Calculating Effects with Linear Models

BIOE 498/598 PJ

Spring 2022

An additive model of effects and interactions

For our reactor example with factors T, C, and K:

$$\begin{aligned}\mathsf{yield} = \beta_0 &+ \beta_{\mathsf{T}}\mathsf{T} + \beta_{\mathsf{C}}\mathsf{C} + \beta_{\mathsf{K}}\mathsf{K} \\ &+ \beta_{\mathsf{TC}}\mathsf{TC} + \beta_{\mathsf{TK}}\mathsf{TK} + \beta_{\mathsf{CK}}\mathsf{CK} \\ &+ \beta_{\mathsf{TCK}}\mathsf{TCK}\end{aligned}$$

where each regression coefficient β_i is half of the ith effect:

$$\beta_{\mathsf{T}} = \frac{ME(\mathsf{T})}{2}, \ldots, \beta_{\mathsf{TCK}} = \frac{Int(\mathsf{TCK})}{2}$$

and the intercept β_0 is the mean of all the responses.

For the reactor example

Effect	**Size**	β
intercept	64.25	64.25
T	23	11.5
C	−5	−2.5
K	1.5	0.75
TC	1.5	0.75
TK	10	5
CK	0	0
TCK	0.5	0.25

$$\begin{aligned}\text{yield} = 64.25 &+ 11.5\,\mathsf{T} - 2.5\mathsf{C} + 0.75\,\mathsf{K}\\ &+ 0.75\,\mathsf{TC} + 5\,\mathsf{TK} + 0\,\mathsf{CK}\\ &+ 0.25\,\mathsf{TCK}\end{aligned}$$

For the reactor example

Effect	Size	β
intercept	64.25	64.25
T	23	11.5
C	−5	−2.5
K	1.5	0.75
TC	1.5	0.75
TK	10	5
CK	0	0
TCK	0.5	0.25

$$\begin{aligned}\text{yield} = 64.25 &+ 11.5\,\mathsf{T} - 2.5\mathsf{C} + 0.75\,\mathsf{K}\\ &+ 0.75\,\mathsf{TC} + 5\,\mathsf{TK} + 0\,\mathsf{CK}\\ &+ 0.25\,\mathsf{TCK}\end{aligned}$$

What is the yield for the treatment T $= +$, C $= -$, K $= +$?

For the reactor example

Effect	Size	β
intercept	64.25	64.25
T	23	11.5
C	−5	−2.5
K	1.5	0.75
TC	1.5	0.75
TK	10	5
CK	0	0
TCK	0.5	0.25

$$\begin{aligned}\text{yield} = 64.25 &+ 11.5\,\mathsf{T} - 2.5\mathsf{C} + 0.75\,\mathsf{K}\\ &+ 0.75\,\mathsf{TC} + 5\,\mathsf{TK} + 0\,\mathsf{CK}\\ &+ 0.25\,\mathsf{TCK}\end{aligned}$$

What is the yield for the treatment T $= +$, C $= -$, K $= +$?

$$\begin{aligned}\text{yield} &= 64.25 + 11.5(1) - 2.5(-1) + 0.75(+1)\\ &\quad + 0.75(1)(-1) + 5(1)(1) + 0(-1)(1)\\ &\quad + 0.25(1)(-1)(1)\\ &= 64.25 + 11.5 + 2.5 + 0.75 - 0.75 + 5 + 0 - 0.25\\ &= 83\end{aligned}$$

Why are the coefficients half of the effect sizes?

Imagine the simplest model with one factor T.

$$y = \beta_0 + \beta_{\mathsf{T}}\mathsf{T}$$

Why are the coefficients half of the effect sizes?

Imagine the simplest model with one factor T.

$$y = \beta_0 + \beta_{\mathsf{T}} \mathsf{T}$$

Remember the definiton of the main effect of T:

$$ME(\mathsf{T}) = \bar{y}(\mathsf{T}+) - \bar{y}(\mathsf{T}-)$$

Why are the coefficients half of the effect sizes?

