# KUEHL - CHAPTER 2 (Completely Randomized Design)

## (2.1) ASSEMBLING THE RESEARCH DESIGN

- RESEARCH HYPOTHESIS (STATEMENT REGARDING RELATIONSHIP BETWEEN TREATMENTS (CONDITIONS) AND OUTCOMES (RESPONSES)
- TREATMENT DESIGN - IDENTIFICATION OF SPECIFIC TREATMENTS (CONDITIONS) TO INCLUDE IN EXPERIMENT/STUDY. INCLUDES CONTROL GROUP(S) WHEN APPOPRIATE.
- EXPERIMENTAL DESIGN - SELECTION OF EXPERIMENTAL UNITS, REPLICATE (SAMPLE) SIZES, ASSIGNMENT OF TRTS TO UNITS.
- OBSERVATIONAL STUDY DESIGN - Choice of Sampling Units to be observed in Study.

## (2.2) Randomization

Once experimental units have been selected for controlled experiment, treatments are assigned @ random to units. Can choose names out of hat, utilize random number tables, or computer random generators.

EXAMPLE $t=3$ TRTS, $r=4$ reps/TRT => $n=12$ EXP. UNITS

| UNIT | RANDOM 3 DIGIT # (Table XII, P.624) | RANK | TRT |
|---|---|---|---|
| 1 | 400 | 8 | B |
| 2 | 040 | 3 | A |
| 3 | 277 | 5 | B |
| 4 | 030 | 1 | A |
| 5 | 037 | 2 | A |
| 6 | 328 | 6 | B |
| 7 | 137 | 4 | A |
| 8 | 909 | 12 | C |
| 9 | 623 | 10 | C |
| 10 | 490 | 9 | C |
| 11 | 362 | 7 | B |
| 12 | 689 | 11 | C |

1. 3 Digits used in place of 2 to avoid "tie" which would occur with the 2 03's (4,5)
2. RANK THE RANDOM DIGITS FROM SMALLEST TO LARGEST
3. ASSIGN UNITS w/ 4 Smallest RANDOM DIGITS TO TRT A, NEXT 4 TO B, LARGEST 4 TO C.

TRT: A → UNIT: 4, 5, 2, 7

TRT: B → UNIT: 3, 6, 11, 1

TRT: C → UNIT: 10, 9, 12, 8

OBSERVATIONAL STUDIES – Probability Samples should be taken from the existing populations in a similar manner.

## 2.3 PREPARATION OF DATA FILES FOR THE ANALYSIS

- EXPERIMENTAL (OBSERVATIONAL) UNITS NEED TO BE IDENTIFIED BY:
  1. ID: (Number, NAME, SSN, or any other code)
  2. TREATMENT (LABEL, or NAME or NUMBER)
  3. RESPONSE (NUMERIC)

### SAS CODE TO CREATE DATA SET

```
DATA ONE;       /* NAMES DATA SET "ONE" */
INPUT SUBJECT  TRT $  Y;  /* 3 VARIABLES:
                             SUBECT (NUMERIC)
                             TRT (CHARACTER)
                             Y (NUMERIC) */
CARDS;   /* DATA LINES BEGIN */
4   A   Y11
5   A   Y12
2   A   Y13
7   A   Y14
3   B   Y21
6   B   Y22
11  B   Y23
1   B   Y24
10  C   Y31
9   C   Y32
12  C   Y33
8   C   Y34
;              /* END OF DATA */.
```

NOTE: Y's will be actual numbers. $Y_{ij} \equiv j^{th}$ rep for TRT $i$

# 2.4 STATISTICAL MODEL FOR CRD

## ① CELL (TREATMENT) MEANS MODEL

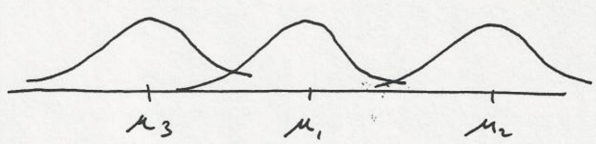
$$y_{ij} = \mu_i + e_{ij} \qquad \underbrace{i = 1, \ldots, t}_{\text{TREATMENTS}} \qquad \underbrace{j = 1, \ldots, r}_{\text{REPLICATES}}$$

- $y_{ij} \equiv j^{th}$ obs. from $i^{th}$ TRT. GROUP
- $\mu_i \equiv$ Mean for the $i^{th}$ TRT POPULATION
- $e_{ij} \equiv$ EXPERIMENTAL ERROR

$E[e_{ij}] = 0 \qquad V[e_{ij}] = \sigma^2 \qquad (\Rightarrow \sigma_1^2 = \cdots = \sigma_t^2)$

Inferences are based on further assumptions of <u>independence</u> and <u>normality</u> of errors.

$e_{ij} \sim NID(0, \sigma^2)$

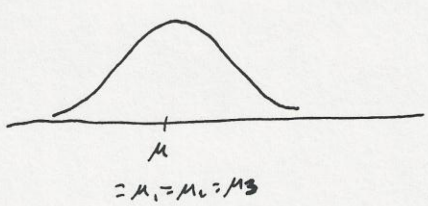



ONE POSSIBLE RESEARCH HYPOTHESIS

## MODEL UNDER HYPOTHESIS OF NO TRT DIFFERENCES

$$y_{ij} = \mu + e_{ij} \qquad (\mu = \mu_1 = \cdots = \mu_t)$$

$e_{ij} \sim NID(0, \sigma^2)$



## GENERAL LINEAR MODEL

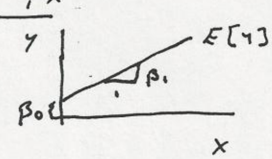
Response ~~or~~ Dependent Variable $\equiv Y$

Predictor or Explanatory or Independent Variables $\equiv X_1, \ldots, X_K$

$$Y = \underbrace{\beta_0 + \beta_1 X_1 + \cdots + \beta_K X_K}_{\text{Systematic relation between } Y \text{ \& } X_1,\ldots,X_K} + \underbrace{e}_{\text{Experimental Error}}$$

### Linear relation between $E[Y]$ & $X$

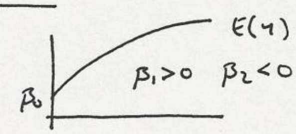
$$Y = \beta_0 + \beta_1 X + e$$




(next semester)

### QUADRATIC RELATION BETWEEN $E[Y]$ & $X$

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + e$$




(next semester)

### Indicator Variables to model categorical variables (TRTS)

$$X_1 = \begin{cases} 1 & \text{if TRT A} \\ 0 & \text{o.w.} \end{cases}$$

$$X_2 = \begin{cases} 1 & \text{if TRT B} \\ 0. & \text{o.w.} \end{cases} \qquad \beta_0 = 0$$

$$X_3 = \begin{cases} 1 & \text{if TRT C} \\ 0 & \text{ow} \end{cases}$$

$$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + e$$

TRT A: $Y = \beta_1(1) + \beta_2(0) + \beta_3(0) + e = \beta_1 + e$ $\quad (\beta_1 = \mu_1)$

TRT B: $Y = \beta_1(0) + \beta_2(1) + \beta_3(0) + e = \beta_2 + e$ $\quad (\beta_2 = \mu_2)$

TRT C: $Y = \beta_1(0) + \beta_2(0) + \beta_3(1) + e = \beta_3 + e$ $\quad (\beta_3 = \mu_3)$

## Model with Treatments and Covariate

$$y_{ij} = \mu_i + \beta x_{ij} + e_{ij}$$

Response depends on treatment and level of covariate of the experimental unit.

