# Partitioning $SS[Tot]$ in $a \times b$ design (Two-way ANOVA)

$$\boxed{Y_{ijk} \quad = \quad \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + E_{ijk}}$$

$$(i = 1, 2, \ldots, a \text{ and } j = 1, 2, \ldots, b \text{ and } k = 1, 2, \ldots, n)$$

Deviations:

total : $y_{ijk} - \overline{y}_{\ldots}$

due to level $i$ of factor A: ________________

due to level $j$ of factor B: ________________

due to levels $i$ of factor A and $j$ of factor B after subtracting main effects:

$$\bar{y}_{ij\cdot} - \bar{y}_{\ldots} -$$

$$SS[Tot] = \sum_i \sum_j \sum_k (y_{ijk} - \qquad\qquad )^2 = \sum_i \sum_j \sum_k (\overline{y}_{ij\cdot} - \qquad\qquad )^2 + \sum_i \sum_j \sum_k (y_{ijk} - \qquad\qquad )^2$$

$$SS[A] = \sum_i \sum_j \sum_k ( \qquad\qquad )^2, \quad SS[B] = \sum_i \sum_j \sum_k ( \qquad\qquad )^2$$

$$SS[AB] = \sum_i \sum_j \sum_k ( \qquad\qquad\qquad )^2, SS[E] = \sum_i \sum_j \sum_k ( \qquad\qquad )^2$$

# ANOVA for two-factor crossed design

$$y_{ijk} = \bar{y}_{\cdot\cdot\cdot} + (\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot\cdot\cdot}) + (\bar{y}_{\cdot j\cdot} - \bar{y}_{\cdot\cdot\cdot}) + (\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdot\cdot\cdot}) + y_{ijk} - \bar{y}_{ij\cdot}$$

$$y_{ijk} - \bar{y}_{\cdot\cdot\cdot} = (\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot\cdot\cdot}) + (\bar{y}_{\cdot j\cdot} - \bar{y}_{\cdot\cdot\cdot}) + (\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdot\cdot\cdot}) + y_{ijk} - \bar{y}_{ij\cdot}$$

Square both sides, sum over $i, j, k$, and the $\times$-products vanish.

$$\begin{aligned} SS(Tot) &= SS(Trt) + SS(\qquad\qquad) \\ SS(Trt) &= SS(A) + \qquad\qquad + \end{aligned}$$

Analysis of replicated two (or more) factor designs often proceed according to the following steps:

1. Check for interaction
    1. If no interaction, analyze main effects
    2. If interaction, analyze simple effects

## $a \times b$ example continued

Test for interaction effect in $2 \times 2$ generalizes to $a \times b$:

$$H_0 : (\alpha\beta)_{ij} \equiv 0 \text{ vs. } H_1 : (\alpha\beta)_{ij} \neq 0 \text{ for some } i, j$$

$$F = \frac{MS[AB]}{MS[E]}$$

on $(a-1)(b-1)$ and $N - ab$ numerator, denominator $df$.

$$SS[AB] = n \sum_{i=1}^{3} \sum_{j=1}^{3} (\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdot\cdot\cdot})^2 = 0.597$$

$$F = \frac{.597/2}{0.025} = 11.96$$

which is highly significant ($p = 0.0014$) on 2,12 $df$.

We could proceed to test for main effects, but we won't.

Q: Why not?

A: Because effect of one factor depends on the level of the other factor, it might not make sense to talk about main effects.

If one insists on main effects, the appropriate $F$-ratios are

$$F_A = \frac{SS[A]/(a-1)}{MS[E]} \text{ on } a-1, N-ab \;\; df$$

$$F_B = \frac{SS[B]/(b-1)}{MS[E]} \text{ on } b-1, N-ab \;\; df$$

but the significance of the interaction effect suggests that the effect of one factor, say $A$, differs across levels of the other factor. A test for the main effect of $A$ is based on the effect of $A$ after *averaging over levels of* $B$. (Draw a picture.)