Imagine the simplest model with one factor T.

$$y = \beta_0 + \beta_{\mathsf{T}}\mathsf{T}$$

Remember the definiton of the main effect of T:

$$ME(\mathsf{T}) = \bar{y}(\mathsf{T}+) - \bar{y}(\mathsf{T}-)$$

From our model

$$\begin{aligned}\bar{y}(\mathsf{T}+) &= \beta_0 + \beta_{\mathsf{T}}(+1)\\ &= \beta_0 + \beta_{\mathsf{T}}\end{aligned}$$

$$\begin{aligned}\bar{y}(\mathsf{T}-) &= \beta_0 + \beta_{\mathsf{T}}(-1)\\ &= \beta_0 - \beta_{\mathsf{T}}\end{aligned}$$

Why are the coefficients half of the effect sizes?

Imagine the simplest model with one factor T.

$$y = \beta_0 + \beta_{\mathsf{T}}\mathsf{T}$$

Remember the definiton of the main effect of T:

$$ME(\mathsf{T}) = \bar{y}(\mathsf{T}+) - \bar{y}(\mathsf{T}-)$$

From our model

$$\begin{aligned}\bar{y}(\mathsf{T}+) &= \beta_0 + \beta_{\mathsf{T}}(+1)\\ &= \beta_0 + \beta_{\mathsf{T}}\end{aligned}$$

$$\begin{aligned}\bar{y}(\mathsf{T}-) &= \beta_0 + \beta_{\mathsf{T}}(-1)\\ &= \beta_0 - \beta_{\mathsf{T}}\end{aligned}$$

So according to the model,

$$\begin{aligned}ME(\mathsf{T}) &= \bar{y}(\mathsf{T}+) - \bar{y}(\mathsf{T}-)\\ &= (\beta_0 + \beta_{\mathsf{T}}) - (\beta_0 - \beta_{\mathsf{T}})\\ &= 2\beta_{\mathsf{T}}\end{aligned}$$

Advantages of estimating effects with linear models

- Easier calculation
- Statistical significance for coefficients (coming soon!)
- Predictions for untested treatments
- Extrapolation beyond the design space (for quantitative factors)

Step 1: Load the data

```r
pilot_data <- read.csv("PilotPlantDesign.csv")
pilot_data
```

```
##   run  T  C  K yield
## 1   1 -1 -1 -1    60
## 2   2  1 -1 -1    72
## 3   3 -1  1 -1    54
## 4   4  1  1 -1    68
## 5   5 -1 -1  1    52
## 6   6  1 -1  1    83
## 7   7 -1  1  1    45
## 8   8  1  1  1    80
```

Step 2: Look at the data

We can visualize factorial designs with a factor-and-response plot, or farplot.

First, we need to install the `doetools` package. If you haven't already, install the `devtools` package:

```
install.packages("devtools")
```

You only need to run this once this semester.

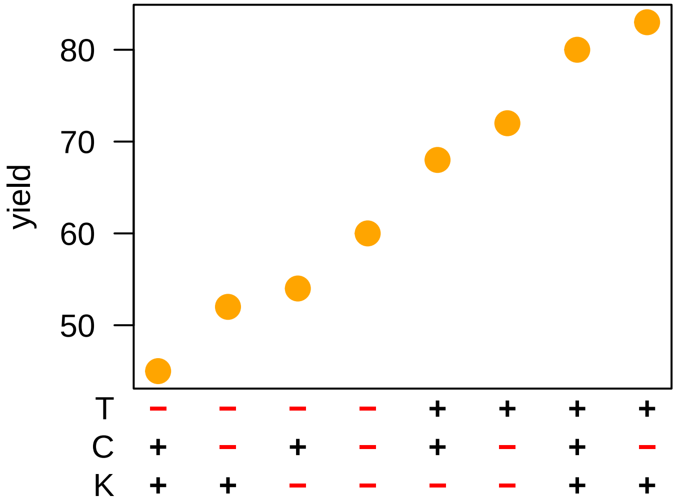
Now you can use `devtools` to install `doetools`:

```
devtools::install_github("jensenlab/doetools")
```

You should re-install this package before every assignment in case we add anything to the package during the semester.

