r$

- $(N - r)MS_{\text{error}}/\sigma^2 \sim \chi^2_{N-r}$
- $(r - 1)MS_{\text{model}}/\sigma^2 \sim \chi^2_{r-1}$
- MS_{error} and MS_{model} are independent.

$$F = \frac{((r-1)MS_{\text{model}}/\sigma^2)/(r-1)}{((N-r)MS_{\text{error}}/\sigma^2)/(N-r)} = \frac{MS_{\text{model}}}{MS_{\text{error}}}$$

has a central F-distribution with $r - 1$ numerator and $N - r$ denominator degrees of freedom.

Hypothesis Test

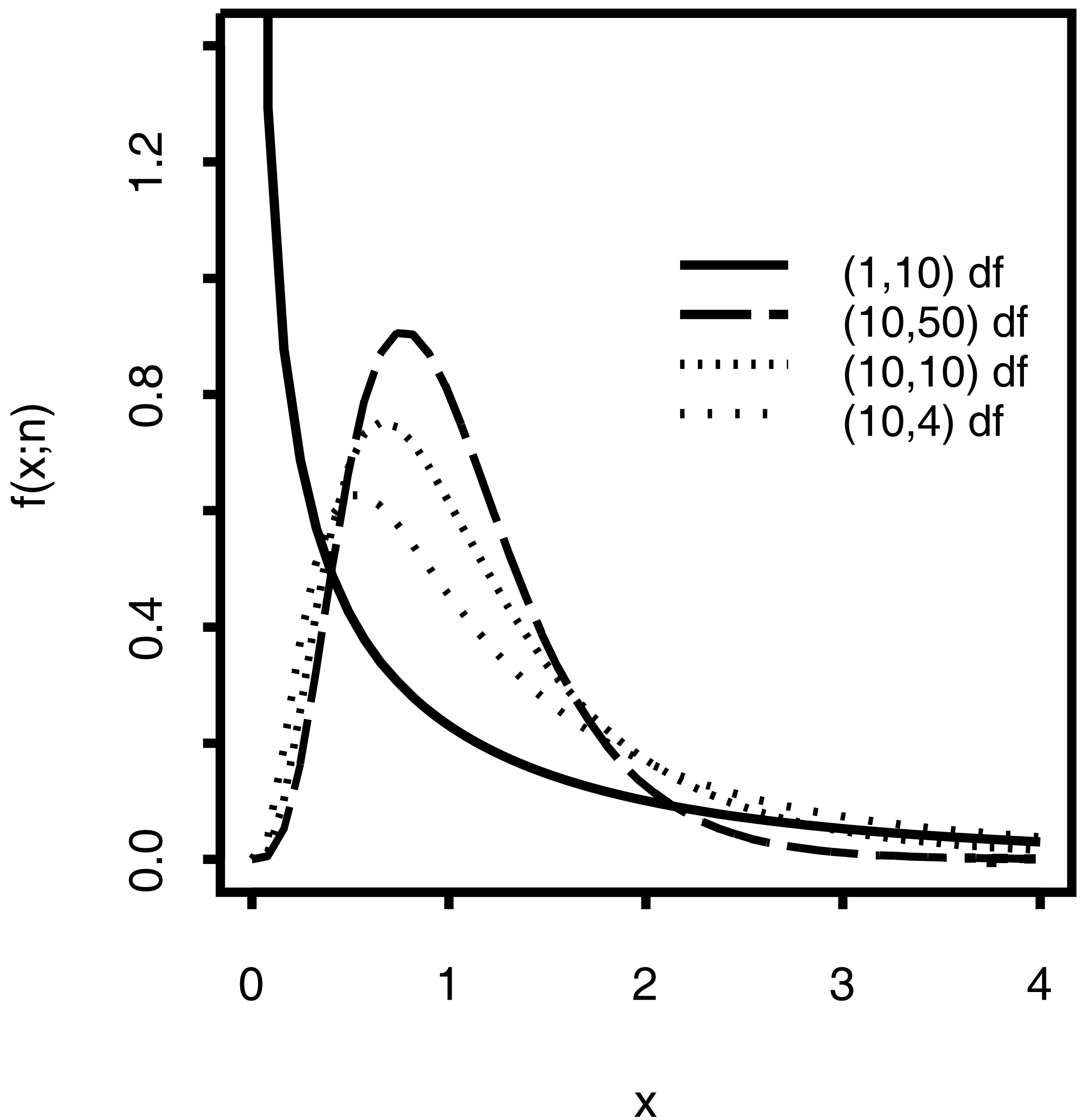
- $H_o : \mu_1 = \mu_2 = \cdots = \mu_r$
- H_a : at least one μ_i is different $i = 1, \ldots, r$
- Test Statistic:

