# STAT 5023: STAT FOR EXPERIMENTORS II CHEAT SHEET

```
DM 'LOG; CLEAR; ODSRESULTS; CLEAR;' ;
DATA one; INPUT lake oxygen @@; /*O2 in 3 lakes */
DATALINES; 1 0 1 2 1 3 2 1 2 4 2 2 3 14 3 26 3 25;
/* Horizontal Plot ; */
PROC PLOT DATA = one VPERCENT = 25;
PLOT lake * (Oxygen a b c) / VREF = 1 2 3;
/* *** Boxplot ***; */
PROC BOXPLOT DATA = one; PLOT lake*oxygen;
/* T-Test ; */
PROC TTEST DATA = one; CLASS = lake; VAR oxygen;
TITLE2 'Recover the t-statistics';
/* Levene's and Bartlett HOV Tests ; */
PROC PRINT DATA = one; /* Print Data */
TITLE 'ANOVA and Levene's & Bartlett Test Example':
PROC GLM DATA = one PLOTS = (BOXPLOT RESIDUALS DIAGNOSTICS
RESIDUALPANEL); CLASS lake;
a = log (oxygen +1); b = sqrt (oxygen); c = oxygen*oxygen; /* DATA
TRANSFORMATION. Replace in MODEL as needed */
MODEL oxygen = lake;
MEANS lake / HOVTEST = LEVENE (TYPE = ABS) HOVTEST = BARTLETT;
/*Notice: LEVENE & BARTLETT Tests*/
/* Mixed Procedure and Satterthwaite DF Test ; */
PROC MIXED DATA = ONE PLOTS = RESIDUALPANEL;
CLASS lake;
MODEL oxygen = lake /DDFM = SATTERTH;
REPEATED / GROUP = lake;
LSMEANS lake / STDERR CLDIFF ALPHA = .01;
TITLE 'A second alternative when the variances are unequal';
TITLE2 'Using the MIXED procedure';
```

**EXAM I 2019 SAS CODE:**

**DATA** exam1; **INPUT** pressure damage @@; **DATALINES;** 30 27 35 23 … 50 20; **PROC GLM DATA** exam1; **CLASS** pressure; **MODEL** Damage = pressure; **LSMEANS** pressure **/STDERR CLDIFF ALPHA = 0.01; CONTRAST** 'Problem 2e' pressure -1 0 0.5 0.5; **RUN: QUIT;**

**EXAM II 2019 SAS CODE:**

**DATA** exam2; **INPUT** pepper technique $ Quality; **DATALINES;** 1 canned 80 1 uncanned 91 … 4 canned 85 4 uncanned 82; **PROC GLM DATA** = exam2; **CLASS** pepper technique; **MODEL** quality = pepper technique pepper*technique; **MEANS** pepper*technique **/ LSD**; **LSMEANS** technique **/STDERR PDIFF ADJUST = T SLICE** = technique; **RUN; QUIT;**

**EXAM III 2019 SAS CODE:**

**DATA** exam3; **INPUT** speed temperature product $ burn pressure; **DATALINES;** 15 50 A 1 43; **PROC MIXED DATA** = exam3; **CLASS** speed temperature product burn; **MODEL** pressure = speed | temperature | product / **DDFM = KR**; **RANDOM** burn; **LSMEANS** speed*temperature*product **/ADJUST = TYKEY PDIFF; RUN; QUIT;**

**W1P1: Review: Mean** = $\Sigma$ $X_i$/n. **Standard Deviation** (SD) = $\sqrt{(\Sigma(X\text{-}X_{bar})^2/(n\text{-}1))}$. **Variance** ($\sigma$) = $SD^2$ & **Standard Error** = $SD/\sqrt{n}$ = $\sqrt{(Variance/n)}$. **$H_o$** = Null & **$H_A$ or $H_1$** = Alternative Hypothesis. **Type I Error:** $H_o$ is true & reject $H_o$. **Type II Error:** $H_o$ is false & failed to reject $H_o$. **How to look for p-value: Steps:** 1) Calculate F-stat ($F_{calc}$) in AOV. 2) Find 5 $F_{cric(\alpha, df1, df2)}$ values varying $\alpha$ value at 0.1, 0.05, 0.025, 0.01, 0.005. Find range of $F_{cric}$ where $F_{calc}$ falls.

**Single Normal Population, Random Sample (T-test)**

Measure Mean ($\bar{Y}$) and variance ($S^2$) from sample of n.
**Hypothesis:** $H_o$: $\mu = \mu_o$ & $H_A$: $\mu \neq \mu_o$
**Test Stat:** t-test = $(Y_{bar} - \mu_o)/$St. Err. of Mean = $(\bar{Y} - \mu_o)/(s/\sqrt{n}) = (Y_{bar} - \mu_o)/\sqrt{(MS\ Error/n)}$
**Conclusion:** Reject $H_o$ if $|t| \geq t_{\alpha/2, df}$ ; **df** = n-1
**Confidence Interval (CI):** (1-$\alpha$)100% CI for $\mu$: $\bar{Y} \pm t_{\alpha/2, df} * $ St Err (S) of Mean ($\bar{Y}$.) **Standard Error of Mean (S)** = $\sqrt{(MSE_{AOV}/n)}$

**Two Norm. Popl$^n$, Eq. Var$^{iances}$, $\sigma_1^2 = \sigma_2^2$, Rand. Sample**

Measure Means ($\bar{Y}_1$ & $\bar{Y}_2$) and variances ($S_1^2$ & $S_2^2$) from samples $n_1$ & $n_2$ respectively. **Hypothesis:** $H_o$: $\mu_1 = \mu_2$ & $H_A$: $\mu_1 \neq \mu_2$
**Pooled Variance Estimate of $\sigma_2^2$ ($S_p^2$)** = **$MSE_{AOV}$** = $((n_1 -1)*S_1^2 + (n_2 - 1)*S_2^2)/ (n_1 + n_2 -2)$ (S = Standard Error)
**Test Statistics:** t-test = $(Y_{1bar} - Y_{2bar})/\sqrt{(S_p^2 (1/n_1 + 1/n_2))}$ = **$(Y_{1bar} - Y_{2bar})/\sqrt{(MSE (1/n_1 + 1/n_2))}$** = $(\bar{Y}_1 - \bar{Y}_2)/ S_{(Y1bar - Y2bar)}$ with df = $n_1 + n_2 - 2$. **Conclusion:** Reject $H_o$ if $|t| \geq t_{\alpha/2, df}$ ;
**Confidence Interval (CI):** (1-$\alpha$) 100% CI for $\mu$: $(\bar{Y}_1 - \bar{Y}_2) \pm t_{\alpha/2, df} * S_{(Y1bar - Y2bar)}$
**St. Err. of Mean** ($S_{(Y1bar - Y2bar)}$) = $\sqrt{(MSE (1/n_1 + 1/n_2))}$

**Two N. Popl$^n$, $\sigma_1^2 \neq \sigma_2^2$, and Ind. & Ran$^{dom}$ Samples** (SAS):

**Test Statistics**: $(\bar{Y}_1 - \bar{Y}_2)/\sqrt{(( S_1^2/n_1) + ( S_2^2/n_2))}$ with df = $(( S_1^2/n_1) + ( S_2^2/n_2))/ (((S_1^2/n_1)/( n_1 -1)) + (( S_2^2/n_2) /( n_2 -1)))$ (Also called Satterthwaite df).
**Confidence Interval (CI):** (1-$\alpha$) 100% CI for $\mu_1$- $\mu_1$: $(\bar{Y}_1 - \bar{Y}_2) \pm t_{\alpha/2, df} * S_{(Y1bar - Y2bar)}$
**Standard Error Difference of Means ($S_{(Y1bar - Y2bar)}$)** = $\sqrt{((MSE_{AOV}/n_1) + (MSE_{AOV}/n_2))}$
**Conclusion:** Reject $H_o$ if $|t| \geq t_{\alpha/2, df}$; df is calculated.

**HOV TESTS: How to know variances are equal ($\sigma_1^2 = \sigma_2^2$).**
**Hypothesis:** $H_o$: **$\sigma_1^2 = \sigma_2^2$** & $H_A$: **$\sigma_1^2 \neq \sigma_2^2$**
**Test Statistics:** F = $S_1^2/ S_2^2$; df = num. $n_1 - 1$ & den. $n_2 - 1$
**Conclusion:** $F \leq 1/F_{\alpha/2, n1-1, n2-2}$ OR $F_{calc} \geq F_{\alpha/2, n1-1, n2-2}$
**Two N. Popl$^n$, $\sigma_1^2 \neq \sigma_2^2$, and Dep$^{endent}$, Random Samples**:
**Hypothesis:** $H_o$: $\mu_1 = \mu_2$ & $H_A$: $\mu_1 \neq \mu_2$
**Test Statistics:** t-test = $(\bar{Y}_1 - \bar{Y}_2)/(S_d/\sqrt{n})$; $S_d$ = Standard deviation of all differences across the pairs.

**ANOVA Table and ANOVA Manual Computation:**

One treatment (T) with i levels.
**Assumptions:** $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$, iid and random sample.
**Hypothesis:** $H_o$: $\mu_1 = \mu_2 = \ldots = \mu_t$ & $H_A$: at least one is different

| Source | df | SS | MS | F |
|---|---|---|---|---|
| **Trt.** | t-1 | SSTrt | SSTrt/(t-1) | MSTrt/MSE |
| **Error** | n. – t | SSE | SSE/(n.-t) | |
| **Total** | n. – 1 | TSS | | |
| **Conclusion = Reject $H_o$ if $F_{calc} \geq F_{(\alpha, t-1, n.-t)}$** | | | | |

| **ANOVA Manual Calculation Example** | | | | |
|---|---|---|---|---|
| | **Trt. = Variety** | | | |
| | 1 | 2 | 3 | |
| | 36 | 57 | 50 | |
| | 33 | 53 | 41 | |
| | 48 | 43 | 47 | |
| | | 54 | 42 | |
| | | 48 | | |
| **$Y_{i.}$** (Sum) | 117 | 255 | 180 | $Y_{..}$ = 552 |
| **$n_i$** (Count) | 3 | 5 | 4 | n. = 12 |
| **Mean, $\bar{Y}_{i.} = Y_{i.}/ n_i$** | 39 | 51 | 45 | $\bar{Y}_{..}$ = 46 |
| **$S_i^2$ =** Sum $(Y_{i.} - \bar{Y}_{i.})^2/( n_i - 1)$ | 63 | 30.5 | 18 | |

$S_1^2 = ((36\text{-}39)^2 + (33\text{-}39)^2 + (48\text{-}39)^2)/ (3\text{-}1) = 63$
$\sum(Y_{ij}^2) = 36^2 + 57^2 + 50^2 + \ldots. + 54^2 + 42^2 + 48^2 = 25{,}970$
**Total Sum of Squares (TSS)** = $25970 - 552^2/12 = 578$
**SSTrt** = $(117^2/3 + 255^2/5 + 180^2/4) - 552^2/12 = 276$
**SSTrt** = $3(39\text{-}46)^2 + 5(51\text{-}46)^2 + 4(45\text{-}46)^2 = 276$
**SSE** = 2(63) + 4(30.5) + 3(118) = 302; **SSE** = 785 – 276 = 203
**MSE** = Pooled Variance ($\sigma_2^2$) Estimate ($S_p^2$) for two sample.

| **ANOVA or AOV Table** | | | | |
|---|---|---|---|---|
| **Source** | df | SS | MS | F |
| **Variety** | 2 | 276 | 276/2 = 138 | 4.1126 |
| **Error** | 9 | 302 | 203/9 = 33.556 | |
| **Total/Adj** | 11 | 578 | | |
| $F_{calc}$ = 8.1576 & $F_{(0.05, 2, 9)}$ from table = 4.26. 0.05 < p < 0.10. (p = 0.0538; Casio fx-9750GII) **Conclusion:** Failed to reject $H_o$ and yield does not differs across three variety of wheat ($\alpha$ = 0.05, $F_{2,9}$ = 4.1126, 0.05 < p < 0.10 (p =0.0538) | | | | |

**Standard Error of Means** = $\sqrt{(MSE/n_i)} = \sqrt{(33.5556/3)}$ for Variety 1 & $\sqrt{(33.5556/4)}$ for variety 3.
**Confidence Interval (CI)** (One population)**:** (1-$\alpha$)100% CI for $\mu$: $Y_{bar} \pm t_{\alpha/2, df} * S_{(Ybar\ or\ Mean)} = Y_{bar} \pm t_{\alpha/2, df} *\sqrt{(MSE/n_i)}$
**St. Err. of Mean** ($S_{(Y1bar - Y2bar)}$) = $\sqrt{(MSE (1/n_1 + 1/n_2))}$
**Confidence Interval (CI) (Two population):** (1-$\alpha$) 100% CI for $\mu_1$- $\mu_1$: $(\bar{Y}_1 - \bar{Y}_2) \pm t_{\alpha/2, df} * S_{(Y1bar - Y2bar)} = (\bar{Y}_1 - \bar{Y}_2) \pm t_{\alpha/2, df} * \sqrt{(MSE (1/n_1 + 1/n_2))}$

**Completely Randomized Design (CRD)** Assumptions: $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$, iid and random sample. **Experimental error** is the measure of

variation among experimental units treated alike. In other words, even though subjects are treated with same treatments in one level, their performances differ due to factors uncontrolled.

