GSERM 2022
Regression for Publishing

June 14, 2022

Parameter Invariance

Implicit in

$$Y = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

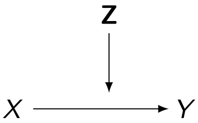
is that

$$\frac{\partial E(Y)}{\partial X_k} = \beta_k \; \forall \text{ values of } X_k, X_\ell, k \neq \ell.$$

Conceptually: *The marginal association between Y and every X is identical for all values of* $\mathbf{X}$.

Moderators

Moderating variable Z:



Intuition: *The marginal association between X and Y varies with / depends on the value(s) of Z.*

Moderating variables imply <u>interactive</u> models.

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{1i} X_{2i} + u_i$$

$$\begin{aligned} \mathrm{E}(Y_i) &= \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{1i} X_{2i} \\ &= \beta_0 + \beta_2 X_{2i} + (\beta_1 + \beta_3 X_{2i}) X_{1i} \\ &= \beta_0 + \beta_2 X_{2i} + \psi_1 X_{1i} \end{aligned}$$

where $\psi_1 = \beta_1 + \beta_3 X_{2i}$. This means that the *marginal effect*:

$$\frac{\partial \mathrm{E}(Y_i)}{\partial X_1} = \beta_1 + \beta_3 X_{2i}.$$

Similarly:

$$\begin{aligned} \mathrm{E}(Y_i) &= \beta_0 + \beta_1 X_{1i} + (\beta_2 + \beta_3 X_{1i}) X_{2i} \\ &= \beta_0 + \beta_1 X_{1i} + \psi_2 X_{2i} \end{aligned}$$

which implies:

$$\frac{\partial \mathrm{E}(Y_i)}{\partial X_2} = \beta_2 + \beta_3 X_{1i}.$$

"Direct Effects"

If $X_2 = 0$, then:

$$\begin{aligned} E(Y_i) &= \beta_0 + \beta_1 X_{1i} + \beta_2(0) + \beta_3 X_{1i}(0) \\ &= \beta_0 + \beta_1 X_{1i}. \end{aligned}$$

Similarly, for $X_1 = 0$:

$$\begin{aligned} E(Y_i) &= \beta_0 + \beta_1(0) + \beta_2 X_{2i} + \beta_3(0) X_{2i} \\ &= \beta_0 + \beta_2 X_{2i} \end{aligned}$$

In most instances, the quantities we care about are not β_1 and β_2, but rather ψ_1 and ψ_2.

Inference

Point estimates:

$$\hat{\psi}_1 = \hat{\beta}_1 + \hat{\beta}_3 X_2$$

and

$$\hat{\psi}_2 = \hat{\beta}_2 + \hat{\beta}_3 X_1.$$

For variance, recall that:

$$\mathrm{Var}(a + bZ) = \mathrm{Var}(a) + Z^2 \mathrm{Var}(b) + 2Z\mathrm{Cov}(a, b)$$

Inference

Means that:

$$\widehat{\mathsf{Var}(\hat{\psi}_1)} = \widehat{\mathsf{Var}(\hat{\beta}_1)} + X_2^2\widehat{\mathsf{Var}(\hat{\beta}_3)} + 2X_2\widehat{\mathsf{Cov}(\hat{\beta}_1, \hat{\beta}_3)}.$$

and

$$\widehat{\mathsf{Var}(\hat{\psi}_2)} = \widehat{\mathsf{Var}(\hat{\beta}_2)} + X_1^2\widehat{\mathsf{Var}(\hat{\beta}_3)} + 2X_1\widehat{\mathsf{Cov}(\hat{\beta}_2, \hat{\beta}_3)}.$$