$$F = \frac{MS_{\text{model}}}{MS_{\text{error}}}$$

- P-value:

$$P(F_{r-1,N-r} > F)$$

Densities for Central F Distributions



Doughnut Example (S&C pp 218-219)

- An experiment was conducted to determine the best type of oil for frying donuts.
  - For anything fried, the goal is to use the oil to heat the food, but not have the food absorb the oil.

- Treatments: Four types of oil for cooking donuts

- Experimental units: batches of donuts

- Randomization: Assignment of batches to cooking oils

- Measured response:

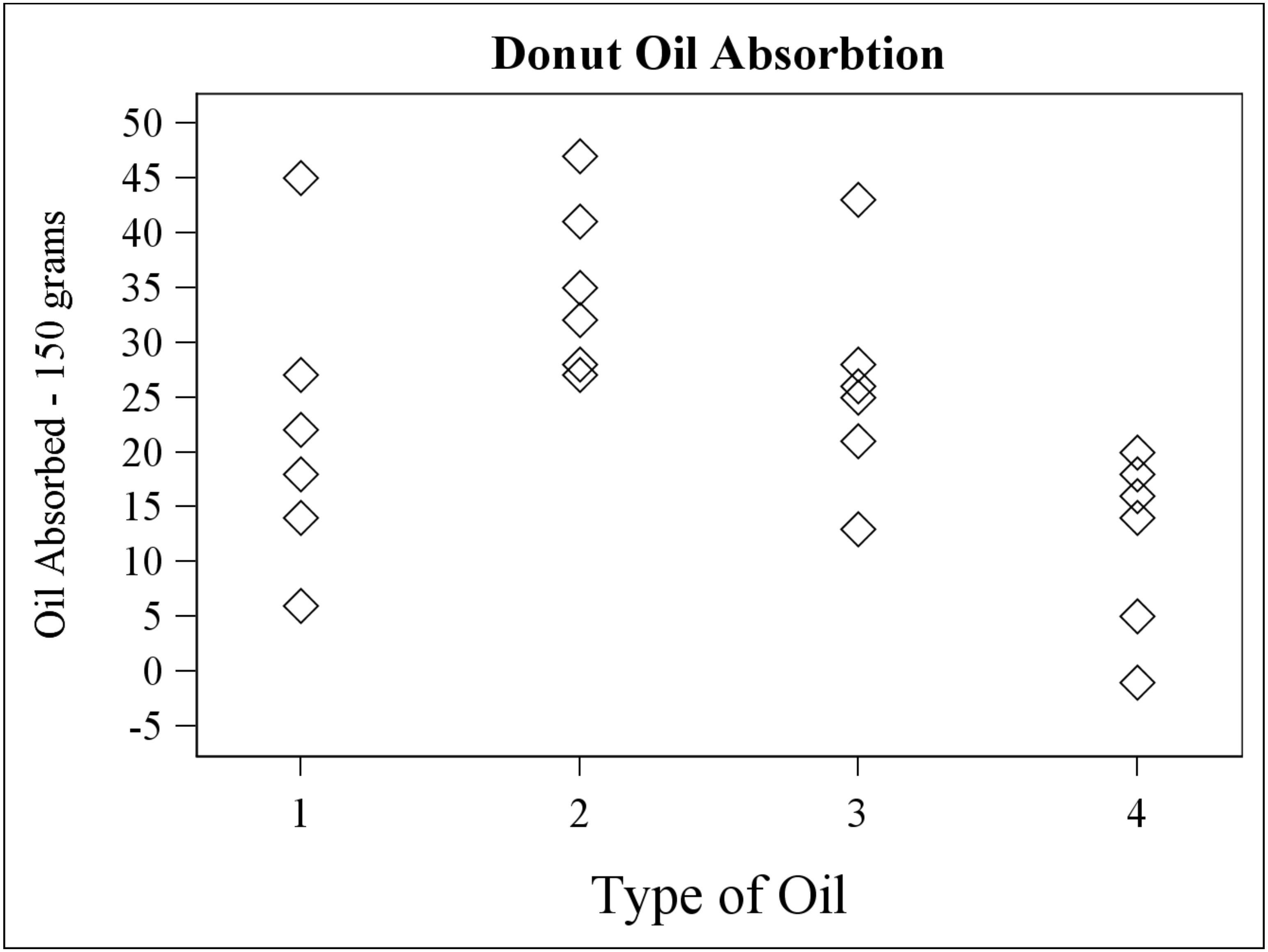
  $\boldsymbol{Y_{ij}}$ = (grams of oil absorbed when one batch is cooked) -150 grams