**Statistical Models for Various Expt. Designs**

**1) Statistical Means Model:** $Y_{ij} = \mu_i + \varepsilon_{ij}$

$E(Y_{ij}) = E(\mu_i) + E(\varepsilon_{ij}) = E(\mu_i) + 0 = \mu_i$

$Var(Y_{ij}) = Var(\mu_i) + Var(\varepsilon_{ij}) = 0 + \sigma_\varepsilon^2 = \sigma_\varepsilon^2$

**2) Statistical Treatment Effect Models:** $Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$

$E(Y_{ij}) = E(\mu) + E(\tau_i) + E(\varepsilon_{ij}) = \mu + \tau_i + 0 = \mu + \tau_i$

$Var(Y_{ij}) = Var(\mu) + Var(\tau_i) + Var(\varepsilon_{ij}) = 0 + 0 + \sigma_\varepsilon^2 = \sigma_\varepsilon^2$

Where $E(Y_{ij})$ = Expectation (mean) of $Y_{ij}$

i = 1, 2, …, t and j = 1, 2, … $n_j$

$Y_{ij}$ = jth observation of treatment level i

$\tau_i$ = treatment level i effect. This is also fixed.

$\varepsilon_{ij}$ = random error due to expt. unit j in trt. level i.

**W1P2: Homogeneity of Variances (HOVs):**

**$H_o$:** $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_t^2 = \sigma_\varepsilon^2$ & **$H_a$:** At least one is different.

1) **Hartley's F-max Test:** requires $n_1 = n_2 = \ldots = n_t$

**Test Statistics: F-max = $S^2_{max}/ S^2_{min}$**

**Conclusion:** Reject $H_o$ if $F\text{-max}_{calc} \geq F\text{-max}_{\alpha, n-1}$; n is common sample. Table A5 for Critical Values for F-max test.

**2) Bartlette's Test:** $\chi^2$-test. **Decision:** Reject $H_o$: $\chi^2_{calc} \geq \chi^2_{\alpha, t-1}$

**3) Levene's Test:** AOV on absolute residuals. $Z_{ij} = |Y_{ij} - Y_{bar}|$ **Decision:** Reject $H_o$ if $F \geq F_{\alpha, t-1, n.-t}$

**4) $\chi^2$Likelihood Ratio Test** ($\chi^2$ LRT)**:** Decision: Reject $H_o$: $\chi^2_{calc} \geq \chi^2_{\alpha, t-1}$

If HOVs are significant, transform data, use non-parametric tests such as Kruskal-Wallis test. If $H_o$: $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_t^2 = \sigma_\varepsilon^2$ is not rejected, use usual ANOVA on $Y_{ij}$'s. MIXED Procedure in SAS if data is non-normal.

**W1P3: LINEAR CONTRAST: Difference is there, but where?**

Random samples $Y_1, Y_2, \ldots Y_n$ from a population with vari. $\sigma^2$

$Var(\sum Y_i) = Var(\sum_{i=1}^{n} \sigma^2) = \sum_{=0}^{n} \sigma^2 = n*\sigma^2$

$Var(\sum a_i Y_i) = \sum a_i^2 Var(Y_i) = \sigma 2 \sum a_i^2$

$Var(\bar{Y}) = Var(\sum_{i=1}^{n} \frac{1}{n^2} Y_i) = (\sum_{i=1}^{n} \frac{1}{n^2}) Var(Y_i) = \frac{1}{n^2} n \sigma^2 = \sigma^2/n$

If $Y_{ij} = \mu_i + \varepsilon_{ij}$ then and $Var(\varepsilon_{ij}) = \sigma^2$, then $Var(Y_{ij}) = \sigma_t^2$.

Above conditions holds true for normal and non-normal distribution.

Let, $\hat{L} = \sum a_i \bar{Y}_{i.}$ where $a_i$ is a constant, L estimates $\sum a_{i\,i.} \mu_{i.}$

**Var($\hat{L}$), ($S^2_{Lhat}$)** = $\sigma_\varepsilon^2 * \sum_{i=1}^{n} (\frac{ai^2}{ni}) = MSE * \sum_{i=1}^{n} (\frac{ai^2}{ni})$

**Standard Error of $\hat{L}$ ($S_{Lhat}$)** = $\sqrt{Var(\hat{L})} = \sqrt{(MSE * \sum_{i=1}^{n} (\frac{ai^2}{ni}))}$

$\hat{L} \sim N(\hat{L}, Var(\hat{L}))$ ➔ $\sum a_i \bar{Y}_{i.} \sim N(\sum a_i \bar{\mu}_{i.}, \sqrt{(MSE * \sum_{i=1}^{n} (\frac{ai^2}{ni}))}$

**T-test and F-test on $\hat{L}$:**

**$H_o$: L=** $\sum a_{i.} \mu_{i.} = 0$ & **$H_a$: L=** $\sum a_{i.} \mu_{i.} \neq 0$ **(Mean Model)**

**$H_o$: L=** $\sum a_{i\,i.} \tau_{i.} = 0$ & **$H_a$: L=** $\sum a_{i\,i.} \tau_{i.} \neq 0$ **(Effect Model)**

**Test Stat.: T-test** = $\hat{L}$/St. Err. = $\sum a_i \bar{Y}_{i.} / \sqrt{(MSE * \sum_{i=1}^{n} (\frac{ai^2}{ni}))}$

**Conclusion:** Reject if $|t| \geq t_{\alpha/2, df\ of\ err}$

**Test Statistics: F-test** = (SSL/1)/MSE; **SSL** = $\bar{L}^2 / \sum_{i=1}^{n} (\frac{ai^2}{ni})$

**Conclusion:** Reject if $F_{calc} \geq F_{\alpha, 1, df\ of\ MSE}$

**Contrast Comparison:**

| Trts = 5 ➔ Reps = 10 | Trt 1 (Cont) | Trt 2 (G) | Trt 3 (F) | Trt 4 (GF) | Trt 5 (S) |
|---|---|---|---|---|---|
| 10 reps. | $\mu_1$ | $\mu_2$ | $\mu_3$ | $\mu_4$ | $\mu_5$ |
| $\sum Y_{ij}$ | 701 | 593 | 582 | 580 | 641 |
| $\bar{Y}$ (Mean) | 70.1 | 59.3 | 58.2 | 58 | 64.1 |
| MSE = 5.4556 & Err df = 45 from ANOVA Table | | | | | |

**Five treatments with five $\mu_s$. Q: Does $\mu_1$ differs from rest.**

**$H_o$:** $\mu_1 = \mu_2 + \mu_3 + \mu_4 + \mu_5$ or $\mu_1 - (\mu_2 + \mu_3 + \mu_4 + \mu_5) = 0$

**$H_a$:** $\mu_1 \neq \mu_2 + \mu_3 + \mu_4 + \mu_5$ or $\mu_1 - (\mu_2 + \mu_3 + \mu_4 + \mu_5) \neq 0$

$L_1 = \mu_1 - (\mu_2 + \mu_3 + \mu_4 + \mu_5)/4 = 0$; $a_1 = 1, a_2 = a_3 = a_4 = a_5 = ¼$

$\hat{L} = \sum a_i \bar{Y}_i$ :: $\hat{L} = 70.1*1 - (59.20 + 58.20 + 58.0 + 64.10)/4 = 10.2$

**St. Err. of $\hat{L}$** = $\sqrt{(MSE * \sum_{i=1}^{n} (\frac{ai^2}{ni}))} = \sqrt{(5.4556 * (1/10) * (1^2 + 4(1/4)^2)} = 0.8253$. **SSL** = $\bar{L}^2 / \sum_{i=1}^{n} (\frac{ai^2}{ni})$ $n_i$ = # of reps or obs.

**T-test:** $\hat{L}$/ St. Err. of $\hat{L}$ ➔ $t_{calc\ 45} = 10.2/0.8258 = 12.3516$ & $t_{cric\ (0.025, 45)} = 2.0141$ **Conclusion:** $H_o$ Rejected (treatment differs) as $|t_{calc}| \geq t_{cric\ (0.025, 45)}$. ($\alpha = 0.005$, $t_{calc\ 45} = 12.35$, $p < 0.001$).

**F-test:** (SSL/1)/MSE; **SSL** = $10.2^2/((1/10)*1.25) = 832.32$ $F_{calc} = (832.32/1)/5.4556 = 152.5625$. $F_{cric(0.05, 1, 45)} = 4.06$

**Conclusion:** $H_o$ Rejected (treatment differs) (Same)

**Confidence Interval** (One Population)**:** $\hat{L} \pm t_{\alpha/2, df,\ St.\ Err.\ of\ Lhat}$

**Confidence Interval** (Two Population)**:** $\hat{L} \pm t_{\alpha/2, df,\ St.\ Err.\ of\ Lhat} = (a_1\bar{Y}_1 - a_2\bar{Y}_2) \pm t_{\alpha/2, df} * \sqrt{(MSE\ (1/n_1 + 1/n_2))}$.

**Orthogonality** ensures no overlapping information among the contrasts: $\hat{L_1} = \sum a_i \bar{Y}_{i.}$ & $\hat{L_2} = \sum b_i \bar{Y}_{i.}$ are two contrast estimates.

**$\hat{L_1}$ & $\hat{L_2}$ orthogonal, if** $\sum_{i=1}^{t} a_1\ b_1 = a_1b_1 + a_2b_2 + \ldots + a_tb_t = 0$

Eg: **A.** $\mu_1 - \mu_2$ ➔ $a_1 = 1, a_2 = -1, a_3 = 0$, **B.** $\mu_1 - \mu_3$ ➔ $a_1 = 1, a_2 = 0\ a_3 = -1$. **C.** $(\mu_1 - \mu_3)/2 - \mu_2$. ➔ $a_1 = ½, a_2 = -1\ a_3 = 1/2$ **Check: 1) A & B** orthogonal: (1)(1) + (-1)(0)+ (0)(-1) ≠ 0. So not orthogonal. **2) B & C** Orthogonal: (1)(1/2) + (0)(-1)+(-1)(1/2) = 0 is orthogonal.

**Mutually Orthogonal** if all pairs of contrasts in the set are orthogonal.

**Note:** if sample sizes are same, then for T treatment levels, t-1 orthogonal contrast can be found such that $\sum_{i=1}^{t} SSL$ = SSTrt. For t treatments, there are many mutually orthogonal contrasts sets.

**W1P4: Simultaneous Contrast: Hypothesis: $H_o$:** $\mu_2 = \mu_3 = \mu_4 = \mu_5$ & **$H_A$:** at least $\mu_i$ is different for I ≥ 2. ➔ (Sum of Squares for $H_o$) $SSH_o$ requires Matrix Calculation; USE SAS. $SSH_o$ df = minimum number of contrasts required to express the simultaneous equality of means in $H_o$

**Test Statistics:** F = $(SSH_o/3)/MSE$ (3 because df = 3 of $H_o$)

**Test Statistics:** F = $(SSH_o/df_{H0})/MSE$

**Conclusion:** Reject $H_o$ if $F_{calc} \geq F_{cric\ (\alpha,\ df_{H0},\ df\ of\ MSE)}$

**Note:** No T-test in Simultaneous Contrast

**/* CONTRAST SAS CODE*/**

```
DM 'LOG; CLEAR; ODSRESULTS; CLEAR;';
TITLE 'SAS FOR CONTRAST: LINEAR AND SIMULTANEOUS';
TITLE2 'S = SUCROSE, G = GLUCOSE F = FRUCTOSE';
DATA one;
INPUT trt $ trt2 length @@;
DATALINES; C 1 75 F 3 58 FG 4 58 G 2 57 S 5 62;
PROC GLM DATA = one PLOTS = (RESIDUALS DIAGNOSTICS);
CLASS trt;
MODEL length = trt;
/* CONTRAST: LINEAR CONTRAST */
MEANS trt / HOVTEST = LEVENE (TYPE = ABS);
CONTRAST 'Control vs Others'        trt 4 -1 -1 -1 -1;
CONTRAST 'Pure vs Mixed'            trt 0 1 -2 1 0;
CONTRAST '12 C vs 6 C'              trt 0 1 0 1 -2;
CONTRAST 'Glucose vs Fructose'      trt 0 1 0 -1 0;
ESTIMATE '1 Control vs Others'      trt 4 -1 -1 -1 -1;
/* estimate 4 x control - sugar total */
ESTIMATE '2 Control vs Others'      trt 1 -.25 -.25 -.25 -.25;
/* estimates control mean - sugar mean */
ESTIMATE '3 Control vs Others'      trt 4 -1 -1 -1 -1/DIVISOR=4;
ESTIMATE 'Pure vs Mixed'            trt 0 1 -2 1 0;
ESTIMATE '12 C vs 6 C'              trt 0 1 0 1 -2;
ESTIMATE 'Glucose vs Fructose'      trt 0 1 0 -1 0;
TITLE 'Tests and Estimates of Linear Contrasts';
TITLE2 'Note the order of the treatments when contrasts are being written';
RUN; QUIT;
```