Types of Interactions: Dichotomous Xs

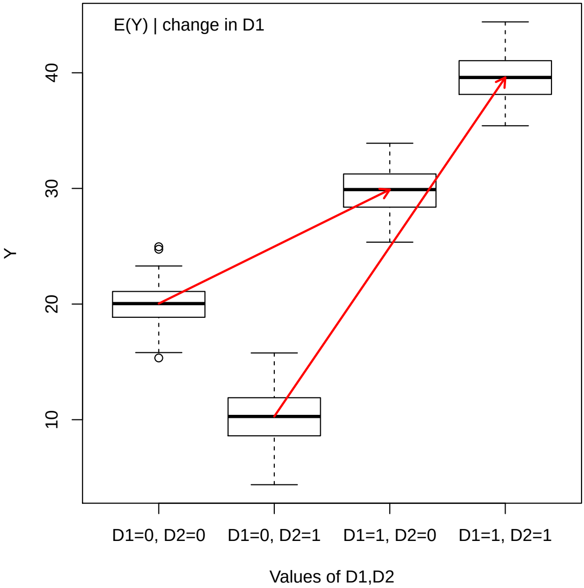
For

$$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 D_{1i} D_{2i} + u_i$$

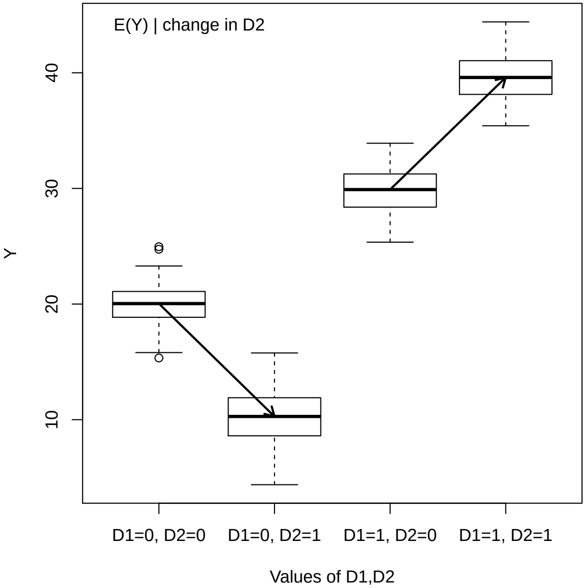
we have:

$$\begin{aligned}
\mathrm{E}(Y|D_1 = 0, D_2 = 0) &= \beta_0 \\
\mathrm{E}(Y|D_1 = 1, D_2 = 0) &= \beta_0 + \beta_1 \\
\mathrm{E}(Y|D_1 = 0, D_2 = 1) &= \beta_0 + \beta_2 \\
\mathrm{E}(Y|D_1 = 1, D_2 = 1) &= \beta_0 + \beta_1 + \beta_2 + \beta_3
\end{aligned}$$

Values of $E(Y)$ for Changes in D_1



Values of $E(Y)$ for Changes in D_2



$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 D_i + \beta_3 X_i D_i + u_i$$

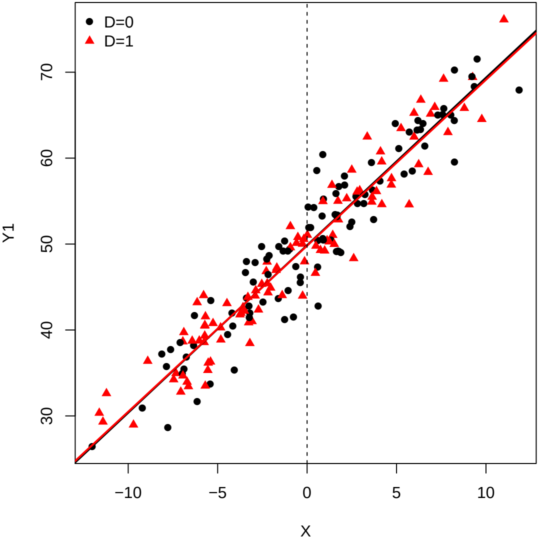
gives:

$$\begin{aligned} \mathrm{E}(Y|X, D=0) &= \beta_0 + \beta_1 X \\ \mathrm{E}(Y|X, D=1) &= (\beta_0 + \beta_2) + (\beta_1 + \beta_3) X \end{aligned}$$

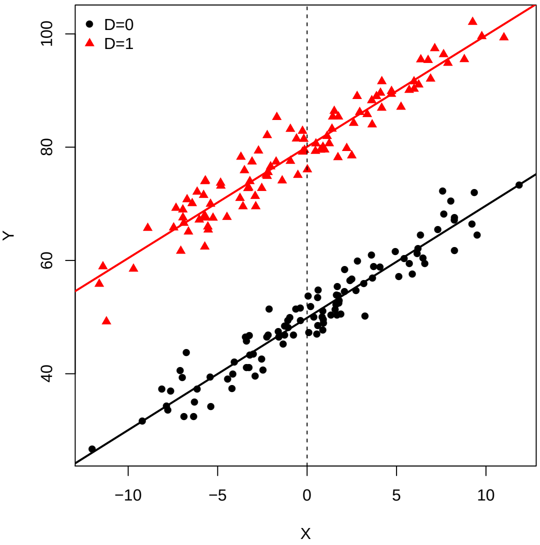
Four possibilities:

- $\beta_2 = \beta_3 = 0$
- $\beta_2 \neq 0$ and $\beta_3 = 0$
- $\beta_2 = 0$ and $\beta_3 \neq 0$
- $\beta_2 \neq 0$ and $\beta_3 \neq 0$

