THREE-WAY ANOVA MODELS (CHAPTER 7)

Completely randomized design for an experiment with three treatment factors A, B and C.

Every combination of levels of A, B and C is observed (so factors are *crossed*).

Notation:

A has a levels, coded 1, 2, … , a
B has b levels, coded 1,2, … , b
C has c levels, coded 1, 2, … , c
v = total number of treatments (= abc)

Example: Pollution noise data (http://lib.stat.cmu.edu/DASL/Datafiles/airpullutionfiltersdat.html)

Data presented by Texaco, Inc. in 1973 to support their claim that the Octel pollution filter was at least equal in noise reduction as standard silencers.

Variables:

NOISE = Noise level reading in decibels (response)

SIZE = Vehicle size: 1 small, 2 medium, 3 large

TYPE: 1 standard silencer, 2 Octel filter

SIDE : 1 right side of car, 2 left side of car

All combinations of size, type, and side were observed (12 treatments in all).

Models: Y_{ijkt} denotes the random variable giving the response for observation t of the treatment at level i of A, level j of B, and level k of C. (r_{ijk} = number of observations at level i of A, level j of B, and level k of C.)

1. *Cell-means model*:

   $Y_{ijkt} = \mu + \tau_{ijk} + \varepsilon_{ijkt}$

   The ε_{ijkt} are independent random variables.

   Each $\varepsilon_{ijkt} \sim N(0, \sigma^2)$

2. *Main effects model* (also known as the *three-way additive model*):

   $Y_{ijkt} = \mu + \alpha_i + \beta_j + \gamma_k + \varepsilon_{ijkt}$

   The ε_{ijkt} are independent random variables.

   Each $\varepsilon_{ijkt} \sim N(0, \sigma^2)$

3. The *three-way analysis of variance model* (also known as the *three-way complete model*):

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

   The ε_{ijkt} are independent random variables.

   Each $\varepsilon_{ijkt} \sim N(0, \sigma^2)$

Each term $(\alpha\beta\gamma)_{ijk}$ is called a *three-way interaction term*.

This model is equivalent to the cell-means model – i.e., gives the same fits and residuals.

4. Various other models lying between the main effects model and the complete model.

As with two-way models, it is good practice to work only with *hierarchical models* – that is, if an interaction term is included in the model, all "subterms" should be included – e.g., if the three-way interaction term is included, then the complete model should be included.

What does three-way interaction mean?

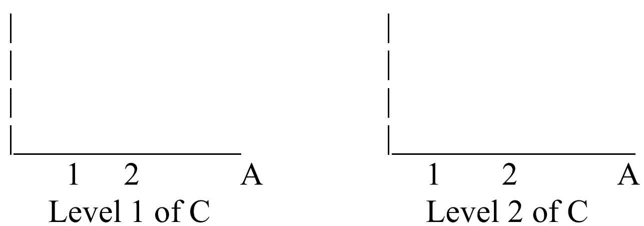
Intuitively, we want it to mean that the interaction between two factors depends on the level of the third factor.

However, our model is a linear model, so we can capture only certain types of "dependence on level."

The following examples of models *without* a three-way interaction term illustrate the possibilities for having no three-way interaction term:

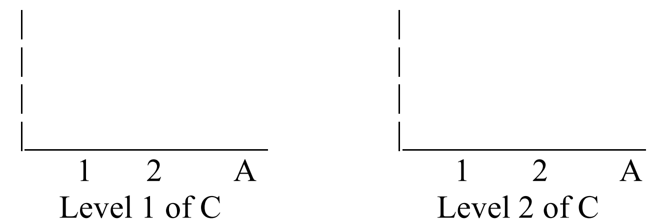
1. $Y_{ijkt} = \beta_j + (\beta\gamma)_{jk} + \varepsilon_{ijkt}$, for i, j, k = 1,2, where $\beta_2 = 1$, $(\beta\gamma)_{22} = 1$, and all other parameters are 0.

A, B interaction plots for the two levels of C:



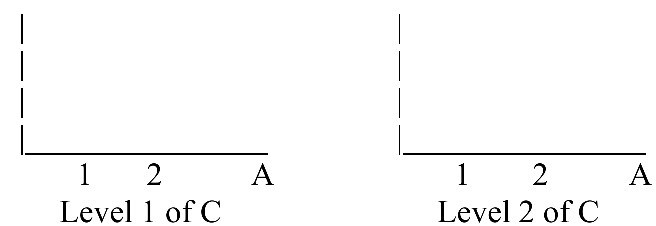
2. $Y_{ijkt} = \beta_j + (\alpha\gamma)_{ik} + \varepsilon_{ijkt}$, for i, j, k = 1,2, where $\beta_2 = 1$, $(\alpha\gamma)_{22} = 1$, and all other parameters are 0.