**/* CONTRAST: SIMULTANEOUS CONTRAST */**

```
/* The Following Three Contrasts are Orthogonal and the sum of squares from
each to test. H0: F = FG = G = S */
CONTRAST 'F = FG'                   trt 0 1 -1 0 0;
CONTRAST 'F + FG = 2G'              trt 0 1 1 -2 0;
CONTRAST 'F + FG + G = 3S'          trt 0 1 1 1 -3;
/* Note: Following contrasts imply H0: F = FG = G = S when the contrasts are
simultaneously true, but the Sums of Squares do not add to SSH0. */
CONTRAST 'F = FG'                   trt 0 1 -1 0 0;
CONTRAST 'F = G'                    trt 0 1 0 -1 0;
CONTRAST 'F = S'                    trt 0 1 0 0 -1;
/* Either of the above sets can be tested simultaneously to obtain the test of H0:
F = FG = G = S. Notice COMMA */
CONTRAST '1. F = FG = G = S '       trt 0 1 -1 0 0,
/* [Contrast: Row Sum = 0] */       trt 0 1 1 -2 0,
                                    trt 0 1 1 1 -3;
CONTRAST '2. F = FG = G = S '       trt 0 1 -1 0 0,
```

```
/*[Comma: Test simultaneously] */        trt 0 1 0 -1 0,
                                         trt 0 1 0 0 -1;
/* Many contrast sets can test above H0 simultaneously */
TITLE 'Simultaneous Contrasts'; RUN; QUIT;
```

**TREND ANALYSIS (Orthogonal Polynomial Contrasts)**
**Hypothesis:** $H_o$: $\mu_1 = \mu_2 = \mu_3$ & $H_A$: at least one is different
**Test Statistics: (F-test)** = $SS/df_{Ho}/MSE$
**Decision Rule:** Reject $H_o$ if $F_{\alpha, df_{Ho}, MSE} \geq F_{calc}$

**Simple Linear Regression:** $Y_i = \beta_o + \beta_1 X_1 + \varepsilon_i$ Where, $\beta_o$ = Y-intercept & $\beta_1$ is slope.
**Polynomial Regression:**
**Quadratic Trend:** $Y_i = \beta_o + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_3 + \varepsilon_i$
**Cubic Trend:** $Y_i = \beta_o + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + \varepsilon_i$
**Quartic Trend:** $Y_i = \beta_o + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + \beta_3 X_i^4 + \varepsilon_i$

**/* TREND ANALYSIS SAS CODE */**

```
DATA one;
INPUT display sales @@; /* Treatment = display */
DATALINES; 10 0.778 15 .665 20 .973 25 1.003 30 1.125;
PROC GPLOT DATA=one;
PLOT sales*display / HAXIS = 10 15 20 25 30;
SYMBOL1 VALUE = TRIANGLE CV = BLUE I = NONE;
TITLE 'Trend Analysis Example. GPLOT for Graphing';
PROC SGPLOT DATA = one;
SCATTER Y = sales X = display;
TITLE2 'SGPLOT Option for Statistical Graphing';
PROC GLM DATA = one;
CLASS display;
MODEL sales = display;
TITLE2;
/* Trend Values comes from Orthogonal Polynomials Table P. 740 */
CONTRAST 'Linear Trend'                 display   -2 -1 0 1 2;
CONTRAST 'Quadratic Trend'              display    2 -1 -2 -1 2;
CONTRAST 'Cubic Trend'          display   -1 2 0 -2 1;
CONTRAST 'Quartic or LOF'       display    1 -4 6 -4 1;
LSMEANS display / STDERR; RUN; QUIT;
```

**W2P1: Multiple Comparisons (Post-Hoc Comparisons)**: **If means are different, which mean is different?**
**Experimentwise error rate ($\alpha_e$):** Probability of at least one Type I error in all test performed. **Comparisonwise error rate ($\alpha_c$):** Probability of a Type I error in a single test of hypothesis.

**(Fishers LSD) Least Significance Difference: when ANOVA F is significant,** does not control experiment wise error rate.
To compare group i and i', compute **LSD** = $t_{\alpha/2, df of err} * \sqrt{(MSE\ (1/n_1 + 1/n_2))}$. If $|\mathbf{Y_i} - \bar{Y}_{I'}| \geq$ LSD, then Conclude $\mu_i \neq \mu_{i'}$
**If $n_i = n$ for all i, then LSD** = $t_{\alpha/2, df of MSE} * \sqrt{(MSE\ (2/n))}$
**(1-α) 100% CI** for $\mu_i = \mu_{i'}$ is $(\bar{Y}_i - \bar{Y}_{I'}) \pm$ LSD
**Hypothesis:** $H_0$: $\mu_i - \mu_{i'} = 0$ $i \neq i'$ $H_a$: $\mu_i - \mu_{i'} \neq 0$
**Test Statistics: $t = L_{hat}/S_{Lhat} = (\mathbf{Y_i} - \bar{Y}_{I'})/\sqrt{(MSE\ (1/n_1 + 1/n_2))}$**
**Decision Rule:** Reject $H_o$ if $|t| \geq t_{\alpha/2, df of err}$ Or $|\mathbf{Y_i} - \bar{Y}_{I'}| \geq$ LSD.
Computation of Test Statistics (DO THE MATH AGAIN)

**Scheffe Method:** Allows **tests and CI's for all types of contrasts** and control experimentwise error rate. SAS: Minimum Significance Difference (MSD).
**Hypothesis:** $H_0$: $\mu_i - \mu_{i'} = 0$ $i \neq i'$ $H_a$: $\mu_i - \mu_{i'} \neq 0$ OR
**Hypothesis:** $H_0$: $\mathbf{L}\left(\sum a_i \mu_i\right) = 0$ $H_a$: $\mathbf{L}\left(\sum a_i \mu_i\right) \neq 0$
**Test Stat:** Scheffe Value = $\sqrt{*((\#trt-1)(F_{\alpha, t-1, df of err}))} * \sqrt{(MSE * \sum_{i=1}^{n} (\frac{a_i^2}{n_i}))}$
**Conclusion:** If $|\hat{L}_t| \geq$ Scheffe Value, Reject $H_o$
**(1-α) 100% Scheffe CI for L $\left(\sum a_i \mu_i\right)$** = $\hat{L} \pm$ Scheffe Value.
Computation of Test Statistics: (DO THE MATH AGAIN)

**(HSD) Tukey Honest Significant Difference Method:** Application to all pairwise difference. Controls experimentwise error rate.
**Hypothesis:** $H_0$: $\mu_i - \mu_{i'} = 0$ $i \neq i'$ $H_a$: $\mu_i - \mu_{i'} \neq 0$ OR
**Hypothesis:** $H_0$: $\mathbf{L}\left(\sum a_i \mu_i\right) = 0$ $H_a$: $\mathbf{L}\left(\sum a_i \mu_i\right) \neq 0$
**Test Stat: Tukey HSD** = $q_{\alpha, \#\ of\ trt, df of err} * \sqrt{(MSE/n)}$ assume $n_i = n\ \forall$ i.
Critical values of $q_{\alpha, \#\ of\ trts, df of err}$ is in table A7 of text book.
**Conclusion:** Reject $H_o$ if $(Y_i - \bar{Y}_{I'})_{calc} \geq HSD_{cric}$
**100% Tukey CI for $\mu_i - \mu_{i'}$** = $(Y_i - \bar{Y}_{I'}) \pm$ HSD

**Tukey Method with Unequal Sample Size: 1) Use** $\sqrt{(1/2(1/n_i + 1/n_{i'})}$ in place of $1/\sqrt{n}$ in HSD. HSD = $q_{\alpha, t, df of err} \sqrt{MSE} * \sqrt{(1/2(1/n_i + 1/n_{i'})}$ . **2)** Let $n = \min(n_i, n_{i'})$ ➔ Standard Error is bigger and 3) let $n = 1/(\frac{1}{t}\sum_{i=1}^{t} \frac{1}{n_i})$

**Summary: Fischers LSD:** Pairwise comparison only and experiment wise error is not controlled. **Scheffe's Method:** Very conservative (finds fewer differences), pairwise comparison and general contrasts, not recommended for orthogonal contrasts,, useful and recommended for contrasts and suggested by data. **Tukey's Method:** Pairwise comparison method only, not as conservative as Scheffe's method. All three have CI method. **Note:** LSD < Tukey < Shefee's Method.

**MULTIPLE COMPARISON: POST HOC TEST:**

```
DATA one; INPUT trt $ wear @@;
DATALINES; A 9.16 B 11.95 C 11.47 D 11.35;
PROC GLM DATA=one PLOTS=NONE;
CLASS trt; /* Class is for Treatment */
MODEL wear = trt;
MEANS trt / < OPTIONS >;
RUN; QUIT; Note: <OPTIONS>: ADJUST = T or SCHEFFE or TUKEY or , BON,
DUNNETT, DUNNETTL, LINES, CLDIFF, CLM, ALPHA = P, SLICE
```

**W2P2 & W2P3: FACTORIAL EXPERIMENTS:** Used for two treatments with their own # of levels of treatment. **Advantages:** Same number of observations in each pairwise differences, computing first treatment (A) cancels the effect of second treatment (B), more info few subjects, combined effects. **Interaction:** The difference in mean between levels of 1st treatment is not consistent across the levels of the 2nd treatment.
**Statistical Models for Factorial Expt. Designs**
**Treatment Structure:**
a*b factorial ➔ ab treatment combinations.
a*b*c factorials. ➔ abc treatment combinations.
**Design Structure:** CRD with $n_{ij}$ replications (a*b factorial)
**Statistical Means Model:** $Y_{ijk} = \mu_{ij} + \varepsilon_{ijk}$
**Statistical Treatment Effect Models** $Y_{ikj} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk}$
Where: $Y_{ijk}$ = $k^{th}$ observation level $A_iB_j$
i (trtA) = 1, 2, …, a, j (trtB) = 1, 2, … b & k (obs)= 1, 2, …, $n_{ij}$
$\mu_{ij} = A_iB_j$ mean.
$\mu$ = Overall mean.
$\alpha_i$ = Treatment 'A' level i effect.
$\beta_j$ = Treatment 'B' level j effect.
$\alpha\beta_{ij}$ = Interaction effect of $A_iB_j$
$\varepsilon_{ijk}$ = random error due to $k^{th}$ expt. unit of $A_iB_j$. $\varepsilon_{ijk} \sim N(0, \sigma_\varepsilon^2)$

**For a*b Factorial in a CRD, the source of variation and df are:**

| | | Treatment A | | |
|---|---|---|---|---|
| | | **A1** | **…** | **Aa** |
| Treatment B | **B1** | $\mu_{11}$ | … | $\mu_{a1}$ |
| | **…** | … | … | … |
| | **Bb** | $\mu_{1b}$ | … | $\mu_{ab}$ |

**ANOVA Table and Hypothesis:**
**Hypothesis:** Mean Model:
$H_o$: $\mu_{bar1.} = \ldots = \mu_{bara.}$ & $H_A$: at least one $\mu_{bar1.}$is different
$H_o$: $\mu_{bar.2} = \ldots = \mu_{bar.b}$ & $H_A$: at least one $\mu_{bar.2}$ is different
$H_o$: $\mu_{ij} - \mu_{i'j} = \mu_{ij'} - \mu_{i'j'}$ & $H_A$: at least one $\mu_{ij}$ is different where $i \neq i'$ & $j \neq j'$

**Hypothesis:** Effects Model:
$H_o$: $\alpha_1 = \ldots = \alpha_a$ & $H_A$: at least one $\alpha_a$ is different
$H_o$: $\beta_1 = \ldots = \beta_b$ & $H_A$: at least one $\beta_b$ is different
$H_o$: No Interaction, & $H_A$: There is interaction.

| Source | df | SS | MS | F |
|---|---|---|---|---|
| **Trt. A** | a-1 | $SS_A$ | $MS_A = SS_A/df$ | $MS_A/MSE$ |
| **Trt. B** | b -1 | $SS_B$ | $MS_B = SS_b/df$ | $MS_B/MS_E$ |
| **Interact.** | (a-1) *(b-1) | $SS_{AB}$ | $MS_{AB} = SS_{AB}/df$ | $MS_{AB}/MS_E$ |
| **Error** | ab(n-1) | $SS_E$ | $MS_E =$ | |

| | | | $SS_E/df$ |
|---|---|---|---|
| **Total** | abn – 1 | TSS | |
| **Conclusion = Reject $H_o$ if $F_{calc} \geq F_{(\alpha, t-1, abn-1)}$** | | | |
| **P-value range ➔** vary α find range of α where $F_{(\alpha, t-1, df(MSE))}$ fits in. | | | |

**Note:** If interaction is significant, main effects are dependent. So, compare means across treatment combinations. If there is no significant interaction, interpret main effects.

