MLR: miscellaneous practical tools

Author: Nicholas G Reich, Jeff Goldsmith

*This material is part of the **statsTeachR** project*

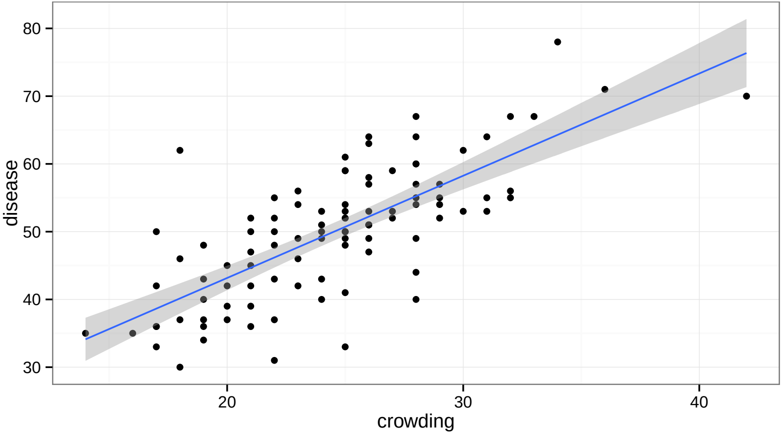
Today's Lecture

A few miscellaneous but important building blocks for regression

- Advanced residual plots
- Interaction models
- Transformations of predictors

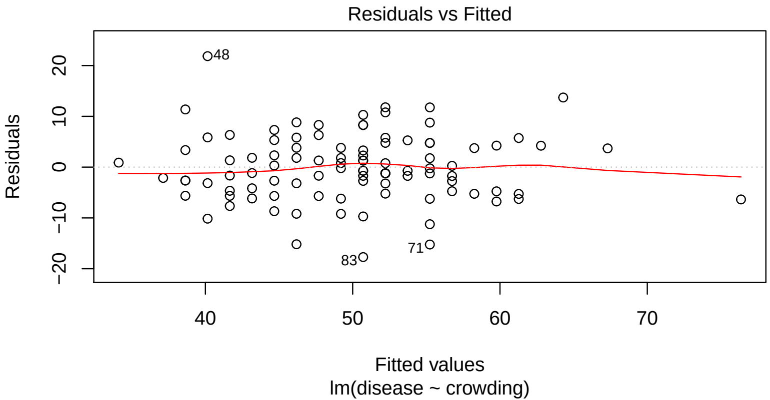
Typical regression plot: fitted line

```
qplot(crowding, disease, geom=c("point", "smooth"),
      method="lm", data=data)
```



Typical residual plot: fitted vs. residuals

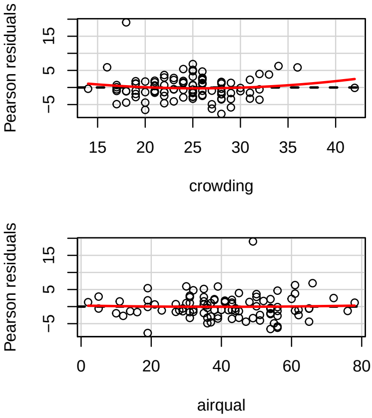
```
slr1 <- lm(disease ~ crowding, data=data)
plot(slr1, which=1)
```



But this is more complicated with MLR: how do we visualize adjusted multivariable relationships?

Predictor vs. residual plots

```
library(car)
mlr1 <- lm(disease ~ crowding + education + airqual, data=data)
residualPlots(mlr1, tests=FALSE)
```



Checking model structure: adjusted variable plots!

- You can plot residuals against each of the predictors, or plot outcomes against predictors, BUT...
- Keep in mind the MLR uses adjusted relationships; scatterplots don't show that adjustment!

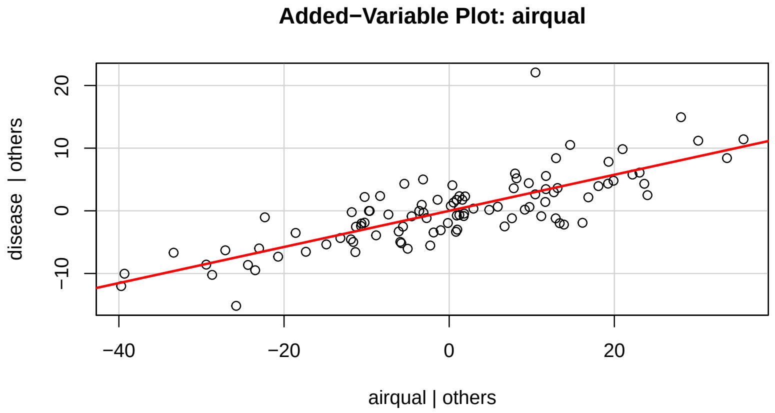
Adjusted variable plots (partial regression plots, added variable plots) can be useful.