## $a \times b$ designs

Yields on 36 tomato crops from balanced, complete, crossed design with $a = 3$ varieties $(A)$ at $b = 4$ planting densities $(B)$ :

| Variety | Density $k/hectare$ | Sample | | |
|---|---|---|---|---|
| 1 | 10 | 7.9 | 9.2 | 10.5 |
| 2 | 10 | 8.1 | 8.6 | 10.1 |
| 3 | 10 | 15.3 | 16.1 | 17.5 |
| 1 | 20 | 11.2 | 12.8 | 13.3 |
| 2 | 20 | 11.5 | 12.7 | 13.7 |
| 3 | 20 | 16.6 | 18.5 | 19.2 |
| 1 | 30 | 12.1 | 12.6 | 14.0 |
| 2 | 30 | 13.7 | 14.4 | 15.4 |
| 3 | 30 | 18.0 | 20.8 | 21.0 |
| 1 | 40 | 9.1 | 10.8 | 12.5 |
| 2 | 40 | 11.3 | 12.5 | 14.5 |
| 3 | 40 | 17.2 | 18.4 | 18.9 |

Statistical model?

$$Y_{ijk} =$$

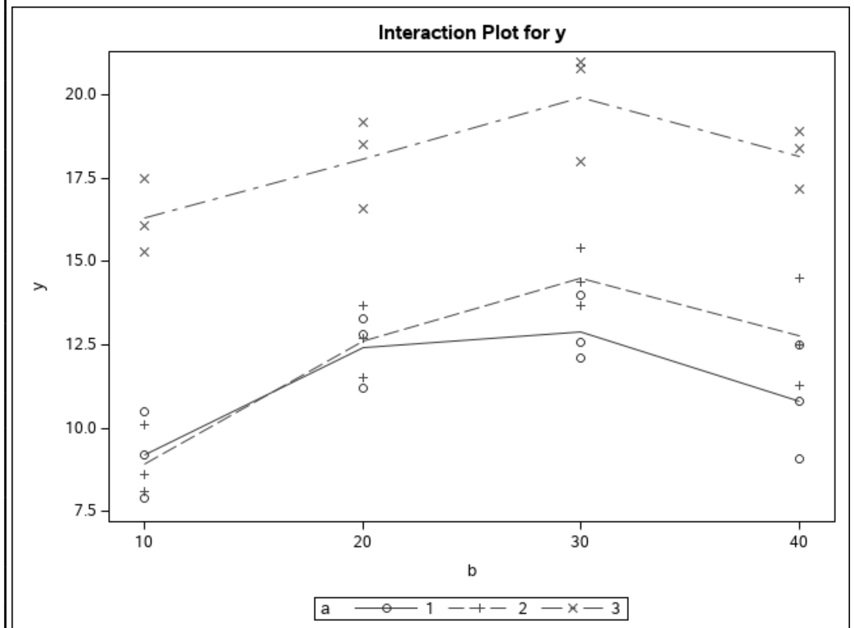
# ANOVA table

```
                         The SAS System                                  1
                        The GLM Procedure

                Class           Levels    Values
                a                    3    1 2 3
                b                    4    10 20 30 40

                                      Sum of
Source                      DF       Squares   Mean Square  F Value  Pr > F

Model                       11   422.3155556    38.3923232    24.22  <.0001
Error                       24    38.0400000     1.5850000
Corrected Total             35   460.3555556

Source                      DF     Type I SS   Mean Square  F Value  Pr > F
a                            2   327.5972222   163.7986111   103.34  <.0001
b                            3    86.6866667    28.8955556    18.23  <.0001
a*b                          6     8.0316667     1.3386111     0.84  0.5484
```

| Level of a | N | y Mean | y Std Dev |
|---|---|---|---|
| 1 | 12 | 11.3333333 | 1.88309867 |
| 2 | 12 | 12.2083333 | 2.34887142 |
| 3 | 12 | 18.1250000 | 1.73369023 |