$\overline{Yi..} = \overline{\mu i.} = \sum_{j=1}^{b} Ybarij./b$ Std. Err of $\overline{Yi..} = \sqrt{(MSE/nb)}$ (Trt A)

$\overline{Y.j.} = \overline{\mu.j} = \sum_{i=1}^{a} Ybarij./a$ Std. Err of $\overline{Yi..} = \sqrt{(MSE/na)}$ (Trt B)

**Combination Means (LSD)** = $\sqrt{(MSE*(2/n))}$

**Contrast and Comparison:** If ANOVA test for interaction is significant, then compare means among treatment combinations.
We assume $n_{ij} = n$ for all i & j.
**Hypothesis:** $H_0$: **L (** $\sum a_{ij}\mu_{ij}) = 0$ $H_a$: **L (** $\sum a_{ij}\mu_{ij}) \neq 0$
$\hat{L} = \sum a_{ij}\bar{Y}_{ij}$ estimates **(** $\sum a_{ij}\mu_{ij}$) with
**Standard Error of $\hat{L}$ ($S_{Lhat}$)** = $\sqrt{}$ Var($\hat{L}$) = $\sqrt{}$ (MSE*$\sum_{i,j} \frac{a_{ij}^2}{n}$))
**Test Statistics:**
T-test = $L_{hat}$/Sd. Err fo $L_{hat}$
F-stat = (SSL/1)/MSE
**SSL** = $\bar{L}^2/\sum_{i=1}^{n}(\frac{aij^2}{n})$ n = # of observation.

## COMPARE AND CONTRAST: FACTORIAL TREATMENT STRUCTURE

**Fischer LSD** = $t_{\alpha/2, df\ of\ err} * S(\bar{Y}_i - \bar{Y}_{i'}) = t_{\alpha/2, df\ of\ err} * \sqrt{(MSE\ (2/n))}$.
Where i ≠ i' Or j ≠ j' or Both.
**If $n_i = n$ for all i, then LSD** = $t_{\alpha/2, df\ of\ MSE} * \sqrt{}$ (MSE (2/n))
**(1-α) 100% CI** for $\mu_i = \mu_{i'}$ is $(\bar{Y}_i - \bar{Y}_{i'}) \pm$ LSD
**Decision Rule:** $|\mathbf{Y_i - \bar{Y}_{i'}}| \geq$ LSD then $A_iB_j$ and $A_{i'}B_{j'}$ have different means.

**Tukey HSD: Test Stat: Tukey HSD** = $q_{\alpha, \#\ of\ trt, df\ of\ err} * \sqrt{(MSE/n)}$ assume $n_i = n\ \forall$ i. Critical values of $q_{\alpha, \#\ of\ trts, df\ of\ err}$ is in table A7 of text book.
**Conclusion:** Reject $H_o$ if $(Y_i - \bar{Y}_{i'})_{calc} \geq HSD_{cric}$ then $A_iB_j$ and $A_{i'}B_{j'}$ have different means.
**100% Tukey CI for $\mu_i - \mu_{i'}$** = $(Y_i - \bar{Y}_{i'}) \pm$ HSD

**Scheffe Method:**
**Scheffe Value** = $\sqrt{((ab-1)(F_{\alpha, t-1, df\ of\ err}))} *$ St. Err. of $\hat{L}$
= $\sqrt{((ab-1)(F_{\alpha, t-1, df\ of\ err}))} * \sqrt{(MSE*\sum_{i=1}^{n}(\frac{a_i^2}{n_i}))}$
**Conclusion:** If $|\hat{L}_t| = |\sum a_i\bar{Y}_i| \geq$ Scheffe Value, Reject $H_o$
**(1 - α) 100% Scheffe CI for L (** $\sum a_i\mu_i$**)** = $\hat{L} \pm$ Scheffe Value.

| Interaction is not significant ➔ Compare means for sig. main effects. | |
|---|---|
| Treat A has 'a' levels and n*a observations per level | Treatment B has 'b' levels and n*b observations per level. |
| $H_0$: **L = (** $\sum a_i\mu_{i.}) = 0$ | $H_0$: **L = (** $\sum b_j\mu_{.j}) = 0$ |
| $\widehat{La} = \sum a_i\bar{Y}_{i..}$ & St. Er. of $\hat{L}$ = $\sqrt{\{MSE*(1/nb)*\Sigma a_i^2\}}$ | $\widehat{Lb} = \sum b_j\bar{Y}_{.j.}$ & St. Er. of $\hat{L}$ = $\sqrt{\{MSE*(1/na)*\Sigma b_j^2\}}$ |
| **Fischer $LSD_A$** = $t_{\alpha/2, df\ of\ err} * \sqrt{(MSE\ (2/nb))}$. | **Fischer $LSD_B$** = $t_{\alpha/2, df\ of\ err} * \sqrt{(MSE\ (2/na))}$. |
| **Tukey $HSD_A$**= $q_{\alpha, a, df\ of\ err} * \sqrt{(MSE/bn)}$ | **Tukey $HSD_B$**= $q_{\alpha, b, df\ of\ err} * \sqrt{(MSE/an)}$ |
| **Scheffe$_A$** = $\sqrt{((a-1)(F_{\alpha, a-1, df\ of\ err}))}$* St. Err. of $\hat{L}_a$ | **Scheffe$_B$** = $\sqrt{((b-1)(F_{\alpha, a-1, df\ of\ err}))}$* St. Err. of $\hat{L}_b$ |

**Interaction Plots or Profile Plots:** diagnostic tool used to access interaction between treatments. Illustrates consistency or lack of consistency of the simple effects. Departure from parallelism illustrates interaction. Do formal F-test to determine interaction.

## TEST OF SIGNIFICANT INTERACTION in FACTORIAL DESIGN

Study: Influence of **time of bleeding & diethylstilbestrol (estrogen)** on **plasma phospholipid** in lambs. **Five lambs** were assigned at random to each of the **four treatment groups**; treatment combinations: **morning and afternoon times** of bleeding **with & without estrogen** treatment. **Data:**

| Treatment Groups | | | | |
|---|---|---|---|---|
| AM (1) | | (b) | PM (2) | |
| Control (1) | Treated (2) | (a) | Control (1) | Treated (2) |
| 8.53 | 17.53 | (y) | 39.14 | 32.00 |
| 20.53 | 21.07 | (y) | 26.20 | 23.80 |
| 12.53 | 20.80 | (y) | 31.33 | 28.87 |
| 14.00 | 17.33 | (y) | 45.80 | 25.06 |
| 10.80 | 20.07 | (y) | 40.20 | 29.33 |

```
DM 'LOG; CLEAR; ODSRESULTS; CLEAR; ';
TITLE; DATA one; INPUT a b y @@;
LABEL A='Treatment' B='Time' Y='Phospholipid';
DATALINES; 1 1 8.53 ... 1 2 17.53 ... 2 1 39.14 ... 2 2 32.00;
PROC PRINT DATA = one; RUN;
PROC GLM DATA = one PLOTS (ONLY) = INTPLOT ;
CLASS a b; /* Treatments. Changing order changes Interaction */
MODEL y = a b a*b;

/* Contrast Not Important for Class */
CONTRAST 'A1-A2 WITHIN CONTROL'      a 1 -1 a*b 1 0 -1 0;
CONTRAST 'A1-A2 WITHIN TREATED'      a 1 -1 a*b 0 1 0 -1;
CONTRAST 'B1-B2 @ AM'                b 1 -1 a*b 1 -1 0 0;
CONTRAST 'B1-B2 @ PM'                b 1 -1 a*b 0 0 1 -1;

MEANS a*b/LSD LINES; /*Main Effects of a & b*/
LSMEANS a*b / STDERR ADJUST=T PDIFF LINES SLICE = (a b);
TITLE ' DATA FROM TABLE 15.2 -- STEEL, TORRIE, AND DICKEY';
/* GENERATE INTERACTION OR PROFILE PLOT; */
PROC SORT DATA=one; BY a b;
PROC MEANS DATA=one NOPRINT;  BY a b; VAR y;
OUTPUT OUT = two MEAN = abmean; /*Gives dataset that have treatment
combination means. Dataset 'two' used below */
PROC GPLOT DATA=two; PLOT abmean*a=b;
/* # of SYMBOL LINES = # OF LEVELS OF B*/
SYMBOL1 V=TRIANGLE L=1 I=JOIN CV=BLUE   H=1.5; /*Lvl B1*/
SYMBOL2 V=CIRCLE   L=3 I=JOIN CV=ORANGE H=1.5; /*Lvl B2*/
TITLE 'INTERACTION PLOT 1';
PROC GPLOT DATA=two; PLOT abmean*b=a;
/* # of SYMBOL LINES = # OF LEVELS OF A*/
SYMBOL1 V=STAR L=1 I =JOIN CV=PURPLE H=1.5 W=2; /*Lvl A1*/
SYMBOL2 V=# L=8 I =JOIN CV=RED    H=1.5 W=2; /*Lvl A2*/
SYMBOL3 V=DOT  L=42 I=JOIN CV=BLACK H=1.5 W=2;
TITLE 'INTERACTION PLOT 2'; RUN; QUIT;
```

## W2P4: THREE WAY FACTORIAL EXPERIMENT

A has a # levels, B has b # levels and C has c # levels. .
**Factorial Treatment Structure** = a*b*c
**Source of Variances in ANOVA**

| Source | df | Coefficients | Definitions |
|---|---|---|---|
| Trt. A | (a-1) | $\alpha_i$ | $A_i$ Main Effect |
| Trt. B | (b-1) | $\beta_j$ | $\beta_j$ Main Effect |
| Trt. C | (c-1) | $\boldsymbol{\delta}_k$ | $C_k$ Main Effect |
| Interact. AB | (a-1)(b-1) | $\alpha\beta_{ij}$ | $A_iB_j$ Interaction |
| Interact. AC | (a-1)(c-1) | $\alpha\boldsymbol{\delta}_{ik}$ | $A_iC_k$ Interaction |
| Interact. BC | (b-1)(c-1) | $\beta\boldsymbol{\delta}_{jk}$ | $B_jC_k$ Interaction |
| Inter. ABC | (a-1)(b-1)(c-1) | $\alpha_i\beta_j\boldsymbol{\delta}_k$ | $A_iB_jC_k$ Interaction |
| Exp. Error | abc(n-1) | $\varepsilon_{ijkl}$ | Random Error |
| Total | abcn - 1 | | $\varepsilon_{ijk} \sim N(0, \sigma_\varepsilon^2)$ iid. |
| Each of the treatments have n number of replications. i.e. $n_{ijk} = n$ | | | |

**Means Model:** $Y_{ijkl} = \mu_{ijk} + \varepsilon_{ijkl}$
**Treatment Effect Models:**
$Y_{ikjl} = \mu + \alpha_i + \beta_j + \boldsymbol{\delta}_k + \alpha\beta_{ij} + \alpha_i\boldsymbol{\delta}_k + \beta_j\boldsymbol{\delta}_k + \alpha_i\beta_j\boldsymbol{\delta}_k + \varepsilon_{ijkl}$
Where: $Y_{ijkl}$ = $l^{th}$ observation level $A_iB_jC_k$
i (Trt A) = levels 1, 2, …, a, j (Trt B) = levels 1, 2, … b & k (Trt C) = levels 1, 2, …, c & l (# Obs or # Replications) = 1, 2, …, $n_{ijk}$
$\mu_{ijk} = A_iB_jC_k$ mean.
μ = Overall mean.
$\alpha_i$, $\beta_j$, $C_k$ = Treatment 'A' level i effect, Treatment 'B' level j effect. & Treatment 'C' level k effect respectively.
$\alpha\beta_{ij}$, $\alpha\boldsymbol{\delta}_{ik}$, $\beta\boldsymbol{\delta}_{jk}$, $\alpha_i\beta_j\boldsymbol{\delta}_k$: Interaction effects $A_iB_j$ $A_iC_k$ & $B_jC_k$ $A_iB_jC_k$.
$\varepsilon_{ijkl}$ = random error due to $k^{th}$ expt. unit of $A_iB_j$. $\varepsilon_{ijk} \sim N(0, \sigma_\varepsilon^2)$.

To test ABC interaction, replicate each combinations. If replication is not possible, use higher order interaction as an estimation of experiment error but as last resort and with caution.
**Interpretation: If three-way interaction is significant**, then make comparisons among $A_iB_jC_k$ treatment combinations. If three way-way interaction is not significant, but **at least one of the two-way interactions is significant**, then compare means among the affected treatments combinations. **If none of the interactions is significant,** then explain significant main effects.