Now, let's visualize the data

```
library(doetools)
farplot(pilot_data, response="yield", factors=c("T", "C", "K"))
```



Step 3: Fit a linear model

```
model <- lm(yield ~ T * C * K, data=pilot_data)
show_effects(model)
```

```
## (Intercept)    64.25
##           T    11.5
##           C    -2.5
##           K      .75
##         T:C      .75
##         T:K     5.
##         C:K      .
##       T:C:K      .25
```

Fitting linear models with `lm`

The standard call to `lm` is

```
model <- lm(<formula>, data=<dataframe>)
```

where `<formula>` takes the form

```
response ~ effects
```

Fitting linear models with `lm`

The standard call to `lm` is

```
model <- lm(<formula>, data=<dataframe>)
```

where `<formula>` takes the form

```
response ~ effects
```

Formulas don't include the coefficients, only the effects. For example the model

$$\text{yield} = \beta_0 + \beta_{\mathsf{T}}\mathsf{T} + \beta_{\mathsf{K}}\mathsf{K} + \beta_{\mathsf{TK}}\mathsf{TK}$$

would be specified

```
yield ~ 1 + T + K + T:K
```

Fitting linear models with `lm`

The standard call to `lm` is

```
model <- lm(<formula>, data=<dataframe>)
```

where `<formula>` takes the form

```
response ~ effects
```

Formulas don't include the coefficients, only the effects. For example the model

$$\text{yield} = \beta_0 + \beta_{\mathsf{T}}\mathsf{T} + \beta_{\mathsf{K}}\mathsf{K} + \beta_{\mathsf{TK}}\mathsf{TK}$$

would be specified

```
yield ~ 1 + T + K + T:K
```

or, since R assumes the model has an intercept, we can omit the `1`.

```
yield ~ T + K + T:K
```

More about formulas

The * operator is a shortcut for adding main effects and interactions.

```
yield ~ T + K + T:K
```

is equivalent to

```
yield ~ T*K
```

More about formulas

The * operator is a shortcut for adding main effects and interactions.

```
yield ~ T + K + T:K
```

is equivalent to

```
yield ~ T*K
```

Our complete model

```
yield ~ T + C + K + T:C + T:K + C:K + T:C:K
```

can be written

```
yield ~ T*C*K
```

More about formulas

The * operator is a shortcut for adding main effects and interactions.

```
yield ~ T + K + T:K
```

is equivalent to

```
yield ~ T*K
```

Our complete model

```
yield ~ T + C + K + T:C + T:K + C:K + T:C:K
```

can be written

```
yield ~ T*C*K
```

The effects in the formula must match column names in the data frame.
R looks up the effects in the data frame to construct the *model matrix*.

Why do we prefer coded factors?

```r
pilot_planning <- read.csv("PilotPlantPlanning.csv")
pilot_planning
```

```
##   run   T  C K yield
## 1   1 160 20 A    60
## 2   2 180 20 A    72
## 3   3 160 40 A    54
## 4   4 180 40 A    68
## 5   5 160 20 B    52
## 6   6 180 20 B    83
## 7   7 160 40 B    45
## 8   8 180 40 B    80
```

```r
model_uncoded <- lm(yield ~ T*C*K, data=pilot_planning)
```

Comparing coded vs. uncoded models

```
show_effects(model) # coded
```

```
## (Intercept)    64.25
##           T    11.5
##           C    -2.5
##           K      .75
##         T:C      .75
##         T:K     5.
##         C:K      .
##       T:C:K      .25
```

```
show_effects(model_uncoded)
```

```
## (Intercept)    -14.
##           T       .5
##           C     -1.1
##          KB   -143.
##         T:C       .005
##        T:KB       .85
##        C:KB     -.85
##      T:C:KB       .005
```

Comparing coded vs. uncoded models

```
show_effects(model) # coded
```

```
## (Intercept)    64.25
##           T    11.5
##           C    -2.5
##           K      .75
##         T:C      .75
##         T:K     5.
##         C:K      .
##       T:C:K      .25
```

```
show_effects(model_uncoded)
```

```
## (Intercept)    -14.
##           T       .5
##           C     -1.1
##          KB   -143.
##         T:C       .005
##        T:KB       .85
##        C:KB      -.85
##      T:C:KB       .005
```

```
show_effects(model_fahr)
```

```
## (Intercept)    -22.88889
##      T_fahr       .27778
##           C     -1.18889
##          KB   -158.11111
##    T_fahr:C       .00278
##   T_fahr:KB       .47222
##        C:KB      -.93889
## T_fahr:C:KB       .00278
```