A, B interaction plots for the two levels of C:



3. $Y_{ijkt} = \beta_j + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + \varepsilon_{ijkt}$, for i, j, k = 1,2, where $\beta_2 = 1$, $(\alpha\gamma)_{22} = 1$, $(\beta\gamma)_{22} = 1$and all other parameters are 0.

A, B interaction plots for the two levels of C:



4. $Y_{ijkt} = (\alpha\beta)_{ij} + (\beta\gamma)_{jk} + \varepsilon_{ijkt}$, for i, j, k = 1,2, where, $(\alpha\beta)_{22} = 1$, $(\beta\gamma)_{22} = 1$, and all other parameters are 0.

2 2 A

Level 1 of C

2 2 A

Level 2 of C

Note that in each of the first three cases, when we moved from level 1 of C to level 2 of C, we retained the property of having parallel lines (that is, of having no two-way interaction), although the distance between the lines and the slopes of the lines could change.

What is the same and what is different about the fourth model?

Now modify the first model by adding various three-way interaction terms:

1'. $Y_{ijkt} = \beta_j + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \varepsilon_{ijkt}$, for i, j, k = 1,2, where $\beta_2 = 1$, $(\beta\gamma)_{22} = 1$, $(\alpha\beta\gamma)_{222} = 1$, and all other parameters are 0.

A, B interaction plots for the two levels of C:

1 2 A

Level 1 of C

1 2 A

Level 2 of C

1''. $Y_{ijkt} = \beta_j + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \varepsilon_{ijkt}$, for i, j, k = 1,2, where $\beta_2 = 1$, $(\beta\gamma)_{22} = 1$, $(\alpha\beta\gamma)_{222} = -2$, and all other parameters are 0.

A, B interaction plots for the two levels of C:

1 2 A

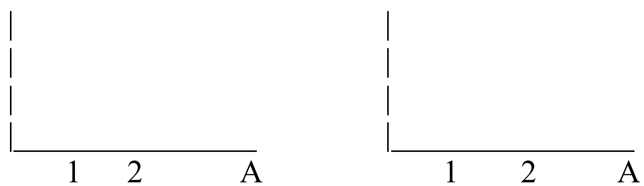
Level 1 of C

1 2 A

Level 2 of C

1'''. $Y_{ijkt} = \beta_j + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \varepsilon_{ijkt}$, for i, j, k = 1,2, where $\beta_2 = 1$, $(\beta\gamma)_{22} = 1$, $(\alpha\beta\gamma)_{212} = 2$, $(\alpha\beta\gamma)_{222} = -2$, and all other parameters are 0.

A, B interaction plots for the two levels of C:



Level 1 of C Level 2 of C

The above examples suggest: "no three-way interaction" (as measured by the absence of all three-way interaction terms) means that *the difference in slopes is independent of the level of C.*

Looking at this more generally (algebraically):

The difference in the slopes in the two interaction plots above are

$$(\tau_{211} - \tau_{111}) - (\tau_{221} - \tau_{121}) \text{ and } (\tau_{212} - \tau_{112}) - (\tau_{222} - \tau_{122}).$$

If there are no three-way interaction terms, then the first difference in slopes is

$$\begin{aligned}&\{[\alpha_2 + \beta_1 + (\alpha\beta)_{21} + (\alpha\gamma)_{21} + (\beta\gamma)_{11}] \\ &\qquad - [\alpha_1 + \beta_1 + (\alpha\beta)_{11} + (\alpha\gamma)_{11} + (\beta\gamma)_{11}]\} \\ &- \{[\alpha_2 + \beta_2 + (\alpha\beta)_{22} + (\alpha\gamma)_{21} + (\beta\gamma)_{21}] \\ &\qquad - [\alpha_1 + \beta_2 + (\alpha\beta)_{12} + (\alpha\gamma)_{11} + (\beta\gamma)_{21}]\end{aligned}$$

$$= [(\alpha\beta)_{21} - (\alpha\beta)_{11}] - [(\alpha\beta)_{22} - (\alpha\beta)_{12}],$$

and the second difference in slopes is exactly the same, since it differs only in the gamma subscripts, but all terms with gamma subscripts have cancelled out.