## SAS CODE for a*b*c (Three Way) Factorial Treatments in a CRD:

```
DATA one; INPUT a b c y; DATALINES;
1 1 1 # .... a b c # ; PROC GLM DATA = one; CLASS a b c;
MODEL y = a b c a*b a*c b*c a*b*c; /*MODEL y = a| b| c|; */
LSMEANS sig effects / LINES PDIFF STDERR SLICE ADJUST = T; RUN;
```

| | HEAT Treatment W | | | | | HEAT Treatment L | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Machines | | | | $Y_{ijkl}$ = | Machines | | | |
| Time | A | B | C | D | Response | A | B | C | D |
| 8:00 AM | 6 | 7 | 1 | 6 | $Y_{ijkl}$ | 4 | 6 | -1 | 4 |
| 11:00 AM | 6 | 8 | 3 | 7 | $Y_{ijkl}$ | 3 | 6 | 2 | 9 |
| 3:00 PM | 5 | 10 | -1 | 10 | $Y_{ijkl}$ | 6 | 8 | 0 | 4 |

```
TITLE '/* 3-WAY ANOVA */ Three-way Factorial Book Example. 9.6';
DATA three; INPUT Heat $ Time $ Machine $ Response @@;
DATALINES ; W 8:00AM A 6 ... W 11:00AM A 6 ... L 3:00PM D 4;
PROC GLM DATA=three PLOTS (ONLY)=(RESIDUALS DIAGNOSTICS);
CLASS time heat machine; MODEL response = time | heat | machine;
LSMEANS time heat machine / STDERR PDIFF ADJUST=TUKEY;
/* Possible SLICES when a 3-way interaction is significant */
LSMEANS time * heat * machine / SLICE=(time*heat time*machine
machine) ; ODS OUTPUT LSMEANS=plotdata;
PROC SORT DATA=plotdata; BY heat; WHERE EFFECT =
"Time_Heat_Machine"; AXIS1 ORDER = -1 TO 9 BY 1;
PROC GPLOT DATA=plotdata; BY heat;
WHERE EFFECT = "Time_Heat_Machine";
PLOT responseLSMEAN * machine = time / VAXIS = AXIS1;
FORMAT responseLSMEAN 4.2; /*optional statement */
SYMBOL1 VALUE=CIRCLE CV=BLACK L=1 W=2 I=JOIN H=1.5;
SYMBOL2 VALUE=DOT CV=BLACK L=3 W=2 I=JOIN H=1.5;
SYMBOL3 VALUE=SQUARE CV=BLACK L=8 W=2 I=JOIN H=1.5;
PROC SGPANEL DATA = plotdata; PANELBY heat;
SERIES Y = responseLSMEAN X = machine / GROUP = time; RUN;
```

## W3P1: FIXED AND RANDOM EFFECTS

**One-way Treatment (t levels) Str. in CRD (Trt. is Fixed)**

(Mean or Effect Models) **Model**: $Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$; ($n_{ij} = n, \forall\ i, j$)
$\mu$ (overall mean) & $\tau_i$ are fixed & random $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$

**Hypothesis:** $H_o$: $\tau_1 = \tau_1 = \ldots = \tau_t$ VS $H_a$: There is treatment effect.
**Decision:** If $H_a$ is true, $EMS_{Treatment} > EMS_{Error}$.

| Error SOURCE | df | (EMS) Expected Mean Square |
|---|---|---|
| Treatment (Trt) | t-1 | $\sigma_\varepsilon^2 + n^* \sum_{i=1}^{t} (\frac{\tau i^2}{t-1})$ |
| Exp. Error | n. – t | $\sigma_\varepsilon^2$ = E(MSE) |
| Total | n. - 1 | |

If $H_a$ is true, continue contrast and treatment level mean estimates.

**One-way Treatment (t-levels) Str. in CRD (Trt. is Random)**

**Only Effect Model**: $Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$; ($n_{ij} = n, \forall\ i, j$) & $\mu$ (fix) & $\tau_i \sim N(0, \sigma_\tau^2)$ & $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$ are random.

**Hypothesis: Hypothesis:** $H_o$: $\sigma_\tau^2 = 0$ & $H_a$: $\sigma_\tau^2 \neq 0$
**Decision:** If $H_a$ is true, $EMS_{Treatment} > EMS_{Error}$.

| Error Source | Df | (EMS) Expected Mean Square |
|---|---|---|
| Treatment (Trt) | t-1 | $\sigma_\varepsilon^2 + n^* \sigma_\tau^2$ |
| Exp. Error | t(n – 1) | $\sigma_\varepsilon^2$ = E(MSE) |
| Total | tn – 1 | |

If $H_o$ is true, all levels of the random factors are same. If $F_{Trt}$ is significant, estimate $\sigma_\tau^2$

$\sigma_{\varepsilon hat}^2$ = MSE is the point estimate of $\sigma_\varepsilon^2$. $MS_{TRT}$ is a point estimate of $\sigma_\varepsilon^2 + n^* \sigma_\tau^2$. Let $\sigma_{\varepsilon hat}^2$ denote the point estimate of $\sigma_\tau^2$. Method of Moments Estimation of $\sigma_\tau^2$: Set EMS equal to the calculated MS and Solve for $\sigma_{hat\tau}^2$.
➔ $\sigma_{hat\tau}^2 = (MS_{Trt} - \sigma_{\varepsilon hat}^2)/n$

Method of moments method is relatively simple but to avoid negative variance estimates, we can use Restricted Maximum Likelihood (RLEM) using SAS/MIXED. We use MIXED procedure. Rule, whenever **analyzing data with random effect** (=! error), use **PROC MIXED or PROC GLIMMIX** instead of PROC GLM

**Two-way Factorial (a*b) Str. in CRD (A & B Fixed)**

**Effect Model**: $Y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk}$; ($n_{ij} = n, \forall\ i, j$) & $\mu$ (overall mean) & $\varepsilon_{ijk} \sim N(0, \sigma_\varepsilon^2)$ are random.

| Err. Source | Df | (EMS or MS) Exp. Mean Square | F | $H_o$ |
|---|---|---|---|---|
| Trt. A | a-1 | $\sigma_\varepsilon^2 + nb^* \sum \frac{\alpha_i^2}{(a-1)}$ | $\frac{MS_A}{MSE}$ | $\alpha_1 = \ldots = \alpha_a$ |
| Trt B | b-1 | $\sigma_\varepsilon^2 + na^* \sum \frac{\beta_j^2}{(b-1)}$ | $\frac{MS_B}{MSE}$ | $\beta_1 = \ldots = \beta_b$ |
| A*B | (a-1)(b-1) | $\sigma_\varepsilon^2 + na^* \sum \frac{\alpha\beta_{ij}^2}{(a-1)(b-1)}$ | $\frac{MS_{AB}}{MSE}$ | No interaction |
| Exp. Err. | ab(n – 1) | $\sigma_\varepsilon^2$ = E(MSE) = MSE | | |
| Total | abn – 1 | | | |

**Two-way Factorial (a*b) Str. CRD (A & B Both Random)**

**Effect Model**: $Y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk}$; ($n_{ijk} = n, \forall\ i, j$), k = 1, 2, … , n & $\mu$ is global or overall mean. **Random Component:** $\alpha_i \sim N(0, \sigma_\alpha^2)$, $\beta_i \sim N(0, \sigma_\beta^2)$, $\alpha\beta_{ij} \sim N(0, \sigma_{\alpha\beta}^2)$ & $\varepsilon_{ijk} \sim N(0, \sigma_\varepsilon^2)$

| Err. Source | Df | (EMS/ MS) Exp. Mean Sq. | F | $H_o$ |
|---|---|---|---|---|
| Trt. A | a-1 | $\sigma_\varepsilon^2 + n\sigma_{\alpha\beta}^2 + nb\sigma_\alpha^2$ | $F_{(a-1),(a-1)(b-1)} = \frac{MS_A}{MS_{AB}}$ | $\sigma_{\alpha\beta}^2 = 0$ |
| Trt B | b-1 | $\sigma_\varepsilon^2 + n\sigma_{\alpha\beta}^2 + na\sigma_\beta^2$ | $F_{(b-1),(a-1)(b-1)} = \frac{MS_B}{MS_{AB}}$ | $\sigma_\beta^2 = 0$ |
| A*B | (a-1)(b-1) | $\sigma_\varepsilon^2 + n\sigma_{\alpha\beta}^2$ | $F_{(a-1)(b-1),\, ab(n-1)} = \frac{MS_{AB}}{MSE}$ | No interaction |
| Exp. Err. | ab(n – 1) | $\sigma_\varepsilon^2$ = MSE | | |
| Total | abn – 1 | | | |

If any of the effects are significant, we are interested in estimating the significant variance components.

**Two-way Factorial (a*b) Str. in CRD (Mixed Model: A Fixed & B Random)** **Effect Model**: $Y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk}$; ($n_{ijk} = n, \forall\ i, j$), k = 1, …, n & $\mu$ is overall mean.
**Random:** $\beta_i \sim N(0, \sigma_\beta^2)$, $\alpha\beta_{ij} \sim N(0, \sigma_{\alpha\beta}^2)$ & $\varepsilon_{ijk} \sim N(0, \sigma_\varepsilon^2)$

| Err. Source | Df | (EMS/ MS) Exp. Mean Sq. | F | $H_o$ |
|---|---|---|---|---|
| Trt. A | a-1 | $\sigma_\varepsilon^2 + n\sigma_{\alpha\beta}^2 + nb^* \sum \frac{\alpha_i^2}{(a-1)}$ | $F_{(a-1),(a-1)(b-1)} = \frac{MS_A}{MS_{AB}}$ | $\alpha_1 = \ldots = \alpha_a$ |
| Trt B | b-1 | $\sigma_\varepsilon^2 + n\sigma_{\alpha\beta}^2 + na\sigma_\beta^2$ | $F_{(b-1),(a-1)(b-1)} = \frac{MS_B}{MS_{AB}}$ | $\sigma_\beta^2 = 0$ |
| A*B | (a-1)(b-1) | $\sigma_\varepsilon^2 + n\sigma_{\alpha\beta}^2$ | $F_{(a-1)(b-1),\, ab(n-1)} = \frac{MS_{AB}}{MSE}$ | $\sigma_{\alpha\beta}^2 = 0$ i.e. No interaction |
| Exp. Err. | ab(n – 1) | $\sigma_\varepsilon^2$ = MSE | | |
| Total | abn – 1 | | | |

Unequal sample size complicates fixed and random model F-tests.

For factorial treatment structures, if all interacting factors are fixed, the interaction are also fixed. If any one of the factor is random, the interaction term with the random factor is also random. Eg. In A*B*C, if A is fixed, B & C are random then only main effect A is fixed but B, C, AB, AC, BC, ABC and error terms are random.

## VARIANCE COMPONENT ESTIMATION (USING MIXED MODEL):

**Method 1:** MSE = $\sigma_\varepsilon^2$; & $MS_{AB} = \sigma_\varepsilon^2 + n\sigma_{\alpha\beta}^2$
$MS_B = \sigma_\varepsilon^2 + n\sigma_{\alpha\beta}^2 + na\sigma_\beta^2 = MS_{AB} + na\sigma_\beta^2$ ➔ $\sigma_\beta^2 = (MS_B - MS_{AB})/n$
**Method 2:** PROC MIXED or PROC GLMIX for REML Estimation (SAS).

**Example:** Design Structure: Completely Randomized Design (CRD) Treatment Structure: 3*4 factorial. There are three reps **(n = 3)** of each treatments combinations. Given the following ANOVA table, compute EMS expressions for each scenario below and compute the method of moments variance component estimates for each scenario.

| Source | Df | SS | MS |
|---|---|---|---|
| N (a = 3) | 2 | 729.2232 | 364.6116 |
| P (b = 4) | 3 | 4969.7302 | 1656.5767 |
| NP | 6 | 560.4031 | 93.4005 |
| Error | 24 | 485.3776 | 20.2241 |
| Total | 35 | 674.7343 | |

Scenario 1: N is Random and P is fixed:

| Source | EMS = MS | Hypothesis | Correct F-value |
|---|---|---|---|
| N (Random) | $\sigma_\varepsilon^2 + n\sigma_{NP}^2 + nP\sigma_N^2$ $= \sigma_\varepsilon^2 + 3\sigma_{NP}^2 + 12\sigma_N^2$ | $H_o$: $\sigma_N^2 = 0$ $H_A$: $\sigma_N^2 \neq 0$ | 364.6116/93.4005 = 3.9037 |
| P (Fix) | $\sigma_\varepsilon^2 + n\sigma_{NP}^2 + nN^* \sum \frac{\alpha_i^2}{(a-1)}$ $= \sigma_\varepsilon^2 + 3\sigma_{NP}^2 + 9/3\Sigma P^2$ | $H_o$: $P_1 = P_2 = P_3 = P_4$ $H_a$: At least one is different. | 1656.5767/93.4005 = 17.7363 |

| | | | |
|---|---|---|---|
| NP (Random) | $\sigma_\varepsilon^2 + n\sigma_{NP}^2 = \sigma_\varepsilon^2 + 3\sigma_{NP}^2$ | $H_o: \sigma^2_{NP} = 0$ <br> $H_A: \sigma^2_{NP} \neq 0$ | 93.4005/20.2241 = 4.6183 |
| Error (R) | $\sigma_\varepsilon^2$ = MSE = 20.2241 | | |
| **Variance Component Estimation:** | | $\sigma_\varepsilon^2$ = **MSE** = 20.2241 | |
| $MS_{NP} = 93.4005 = 20.2241 + 3\sigma^2_{NP}$ | | ➔ | $\sigma^2_{NP} = 24.3921$ |
| $MS_N = 20.2241 + 3(24.3921) + 12\sigma^2_N$ | | ➔ | $\sigma^2_N = 22.6001$. |