(will cover this next semester)

## 2.5 Least Squares Estimates of Model Parameters

Full Model: $y_{ij} = \mu_i + e_{ij}$ $\quad i = 1, \dots, t$, $\; j = 1, \dots, r$

### L.S. Estimates of $\{\mu_i\}$

$e_{ij} = y_{ij} - \mu_i$ $\quad i = 1, \dots, t$, $\; j = 1, \dots, r$

Choose $\{\hat{\mu}_i\}$ as values of $\{\mu_i\}$ that minimize $Q = \sum_i \sum_j e_{ij}^2$

$$Q = \sum_i \sum_j (e_{ij}^2) = \sum_i \sum_j (y_{ij} - \mu_i)^2$$

$$\frac{\partial Q}{\partial \mu_1} = \sum_j (y_{1j} - \mu_1)(2)(-1) + \sum_{i=2}^{t} \sum_{j=1}^{r} 0 = 0$$

$$\Rightarrow -2 \sum_j (y_{1j} - \hat{\mu}_1) = 0$$

$$\Rightarrow \sum_{j=1}^{r} y_{1j} - r\hat{\mu}_1 = 0$$

$$\Rightarrow \sum_{j=1}^{r} y_{1j} = r\hat{\mu}_1 \Rightarrow \boxed{\hat{\mu}_1 = \frac{\sum_{j=1}^{r} y_{1j}}{r} = \bar{y}_{1\cdot}}$$

By similar argument: $\hat{\mu}_i = \bar{y}_{i\cdot}$ $\quad i = 1, \dots, t$

## Dot Notation

$$y_{i.} = \sum_{j=1}^{r} y_{ij} \qquad (\text{DOT} \Rightarrow \text{Summing over subscript})$$

$$\bar{y}_{i.} = \frac{\sum y_{ij}}{r} \qquad (\text{DOT w/ BAR} \Rightarrow \text{MEAN OVER SUBSCRIPT})$$

## Error Sum of Squares for Full Model

$$SSE_f = \sum_i \sum_j (y_{ij} - \hat{\mu}_i)^2 = \sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2 = \sum_i \sum_j \hat{e}_{ij}^2$$

NOTE: $s_i^2 = \dfrac{\sum_{j=1}^{r}(y_{ij} - \bar{y}_{i.})^2}{r-1} \Rightarrow \sum_{j=1}^{r}(y_{ij} - \bar{y}_{i.})^2 = (r-1)s_i^2$

$$\Rightarrow SSE_f = \sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2 = (r-1)\sum_{i=1}^{t} s_i^2$$

Assuming $\sigma_1^2 = \dots = \sigma_t^2 = \sigma^2$:

$$s^2 = \frac{(r-1)\sum_{i=1}^{t} s_i^2}{t(r-1)} = \frac{\sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2}{t(r-1)} = \frac{SSE_f}{t(r-1)}$$

is an unbiased estimate of $\sigma^2$

REDUCED MODEL: $y_{ij} = \mu + e_{ij} \Rightarrow e_{ij} = y_{ij} - \mu$

$$Q = \sum_i \sum_j (e_{ij})^2 = \sum_i \sum_j (y_{ij} - \mu)^2$$

Choose $\hat{\mu}$ to minimize Q.

$$\frac{\partial Q}{\partial \hat{\mu}} = 2\sum_i \sum_j (y_{ij} - \mu)(-1) = 0$$

$$\Rightarrow \sum_i \sum_j y_{ij} - rt\hat{\mu} = 0 \Rightarrow \boxed{\hat{\mu} = \frac{\sum\sum y_{ij}}{rt} = \frac{y_{..}}{rt} = \bar{y}_{..}}$$

## Error Sum of Squares for Reduced Model

$$SSE_r = \sum_i \sum_j (y_{ij} - \hat{\mu})^2 = \sum_i \sum_j (y_{ij} - \bar{y}_{..})^2$$

## 2.6 Sums of Squares for Important Sources of Variation

Kuehl's Data from Chapter 2 $(t=4,\ r=3)$

TRT 1: COMMERCIAL: $y_{1.} = 7.66 + 6.98 + 7.80 = 22.44$ $\quad \bar{y}_{1.} = 7.48 = \hat{\mu}_1$

TRT 2: VACUUM: $y_{2.} = 5.26 + 5.44 + 5.80 = 16.50$ $\quad \bar{y}_{2.} = 5.50 = \hat{\mu}_2$

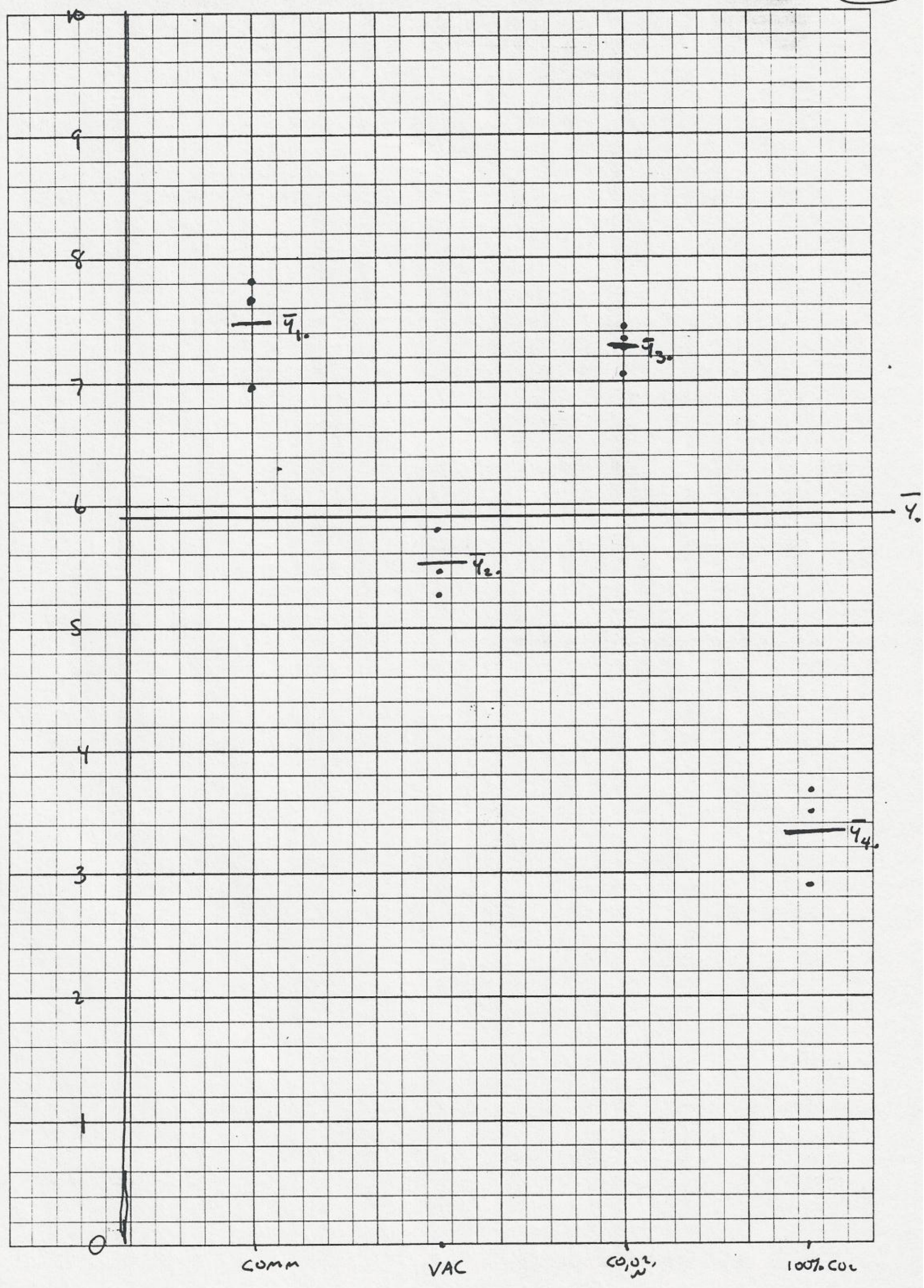
TRT 3: $CO, O_2, N$: $y_{3.} = 7.41 + 7.33 + 7.04 = 21.78$ $\quad \bar{y}_{3.} = 7.26 = \hat{\mu}_3$