| Level of b | N | y Mean | y Std Dev |
|---|---|---|---|
| 10 | 9 | 11.4777778 | 3.75458978 |
| 20 | 9 | 14.3888889 | 2.96835158 |
| 30 | 9 | 15.7777778 | 3.36480972 |
| 40 | 9 | 13.9111111 | 3.53250777 |

| Level of a | Level of b | N | y Mean | y Std Dev |
|---|---|---|---|---|
| 1 | 10 | 3 | 9.2000000 | 1.30000000 |
| 1 | 20 | 3 | 12.4333333 | 1.09696551 |
| 1 | 30 | 3 | 12.9000000 | 0.98488578 |
| 1 | 40 | 3 | 10.8000000 | 1.70000000 |
| 2 | 10 | 3 | 8.9333333 | 1.04083300 |
| 2 | 20 | 3 | 12.6333333 | 1.10151411 |
| 2 | 30 | 3 | 14.5000000 | 0.85440037 |
| 2 | 40 | 3 | 12.7666667 | 1.61658075 |
| 3 | 10 | 3 | 16.3000000 | 1.11355287 |
| 3 | 20 | 3 | 18.1000000 | 1.34536240 |
| 3 | 30 | 3 | 19.9333333 | 1.67729942 |
| 3 | 40 | 3 | 18.1666667 | 0.87368949 |



A conventional look at main effects is just to make pairwise comparisons among marginal means, after averaging over other factors. Pairwise comparisons of density means using Tukey's procedure with $\alpha = 0.05$ are given below. (Use `means b/tukey;`. to obtain the output.)

```
                          The GLM Procedure

            Tukey's Studentized Range (HSD) Test for y

  NOTE: This test controls the Type I experimentwise error rate, but it
        generally has a higher Type II error rate than REGWQ.

            Alpha                                      0.05
            Error Degrees of Freedom                     24
            Error Mean Square                         1.585
            Critical Value of Studentized Range     3.90126
            Minimum Significant Difference           1.6372


       Means with the same letter are not significantly different.


          Tukey Grouping          Mean       N     b

                    A          15.7778       9    30
                    A
                B   A          14.3889       9    20
                B
                B              13.9111       9    40

                    C          11.4778       9    10
```

$q = \text{qtukey}(.95, 4, 24)$

$q = 3.9$

$HSD = 3.9\sqrt{\frac{1.58}{9}}$

## A three-factor example

tanks of shrimp

In a balanced, complete, crossed design, $N = 36$ ~~shrimp~~ were randomized to $abc = 12$ treatment combinations from the factors below:

A1: Temperature at $25^o$ C
A2: Temperature at $35^o$ C
B1: Density of shrimp population at 80 shrimp/$40l$
B2: Density of shrimp population at 160 shrimp/$40l$
C1: Salinity at 10 units
C2: Salinity at 25 units
C3: Salinity at 40 units

The response variable of interest is weight gain $Y_{ijkl}$ after four weeks.