**Scenario 2: N and P both are random:**

| Source | EMS = MS | Hypothesis | Correct F-value |
|---|---|---|---|
| N (Random) | $\sigma_\varepsilon^2 + n\sigma^2_{NP} + nP\sigma^2_N = \sigma_\varepsilon^2 + 3\sigma^2_{NP} + 12\sigma^2_N$ | $H_o: \sigma_N^2 = 0$ <br> $H_A: \sigma_N^2 \neq 0$ | 364.6116/93.4005 = 3.9037 |
| P (Random) | $\sigma_\varepsilon^2 + n\sigma^2_{NP} + nN\sigma^2_P = \sigma_\varepsilon^2 + 3\sigma^2_{NP} + 12\sigma^2_P$ | $H_o: \sigma_P^2 = 0$ <br> $H_A: \sigma_P^2 \neq 0$ | 1656.5767/93.4005 = 17.7363 |
| NP (Random) | $\sigma_\varepsilon^2 + n\sigma^2_{NP} = \sigma_\varepsilon^2 + 3\sigma^2_{NP}$ | $H_o: \sigma^2_{NP} = 0$ <br> $H_A: \sigma^2_{NP} \neq 0$ | 93.4005/20.2241 = 4.6183 |
| Error (R) | $\sigma^2_\varepsilon$ = MSE = 20.2241 | | |
| **Variance Component Estimation:** | | $\sigma_\varepsilon^2$ = **MSE** = 20.2241 | |
| $MS_{NP} = 93.4005 = 20.2241 + 3\sigma^2_{NP}$ | | ➔ | $\sigma^2_{NP} = 24.3921$ |
| $MS_N = 20.2241 + 3(24.3921) + 12\sigma^2_N$ | | ➔ | $\sigma^2_N = 22.6001$ |
| $MS_P = 20.2241 + 3(24.3921) + 12(\sigma^2_P)$ | | ➔ | $\sigma^2_P = 173.6863$ |

## W3P2: (RCBD) RANDOMIZED COMPLETE BLOCK A*B Fact DESIGN

**Expt. unit:** In RCBD, treatment levels may be grouped or blocked to account for outside influences. Each block has random t-treatments levels. 'n' blocks implies that there are 'n replications of each treatment level'. Block j affects each treatment levels the same way. Block and treatment do not interact. **Design Structure:** RCBD with n blocks. **Treatment Structure:** Two-way, t-levels. **Stat. Model:** $Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$; $\tau_i$ is trt effect, $\beta_j$ is block effect $\beta_j \sim N(0, \sigma^2_{BLOCK})$ & $\varepsilon_{ij} \sim N(0, \sigma^2_\varepsilon)$.

| Source | Df | EMS | Hypothesis | Test Stat |
|---|---|---|---|---|
| Block | n -1 | $\sigma_\varepsilon^2 + ab*\sigma^2_{BLOCK}$ | $H_o: \sigma^2_{BLOCK} = 0$ | $F \sim F_{(t-1),\ dferr}$ |
| Trt. A | a-1 | $\sigma_\varepsilon^2 + nb*\sum \frac{\alpha_i^2}{(a-1)}$ | $H_o: \alpha_1 = \ldots = \alpha_a$ | |
| Trt. B | b-1 | $\sigma_\varepsilon^2 + na*\sum \frac{\beta_i^2}{(b-1)}$ | $H_o: \beta_1 = \ldots = \beta_b$ | |
| Int. AB | (a-1)(b-1) | $\sigma_\varepsilon^2 + na*\sum \frac{\alpha\beta_{ij}^2}{(a-1)(b-1)}$ | $H_o: \alpha\beta_{ij} - \alpha\beta_{i'j} = \alpha\beta_{ij'} - \alpha\beta_{i'j'}$ $\forall\ i \neq i'$. | |
| Error | (n-1)(ab-1) | $\sigma_\varepsilon^2$ | $H_o: \sigma_\varepsilon^2 = 0$ | |
| Total | nab-1 | | | |

**Standard Error of $\overline{Yi.}$ = ($S\overline{Yi.}$) =** $\sqrt{((\sigma_\varepsilon^2 + \sigma_{BLOCK}^2)/n)}$

**Standard Error of $\overline{Yi.} - \overline{Yi'.}$ = $S(\overline{Yi.} - \overline{Yi'.})$ =** $\sqrt{(2\sigma_\varepsilon^2/n)}$

**Note:** Contrast, comparison, post hoc tests (LSD, Scheffe & Tukey) all applies.

$\hat{L} = \sum a_i \bar{\mu}_{i\ hat}$ **Estimated Standard Error of $L_{hat}$ =** $\sqrt{(\ \sigma_{\varepsilon hat}{}^2 * \sum_{i=1}^{n} (\frac{ai^2}{ni})}$

### General SAS CODE for a*b*c RCBD Experimental Design

DATA name; INPUT trt1 trt2 trt3 block response;
DATALINES; ..... ; PROC MIXED DATA = name PLOT = RESIDUALPANEL; CLASS a b c block; MODEL response = a | b| c / DDFM = SATTERTH ( or KR); RANDOM block; LSMEANS trt1 trt2 trt3/ < OPTIONS> CONTRAST 'label ' effect ; /* List effects based on significance */ ESTIMATE ' label' effect / <OPTIONS: DIVISOR CL>;

## W3P3: LATIN SQUARE DESIGN (LSD)

Double blocking factors: row and column. Needed: # of rows = # of columns = # of treatments = t. One Row, One Column One Treatment.

**5*5 Latin Square Design:**

| | | | | | |
|---|---|---|---|---|---|
| 2 | 5 | 4 | 1 | 3 | |
| 5 | 1 | 2 | 3 | 4 | |
| 4 | 3 | 1 | 5 | 2 | |
| 1 | 4 | 3 | 2 | 5 | |
| 3 | 2 | 5 | 4 | 1 | |

**4*4 L.S. Design:**

| | | | |
|---|---|---|---|
| 1 | 3 | 2 | 4 |
| 3 | 1 | 4 | 2 |
| 4 | 2 | 1 | 3 |
| 2 | 4 | 3 | 1 |

### Single Latin Square Design:

**Model**: $Y_{ij(k)} = \mu + \boldsymbol{\delta}_i + \lambda_j + \tau_k + \varepsilon_{ijk}$; Row: $\boldsymbol{\delta}_i \sim N(0, \sigma_\delta^2)$, Column: $\lambda_j \sim N(0, \sigma_\lambda^2)$& $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$ are random.

| Source | Df | EMS | Null Hypo. | Test Stat. |
|---|---|---|---|---|
| Row | (r-1) | $\sigma_\varepsilon^2 + \sigma_{row}^2$ | N/A | |
| Column | (c-1) | $\sigma_\varepsilon^2 + \sigma_{col}^2$ | N/A | |
| Treatment | (t-1) | $\sigma_\varepsilon^2 + t*\sum\tau_i^{2/(t-1)}$ | $H_o$: $\tau_1 = \ldots = \tau_i$ | $F_{trt} \sim F_{(t-1),\ (t-1)(t-2)}$ |
| Expt. Error | (t–1)(t-2) | $\sigma_\varepsilon^2$ | | |
| Total | ($t^2$ – 1) | | | |

**SAS Code:**

DATA latin1; INPUT row column trt response; DATALINES; ...;
PROC MIXED DATA = latin1 PLOTS = RESIDUALPANEL; CLASS row column trt; MODEL response = trt/ DDFM = SATTERTH; RANDOM row column; CONTRAST 'label' TRT $a_k$ values; ESTIMATE 'label' TRT $a_k$ values; LSMEANS trt / <OPTIONS> ; RUN; QUIT;

**Standard error of the Treatment K mean** = $\sqrt{((\sigma_{row}^2/t) + (\sigma_{col}^2/t) + (\sigma_\varepsilon^2/t))}$

**Variance of the difference between means** = $2\sigma_\varepsilon^2/t = 2MSE/t$

**Confidence Interval (CI) (Two population):** (1-α) 100% CI for $\tau_{kbar} - \tau_{k'bar}$: $(\bar{Y}_k - \bar{Y}_{k'}) \pm t_{\alpha/2,\ df} * S_{(Y1bar\ -\ Y2bar)} = (\bar{Y}_1 - \bar{Y}_2) \pm t_{\alpha/2,\ df} * \sqrt{(2MSE/t)}$

**Pairwise Comparison same as LSD, HSD, Scheffes**

### More than One Square Latin Square Design:

**Model**: $Y_{ij(k)} = \mu + S_l + \boldsymbol{\delta}_i + \lambda_j + \tau_k + \varepsilon_{ijkl}$; Row: $\boldsymbol{\delta}_i \sim N(0, \sigma_\delta^2)$, Column: $\lambda_j \sim N(0, \sigma_\lambda^2)$, Square: $S_l \sim N(0, \sigma_s^2)$ & $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$ are random.

| Source | Df |
|---|---|
| Square | (s-1) |
| Row in Sq. | (r-1) |
| Col in Sq. | (c-1) |
| Treatment | (t-1) |
| Expt. Error | s(t–1)(t-2) + (s-1)(t-1) |
| Total | ($t^2$ – 1) |

**SAS CODE for More than One LSD:**
DATA latin2; INPUT row column trt response; DATALINES; ...; PROC MIXED DATA = latin2 PLOTS = RESIDUALPANEL; CLASS square row column trt; MODEL response = trt/ DDFM = SATTERTH; RANDOM square square(row) square(column); CONTRAST 'label' TRT $a_k$ values; ESTIMATE 'label' TRT $a_k$ values; LSMEANS trt / <OPTIONS> ; RUN; QUIT;

## W3P4: NESTED EFFECTS

One whose levels are unique to each level of a second factor. A treatment can nest in another treatment. Experimental unit nest within a treatment level. Sub sampling units nested within experimental units.

**Case 1: CRD and One Way Treatment Structure:** Compare patient treatment at randomly selected hospitals. Select n patients at each of h randomly selected hospitals. Hospital is a random effect and patients are unique to each hospital.

| **Source** (All Random) | **df** | **EMS** |
|---|---|---|
| Hospital | h-1 | $\sigma_\varepsilon^2 + n\ \sigma_H^2$ |
| Exp. Error = Patient within Hospital | h(n-1) | $\sigma_\varepsilon^2$ |
| Total | hn-1 | |

**Case 2: CRD& 2*10 Factorial Treatment Structure:** Two drugs at the same 10 hospitals. Select n patients for drug 1 at each hospital, and n patients for drug 2 at each hospital. Hospital is random effect. Drug is a fixed effect. $\alpha_i$ = drug i effect, i = 1, 2.

| **Source** | **df** | **EMS** |
|---|---|---|
| Hospital (Random) | h-1 = 9 | $\sigma_\varepsilon^2 + n\sigma_{DH}^2 + 2n\sigma_H^2$ |
| Drug (Fixed) | d – 1 = 1 | $\sigma_\varepsilon^2 + n\sigma_{DH}^2 + 10n\Sigma\alpha_i^2/df$ |
| Hospital*Drug (Random) | (h-1)(d-1) = 9 | $\sigma_\varepsilon^2 + n\sigma_{DH}^2$ |
| Exp. Error = Patient within Drug and Hospital (Random) | h(n-1) = 20(n-1) | $\sigma_\varepsilon^2$ |
| Total | hn-1 = 20n - 1 | |

**Case 3: Case 2 Extended:** An average patient response obtained from each drug/hospital combination in case 2.

| **Source** | **Df** | **EMS** |
|---|---|---|
| Hospital (Random) | h-1 = 9 | $\sigma_{DH}^2 + 2\sigma_H^2$ |
| Drug (Fixed) | d – 1 = 1 | $n\sigma_{DH}^2 + 10n\Sigma\alpha_i^2/df$ |
| Hospital*Drug (Random) | (h-1)(d-1) = 9 | $\sigma_{DH}^2$ |
| Exp. Error = Patient within Drug and Hospital | h(n-1) = 0 | 0 |
| Total | hnc-1 = (10)(2)(1) = 19 | |

**Case 4: Experimental Design depends on Context:** Randomly selected 10 hospitals, to test drug 1, another 10 hospitals randomly selected to test drug 2. Each hospital selects a random sample of n patients. **Note:** Patient nested within hospital, and hospital nested within drug. Not all drug and hospital combinations are available.

| **Source** | **df** | **EMS** |
|---|---|---|
| Drug (Fixed) | d – 1 = 1 | $\sigma_\varepsilon^2 + n\sigma_{H(D)}^2 + 10n\Sigma\alpha_i^2/df$ |
| Hospital within Drug (Rand) | d(h-1) = 2(10-1) = 18 | $\sigma_\varepsilon^2 + n\sigma_{H(D)}^2$ |
| Patient within hospital within drug (Rand) | dh(n-1) = 2(10)(n-1) | $\sigma_\varepsilon^2$ |
| Total | dhn-1 = (2)(10)n | |

**Often, Experiments with one-way trt. Str. with t levels, CRD design with n reps & S (# of sample & are equal) subsamples (Hierarchy):**

| **Source** | **Df** | **EMS** |
|---|---|---|
| Treatment (Fixed) | t-1 | $\sigma_\delta^2 + s\sigma_\varepsilon^2 + sn\Sigma\alpha_i^2/df$ |
| Exp. Error = Experiment unit within treatment (Random) | t(n-1) | $\sigma_\delta^2 + \sigma_\varepsilon^2$ |

| | | |
|---|---|---|
| Sub sampling error = sampling within the exp. Unit within trt. (Random) | tn(s-1) | $\sigma_{\delta}^2$ |
| Total | tsn- 1 | |

**STATISTICAL MODEL**

$Y_{ijk} = \mu + \tau_i + \varepsilon_{ij} + \boldsymbol{\delta}_{ij(k)}$

Where: $Y_{ijk}$ = Observation on subsample k (k = 1, 2,..., s) of expt. unit j (j = 1, 2, … , n) in treatment level i.
$\tau_i$ = Treatment level i, i = 1, 2, …, t.
$\mu$ = Overall mean.
$\varepsilon_{ij}$ = random error due to expt. unit j & trt. i. $\varepsilon_{ij} \sim N(0, \sigma_{\varepsilon}^2)$.
$\boldsymbol{\delta}_{ij(k)}$ = Random error due to subsample k in experimental unit j in treatment level i and $\boldsymbol{\delta}_{ij(k)} \sim N(0, \sigma_{\delta}^2)$.