TRT 4: 100% $CO_2$ $\quad y_{4.} = 3.51 + 2.91 + 3.66 = 10.08$ $\quad \bar{y}_{4.} = 3.36 = \hat{\mu}_4$

$$\hat{\mu} = \frac{y_{1.} + y_{2.} + y_{3.} + y_{4.}}{tr} = \frac{22.44 + 16.50 + 21.78 + 10.08}{4(3)}$$

$$= \frac{70.80}{12} = 5.90$$

| | | Full model $y_{ij} = \mu_i + e_{ij}$ | | | Reduced $y_{ij} = \mu + e_{ij}$ | | |
|---|---|---|---|---|---|---|---|
| TRT | $y_{ij}$ | $\hat{\mu}_i$ | $y_{ij} - \hat{\mu}_i$ | $(y_{ij} - \hat{\mu}_i)^2$ | $\hat{\mu}$ | $y_{ij} - \hat{\mu}$ | $(y_{ij} - \hat{\mu})^2$ |
| Commercial (i=1) | 7.66 | 7.48 | 0.18 | .0324 | 5.90 | 1.76 | 3.0976 |
| | 6.98 | 7.48 | −0.50 | .2500 | 5.90 | 1.08 | 1.1664 |
| | 7.80 | 7.48 | 0.32 | .1024 | 5.90 | 1.90 | 3.6100 |
| Vacuum (i=2) | 5.26 | 5.50 | −0.24 | .0576 | 5.90 | −0.64 | 0.4096 |
| | 5.44 | 5.50 | −0.06 | .0036 | 5.90 | −0.46 | 0.2116 |
| | 5.80 | 5.50 | 0.30 | .0900 | 5.90 | −0.10 | 0.0100 |
| $CO, O_2, N$ (i=3) | 7.41 | 7.26 | 0.15 | .0225 | 5.90 | 1.51 | 2.2801 |
| | 7.33 | 7.26 | 0.07 | .0049 | 5.90 | 1.43 | 2.0449 |
| | 7.04 | 7.26 | −0.22 | .0484 | 5.90 | 1.14 | 1.2996 |
| 100% $CO_2$ (i=4) | 3.51 | 3.36 | 0.15 | .0225 | 5.90 | −2.39 | 5.7121 |
| | 2.91 | 3.36 | −0.45 | .2025 | 5.90 | −2.99 | 8.9401 |
| | 3.66 | 3.36 | 0.30 | .0900 | 5.90 | −2.24 | 5.0176 |
| | | — | 0 | $SSE_F = 0.9268$ | — | 0 | $SSE_R = 33.7996$ |


10
9
8
7
6
5
4
3
2
1
0
$\bar{Y}_{1.}$
$\bar{Y}_{3.}$
$\bar{Y}_{..}$
$\bar{Y}_{2.}$
$\bar{Y}_{4.}$
COMM
VAC
$CO, O_2, N$
100% $CO_2$

Reduced Model: $SSE_r = \sum_i\sum_j (y_{ij} - \hat{\mu})^2 = \sum_i\sum_j (y_{ij} - \bar{y}_{..})^2$

FULL MODEL: $SSE_f = \sum_i\sum_j (y_{ij} - \hat{\mu}_i)^2 = \sum_i\sum_j (y_{ij} - \bar{y}_{i.})^2$

DIFFERENCE: $SSE_r - SSE_f = \sum_i\sum_j (y_{ij} - \bar{y}_{..})^2 - \sum_i\sum_j (y_{ij} - \bar{y}_{i.})^2$

$$= \sum_i\sum_j (\bar{y}_{i.} - \bar{y}_{..})^2 \quad (*) \text{ Proof Below}$$

$$= r \sum_i (\bar{y}_{i.} - \bar{y}_{..})^2$$

$\boxed{\text{For Kuehl's data: } r\sum(\bar{y}_{i.} - \bar{y}_{..})^2 = 3\left[(7.48-5.90)^2 + (5.50-5.90)^2 + (7.26-5.90)^2 + (3.36-5.90)^2\right] = 32.87}$

See graph

$SSE_r \equiv$ TOTAL (CORRECTED) SUM OF SQUARES

$SSE_f \equiv$ Error Sum of Squares (when Trts are in model)

$SSE_r - SSE_f \equiv$ TREATMENT SUM OF SQUARES

$SSE_f$ is also known as WITHIN TRT SS

$SSE_r - SSE_f$ " " " BETWEEN " "

SS TOTAL = SS TREATMENT + SS ERROR

= SS ERROR + SS TREATMENTS

$\boxed{\text{Proof}}$

$$y_{ij} - \bar{y}_{..} = (y_{ij} - \bar{y}_{i.}) + (\bar{y}_{i.} - \bar{y}_{..})$$

$$\sum_i\sum_j (y_{ij} - \bar{y}_{..})^2 = \sum_i\sum_j \left[(y_{ij} - \bar{y}_{i.}) + (\bar{y}_{i.} - \bar{y}_{..})\right]^2$$

$$= \underbrace{\sum_i\sum_j (y_{ij} - \bar{y}_{i.})^2}_{①} + \underbrace{\sum_i\sum_j (\bar{y}_{i.} - \bar{y}_{..})^2}_{②} + \underbrace{2\sum_i\sum_j (y_{ij} - \bar{y}_{i.})(\bar{y}_{i.} - \bar{y}_{..})}_{③}$$

TERM 3 $= 2\sum_i (\bar{y}_{i.} - \bar{y}_{..}) \sum_j (y_{ij} - \bar{y}_{i.})^0 \Rightarrow$

$$\Rightarrow \sum_i \sum_j (y_{ij} - \bar{y}_{..})^2 = \sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2 + \sum_i \sum_j (\bar{y}_{i.} - \bar{y}_{..})^2$$

SSTOTAL = SSERROR + SS TREATMENTS

## SUMS OF SQUARES FROM SUFFICIENT STATS $\{n_i\}, \{\bar{y}_i\}, \{s_i\}$

$$SS_{ERROR} = \sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2 = (r-1) \sum_i s_i^2$$

$$SS_{TREATMENTS} = \sum_i \sum_j (\bar{y}_{i.} - \bar{y}_{..})^2 = r \sum_i (\bar{y}_{i.} - \bar{y}_{..})^2$$

## 2.7 Treatment Effects Model

Considers the Effects of Treatments as Deviation from overall mean

$$\mu_i = \bar{\mu}_. + \tau_i \quad \text{where} \quad \bar{\mu}_. = \frac{\sum \mu_i}{t} = \mu$$

$$\mu_i = \mu + \tau_i \qquad \tau_i = \mu_i - \mu$$

NOTE: $\sum_{i=1}^{t} \tau_i = \sum_{i=1}^{t} (\mu_i - \mu) = \sum \mu_i - t\mu = 0$

### Least Squares Estimates

$$\hat{\mu} = \bar{y}_{..} \qquad \hat{\mu}_i = \bar{y}_{i.} \Rightarrow \hat{\tau}_i = \hat{\mu}_i - \hat{\mu} = \bar{y}_{i.} - \bar{y}_{..}$$

## 2.8 Degrees of Freedom

TOTAL (UNCORRECTED) SUM OF SQUARES $= \sum_i \sum_j y_{ij}^2$ has $N = tr$ independent terms and thus $N$ degrees of freedom

Total (CORRECTED) SUM OF SQUARES $= \sum_i \sum_j (y_{ij} - \bar{y}_{..})^2$ has $N = tr$ terms, but $\sum_i \sum_j (y_{ij} - \bar{y}_{..}) = 0$, thus any one term is the negative sum of all over terms
$\Rightarrow N-1$ degrees of freedom

$\rightarrow$

ERROR SUM OF SQUARES $= \sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2$

$$= \sum_{i=1}^{t} \Big[ \underbrace{\sum_{j=1}^{r} (y_{ij} - \bar{y}_{i.})^2}_{} \Big]$$

Term sum to zero (before squaring) $\Rightarrow$ $r-1$ independent terms from treatment $i$

$\sum_{i=1}^{t} (r-1) = rt - t$

$= N - t$

Error degrees of freedom.