Adjusted (or added) variable plots

- Regress y on everything but x_j; take residuals $r_{y|-x_j}$
- Regress x_j on everything but x_j; take residuals $r_{x_j|-x_j}$
- Regress $r_{y|-x_j}$ on $r_{x_j|-x_j}$; slope of this line will match β_j in the full MLR
- Plot of $r_{y|-x_j}$ against $r_{x_j|-x_j}$ shows the "adjusted" relationship
- This figure can be used to diagnose violations of linearity in MLR models.

AV plots

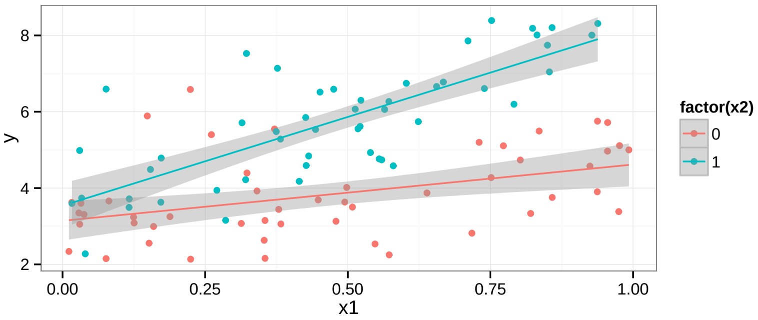
```r
library(visreg)
avPlot(mlr1, variable="airqual")
```



What is interaction?

Definition of interaction

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



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

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

β_3 is the change in the slope of the line that describes the relationship of $y \sim x_1$ comparing the groups defined by $x_2 = 0$ and $x_2 = 1$.
$\beta_1 + \beta_3$ is the expected change in y for a one-unit increase in x_1 in the group $x_2 = 1$.
$\beta_0 + \beta_2$ is the expected value of y in the group $x_2 = 1$ when $x_1 = 0$.

Example interaction model with FEV data

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

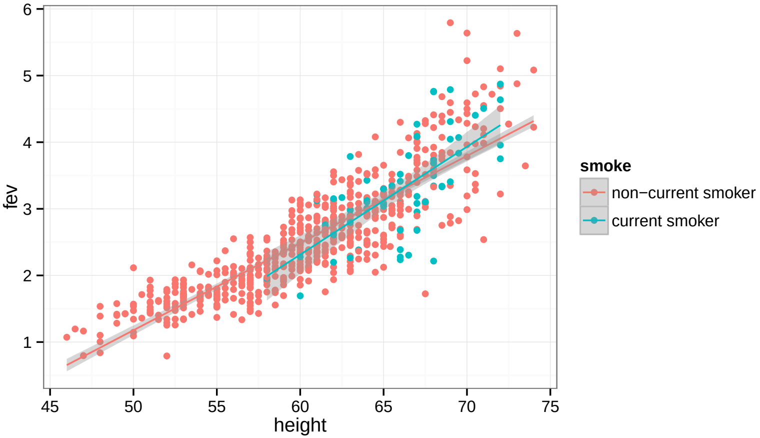
```
mi1 <- lm(fev ~ age + height + smoke + sex, data=FEV)
mi2 <- lm(fev ~ age + height*smoke + sex, data=FEV)
c(AIC(mi1), AIC(mi2))
round(summary(mi2)$coef,2)

## [1] 703.8 700.5
##                             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
## height                          0.10       0.00   21.08     0.00
## smokecurrent smoker            -2.61       1.10   -2.37     0.02
## sexmale                         0.15       0.03    4.43     0.00
## height:smokecurrent smoker      0.04       0.02    2.30     0.02
```

For current smokers, the relationship between height and FEV is stronger than in non-current smokers. In non-current smokers, we observe that a one-unit increase in height is associated with a 0.10 increase in expected FEV. In current smokers, this changes to a 0.14 increase in expected FEV.

