MLR Interactions and variable transformations

a **statsTeachR** resource

Today's Lecture

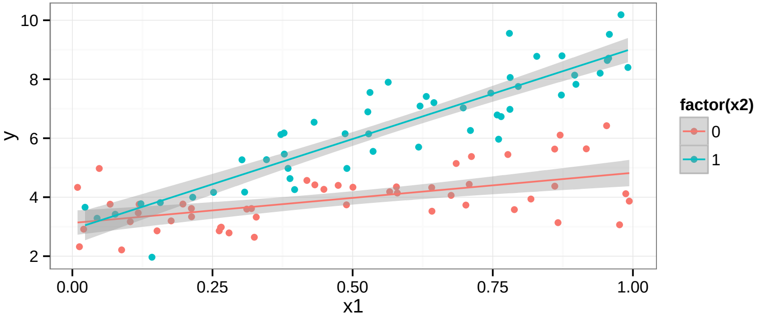
A few important building blocks for regression

- Interaction models
- Transformations of predictors

What is interaction?

Definition of interaction

Interaction occurs when the relationship between two variables depends on the value of a third variable.



[Good overview: KNN pp. 306–313]

Some real world examples?

How to include interaction in a MLR

Model A: $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i$
Model B: $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1} \cdot x_{i2} + \epsilon_i$

Key points

- "easily" conceptualized with 1 continuous, 1 categorical variable
- models possible with other variable combinations, but interpretation/visualization harder
- two variable interactions are considered "first-order" interactions (often used to define a class of models)
- still a **linear** model, but no longer a strictly **additive** model

How to interpret an interaction model

For now, assume x_1 is continuous, x_2 is 0/1 binary.
Model A: $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i$
Model B: $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i1} \cdot x_{i2} + \epsilon_i$

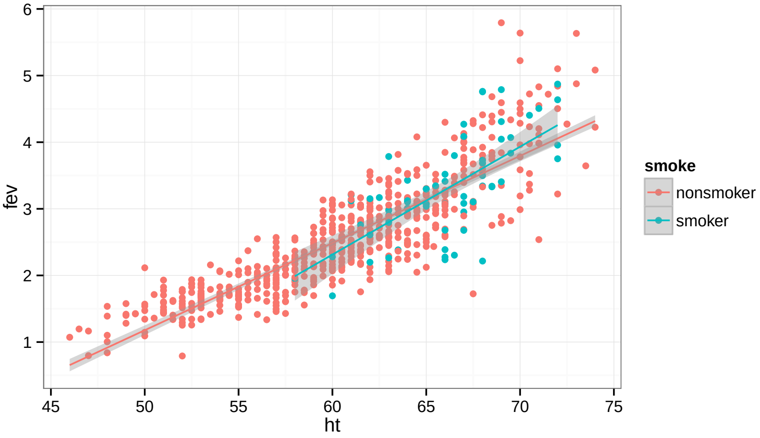
Example interaction model with FEV data

$$fev_i = \beta_0 + \beta_1 age_i + \beta_2 ht_i + \beta_3 sex_{i2} + \beta_4 smoke_i + \beta_5 ht \cdot smoke_i + \epsilon_i$$

```
mi1 <- lm(fev ~ age + ht + sex + smoke, data=dat)
mi3 <- lm(fev ~ age + ht*smoke + sex, data=dat)
c(AIC(mi1), AIC(mi3))

## [1] 703.8 700.5

round(summary(mi3)$coef,2)

##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -4.35       0.23  -19.12     0.00
## age                0.07       0.01    7.17     0.00
## ht                 0.10       0.00   21.08     0.00
## smokesmoker       -2.61       1.10   -2.37     0.02
## sexmale            0.15       0.03    4.43     0.00
## ht:smokesmoker     0.04       0.02    2.30     0.02
```

Example interaction model with FEV data

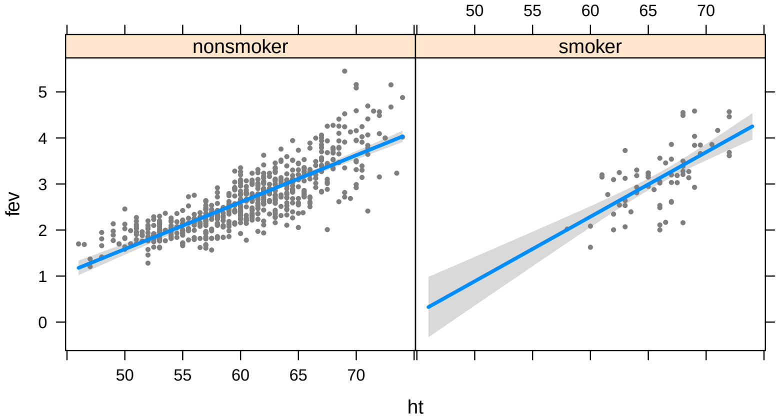
```
qplot(ht, fev, data=dat, color=smoke,
      geom=c("point", "smooth"), method="lm")
```