TREATMENT SUM OF SQUARES $= r \sum_{i=1}^{t} (\bar{y}_{i.} - \bar{y}_{..})^2$

But: $\sum_{i=1}^{t} (\bar{y}_{i.} - \bar{y}_{..}) = 0$ $\Rightarrow$ $t-1$ independent terms

$\Rightarrow$ $t-1$ Treatment degrees of freedom

NOTE: $\underset{df_{TOTAL}}{N-1} = \underset{df_{Error}}{(N-t)} + \underset{df_{TRTS}}{(t-1)}$

## 2.9 Analysis of Variance Table

| Source of Variation | DEGREES OF FREEDOM | SUM OF SQUARES | MEAN SQUARE |
|---|---|---|---|
| TREATMENTS | $t-1$ | SS TREATMENTS | $MST = \frac{SST}{t-1}$ |
| ERROR | $N-t$ | SS ERROR | $MSE = \frac{SSE}{N-t}$ |
| TOTAL (CORRECTED) | $N-1$ | SS TOTAL | |

# COMPUTING EXPECTED MEAN SQUARES

① EXPAND SUMS of SQUARES

Ⓐ $SS_{TREATMENTS} = \sum_i \sum_j (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 = r \sum_i (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2$

$$= r\left\{ \sum_{i=1}^{t} (\bar{y}_{i\cdot}^2) + \sum_{i=1}^{t} \bar{y}_{\cdot\cdot}^2 - 2\sum_i \bar{y}_{i\cdot}\bar{y}_{\cdot\cdot} \right\}$$

$$= r\left\{ \sum_{i=1}^{t} \bar{y}_{i\cdot}^2 + t\bar{y}_{\cdot\cdot}^2 - 2\bar{y}_{\cdot\cdot}\sum_i \bar{y}_{i\cdot} \right\}$$

Aside: $\sum_{i=1}^{t} \bar{y}_{i\cdot} = t\,\bar{y}_{\cdot\cdot}$ $\quad \left( \frac{\sum \bar{y}_{i\cdot}}{t} = \bar{y}_{\cdot\cdot} \right)$

$$r\left\{ \sum_i \bar{y}_{i\cdot}^2 + t\bar{y}_{\cdot\cdot}^2 - 2t\bar{y}_{\cdot\cdot}^2 \right\} = r\left\{ \sum_i \bar{y}_{i\cdot}^2 - t\bar{y}_{\cdot\cdot}^2 \right\}$$

Ⓑ $SS_{ERROR} = \sum_i \sum_j (y_{ij} - \bar{y}_{i\cdot})^2 = \sum_i \sum_j y_{ij}^2 + r\sum_i \bar{y}_{i\cdot}^2 - 2\sum_i \sum_j y_{ij}\bar{y}_{i\cdot}$

$$= \sum_i \sum_j y_{ij}^2 + r\sum_i \bar{y}_{i\cdot}^2 - 2\sum_i \bar{y}_{i\cdot} \sum_j y_{ij} \qquad \left( \sum_j y_{ij} = y_{i\cdot} \right)$$

$$= \sum_i \sum_j y_{ij}^2 + r\sum_i \bar{y}_{i\cdot}^2 - 2\sum_i \bar{y}_{i\cdot}\, y_{i\cdot} \qquad \left( y_{i\cdot} = r\bar{y}_{i\cdot} \right)$$

$$= \sum_i \sum_j y_{ij}^2 + r\sum_i \bar{y}_{i\cdot}^2 - 2r\sum_i \bar{y}_{i\cdot}^2 = \sum_i \sum_j y_{ij}^2 - r\sum_i \bar{y}_{i\cdot}^2$$

$\Rightarrow$ $SS_{TREATMENTS} = r\left\{ \sum_i \bar{y}_{i\cdot}^2 - t\bar{y}_{\cdot\cdot}^2 \right\}$

$SS_{ERROR} = \sum_i \sum_j y_{ij}^2 - r\sum_i \bar{y}_{i\cdot}^2$

(continued)

$$E[Y_{ij}] = \mu_i \quad V[Y_{ij}] = \sigma^2 \quad \text{cov}(Y_{ij}, Y_{i'j'}) = 0 \quad i \neq i' \text{ and/or } j \neq j'$$

## ② Obtain Expectations of Elements

$$V(Y) = E[(Y-\mu)^2] = E[Y^2 + \mu^2 - 2\mu Y]$$

$$= E[Y^2] + \mu^2 - 2\mu E(Y) = E[Y^2] + \mu^2 - 2\mu^2$$

$$= E[Y^2] - \mu^2 \Rightarrow \boxed{E[Y^2] = V[Y] + \mu^2}$$

$E[Y_{ij}] = \mu_i \quad V[Y_{ij}] = \sigma^2$

$$E[\bar{Y}_{i.}] = E\left[\sum_{j=1}^{r} Y_{ij} / r\right] = \frac{1}{r} E\left[\sum_{j=1}^{r} Y_{ij}\right]$$

$$= \frac{1}{r} \sum_{j=1}^{r} E[Y_{ij}] = \frac{1}{r} \sum_{j=1}^{r} \mu_i = \frac{1}{r}(r\mu_i) = \mu_i$$

$$V[\bar{Y}_{i.}] = V\left[\frac{\sum_{j=1}^{r} Y_{ij}}{r}\right] = \frac{1}{r^2} V\left[\sum_{j=1}^{r} Y_{ij}\right]$$

$$= \frac{1}{r^2}\left\{\sum_{j=1}^{r} V(Y_{ij}) + 2\underbrace{\sum_{j=1}^{r-1}\sum_{j'=j+1}^{r}}_{\binom{r}{2} \text{ Pairs } j,j'} \text{Cov}(Y_{ij}, Y_{ij'})\right\}$$

each element 0 for this model

$$= \frac{1}{r^2}\left[\sum_{j=1}^{r} \sigma^2 + 2\binom{r}{2} 0\right] = \frac{r\sigma^2}{r^2} = \frac{\sigma^2}{r}$$

$$E[\bar{Y}_{..}] = E\left[\frac{\sum_{i=1}^{t}\sum_{j=1}^{r} Y_{ij}}{rt}\right] = \frac{1}{rt} E\left[\sum_i\sum_j Y_{ij}\right]$$

$$= \frac{1}{rt}\sum_i\sum_j E[Y_{ij}] = \frac{1}{rt}\sum_i r\mu_i = \frac{r\sum\mu_i}{rt} = \frac{\sum\mu_i}{t} = \bar{\mu}.$$

$$V[\bar{Y}_{..}] = V\left[\frac{\sum\sum Y_{ij}}{rt}\right] = \left[\frac{1}{rt}\right]^2 V\left[\sum_i\sum_j Y_{ij}\right]$$

$$= \left(\frac{1}{rt}\right)^2\left\{\sum_i\sum_j V(Y_{ij}) + 2\underbrace{\sum_{i=1}^{t}\sum_{i'=1}^{t}\sum_{j=1}^{r-1}\sum_{j=j'+1}^{r}}_{t^2\binom{r}{2} \text{ pairs}} \text{Cov}(Y_{ij}, Y_{i'j'})\right\}$$

all elements 0.

→

$$V[\bar{y}_{..}] = \left(\frac{1}{rt}\right)^2 \left[rt\ \sigma^2 + 2t^2(\tfrac{r}{2})(0)\right] = \frac{\sigma^2}{rt}$$

- - - - - - - - - - - - - - - - - -

Summary for fixed Effects model

$E[Y_{ij}] = \mu_i \quad i=1,\dots,t \quad \Rightarrow E[Y_{ij}^2] = \sigma^2 + \mu_i^2$
$V(Y_{ij}) = \sigma^2 \quad j=1,\dots,r$

$E[\bar{Y}_{i.}] = \mu_i \quad i=1,\dots,t \quad \Rightarrow E[\bar{Y}_{i.}^2] = \frac{\sigma^2}{r} + \mu_i^2$
$V[\bar{Y}_{i.}] = \frac{\sigma^2}{r}$

$E[\bar{Y}_{..}] = \bar{\mu}_. = \frac{\sum \mu_i}{t} \quad \Rightarrow \quad E[\bar{Y}_{..}^2] = \frac{\sigma^2}{rt} + \bar{\mu}_.^2$
$V[\bar{Y}_{..}] = \frac{\sigma^2}{rt}$