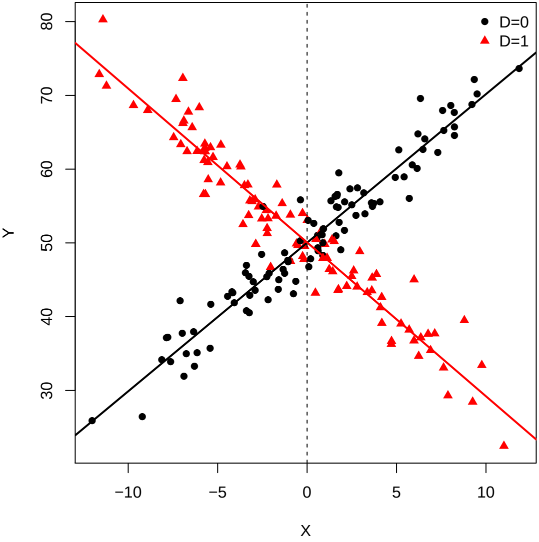
No Slope or Intercept Differences ($\beta_2 = \beta_3 = 0$)



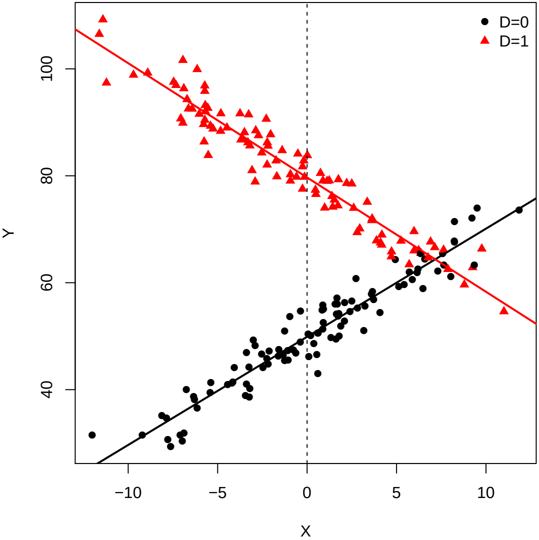
Intercept Shift ($\beta_2 \neq 0$, $\beta_3 = 0$)



Slope Change ($\beta_2 = 0,\ \beta_3 \neq 0$)



Slope and Intercept Change ($\beta_2 \neq 0$, $\beta_3 \neq 0$)



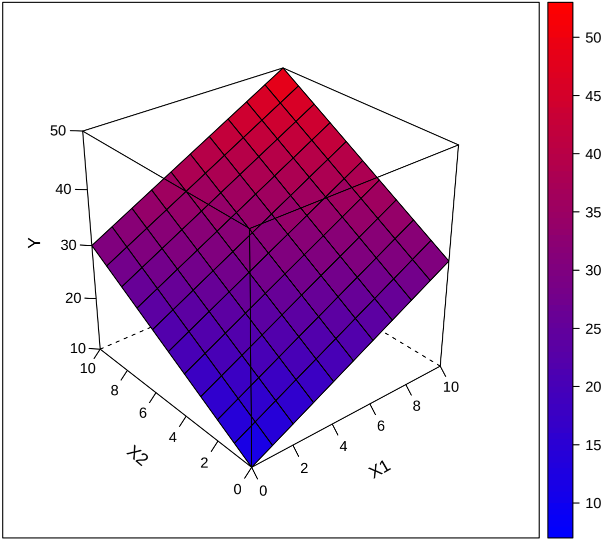
Two Continuous Xs

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{1i} X_{2i} + u_i.$$

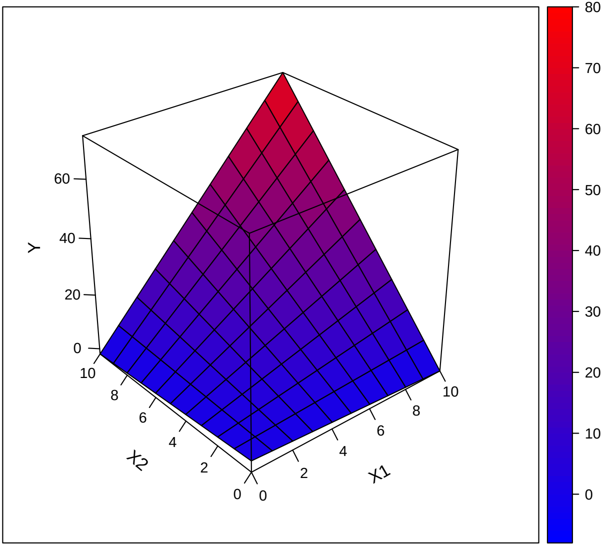
Implies

$$\beta_3 = 0 \; \rightarrow \; \frac{\partial E(Y)}{\partial X_1} = \beta_1 \, \forall \, X_2 \text{ and } \frac{\partial E(Y)}{\partial X_2} = \beta_2 \, \forall \, X_1$$

Two Continuous Variables: No Interactive Effects



Two Continuous Variables: Interaction Present



Quadratic, Cubic, and Other Polynomial Effects

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + ... + \beta_j X_i^j + u_i.$$

In general:

$$\frac{\partial \mathrm{E}(Y)}{\partial X} = \beta_1 + 2\beta_2 X + 3\beta_3 X^2 + ... + j\beta_j X^{j-1}$$