**SAS EXAMPLE FOR NESTED EFFECT**

| Shift | A | B | C |
|---|---|---|---|
| Worker | 1 2 3 4 @@ | 5 6 7 8 @@ | 9 10 11 12 @@ |
| Observed Values | 3 5 0 10 5 7 3 7<br>3 4 3 5 6 4 4 4 4 6 5 7 | 4 5 0 7 3 5 1 6 4 4 2 5<br>3 7 1 10 4 6 1 5 | 14 5 9 9 12 2 5 5<br>10 6 2 9 9 6 6 4 10 3 6 7 |

```
PROC MIXED DATA=one METHOD=TYPE3 COVTEST; /* EMS */
CLASS shift worker; MODEL response = shift; /* Dep = Trt */
RANDOM worker (shift); /* Worker unique to Shift */
LSMEANS shift; ODS HTML SELECT TYPE3; /* TYPE3 = EMS ONLY */
TITLE 'Nested Effects Example';
PROC MIXED DATA=one METHOD = REML COVTEST;
CLASS shift worker; MODEL response = shift / DDFM=SATTERTH;
RANDOM worker (shift); /* Worker within Shift */
LSMEANS shift; RUN; QUIT;
```

**W4P1 & W4P2: DEFINITION OF TERMS**

**Experimental Objective:** Statements of the decision. **Population of Interest:** Population in which decision is applicable. **Response Variable: Decision Variables:** variables that helps to make decision about population. **Ancillary Variables:** similar to control variable. **Treatment Structure:** The category of design that used for the experimentation groups. **Design Structure:** Descriptor for the basic statistical design used to assign experimentation units to experimentation groups. **Randomization Procedure:** Formal process by which the treatment assigned to the experimentation units with specific number of treatment and replications. **Statistical Model:** specifies the formal process by which the treatment assigned to the experimentation units. **Decision process** outlines the result of the statistical analysis and present the conclusions of the experiments.

**SPLIT PLOT DESIGN**

Only design requiring factorial treatment (no one-way trt.) Two sizes plots: whole plots (bigger units, $A_i$) and sub-plots (plots inside whole plot, $B_j$). If A treatments are replicated, all B treatments are also replicated. Each level of A has 'n' replications (experiment units) and called **whole plot analysis**. Each experiment unit A serve as block for levels of B. Subplots B are also the experimental unit for $A_iB_j$ interaction. Subplots analysis includes B and AB interaction. Both effect use the 1/b portion of the whole plot experimental unit. There are two-error variances (treated alike) 1) Whole plot A EV & 2) Subplot B EV.

**Main difference between CRD and RCBD split plot design:** CRD design has main plots separately but RCBD has main plots together but blocked. RCBD is usually common agriculture researches with big plots.

**SPLIT PLOT in Completely Randomized Design (CRD)**

**Treatment Structure:** a*b (Whole plot (A) * Subplot (B)) factorial,
**Design Structure:** Split plot in Completely Randomized Design (CRD), n reps: CRD with n reps is Whole plot design structure in which independent experimental units randomly assigned to each levels of A.

**Statistical Model for a*b factorial Treatment in split-plot CRD**

$Y_{ijk} = \mu + (\alpha_i + \boldsymbol{\delta}_{k(i)}) \sim$ (from whole plot) $+ (\beta_j + \alpha\beta_{ij} + \varepsilon_{ijk}) \sim$ (Sub plot)

Where: $Y_{ijk}$ = $k^{th}$ response of $A_iB_j$
$\mu$ = Overall mean.
$\alpha_i$ & $\beta_j$ = Main Effect 'A' with level i, Main Effect B with level j
& $\alpha\beta_{ij}$, = Interaction effects $A_iB_j$.
i = 1, 2, …, a; j = 1, 2, …, b; & k = 1, 2, …, n.
$\boldsymbol{\delta}_{k(i)}$ = Main plot random error $\boldsymbol{\delta}_{k(i)} \sim N(0, \sigma_{\delta}^2)$.
$\varepsilon_{ijk}$ = Subplot random error; $\varepsilon_{ijk} \sim N(0, \sigma_{\varepsilon}^2)$. $\boldsymbol{\delta}_{k(i)}$ & $\varepsilon_{ijk}$ are ind.

| Error Source | df | (EMS = MS) | F |
|---|---|---|---|
| Whole plot Analysis | | | |
| A | a-1 | $\sigma_{\varepsilon}^2 + b\sigma_{\delta}^2 + na\Sigma\alpha_j^2/df$ | $F_A \sim F_{(a-1),\, a(n-1)}$ |
| Rep within A | a (n-1) | $\sigma_{\varepsilon}^2 + b\sigma_{\delta}^2$ | |
| Subplot Analysis | | | |
| B | (b-1) | $\sigma_{\varepsilon}^2 + na\Sigma\beta_j^2/df$ | $F_B \sim F_{(a-1)(b-1),\, ab(n-1)}$ |
| AB | (a-1)(b-1) | $\sigma_{\varepsilon}^2 + n\Sigma(\alpha\beta)_{ij}^2/df$ | $F_{AB} \sim F_{(a-1)(b-1),\, a(b-1)(n-1)}$ |
| Exp. Err. = B*Rep(A) | a(b-1)(n – 1) | $\sigma_{\varepsilon}^2$ = MSE | |
| Total | abn – 1 | | |

**Decision Rules for F-Statistics:**

**For the A Main Effect:**

**Effect Model:** $\mathbf{H_o}$: $\alpha_1 = \ldots = \alpha_a$ & $\mathbf{H_a}$: There is an A main effect.
**Mean Model:** $H_o$: $\mu_{1.bar} = \ldots = \mu_{a.bar}$ Where, $\mu_{ij} = A_iB_j$ & $\mu_{1.bar} = A_i$ mean.
**Reject $\mathbf{H_o}$** if MSA/WP Error $\geq F_{\alpha,\,(a-1),\,a(n-1)}$

**For the B Main Effect:**

**Effect Model:** $\mathbf{H_o}$: $\beta_1 = \ldots = \beta_b$ & $\mathbf{H_a}$: There is an B main effect.
**Mean Model:** $H_o$: $\mu_{.1bar} = \ldots = \mu_{.bbar}$ Where, $\mu_{.jbar} = B_j$ mean.
**Reject $\mathbf{H_o}$** if MSB/SP Error $\geq F_{\alpha,\,(b-1),\,a(b-1)(n-1)}$

**For the AB Interaction Effect:**

$H_o$: No AB Interaction. $H_a$: AB Interaction.
**Reject $\mathbf{H_o}$** if MSAB/SP Error $\geq F_{\alpha,\,,\,(a-1)(b-1),\,a(b-1)(n-1)}$

**SPLIT PLOT in Completely Randomized Block Design (RCBD)**

**Treatment Structure:** a*b (Whole plot (A) * Subplot (B)) factorial,
**Design Structure:** Split plot in Completely Randomized Design (CRD), n reps: CRD with n reps is Whole plot design structure in which independent experimental units randomly assigned to each levels of A.

**Statistical Model for a*b factorial Treatment in split-plot CRD**

$Y_{ijk} = \mu + (\alpha_i + \rho_k + \boldsymbol{\delta}_{k(i)}) \sim$ (from whole plot) $+ (\beta_j + \alpha\beta_{ij} + \varepsilon_{ijk}) \sim$ (Sub plot)

Where: $Y_{ijk}$ = $k^{th}$ response of $A_iB_j$
$\mu$ = Overall mean.
$\alpha_i$, & $\beta_j$ = Main Effect 'A' with level i, Main Effect B with level j,
& $\alpha\beta_{ij}$, = Interaction effects $A_iB_j$.
$\rho_k$ = Block k effect & $\rho_k \sim N(0, \sigma_{\rho}^2)$.
i = 1, 2, …, a; j = 1, 2, …, b; & k = 1, 2, …, n.
$\boldsymbol{\delta}_{k(i)}$ = Main plot random error & $\boldsymbol{\delta}_{k(i)} \sim N(0, \sigma_{\delta}^2)$.
$\varepsilon_{ijk}$ = Subplot random error; $\varepsilon_{ijk} \sim N(0, \sigma_{\varepsilon}^2)$. $\boldsymbol{\delta}_{k(i)}$ & $\varepsilon_{ijk}$ are ind.

| Error Source | df | (EMS = MS) | F |
|---|---|---|---|
| Whole plot Analysis | | | |
| Block | n-1 | N/A | |
| A | a-1 | $\sigma_{\varepsilon}^2 + b\sigma_{\delta}^2 + nb\Sigma\alpha_j^2/df$ | $F_A \sim F_{(a-1),\, a(n-1)}$ |
| Exp. Err. = Block*A | (a-1) (n-1) | $\sigma_{\varepsilon}^2 + b\sigma_{\delta}^2$ | |
| Subplot Analysis | | | |
| B | (b-1) | $\sigma_{\varepsilon}^2 + na\Sigma\beta_j^2/df$ | $F_B \sim F_{(a-1)(b-1),\, ab(n-1)}$ |
| AB | (a-1)(b-1) | $\sigma_{\varepsilon}^2 + n\Sigma(\alpha\beta)_{ij}^2/df$ | $F_{AB} \sim F_{(a-1)(b-1),\, a(b-1)(n-1)}$ |
| SP Exp. Err. = Block*B + Block*AB | a(b-1)(n – 1) | $\sigma_{\varepsilon}^2$ = MSE | |
| Total | abn – 1 | | |

**Means Comparison of Whole Plots, Sub Plots and Interactions:**

**Whole Plot (A) Means Comparison:**

$\mathbf{\overline{Yi..} - \overline{Yi'..} = (\alpha_i - \alpha_{i'}) + (\alpha\beta_{bar\ i.} - \alpha\beta_{bar\ i'.}) + (\boldsymbol{\delta}_{bar\ i.} - \boldsymbol{\delta}_{bar\ i'.}) + (\varepsilon_{bar\ i..} - \varepsilon_{bar\ i'..})}$
**Variance** $\mathbf{(\overline{Yi..} - \overline{Yi'..}) = 2\sigma_{\delta}^2/n + 2\sigma_{\varepsilon}^2/bn}$
**Std. Error** $\mathbf{S(\overline{Yi..} - \overline{Yi'..}) = \sqrt{(2*\sigma_{\varepsilon}^2 + b\sigma_{\delta}^2/nb)}}$ with Whole plot error df.

**Sub Plot (B) Means Comparison:**

$\mathbf{\overline{Y.j.} - \overline{Y.j'.} = (\beta_j - \beta_{j'}) + (\alpha\beta_{bar\ .j.} - \alpha\beta_{bar\ .j'.}) + (\varepsilon_{bar\ .j.} - \varepsilon_{bar\ .j'.})}$
**Variance** $\mathbf{(\overline{.j.} - \overline{Y.j'.}) = 2\sigma_{\varepsilon}^2/an}$
**Std. Error** $\mathbf{S(\overline{Y.j.} - \overline{Y.j'.}) = \sqrt{(2*\sigma_{\varepsilon}^2/na)}}$ with sub plot error df.

**Whole Plot and Subplot Interaction (AB) Means Comparison:**
$\overline{Yij.} - \overline{Yij'.} = (\beta_{j.} - \beta_{j'.}) + (\alpha\beta_{ij} - \alpha\beta_{ij'}) + (\varepsilon_{\text{bar ij.}} - \varepsilon_{\text{bar ij'.}})$
**Variance** $(\overline{Yij.} - \overline{Yij'.}) = 2\,\sigma_\varepsilon^2/n$
**Std. Error of** $(\overline{Yij.} - \overline{Yij'.}) = S\,(\overline{Yij.} - \overline{Yij'.}) = \sqrt{(2^*\,\sigma_\varepsilon^2/n)}$ with sub plot error df.

**Difference in A at the same or different levels of B** i.e. $(A_iB_j - A_{i'}B_j)$ or $(A_iB_j - A_{i'}B_{j'})$ is estimated by $\sqrt{}$ $[2\{(b\text{-}1)SPError + WPError\}/nb]$ with satterthwaite df = $[(b\text{-}1)SPError + WPError]^2/[[\{(b\text{-}1)SPError\}^2/df_{SPError}] + [WPError^2/df_{SPError}]]$.