- - - - - - - - - - - - - - - - - -

(3) Putting (1) and (2) Together

$$SS_{\text{TREATMENTS}} = r\left\{\sum \bar{y}_{i.}^2 - t\bar{y}_{..}^2\right\} = r\sum \bar{y}_{i.}^2 - rt\,\bar{y}_{..}^2$$

$$\Rightarrow E\{SS_{\text{TREATMENTS}}\} = E\left\{r\sum_i \bar{Y}_{i.}^2 - rt\bar{Y}_{..}^2\right\}$$

$$= r\sum_i E(\bar{Y}_{i.}^2) - rt\,E(\bar{Y}_{..}^2)$$

$$= r\sum_{i=1}^{t}\left(\frac{\sigma^2}{r} + \mu_i^2\right) - rt\left(\frac{\sigma^2}{rt} + \bar{\mu}_.^2\right)$$

$$= \left[t\sigma^2 + r\sum_i \mu_i^2\right] - \left[\sigma^2 + rt\bar{\mu}_.^2\right]$$

$$= \sigma^2(t-1) + r\left[\sum \mu_i^2 - t\bar{\mu}_.^2\right] \quad (*)$$

Aside: $\sum_{i=1}^{t}(\mu_i - \bar{\mu}_.)^2 = \sum_{i=1}^{t}\mu_i^2 - 2\mu_.\sum\mu_i + t\bar{\mu}_.^2$

$$= \sum \mu_i^2 - 2\mu_.(t\mu_.) + t\mu_.^2 = \sum_i \mu_i^2 - t\,\mu_.^2$$

$$\Rightarrow (*) = \boxed{\sigma^2(t-1) + r\sum_{i=1}^{t}(\mu_i - \bar{\mu}_.)^2 = E(SS\ TRTS)}$$

$$SS_{ERROR} = \sum_i \sum_j y_{ij}^2 - r \sum_i \bar{y}_{i.}^2$$

$$\Rightarrow E[SS_{ERROR}] = E\left\{ \sum_i \sum_j y_{ij}^2 - r \sum_i \bar{y}_{i.}^2 \right\}$$

$$= \sum_i \sum_j E(y_{ij}^2) - r \sum_i E[\bar{y}_{i.}^2]$$

$$= \sum_i \sum_j [\sigma^2 + \mu_i^2] - r \sum_i \left[\frac{\sigma^2}{r} + \mu_i^2\right]$$

$$= \left[rt\sigma^2 + r\sum_{i=1}^{t} \mu_i^2\right] - \left[r\left(t\frac{\sigma^2}{r}\right) + r\sum_{i=1}^{t} \mu_i^2\right]$$

$$= (rt - t)\sigma^2 + (r-r)\sum \mu_i^2 = t(r-1)\sigma^2 = \boxed{(N-t)\sigma^2 = E(SSE)}$$

$$MST = \frac{SST}{t-1} \Rightarrow$$

$$E[MST] = \frac{E[SST]}{t-1} = \frac{\sigma^2(t-1) + r\sum_{i=1}^{t}(\mu_i - \bar{\mu}_.)^2}{t-1}$$

$$= \sigma^2 + \frac{r\sum(\mu_{i.} - \bar{\mu}_.)^2}{t-1} = \sigma^2 + r\theta_t^2$$

$$MSE = \frac{SSE}{N-t} \Rightarrow E[MSE] = \frac{E[SSE]}{N-t} = \frac{(N-t)\sigma^2}{N-t} = \sigma^2$$

NOTE THAT $E[MST] \geq E[MSE]$

w/ EQUALITY HOLDING IFF $\sum(\mu_i - \bar{\mu}_.)^2 = 0$

which holds only if $\mu_1 = \mu_2 = \cdots = \mu_t = \mu$

## 2.10 Tests of Hypotheses About Linear Models

$H_0: \mu_1 = \mu_2 = \cdots = \mu_t$ (No differences among trt effects)

$H_A$: Not all $\mu_i$ are equal (Differences exist among trt effects)

**UNDER $H_0$ & $H_A$**

$$\frac{SS_{ERROR}}{\sigma^2} \sim \chi^2_{N-t}$$ (Result from math stat)

**UNDER $H_0$**

$$\frac{SS_{TREATMENTS}}{\sigma^2} \sim \chi^2_{t-1}$$ (Result from math stat)

**UNDER $H_A$**

$$\frac{SS_{TREATMENTS}}{\sigma^2} \sim \chi^2_{t-1}$$ (Non-central)

$SS_{ERROR}$ & $SS_{TREATMENTS}$ ARE INDEPENDENT.

**UNDER $H_0$** $(E[MSE] = E(MST) = \sigma^2)$

$$F_0 = \frac{MST}{MSE} = \frac{SS_{TRTS}/t-1}{SS_{ERROR}/N-t} \sim F_{(t-1,\, N-t)}$$

Num df Denominator df

**UNDER $H_A$** $(E(MST) = \sigma^2 + r\theta_t^2 > \sigma^2 = E(MSE))$

$$F_0 = \frac{MST}{MSE} \sim \text{Noncentral } F\left(t-1,\, N-t,\, \lambda = \frac{r\sum \tau_i^2}{\sigma^2}\right)$$

Noncentrality Para.

Note: Large values of $F_0$ consistent w/ $H_A$.

where $\tau_i = \mu_i - \bar{\mu}.$

(*) NOTE: YOU WILL SEE VARIOUS FORMULATIONS FOR THE NONCENTRALITY PARAMETER (MOST OFTEN BEING THIS VERSION OF $\lambda$ DIVIDED BY 2). THIS FORMULATION IS CONSISTENT W/ THE NC PARAMETER USED BY SAS IN ITS FINV and PROBF FUNCTIONS

EXAMPLE - CONSIDER THE MEAT STORAGE EXAMPLE FROM KUEHL W/ SEVERAL HYPOTHETICAL PARAMETER ARRANGEMENTS. ASSUME $\sigma^2 = 1.0$ (which is much larger than its estimate $s^2$).

$r=3$

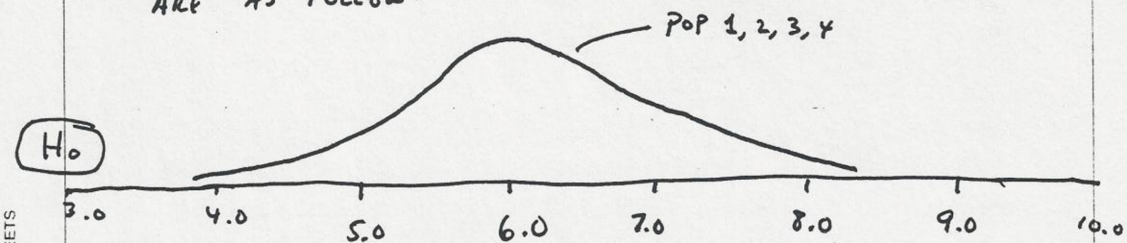
| | $H_0$ | | $H_A^1$ | | $H_A^2$ | |
|---|---|---|---|---|---|---|
| TREATMENT | $\mu_i$ | $\mu_i - \bar{\mu}.$ | $\mu_i$ | $\mu_i - \bar{\mu}.$ | $\mu_i$ | $\mu_i - \bar{\mu}.$ |
| COMMERCIAL (1) | 6 | 0 | 6.91 | 0.91 | 7.12 | 1.12 |
| VACUUM (2) | 6 | 0 | 6.00 | 0 | 4.88 | -1.12 |
| $CO, O_2, N$ (3) | 6 | 0 | 6.00 | 0 | 7.12 | 1.12 |
| 100% $CO_2$ (4) | 6 | 0 | 5.09 | -0.91 | 4.88 | -1.12 |
| | $\bar{\mu}.=6$ | $\Sigma=0$ | $\bar{\mu}.=6$ | $\Sigma=0$ | $\bar{\mu}.=6$ | $\Sigma=0$ |

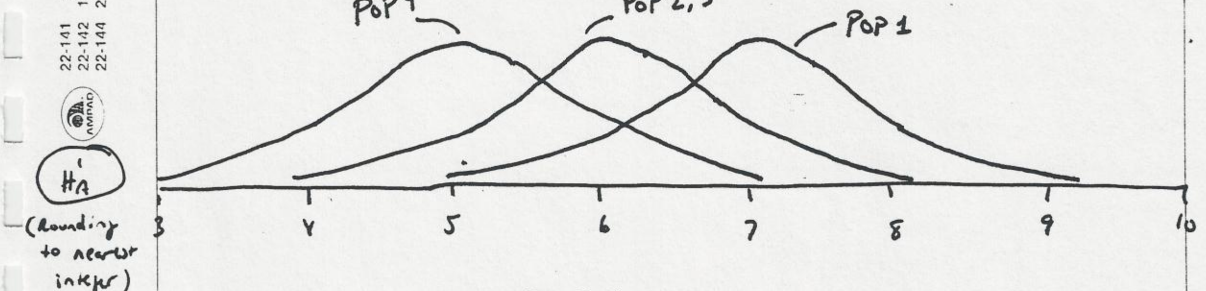
Under $H_0$: $\lambda = \dfrac{3\left[0^2 + \cdots + 0^2\right]}{1} = 0$ (Central F)