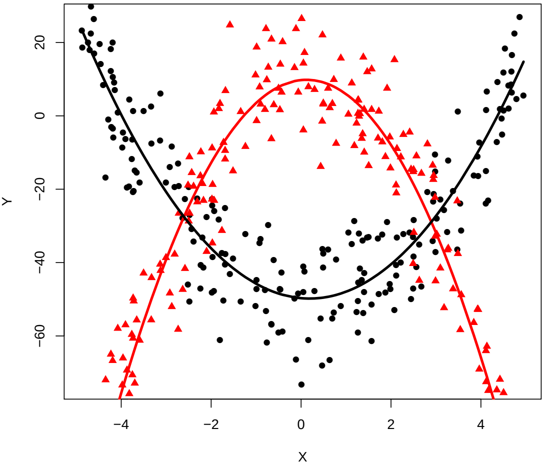
Quadratic case ($j = 2$):

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + u_i.$$

implies

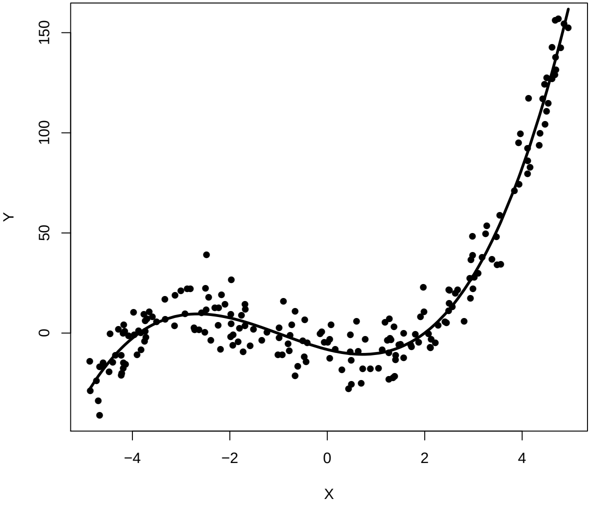
$$\frac{\partial \mathrm{E}(Y)}{\partial X} = \beta_1 + 2\beta_2 X$$

Two Quadratic Relationships



Note: Red line is $Y_i = 10 + 1X_i - 5X_i^2 + u_i$; black line is $Y_i = -50 - 1X_i + 3X_i^2 + u_i$.

Example of a Cubic Relationship



Note: Solid line is $Y_i = -1 + 1X_i - 8X_i^2 + 5X_i^3 + u_i$.

Higher-Order Interactive Models

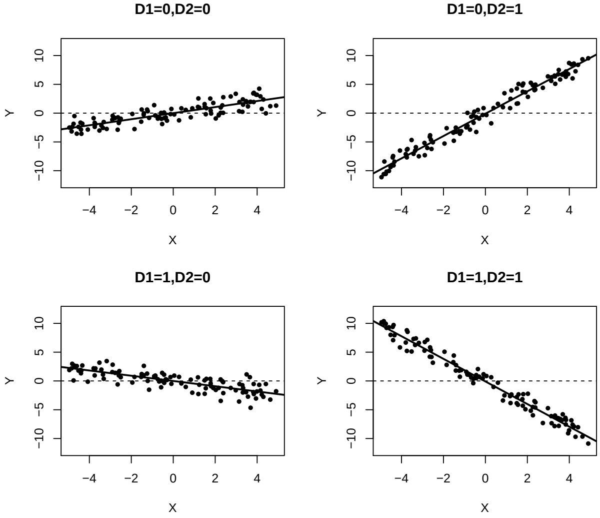
Three-way interaction:

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \\ &\quad \beta_4 X_{1i} X_{2i} + \beta_5 X_{1i} X_{3i} + \beta_6 X_{2i} X_{3i} + \beta_7 X_{1i} X_{2i} X_{3i} + u_i \end{aligned}$$

Special case of dichotomous X_1, X_2:

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_i + \beta_2 D_{1i} + \beta_3 D_{2i} + \\ &\quad \beta_4 X_i D_{1i} + \beta_5 X_i D_{2i} + \beta_6 D_{1i} D_{2i} + \beta_7 X_i D_{1i} D_{2i} + u_i \end{aligned}$$