**Mean Comparisons and Post-hoc Tests for Split Plot Design**

**Fischer $\underline{LSD_A}$** = $t_{\alpha/2,\text{ wp df}} * S(\bar{Y}_i - \bar{Y}_{i'}) = t_{\alpha/2,\text{ wp df}} *$ Std. Error $(\overline{Yi..} - \overline{Yi'..})$
**Fischer $\underline{LSD_B}$** = $t_{\alpha/2,\text{ wp df}} * S(\bar{Y}_i - \bar{Y}_{i'}) = t_{\alpha/2,\text{ sp df}} *$ Std. Error $(\overline{Y.j.} - \overline{Y.j'.})$

**Scheffe A** = $\sqrt{((a\text{-}1)(F_{\alpha,\text{ a-1, Wp df}}))} *$ Std. Error $(\overline{Yi..} - \overline{Yi'..})$
**Scheffe B** = $\sqrt{((b\text{-}1)(F_{\alpha,\text{ b-1, Sp df}}))} *$ Std. Error $(\overline{Y.j.} - \overline{Y.j'.})$

**Tukey $HSD_A$:** = $q_{\alpha,\text{ \# levels a, Wp df}} *$ Std. Error $(\overline{Yi..} - \overline{Yi'..})/\sqrt{2}$
**Tukey $HSD_B$:** = $q_{\alpha,\text{ \# levels b, Sp df}} *$ Std. Error $(\overline{Y.j.} - \overline{Y.j'.})/\sqrt{2}$

**SAS CODE: If design structure is a split plot in a CRD**
**PROC GLMMIX** DATA = splitcad PLOTS = RESIDUALPANEL;
CLASS wptrt sptrt wpexpunit; MODEL response = wptrt sptrt wptrt * sptrt/DDFM = SATTERTH; RANDOM wpexpunit(wptrt); LSMEANS …;
CONTRAST …; ESTIMATE …;

**REPEATED MEASURES OR LONGITUDINAL EXPERIMENTS**
Each experiments unit has more than one measurement taken over time or location. Differs from subsampling since time (or location) is associated with each measure on the experimental unit. Different from split plot because time and location are not random. Also called as LONGITUDINAL EXPERIMENTS (~ TIME SERIES). In repeated measures analysis, not all εijk's are necessarily independent with same variances $\sigma_\varepsilon^2$. $\varepsilon_{ijk}$'s are correlated with each other on the same experimental unit. A corrected repeated measures analysis must consider the correlated responses.

**ONE-WAY AND CRD WITH n REPS AND p REPEATED MEASURES:**
n Subjects randomly assigned to one of the t treatments and responses measured every weeks for 6 weeks:

| Source | df |
|---|---|
| Trt | (t-1) |
| Time | (p -1) |
| Time*Trt | (t-1)(p-1) |
| Error | tp(n-1) |
| Total | tpn – 1 |

**Statistical Model:** $Y_{ijk} = \mu + \tau_i + \pi_j + \tau\pi_{ij} + \varepsilon_{ijk}$
i = 1, 2, …, t; j = 1, 2,…, p & k = 1, 2,… n.

Obs at time j on exp unit k of trt level i = overall mean + trt level I effect + time j effect + trt*time interaction + random error. $\varepsilon_{ijk} \sim N(0, \sigma_\varepsilon^2)$ & iid.

**DATA** one; INPUT trt time expunit response; DATALINES; ;
**PROC MIXED** DATA = one PLOTS = RESIDUALPANEL; CLASS trt time expunit; MODEL response = trt time trt*time / DDFM = SATTERTH; /* DDFM = KR*/ REPEATED time / SUBJECT = expunit(trt) type = <options>;
**PROC GLIMMIX** DATA = one PLOTS = RESIDUALPANEL; CLASS trt time expunit; MODEL response = trt time trt*time / DDFM = SATTERTH; RANDOM RESIDUAL / SUBJECT = expunit(trt) TYPE = <option>; **RUN**; **QUIT**;

**ONE-WAY AND RCBD WITH n REPS AND p REPEATED MEASURES:**
**DATA** two; INPUT trt time block response; DATALINES; ;
**PROC MIXED** DATA=two PLOTS=RESIDUALPANEL; CLASS trt time block; MODEL response = trt time trt*time / DDFM=SATTERTH; RANDOM block; REPEATED / SUBJECT=block*trt TYPE =<option>;
**PROC GLIMMIX** DATA=two PLOTS=RESIDUALPANEL; CLASS trt time block; MODEL response = trt time trt*time / DDFM=KR; RANDOM block; RANDOM RESIDUAL / SUBJECT=block*trt TYPE =<option>;

**a*b Factorial in a CRD with Repeated Measures**
**DATA** three; INPUT trtA trtB time expunit response; DATALINES; ;
**PROC MIXED** DATA = three PLOTS = RESIDUALPANEL; CLASS trtA trtB time expunit; MODEL response = trtA|TrtB|time / DDFM = SATTERTH; /* DDFM = KR*/ REPEATED time/ SUBJECT = expunit(trtA trtB) type = <options>;
**PROC GLIMMIX** DATA = two PLOTS = RESIDUALPANEL; CLASS trt time expunit; MODEL response = trtA|TrtB|time / DDFM = SATTERTH; /* DDFM = KR*/ RANDOM RESIDUAL / SUBJECT = expunit(trtA*trtB) TYPE = <option>;
**RUN**; **QUIT**; /* No REPEATED in GLIMMIX */

**a*b Factorial in a RCBD with Repeated Measures**
**DATA** four; INPUT trtA trtB time block response; DATALINES;;
**PROC MIXED** DATA=four PLOTS=RESIDUALPANEL; CLASS trtA trtB time block; MODEL response = trtA | trtB | time / DDFM=SATTERTH; RANDOM block; REPEATED / SUBJECT = block*trtA*trtB TYPE =<option>;
**PROC GLIMMIX** DATA = four PLOTS = RESIDUALPANEL; CLASS trtA trtB time block; MODEL response = trtA | trtB | time / DDFM=KR; RANDOM block; RANDOM RESIDUAL / SUBJECT=block*trtA*trtB TYPE =<option>;

**Note:** After adopting a correlate structure, add LSMEANS, CONTRAST and ESTIMATE statements in all cases above. **$\underline{\text{Correlation Structures is given by TYPE = <option>}}$** where, options are VC or VC GROUP = time, CS or CSH, AR(1) or ARH(1), TOEP or TOEPH, UN. Covariance structure is determined by tests to compare covariance structures i.e. incorrect covariance structure ➔ incorrect test of effects.

**W3P3: Analysis of Covariance (ANCOVA) ANALYSIS:** The process of fitting a system of lines to data. The line reflect the differences between treatments levels and linear associate with the covariate using indicator variables. Lines may be parallel or non-parallel. Each line models the responses for a particular treatment level.
**Parallel Line Model:** $Y_{ij} = \alpha_i + \beta X_{ij} + \varepsilon_{ij}$; i = 1, 2, …, t; j = 1, 2, …, n
$J^{th}$ response of trt level i = y-intercept for trt level I + slope (same for all trt levels)*covariate value for the jth response in trt level i + random error $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$.
**Non-Parallel Line Model:** $Y_{ij} = \alpha_i + \beta_i X_{ij} + \varepsilon_{ij}$; $\beta_i$ = slope for trt level i.
**Analysis Procedure: Step 1:** Is the covariate necessary? $H_o$: $\beta_1 = \ldots = \beta_t = 0$. $H_a$: At least one slope is different. If $H_o$ not rejected, ➔ perform ANOVA on trt levels. The covariate information is not significant. If $H_o$ is rejected, Step 2. **Step 2:** Are lines parallel: $H_o$: $\beta_1 = \ldots = \beta_t$ $H_a$: At least one slope is different. If $H_o$ not rejected ➔ adopt parallel line model. Treatment levels can be compared at any values of the covariate. This would give the same information as the test of $H_o$: $\alpha_1 = \ldots = \alpha_t$. If $H_o$ is rejected, adopt unequal slope model. Compare the treatment levels at low, medium, and high values of the covariate (or other meaningful covariate values). That is, compare $H_{o:}$ $\alpha_1 + \beta_1 X_0 = \alpha_2 + \beta_1 X_0 = \alpha_t + \beta_t X_0$ where $X_0$ is a specified value.

```
DM 'LOG; CLEAR; ODSRESULTS; CLEAR; ';
SYMBOL;
*clears all plotting symbols and colors previously ;
*used in the current SAS session ;
DATA cov2;
INPUT epro hr ihr @@;
DATALINES;
1 118 56 2 148 60 3 153 56
1 138 59 2 159 62 3 150 58
1 142 62 2 162 65 3 158 61
1 147 68 2 157 66 3 152 64
1 160 71 2 169 73 3 160 72
1 166 76 2 164 75 3 154 75
1 165 83 2 179 84 3 155 82
1 171 87 2 177 88 3 164 86
;
PROC GPLOT DATA=cov2;
PLOT hr*ihr=epro;
SYMBOL1 VALUE=CIRCLE CV=ORANGE I=NONE; *Epro=1 ;
SYMBOL2 VALUE=DIAMOND CV=PURPLE I=NONE; *Epro=2 ;
SYMBOL3 VALUE=SQUARE CV=RED I=NONE; *Epro=3 ;
TITLE 'Analysis of Covariance - Exercise Program Example';
TITLE3 'Visual Examination of Data for a Linear Trend';
PROC GLM DATA=cov2 PLOTS=NONE;
CLASS epro ;
MODEL hr= epro /SOLUTION ;
LSMEANS epro /STDERR ;
TITLE3 'Usual ANOVA';
PROC GLM DATA=cov2 PLOTS=NONE;
CLASS epro;
MODEL hr = epro ihr*epro/NOINT SOLUTION;
TITLE3 'IHR*EPRO - Tests for the Necessity of a Covariate';
PROC GLM DATA=cov2 PLOTS=NONE;
CLASS epro;
MODEL hr = epro ihr ihr*epro;
TITLE3 'IHR*EPRO - Tests whether all slopes are equal.';
PROC GLM DATA=cov2 PLOTS(ONLY)=ANCOVAPLOT ;
CLASS epro;
MODEL hr = epro ihr*epro/solution;
```

```
LSMEANS epro / stderr pdiff; TITLE3 'An unequal slopes model is fit to the
data';
*Compare the slopes;
ESTIMATE 'B1 - B2' ihr*epro 1 -10;
ESTIMATE 'B1 - B3' ihr*epro 10 -1;
ESTIMATE 'B2 - B3' ihr*epro 01 -1;
*Compare the y-intercepts;
ESTIMATE 'Int 1 - Int 2' epro 1 -10;
ESTIMATE 'Int 1 - Int 3' epro 10 -1;
ESTIMATE 'Int 2 - Int 3' epro 01 -1;
*Compare the predicted means at a given value of the covariate;
ESTIMATE '1-2 at 55' epro 1 -10  ihr*epro 55 -550;
ESTIMATE '1-2 at 70' epro 1 -10  ihr*epro 70 -700;
ESTIMATE '1-2 at 85' epro 1 -10  ihr*epro 85 -850;
ESTIMATE '1-3 at 55' epro 10 -1  ihr*epro 550 -55;
ESTIMATE '1-3 at 70' epro 10 -1  ihr*epro 700 -70;
ESTIMATE '1-3 at 85' epro 10 -1  ihr*epro 850 -85;
ESTIMATE '2-3 at 55' epro 01 -1  ihr*epro 055 -55;
ESTIMATE '2-3 at 70' epro 01 -1  ihr*epro 070 -70;
ESTIMATE '2-3 at 85' epro 01 -1  ihr*epro 085 -85;
LSMEANS epro / PDIFF STDERR AT ihr=55 ;
LSMEANS epro / PDIFF STDERR AT ihr=70 ;
LSMEANS epro / PDIFF STDERR AT ihr=85 ;
*Recover the predicted values for the lines in a data table called new.;
OUTPUT OUT=new P=pred;
*Create a graph containing the original data and the predicted lines;
DATA plot;
SET new;
hrate = hr;
epro=epro;
OUTPUT;
hrate = pred;
epro=10*epro;
OUTPUT;
LABEL epro = 'Exercise Program'      hrate = 'Heart Rate';
PROC FORMAT;
VALUE eprofmt  1 = '1 observed'  10 = '1 predicted'  2 = '2 observed'  20 = '2
predicted'   3 = '3 observed'  30 = '3 predicted';
AXIS1 LABEL=(A=90) ;
LEGEND1 ACROSS=3 ;
PROC GPLOT DATA=plot ;
PLOT hrate*ihr = epro / LEGEND=LEGEND1 VAXIS=AXIS1 ;
SYMBOL1 VALUE=CIRCLE CV=ORANGE I=NONE; SYMBOL2 VALUE=DIAMOND
CV=PURPLE I=NONE;
SYMBOL3 VALUE=SQUARE CV=RED I=NONE ;
SYMBOL4 VALUE=NONE CV=ORANGE I=JOIN L=1;
SYMBOL5 VALUE=NONE CV=PURPLE I=JOIN L=5;
SYMBOL6 VALUE=NONE CV=RED I=JOIN L=8;
FORMAT epro eprofmt. ;
TITLE3 'Data and Lines for each Exercise Program';
RUN;  QUIT;
```