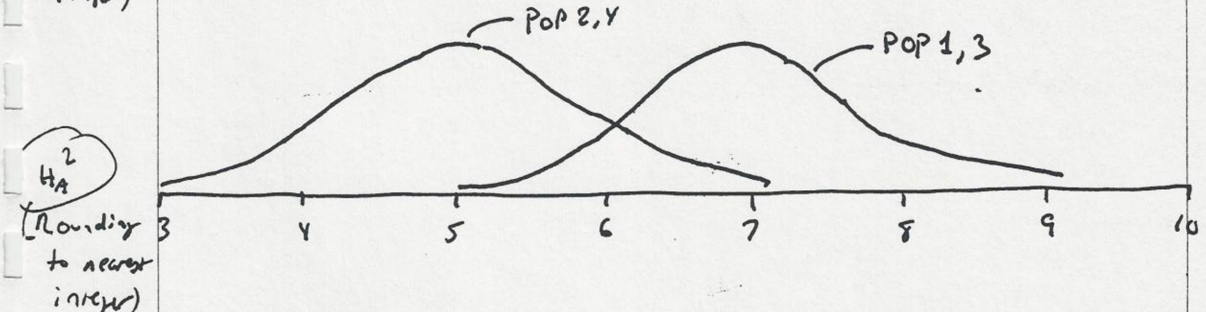
Under $H_A^1$: $\lambda = \dfrac{3\left[0.91^2 + 0 + 0 + (-0.91)^2\right]}{1} = 5.0$

Under $H_A^2$: $\lambda = \dfrac{3\left[(1.12)^2 + (-1.12)^2 + (1.12)^2 + (-1.12)^2\right]}{1} = 15.0$

The distributions of measurements under $H_0$, $H_A^1$, $H_A^2$ ARE AS FOLLOW:




There is more separation in treatment means in $H_A^2$ Than in $H_A^1$. That is the "distance" from $H_A^2$ to $H_0$ is further than the "distance" from $H_A^1$ to $H_0$.

$\lambda$ is a measure of the distance.

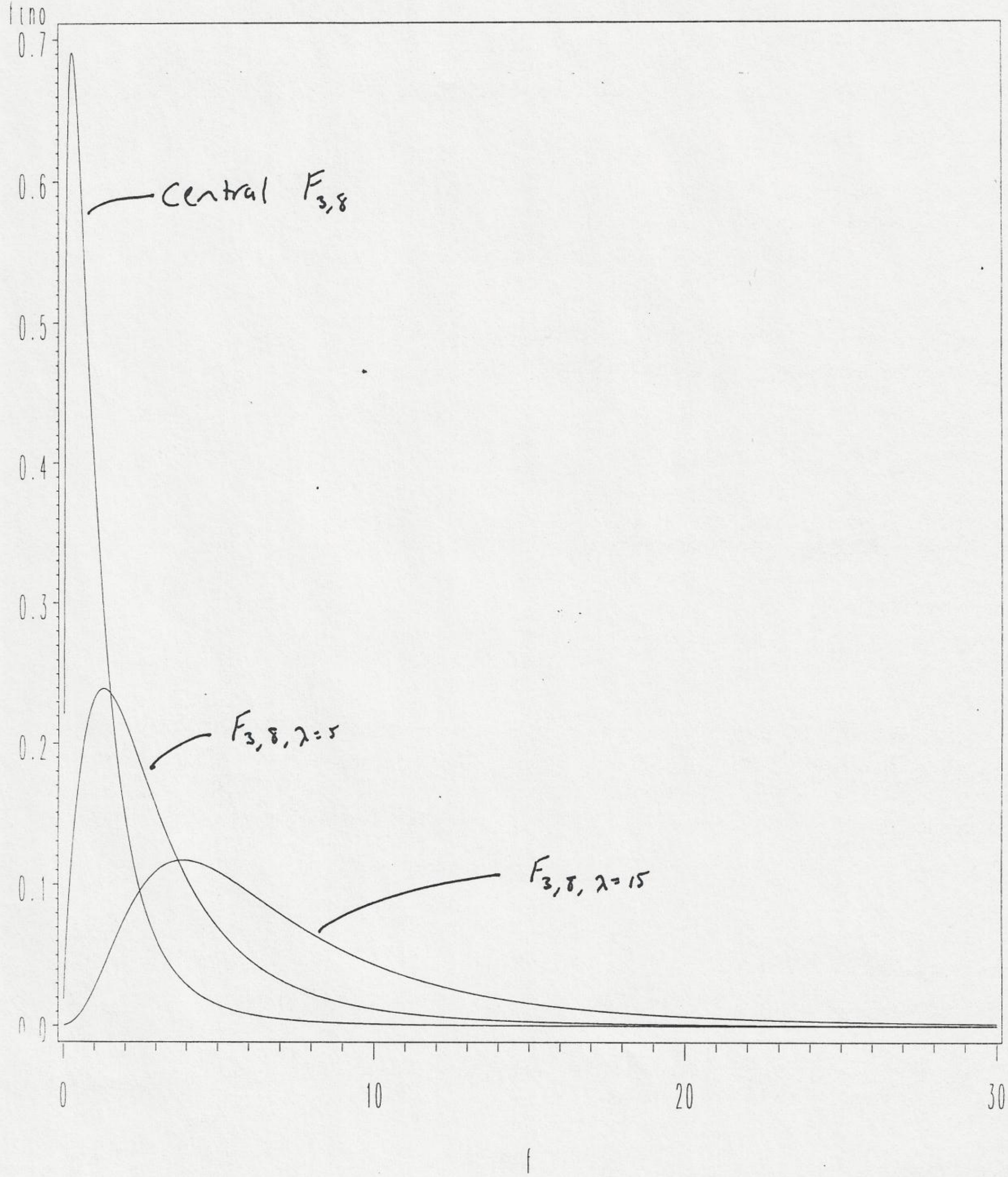
Note: If we took the estimates $\{\hat{\mu}_i\}$, $s^2$ as the true parameter values $\{\mu_i\}$, $\sigma^2$ from Kuehl's example, we'd have.

$$\lambda = \frac{3\left[(7.48-5.90)^2 + (5.503-5.90)^2 + (7.26-5.90)^2 + (3.36-5.90)^2\right]}{0.11585} = 284$$

which is HUGE.

# Distribution of F Under H0, HA1 and HA2





## TEST OF HYPOTHESIS FOR TREATMENT EFFECTS

$H_0: \mu_1 = \mu_2 = \cdots = \mu_t$ (NO "TRT EFFECTS")

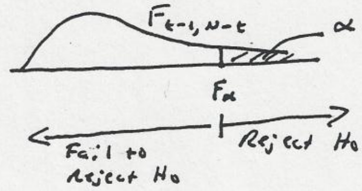
$H_A$: NOT ALL $\mu_i$ are = ("TRT EFFECTS" NOT ALL EQUAL)

T.S. $F_0 = \frac{MST}{MSE}$

UNDER $H_0$: $F_0 \sim F_{t-1, N-t}$

UNDER $H_A$: $F_0 \sim F_{t-1, N-t, \lambda > 0}$

RR: $F_0 \geq F_{\alpha, t-1, N-t}$




$\alpha \equiv$ "size" of the rejection region or test

$= Pr\{\text{Type I ERROR}\}$

$= Pr\{\text{Conclude } H_A \mid H_0 \text{ TRUE}\}$

$\alpha$ IS TYPICALLY ACCEPTED @ .05

## KUEHL'S EXAMPLE $t=4$ STORAGE TYPES, $r=3$ REPS/TYPE

ANOVA

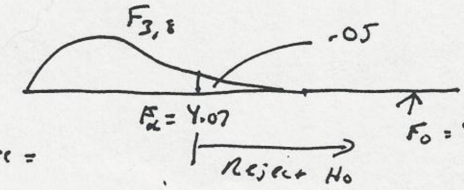
| SOURCE | df | SS | MS | $F_0$ |
|---|---|---|---|---|
| TRTS | 3 | 32.8728 | 10.9576 | 94.54 |
| ERROR | 8 | 0.9268 | .1159 | — |
| TOTAL (CORRECTED) | 11 | 33.7996 | — | — |

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$ (Mean bacteria counts are =)

$H_A$: Not all $\mu_i$ are = (Mean bacteria counts vary among Storage types)

T.S. $F_0 = \frac{MST}{MSE} = 94.54$

RR: $F_0 \geq F_{.05, 3, 8} = 4.07$
($\alpha = .05$)




Reject $H_0$, Conclude NOT ALL $\mu_i$ are =

## 2.11 Significance Testing and Tests of Hypotheses

Alternative to setting up rejection region is to compute a P-value, and make judgment based on it.