Three-Way Interaction: Two Dummy and One Continuous Covariates

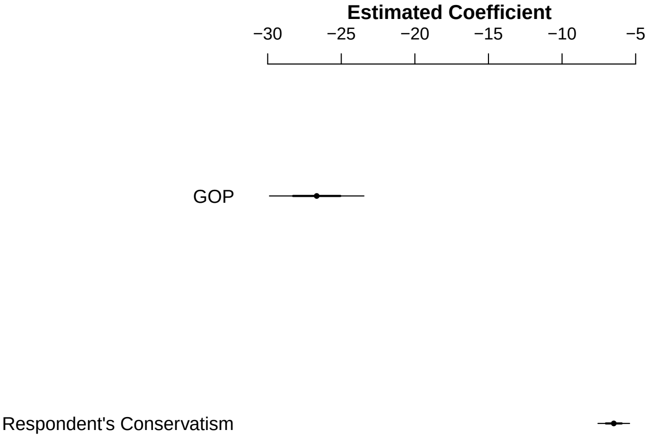


Example: President Clinton's "Thermometer Score"

```
> summary(ClintonTherm)
     caseid      ClintonTherm     RConserv     ClintonConserv
 Min.   :1001   Min.   :  0   Min.   :1.00   Min.   :1.00  
 1st Qu.:1440   1st Qu.: 30   1st Qu.:3.00   1st Qu.:2.00  
 Median :1854   Median : 60   Median :4.00   Median :3.00  
 Mean   :2001   Mean   : 57   Mean   :4.32   Mean   :2.98  
 3rd Qu.:2262   3rd Qu.: 85   3rd Qu.:5.00   3rd Qu.:4.00  
 Max.   :3403   Max.   :100   Max.   :7.00   Max.   :7.00  
      PID            GOP       
 Min.   :1.00   Min.   :0.000  
 1st Qu.:1.00   1st Qu.:0.000  
 Median :2.00   Median :0.000  
 Mean   :2.06   Mean   :0.316  
 3rd Qu.:3.00   3rd Qu.:1.000  
 Max.   :5.00   Max.   :1.000  
```

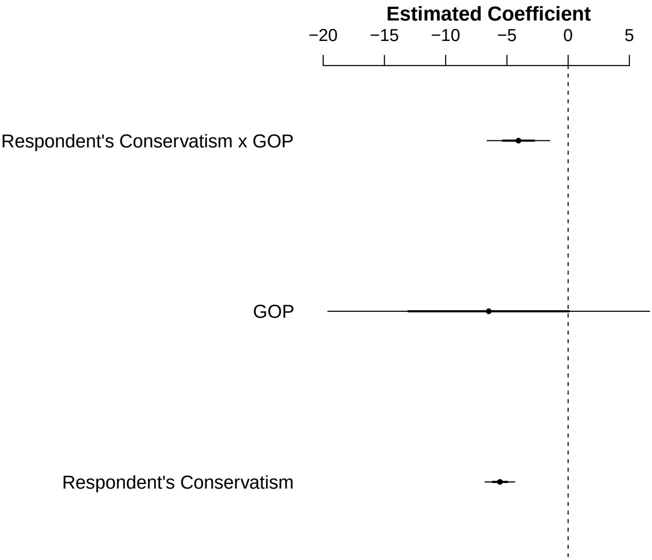
```
> summary(with(ClintonTherm, lm(ClintonTherm~RConserv+GOP)))

Call:
lm(formula = ClintonTherm ~ RConserv + GOP)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  93.4756     2.2278   41.96   <2e-16 ***
RConserv     -6.4866     0.5373  -12.07   <2e-16 ***
GOP         -26.6699     1.6056  -16.61   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 23.65 on 1294 degrees of freedom
Multiple R-squared: 0.3795,Adjusted R-squared: 0.3786
F-statistic: 395.7 on 2 and 1294 DF,  p-value: < 2.2e-16
```

Coefficient Plot: Non-Interactive Model



```
> fit1<-with(ClintonTherm, lm(ClintonTherm~RConserv+GOP+
             RConserv*GOP))
> summary(fit1)

Call:
lm(formula = ClintonTherm ~ RConserv + GOP + RConserv * GOP)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   89.9271     2.4866  36.165  < 2e-16 ***
RConserv      -5.5705     0.6085  -9.154  < 2e-16 ***
GOP           -6.4840     6.5690  -0.987  0.32379
RConserv:GOP  -4.0581     1.2808  -3.168  0.00157 **
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 23.57 on 1293 degrees of freedom
Multiple R-squared: 0.3843,Adjusted R-squared: 0.3829
F-statistic:   269 on 3 and 1293 DF,  p-value: < 2.2e-16
```