### Three-way ANOVA Model:

$$Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \varepsilon_{ijkl}$$

$$i = 1,2 \quad j = 1,2 \quad k = 1,2,3 \quad l = 1,2,3$$

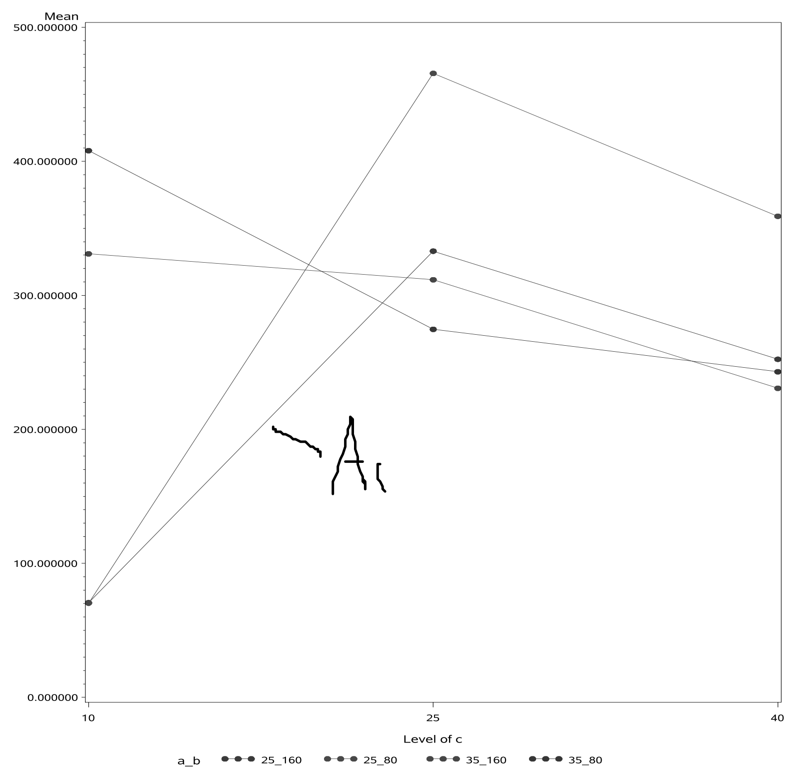
$$E_{ijkl} \overset{iid}{\sim} N(0, \sigma^2)$$

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 11 | 467636.3333 | 42512.3939 | 14.64 | <.0001 |
| Error | 24 | 69690.6667 | 2903.7778 | | |
| Corrected Total | 35 | 537327.0000 | | | |

sqrt(2MS(E)/3) ~= 44

| Source | DF | Type I SS | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| a | 1 | 15376.0000 | 15376.0000 | 5.30 | 0.0304 |
| b | 1 | 21218.7778 | 21218.7778 | 7.31 | 0.0124 |
| a*b | 1 | 8711.1111 | 8711.1111 | 3.00 | 0.0961 |
| c | 2 | 96762.5000 | 48381.2500 | 16.66 | <.0001 |
| a*c | 2 | 300855.1667 | 150427.5833 | 51.80 | <.0001 |
| b*c | 2 | 674.3889 | 337.1944 | 0.12 | 0.8909 |
| a*b*c | 2 | 24038.3889 | 12019.1944 | 4.14 | 0.0285 |

$A_2$




$-A_1$

$\mu[A, BC]$

```
Level of     Level of              -------------y-------------
a            b            N               Mean          Std Dev
25           80           9         298.333333       185.106051
25           160          9         218.666667       128.739077
35           80           9         308.555556        85.475305
35           160          9         291.111111        57.953525

Level of     Level of              -------------y-------------
a            c            N               Mean          Std Dev
25           10           6          70.500000        15.109600
25           25           6         399.333333       114.206246
25           40           6         305.666667        69.987618
35           10           6         369.500000        56.450864
35           25           6         293.166667        45.375838
35           40           6         236.833333        38.096807

Level of     Level of              -------------y-------------
b            c            N               Mean          Std Dev
80           10           6         239.166667       188.065326
80           25           6         370.166667       122.218520
80           40           6         301.000000        77.415761
160          10           6         200.833333       144.240655
160          25           6         322.333333        74.529636
160          40           6         241.500000        32.788718

Level of     Level of     Level of              -------------y-------------
a            b            c            N               Mean          Std Dev
25           80           10           3          70.333333        17.156146
25           80           25           3         465.666667        87.648921
25           80           40           3         359.000000        59.858166
25           160          10           3          70.666667        16.623277
25           160          25           3         333.000000       108.282039
25           160          40           3         252.333333        11.372481
35           80           10           3         408.000000        51.117512
35           80           25           3         274.666667        47.961790
35           80           40           3         243.000000        36.166283
35           160          10           3         331.000000        30.116441
35           160          25           3         311.666667        42.665365
35           160          40           3         230.666667        46.971623
```