P-value = Pr{Obtaining a Test Statistic as consistent or more as the observed test statistic is wrt $H_A$ | $H_o$ is true}

Small values of P-value => that observed data is inconsistent w/ $H_o$ (is consistent w/ $H_A$).

We reject $H_o$ if P-value $\leq \alpha$

Kuehl's example

① Values of $F = \frac{MST}{MSE} \geq F_o = 94.54$ are as consistent or more consistent w/ $H_A$ than what we observed in this experiment.

② When $H_o$ is true the distribution is (central) F w/ $(t-1)$ numerator and $(N-t)$ denominator degrees of freedom.

③ So the area in the upper tail of the $F_{3,8}$ distribution (above 94.54) is the P-value.

## 2.12 STANDARD ERRORS & Confidence Intervals for TRT MEANS

- **L.S. Means**: $\hat{\mu}_i = \bar{Y}_{i.} \qquad i = 1, \dots, t$

If $Y_{i1}, \dots, Y_{ir} \sim NID(\mu_i, \sigma^2)$,

Then:

$$E[\bar{Y}_{i.}] = E\left\{\frac{\sum_{j=1}^{r} Y_{ij}}{r}\right\} = \frac{1}{r} E\left\{\sum_j Y_{ij}\right\}$$

$$= \frac{1}{r}\sum_j E[Y_{ij}] = \frac{1}{r}\sum_j \mu_i = \frac{1}{r}(r\mu_i) = \mu_i$$

$$V[\bar{Y}_{i.}] = V\left\{\frac{\sum_j Y_{ij}}{r}\right\} = \frac{1}{r^2} V\left\{\sum_j Y_{ij}\right\}$$

$$= \frac{1}{r^2}\Big\{\underbrace{\sum_j V(Y_{ij})}_{\substack{r \text{ terms} \\ \text{each} = \sigma^2}} + \underbrace{2\sum\sum_{j<j'} Cov(Y_{ij}, Y_{ij'})}_{\substack{\binom{r}{2} \text{ terms} \\ \text{each} = 0 \\ \text{by independence}}}\Big\}$$

$$= \frac{1}{r^2}\{r\sigma^2 + 0\} = \frac{\sigma^2}{r}$$

$\sigma^2_{\bar{Y}_{i.}} = \dfrac{\sigma^2}{r}$ $\qquad$ $\sigma^2$ unknown $\qquad$ ESTIMATE $= S^2 = MSE$

$S^2_{\bar{Y}_{i.}} = \dfrac{S^2}{r}$ $\qquad$ Estimated Variance of $\bar{Y}_{i.}$

$S_{\bar{Y}_{i.}} = \dfrac{S}{\sqrt{r}}$ $\qquad$ Estimated Std. Error of $\bar{Y}_{i.}$

Note: $\{\bar{Y}_{i.}\}$ and $S^2 = MSE$ are independent

$$\bar{Y}_{i.} \sim N(\mu_i, \frac{\sigma^2}{r}) \Rightarrow \frac{\bar{Y}_{i.} - \mu_i}{\sigma/\sqrt{r}} \sim N(0,1)$$

$$\frac{SSE}{\sigma^2} \sim \chi^2_{N-t} \qquad \text{(From math stat)}$$

Let: $Z \sim N(0,1)$

$V \sim \chi^2_r$

w/ $Z$ & $V$ independent

Then: $T = \frac{Z}{\sqrt{V/r}} \sim t_r$ (Students-t w/ r degrees of freedom)

For our case: ① $Z = \frac{\bar{Y}_{i.} - \mu_i}{\sigma/\sqrt{r}}$

② $V = \frac{(N-t)S^2}{\sigma^2} = \frac{SSE}{\sigma^2}$

③ $r = N-t$

Now: $$T = \frac{Z}{\sqrt{V/r}} = \frac{\left[\frac{\bar{Y}_{i.} - \mu_i}{\sigma/\sqrt{r}}\right]}{\sqrt{\frac{(N-t)S^2}{\sigma^2}/(N-t)}} = \frac{\frac{\bar{Y}_{i.} - \mu_i}{\sigma/\sqrt{r}}}{S/\sigma}$$

$$= \frac{\bar{Y}_{i.} - \mu_i}{\sigma/\sqrt{r}} \cdot \frac{\sigma}{S} = \frac{\bar{Y}_{i.} - \mu_i}{S/\sqrt{r}} \sim t_{N-t}$$

Student's-t distribution w/ $\nu$ degrees of freedom $\nu = 1, 2, \ldots$

$$f(t) = \frac{\Gamma[(\nu+1)/2]}{\sqrt{\pi\nu}\,\Gamma(\nu/2)} \frac{1}{(1+t^2/\nu)^{(\nu+1)/2}} \qquad -\infty < t < \infty$$

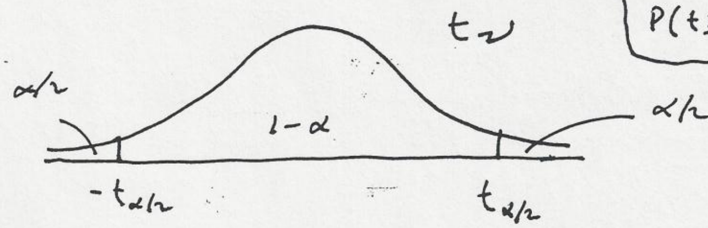
where $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\,dx$

- $\Gamma(\alpha) = (\alpha-1)\,\Gamma(\alpha-1)$ if $\alpha > 1$
- $\Gamma(n) = (n-1)!$ for integer $n$
- $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

- Centered @ 0
- Symmetric
- Mound-shaped
- Limit is $N(0,1)$ as $\nu \to \infty$

$\boxed{\text{Table II. P. 589}}$

Table gives $t_{\alpha,\nu} \ni P(t \geq t_{\alpha,\nu}) = \alpha$




$$P\left(-t_{\frac{\alpha}{2},N-t} \leq \frac{\bar{y}_{i\cdot} - \mu_i}{s/\sqrt{r}} \leq t_{\frac{\alpha}{2},N-t}\right) = 1-\alpha$$

$$\Rightarrow P\left(-t_{\frac{\alpha}{2},N-t} \cdot \frac{s}{\sqrt{r}} \leq \bar{y}_{i\cdot} - \mu_i \leq t_{\frac{\alpha}{2},N-t} \cdot \frac{s}{\sqrt{r}}\right) = 1-\alpha$$

$$\Rightarrow P\left(-\bar{y}_{i\cdot} - t_{\frac{\alpha}{2},N-t} \cdot \frac{s}{\sqrt{r}} \leq -\mu_i \leq -\bar{y}_{i\cdot} + t_{\frac{\alpha}{2},N-t} \cdot \frac{s}{\sqrt{r}}\right) = 1-\alpha$$

$$\Rightarrow P\left(\bar{y}_{i\cdot} + t_{\frac{\alpha}{2},N-t} \cdot \frac{s}{\sqrt{r}} \geq \mu_i \geq \bar{y}_{i\cdot} - t_{\frac{\alpha}{2},N-t} \cdot \frac{s}{\sqrt{r}}\right) = 1-\alpha$$

$\Rightarrow$ $(1-\alpha)100\%$ CI for $\mu_i$

$$\bar{y}_{i.} \pm t_{\frac{\alpha}{2}, N-t} \cdot \frac{S}{\sqrt{r}} \equiv \bar{y}_{i.} \pm t_{\frac{\alpha}{2}, N-t} \cdot S_{\bar{y}_{i.}}$$

Kuehl EXAMPLE (Meat Storage)