Donut Example

The amount of oil absorbed (-150gram) in cooking 24 batches of donuts is given below.

Grams of oil absorbed (-150g) for
different types of cooking oil

Type 1	Type 2	Type 3	Type 4
14	28	25	5
22	41	43	16
18	47	28	-1
27	32	21	14
6	35	13	20
45	27	26	18


Donut Oil Absorbtion
Oil Absorbed - 150 grams
50
45
40
35
30
25
20
15
10
5
0
-5
1
2
3
4
Type of Oil

Donut Example

Summary Statistics

Oil	n_i	$\bar{Y}_{i\cdot}$	S_i^2
Type 1	6	22	178.0
Type 2	6	35	60.4
Type 3	6	26	97.6
Type 4	6	12	67.6

ANOVA table: Total Sums of Squares

The table below provides $\boldsymbol{Y_{ij} - \bar{Y}_{..}}$ where $\boldsymbol{\bar{Y}_{..} = 23.75}$.

Type 1	Type 2	Type 3	Type 4
-9.75	4.25	1.25	-18.75
-1.75	17.25	19.25	-7.75
-5.75	23.25	4.25	-24.75
3.25	8.25	-2.75	-9.75
-17.75	11.25	-10.75	-3.75
21.25	3.25	2.25	-5.75

To find $\boldsymbol{SS_{Total}}$, we square all values in the table and sum them to get 3654.5.
There are 23 degrees of freedom for this SS since we are calculating it by subtracting 24 observation values from the overall mean value.

ANOVA table: Model Sums of Squares

The table below provides $\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot}$

Type 1	Type 2	Type 3	Type 4
-1.75	11.25	2.25	-11.75
-1.75	11.25	2.25	-11.75
-1.75	11.25	2.25	-11.75
-1.75	11.25	2.25	-11.75
-1.75	11.25	2.25	-11.75
-1.75	11.25	2.25	-11.75

To find SS_{Model}, we square all values in the table and sum them to get 1636.5.
There are 3 degrees of freedom for this SS since we are calculating it by subtracting the 4 group means from the overall mean.

ANOVA table: Error Sums of Squares

The table below provides $Y_{ij} - \bar{Y}_{i\cdot}$

Type 1	Type 2	Type 3	Type 4
-8	-7	-1	-7
0	6	17	4
-4	12	2	-13
5	-3	-5	2
-16	0	-13	8
23	-8	0	6

To find SS_{Error}, we square all values in the table and sum them to get 2018.
There are $4 * (6 - 1) = 20$ degrees of freedom for this SS since since it is calculated by subtracting the 6 observations in each of the 4 groups from the corresponding group mean.

ANOVA Table: Donut Example

ANOVA table

source of variation	degrees of freedom	sums of squares	mean square	F
Model	**3**	**1636.5**	**545.5**	**5.41**
Error	**20**	**2018.0**	**100.9**	
total	**23**	**3654.5**		

$$\boldsymbol{F = 5.41 > F_{(3,20),\ .99} = 4.94}$$

From computer output: $\boldsymbol{p}$-value=0.0069

The average amount of absorbed oil is not the same for all four types of oil.