Thus: "No three-way interaction terms" tells us that *the difference in slopes in A,B interaction plots is independent of the level of C.*

In other words: "no three-way interaction terms" tells us that *the difference of the difference of slopes in A,B interaction plots for two levels of C is zero.*

In the *complete* three-way model, the first difference of slopes we calculated above would have the additional terms

$$[(\alpha\beta\gamma)_{211} - (\alpha\beta\gamma)_{111}] - [(\alpha\beta\gamma)_{221} - (\alpha\beta\gamma)_{121}].$$

The second difference of slopes would have the additional terms

$$[(\alpha\beta\gamma)_{212} - (\alpha\beta\gamma)_{112}] - [(\alpha\beta\gamma)_{222} - (\alpha\beta\gamma)_{122}].$$

So if we want to test for no three-way interaction (for just these levels), our null hypothesis will be

$$H_0^{ABC}\colon \{[(\alpha\beta\gamma)_{211} - (\alpha\beta\gamma)_{111}] - [(\alpha\beta\gamma)_{221} - (\alpha\beta\gamma)_{121}]\} - \{[(\alpha\beta\gamma)_{212} - (\alpha\beta\gamma)_{112}] - [(\alpha\beta\gamma)_{222} - (\alpha\beta\gamma)_{122}]\} = 0.$$

For more levels of A, B, and C, the null hypothesis would be

$$H_0^{ABC}\colon \{[(\alpha\beta\gamma)_{i+1,jk} - (\alpha\beta\gamma)_{ijk}] - [(\alpha\beta\gamma)_{i+1,qk} - (\alpha\beta\gamma)_{iqk}]\} - \{[(\alpha\beta\gamma)_{i+1,jr} - (\alpha\beta\gamma)_{ijr}] - [(\alpha\beta\gamma)_{i+1,qr} - (\alpha\beta\gamma)_{iqr}]\} = 0$$

We will return to this later.

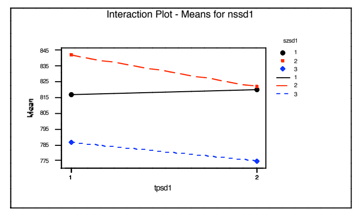
Example: Pollution Noise Data

We can make empirical three-way interaction plots in Minitab as follows:

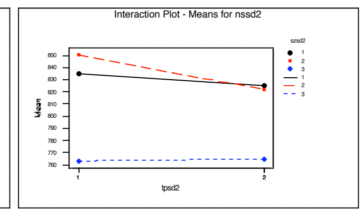
1. Unstack two factors and response by the third factor.

2. Use the unstacked data to form one interaction plot for each level of the third factor.

The interaction plots of type and size for side = 1 and 2, respectively are:



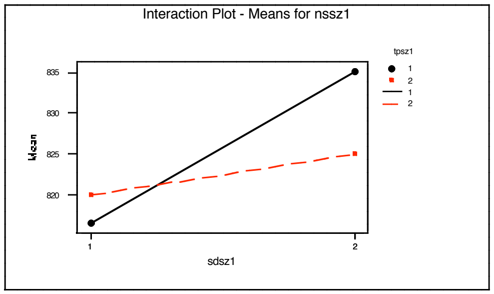
Side = 1



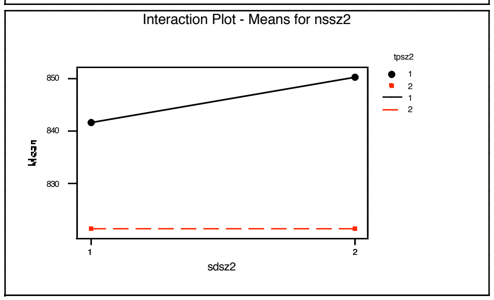
Side = 2

Does this suggest three-way interaction?

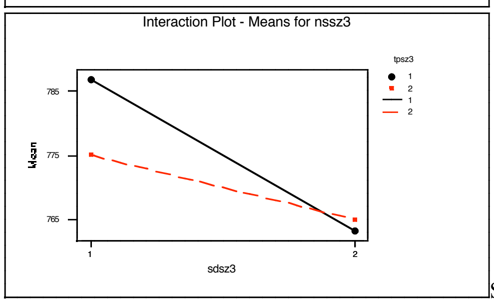
Interaction plots of type and side for size = 1, 2, 3, respectively:



Side = 1



Side = 2



Side = 3