## Interpretation of third order interaction

## Interpretation of second order interaction

$\boxed{1^{st} \text{ order}}$ interaction is between two factors

$\boxed{2^{nd} \text{ order}}$ interaction is between three factors

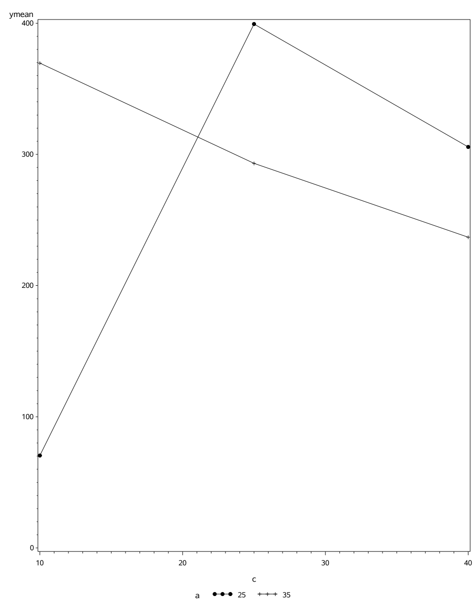
Upon inspection of the interaction plot, what do you see?
What is the primary two-factor/first-order interaction? A*C
Consider the means for low temperature (red and blue). Do you see evidence of BC interaction for temperature is low? Characterize it. Yes. No effect of B when C=1 Neg eff of B when C=2 or C=3
Do ~~you see~~ evidence of BC interaction for temperature is high?
Not so much

If there is a BC interaction at one level of A but not the other, this is a second-order interaction.

Characterization of a three-factor interaction may not be unique. Here we first fixed A, but another analyst might first fix some other factor and characterize factorial effects in a different order.

```
%let d=divisor;   *an example of a macro variable;
data one;
   drop i;           /* a=temp, b=density, c=salinity */
   input a b c @;    * @ hold the line. prevent DATA step from loading ;
   do i=1 to 3;      *   new record when next INPUT encountered;
      input y @;     * @ hold the line;   *love isn't always on time! (Toto) ;
      y0=sqrt(y);
      output;
   end;
   cards;
   25 80 10 86 52 73
   25 80 25 544 371 482
   25 80 40 390 290 397
   25 160 10 53 73 86
   25 160 25 393 398 208
   25 160 40 249 265 243
   35 80 10 439 436 349
   35 80 25 249 245 330
   35 80 40 247 277 205
   35 160 10 324 305 364
   35 160 25 352 267 316
   35 160 40 188 223 281
;
run;

proc glimmix data=one;
   class a b c;
   model y=a|b|c;
   lsmeans a*b*c/slicediff=a*c;
run;
```

```
proc glimmix data=one;
   class a b c;
   model y=a|b|c;
   estimate "temp effect at c=1" a -1 1 a*c  -1 0 0   1 0 0;
   estimate "temp effect at c=2" a -1 1 a*c  0 -1 0   0 1 0;
   estimate "temp effect at c=3" a -1 1 a*c  0 0 -1   0 0 1;
   estimate "avg of temp effects at c=2,3" a -2 2  a*c 0 -1 -1  0 1 1/&d=2;
   estimate "mu[AC1]-.5(mu[AC2]+mu[AC3]) " a*c -2 1 1  2 -1 -1/divisor=2;
run;
```




Estimates

| Label | Estimate | Standard Error | DF | t Value | Pr > \|t\| |
|---|---|---|---|---|---|
| temp effect at c=1 | 299.00 | 31.1115 | 24 | 9.61 | <.0001 |
| temp effect at c=2 | -106.17 | 31.1115 | 24 | -3.41 | 0.0023 |
| temp effect at c=3 | -68.8333 | 31.1115 | 24 | -2.21 | 0.0367 |
| avg of temp effects at c=2,3 | -87.5000 | 21.9992 | 24 | -3.98 | 0.0006 |
| mu[AC1]-.5(mu[AC2]+mu[AC3]) | 386.50 | 38.1037 | 24 | 10.14 | <.0001 |

We've characterized the $A \times C$ interaction. Note $SS(AC)$.