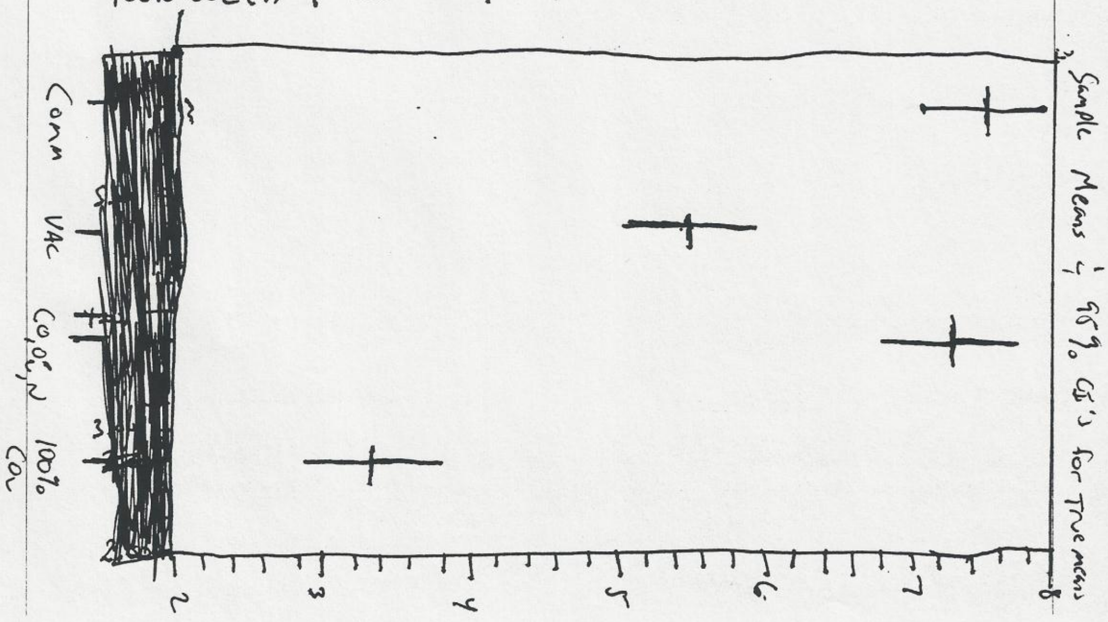
$t = 4$ Methods $\quad r = 3$ reps/Method $\quad S^2 = MSE = 0.116$

$\bar{y}_{1.} = 7.48 \quad \bar{y}_{2.} = 5.50 \quad \bar{y}_{3.} = 7.26 \quad \bar{y}_{4.} = 3.36$

$S_{\bar{y}_{i.}} = \sqrt{\frac{S^2}{r}} = \sqrt{\frac{.116}{3}} = .197 \quad t_{.025, 12-4} = 2.306$

$t_{\frac{\alpha}{2}, N-t} \cdot S_{\bar{y}_{i.}} = 2.306(.197) = .45$

| TRT (i) | $\bar{y}_{i.}$ | 95% CI | | |
|---|---|---|---|---|
| Comm (1) | 7.48 | $7.48 \pm .45$ | $\equiv$ | (7.03, 7.93) |
| VAC (2) | 5.50 | $5.50 \pm .45$ | $\equiv$ | (5.05, 5.95) |
| $CO, O_2, N$ (3) | 7.26 | $7.26 \pm .45$ | $\equiv$ | (6.81, 7.71) |
| 100% $CO_2$ (4) | 3.36 | $3.36 \pm .45$ | $\equiv$ | (2.91, 3.81) |



## 2.13 Unequal Replication of Treatments.

Suppose that for trt $i$, we have $r_i$ observations $i=1,\dots,t$

$$y_{ij} = \mu_i + e_{ij} \qquad i=1,\dots,t \quad j=1,\dots,r_i$$

$$N = \sum_{i=1}^{t} r_i$$

L.S. estimates: $\hat{\mu}_i = \dfrac{y_{i.}}{r_i} = \bar{y}_{i.}$

ANALYSIS OF VARIANCE

$$SSTREATMENTS = \sum_{i=1}^{t}\sum_{j=1}^{r_i}(\bar{y}_{i.} - \bar{y}_{..})^2 = \sum_{i=1}^{t} r_i(\bar{y}_{i.} - \bar{y}_{..})^2$$

$$SSERROR = \sum_{i=1}^{t}\sum_{j=1}^{r_i}(y_{ij} - \bar{y}_{i.})^2 = \sum_{i=1}^{t}(r_i - 1)\frac{\sum_{j=1}^{r_i}(y_{ij} - \bar{y}_{i.})^2}{r_i - 1}$$

$$= \sum_{i=1}^{t}(r_i - 1)\, s_i^2$$

DEGREES OF FREEDOM ARE UNCHANGED ($df_{TRT} = t-1$, $df_{ERR} = N-t$)

$$\bar{\mu}_. = \frac{\sum_{i=1}^{t} r_i \mu_i}{\sum_{i=1}^{t} r_i} = \text{Weighted mean of the TRT MEANS}$$

$$\hat{\bar{\mu}}_. = \frac{\sum_{i=1}^{t} r_i \bar{y}_{i.}}{\sum_{i=1}^{t} r_i}$$

$$E[MSE] = \sigma^2$$

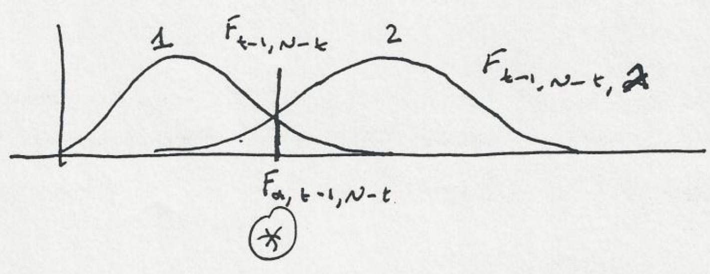
$$E[MST] = \sigma^2 + \frac{\sum_{i=1}^{t} r_i(\mu_i - \bar{\mu}_.)^2}{t-1} = \sigma^2 + \theta_t^2$$

$$S^2_{\bar{Y}_{i.}} = \frac{s^2}{r_i} \Rightarrow S_{\bar{Y}_{i.}} = \frac{s}{\sqrt{r_i}}$$

## 2.14 Choosing the number of Replicates Per Treatment

This method differs from Keuhl's who's uses very difficult to read charts.

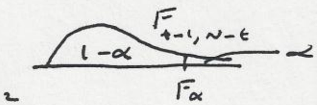
1. Determine a set of parameter values for the $\{\tau_i = \mu_i - \bar{\mu}_.\}$ that is important to detect. These are sometimes referred to as practical or clinically important differences. This is usually done in consultation w/ researcher in their field of expertise. If $\sigma^2$ is unknown you may need $\left\{\frac{\tau_i^2}{\sigma^2}\right\}$
2. Set up $\lambda = r \frac{\sum \tau_i^2}{\sigma^2}$ as a function of $r$, the replicate size
3. Select the <u>power</u> of the test – the prob. you <u>will</u> reject Ho, given this set of $\left\{\frac{\tau_i^2}{\sigma^2}\right\}$ terms. This will ideally be around .80 or higher. As the power increases, so will the necessary number of replicates. Power $= 1-\beta$ where $\beta = Pr\{\text{Type II error}\} = Pr\{\text{Conclude } H_0 \mid H_A \text{ TRUE}\}$.




Goal: Area under curve 2 to right of * to be = Power $= 1-\beta$

## (4) ALGORITHM

- STEP 1: BEGIN CYCLING FROM $r = 1$
- STEP 2: COMPUTE $N = tr$
- STEP 3: Obtain $F_{\alpha, t-1, N-t}$

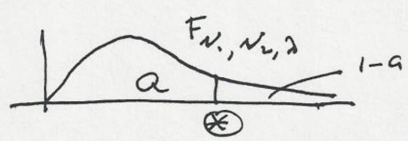



- STEP 4: Compute $\lambda = r \dfrac{\sum \tau_i^2}{\sigma^2}$
- STEP 5: Obtain area under the $F_{t-1, N-t, \lambda}$ distribution above the result from STEP 3.
- STEP 6: If result from STEP 5 $\geq$ Power, Stop
  If result from STEP 5 $<$ Power, THEN:
  RETURN TO STEP 1 AND INCREMENT $r$ by 1.

## USEFUL SAS FUNCTIONS

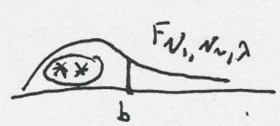
- $\text{FINV}(a, \nu_1, \nu_2, \lambda)$ - Critical Value from F-distribution

  $a \equiv$ area below Critical value $(0 < a < 1)$
  $\nu_1 \equiv$ numerator df $(\nu_1 > 0)$
  $\nu_2 \equiv$ denominator df $(\nu_2 > 0)$
  $\lambda \equiv$ non-centrality parameter (if not given, $\lambda = 0$ assumed) $(\lambda \geq 0)$




(*) $\equiv$ Result from function.

- $\text{PROBF}(b, \nu_1, \nu_2, \lambda)$ - Prob that F falls below $b$.
  $b \equiv$ Positive value




(**) $\equiv$ Result from function