Example interaction model with FEV data

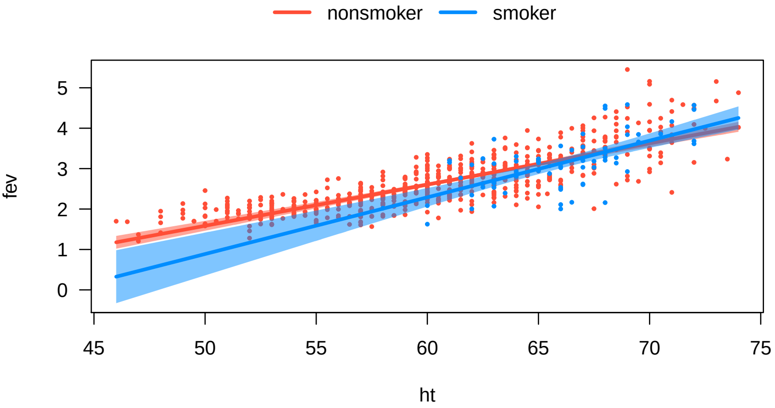
The `visreg` package plots not the data but the partial residuals (a.k.a. the adjusted variable) plot.

```r
require(visreg)
visreg(mi3, "ht", by="smoke")
```



Example interaction model with FEV data

```
visreg(mi3, "ht", by="smoke", overlay=TRUE)
```



Overview of variable transformations

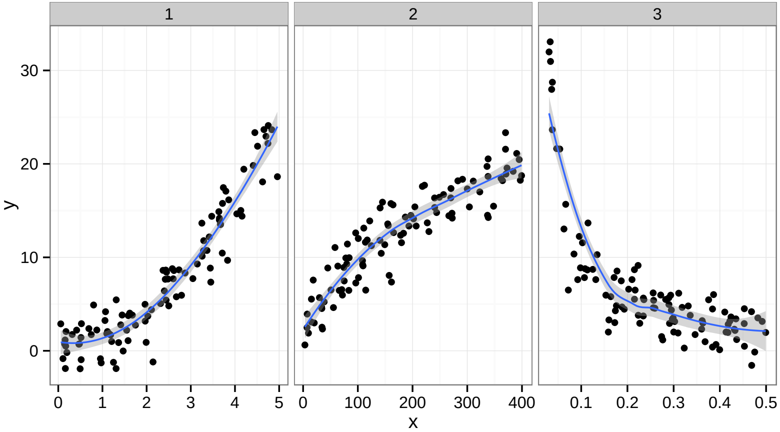
The problems

- Non-linearity between X and $Y \longrightarrow$ transform X
- Skewed distribution of Xs/points with high leverage $\longrightarrow$ transform X
- Non-constant variance $\longrightarrow$ transform Y

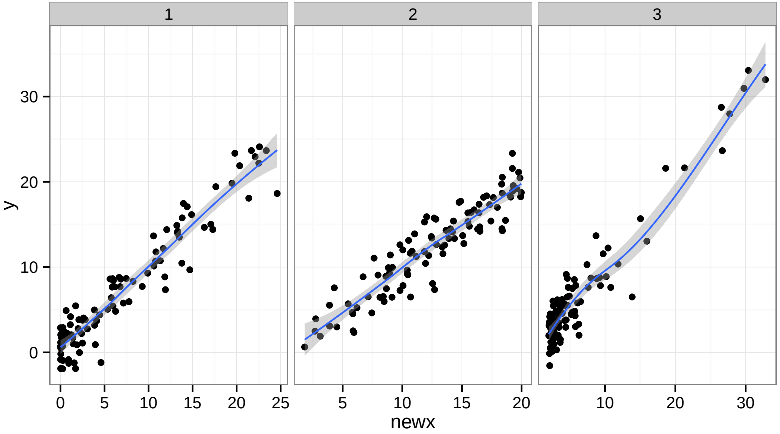
[More info: KNN Ch 3.9, pp. 129–137]

Transforming your X variables

Transforming predictor variables can help with constant-variance non-linear relationships.



Transforming your X variables



β interpretations with transformed Xs

Transforming predictor variables can help with non-linearities, but can make coefficient interpretations hard.

Possible solutions

- Interpret βs qualitatively across a region of interest: "We found strong evidence for a inverse association, where values of Y decreased inversely proportional to X across the observed range (a, b).
- Occasionally, a "one unit change in X" can be meaningful: e.g. $\log_a X$. A one unit change in $\log_a X$ indicates a a-fold increase in X.

β interpretations with transformed Xs

Transforming predictor variables can help with non-linearities, but can make coefficient interpretations hard.

Transforming Ys for non-constant variance

What to do ...

- Nothing; just use least squares and bootstrap
- Use weighted LS, GLS (Methods 3?)
- Use a variance stabilizing transformation
- Consider a generalized linear model (more soon)

Box-Cox Transformations

Outcome is raised to the λ power:

$$y_i^\lambda = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i$$

- Estimate λ, a new parameter, by maximum likelihood.
- Some well-known choices of λ: 2, -1, $1/2$
- By definition, when $\lambda = 0$, we specify $y_i^\lambda = \log_e y_i$

[More detailed info: KNN Ch 3.9, pp. 134–137]

Wrap-up

New instruments for your regression tool-kit

- Interactions and data transformations are common extensions/additions to regression models in practice.
- Both are simple to implement, challenging to interpret correctly!
- But, you may not always need a interpretation, e.g. you might just want a good prediction.