Coefficient Plot: Interactive Model

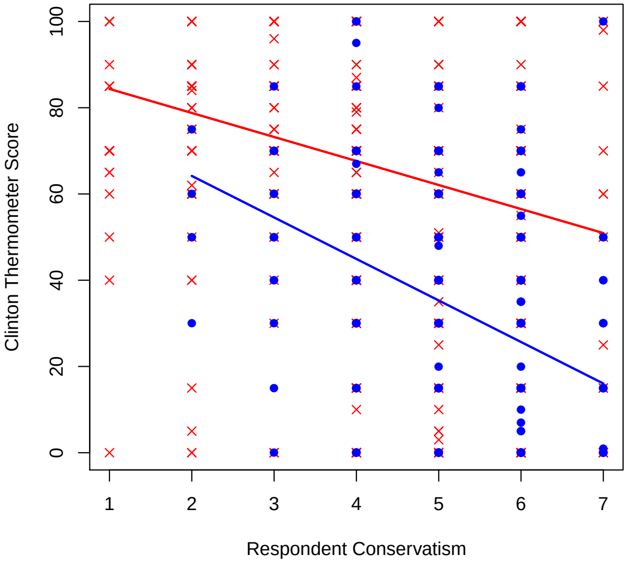


Two Regressions, Sort Of

$$\begin{aligned}
\text{E(Thermometer} \mid \text{Non-GOP)}_i &= 89.9 - 6.5(0) - 5.6(\text{R's Conservatism}_i) \\
&\quad -4.0(0 \times \text{R's Conservatism}_i) \\
&= \mathbf{89.9 - 5.6}(\text{R's Conservatism}_\mathbf{i})
\end{aligned}$$

$$\begin{aligned}
\text{E(Thermometer} \mid \text{GOP)}_i &= [89.9 - 6.5(1)] + [-5.6 - 4.0(1 \times \text{R's Conservatism}_i)] \\
&= \mathbf{83.4 - 9.6}(\text{R's Conservatism}_\mathbf{i})
\end{aligned}$$

Thermometer Scores by Conservatism, GOP and Non-GOP



Interactive Results are (Almost) Identical to Separate Regressions

```
> NonReps<-subset(ClintonTherm,GOP==0)
> summary(with(NonReps, lm(ClintonTherm~RConserv)))

Call:
lm(formula = ClintonTherm ~ RConserv, data = NonReps)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  89.9271     2.4695  36.416   <2e-16 ***
RConserv     -5.5705     0.6043  -9.217   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 23.41 on 885 degrees of freedom
Multiple R-squared: 0.08759,Adjusted R-squared: 0.08656
F-statistic: 84.96 on 1 and 885 DF,  p-value: < 2.2e-16
```

Interactive Results are (Almost) Identical to Separate Regressions

```
> Reps<-subset(ClintonTherm,GOP==1)
> summary(with(Reps, lm(ClintonTherm~RConserv)))

Call:
lm(formula = ClintonTherm ~ RConserv, data = Reps)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   83.443      6.170  13.524  < 2e-16 ***
RConserv      -9.629      1.144  -8.419 6.52e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 23.92 on 408 degrees of freedom
Multiple R-squared: 0.148,Adjusted R-squared: 0.1459
F-statistic: 70.88 on 1 and 408 DF,  p-value: 6.518e-16
```

Discovering $\hat{\psi}_1$ and $\hat{\psi}_2$

For `RConserv`:

$$\begin{aligned}
\text{Clinton Thermometer}_i &= \beta_0 + (\beta_1 + \beta_3 \text{GOP}_i)\text{R's Conservatism}_i + \\
&\quad \beta_2 \text{GOP}_i + u_i \\
&= \beta_0 + \psi_{1i}\text{R's Conservatism}_i + \beta_2 \text{GOP}_i + u_i.
\end{aligned}$$

So:

$$\hat{\psi}_{1i} = \hat{\beta}_1 + \hat{\beta}_3 \times \text{GOP}_i$$

and

$$\hat{\sigma}_{\psi_1} = \sqrt{\widehat{\text{Var}(\hat{\beta}_1)} + (\text{GOP})^2 \widehat{\text{Var}(\hat{\beta}_3)} + 2(\text{GOP})\widehat{\text{Cov}(\hat{\beta}_1, \hat{\beta}_3)}}.$$

Discovering $\hat{\psi}_1$ and $\hat{\psi}_2$

For `GOP`:

$$\begin{aligned}
\text{Clinton Thermometer}_i &= \beta_0 + (\beta_2 + \beta_3 \times \text{R's Conservatism}_i)\text{GOP}_i + \\
&\quad \beta_1(\text{R's Conservatism}_i) + u_i \\
&= \beta_0 + \psi_{2i}\text{GOP}_i + \beta_1(\text{R's Conservatism}_i) + u_i.
\end{aligned}$$

So:

$$\hat{\psi}_{2i} = \hat{\beta}_2 + \hat{\beta}_3 \times (\text{R's Conservatism}_i).$$

and

$$\hat{\sigma}_{\psi_2} = \sqrt{\widehat{\text{Var}(\hat{\beta}_2)} + (\text{R's Conservatism}_i)^2\widehat{\text{Var}(\hat{\beta}_3)} + 2k\widehat{\text{Cov}(\hat{\beta}_2, \hat{\beta}_3)}}.$$