Example interaction model with FEV data

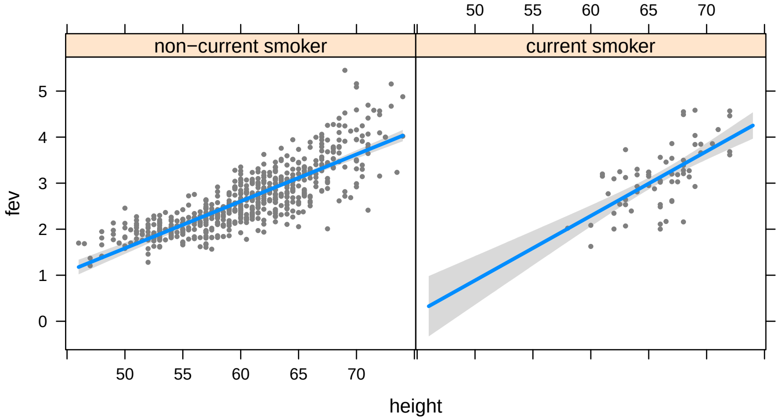
```
qplot(height, fev, data=FEV, color=smoke,
      geom=c("point", "smooth"), method="lm")
```



Example interaction model with FEV data

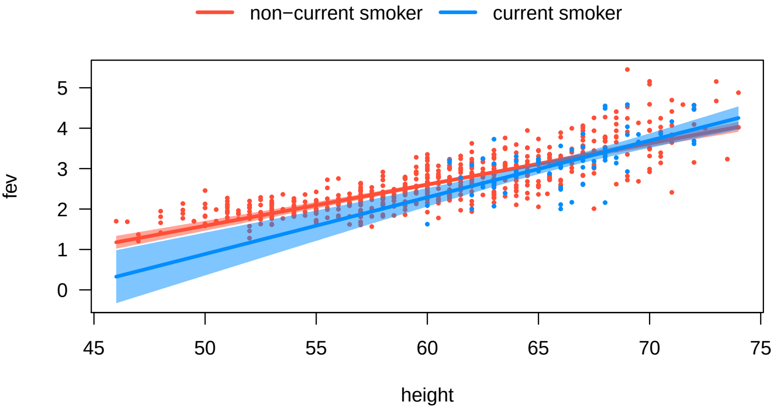
The `visreg` package plots not the data but the partial residuals (a.k.a. the adjusted variable) plot.

```
visreg(mi2, "height", by="smoke")
```



Example interaction model with FEV data

```
visreg(mi2, "height", by="smoke", overlay=TRUE)
```



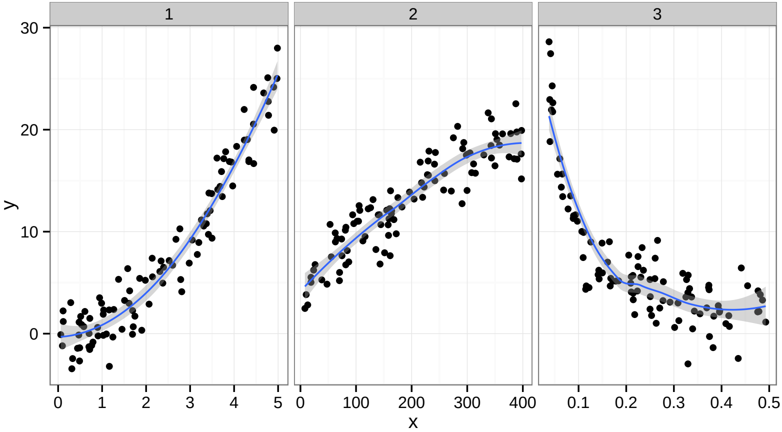
Overview of variable transformations

The problems

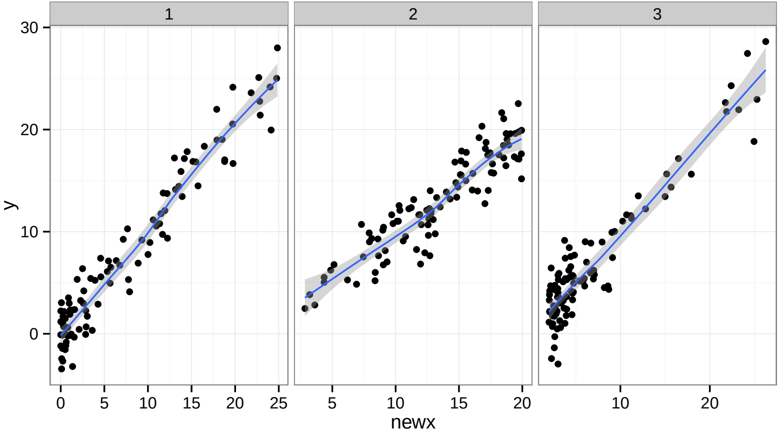
- Non-linearity between X and Y $\longrightarrow$ transform X
- Skewed distribution of Xs/points with high leverage $\longrightarrow$ transform X
- Non-constant variance $\longrightarrow$ transform Y

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 $y_i^{\lambda} = \log_e y_i$

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.