Discovering $\hat{\psi}_1$ and $\hat{\psi}_2$

```
> Psi1<-fit1$coeff[2]+fit1$coeff[4]

> Psi1
 RConserv
-9.628577

> SPsi1<-sqrt(vcov(fit1)[2,2] + (1)^2*vcov(fit1)[4,4] + 2*1*vcov(fit1)[2,4])
> SPsi1
[1] 1.127016

> Psi1 / SPsi1 # <-- t-statistic
 RConserv
-8.543422
```

Discovering $\hat{\psi}_1$ and $\hat{\psi}_2$

```
> # psi_2 | RConserv = 1
> fit1$coeff[3]+(1 * fit1$coeff[4])
      GOP
-10.54208

> sqrt(vcov(fit1)[3,3] + (1)^2*vcov(fit1)[4,4] + 2*1*vcov(fit1)[3,4])
[1] 5.335847

# Implies t is approximately 2
```

```
> # psi_2 | RConserv = 7
> fit1$coeff[3]+(7 * fit1$coeff[4])
      GOP
-34.89045

> sqrt(vcov(fit1)[3,3] + (7)^2*vcov(fit1)[4,4] + 2*7*vcov(fit1)[3,4])
[1] 3.048302

# t is approximately 11
```

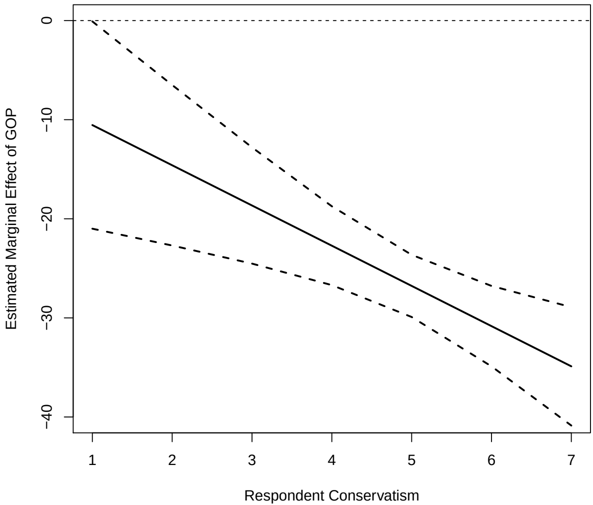
An Easier Way: `linearHypothesis()`

```
> library(car)
> linearHypothesis(fit1,"RConserv+RConserv:GOP")
Linear hypothesis test

Hypothesis:
RConserv  + RConserv:GOP = 0

Model 1: restricted model
Model 2: ClintonTherm ~ RConserv + GOP + RConserv * GOP

  Res.Df    RSS Df Sum of Sq     F    Pr(>F)
1   1294 758714
2   1293 718173  1     40541 72.99 < 2.2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

> # Note: Same as t-test:
> sqrt(72.99)
[1] 8.543419
```

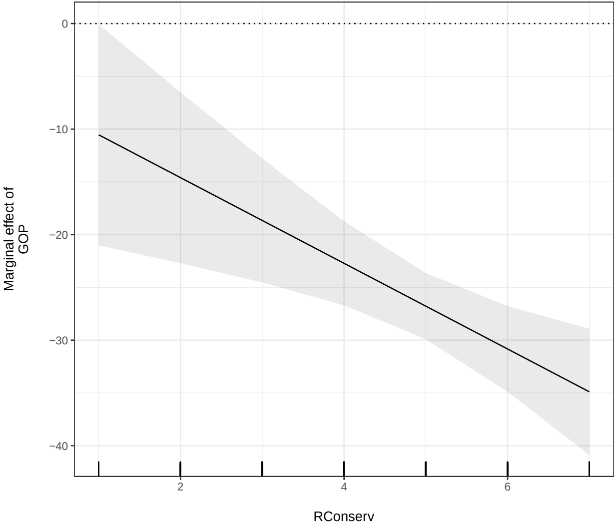
An Easier Way: `linearHypothesis()`

```
> # psi_2 | RConserv = 7:
> linearHypothesis(fit1,"GOP+7*RConserv:GOP")
Linear hypothesis test

Hypothesis:
GOP  + 7 RConserv:GOP = 0

Model 1: restricted model
Model 2: ClintonTherm ~ RConserv + GOP + RConserv * GOP

  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1   1294 790938
2   1293 718173  1     72766 131.01 < 2.2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

```
> ConsSim<-seq(1,7,1)
> psis<-fit1$coeff[3]+(ConsSim * fit1$coeff[4])
> psis.ses<-sqrt(vcov(fit1)[3,3] +
  (ConsSim)^2*vcov(fit1)[4,4] + 2*ConsSim*vcov(fit1)[3,4])

> plot(ConsSim,psis,t="l",lwd=2,xlab="Respondent Conservatism",
  ylab="Estimated Marginal Effect",ylim=c(-40,0))
> lines(ConsSim,psis+(1.96*psis.ses),lty=2,lwd=2)
> lines(ConsSim,psis-(1.96*psis.ses),lty=2,lwd=2)
> abline(h=0,lwd=1,lty=2)
```


Estimated Marginal Effect of GOP
0
−10
−20
−30
−40
1
2
3
4
5
6
7
Respondent Conservatism

Same, Using `plot_me`



Interacting Two Continuous Covariates

```
> fit2<-with(ClintonTherm,
+       lm(ClintonTherm~RConserv+ClintonConserv+RConserv*ClintonConserv))
> summary(fit2)

Call:
lm(formula = ClintonTherm ~ RConserv + ClintonConserv + RConserv *
    ClintonConserv)

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)             119.3515     5.1634  23.115  < 2e-16 ***
RConserv                -19.5673     1.0362 -18.884  < 2e-16 ***
ClintonConserv           -7.9311     1.6477  -4.813 1.66e-06 ***
RConserv:ClintonConserv   3.6293     0.3394  10.695  < 2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 22.03 on 1293 degrees of freedom
Multiple R-squared: 0.4619,Adjusted R-squared: 0.4606
F-statistic:   370 on 3 and 1293 DF,  p-value: < 2.2e-16
```

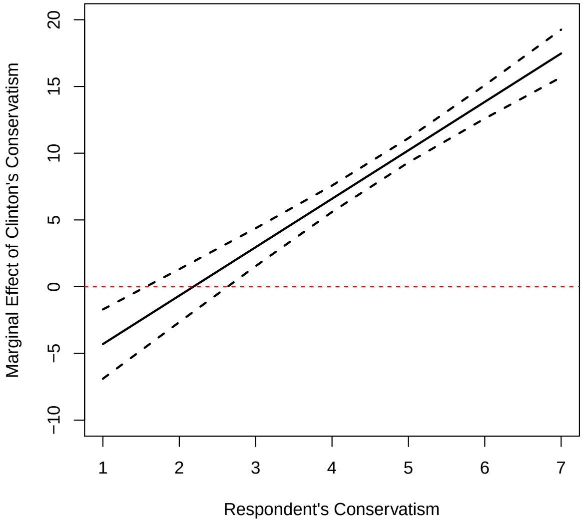
```
> fit2$coef[2]+(1*fit2$coef[4])
 RConserv
-15.93803

> sqrt(vcov(fit2)[2,2] + (1)^2*vcov(fit2)[4,4] + 2*1*vcov(fit2)[2,4])
[1] 0.7439696
```

```
> linearHypothesis(fit2,"RConserv+1*RConserv:ClintonConserv")
Linear hypothesis test

Hypothesis:
RConserv  + RConserv:ClintonConserv = 0

Model 1: restricted model
Model 2: ClintonTherm ~ RConserv + ClintonConserv + RConserv * ClintonConserv

  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1   1294 850442
2   1293 627658  1    222784 458.94 < 2.2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

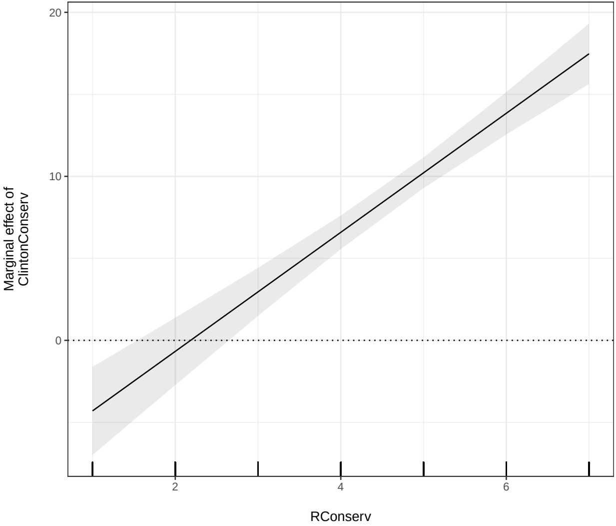
```
> # psi_1 | ClintonConserv = mean
> fit2$coef[2]+((mean(ClintonTherm$ClintonConserv))*fit2$coef[4])

 RConserv
-8.735424

> sqrt(vcov(fit2)[2,2] + (mean(ClintonTherm$ClintonConserv)^2*vcov(fit2)[4,4] +
+                            2*(mean(ClintonTherm$ClintonConserv))*vcov(fit2)[2,4]))

[1] 0.4507971

> pt(((fit2$coef[2]+(2.985*fit2$coef[4])) / sqrt(vcov(fit2)[2,2] +
+      (2.985)^2*vcov(fit2)[4,4] + 2*2.985*vcov(fit2)[2,4])),df=1293)

    RConserv
6.483788e-74

> # psi_2 | RConserv = 1
> fit2$coef[3]+(1*fit2$coef[4])

ClintonConserv
     -4.301803

> # psi_2 | RConserv = 6
> fit2$coef[3]+(6*fit2$coef[4])

ClintonConserv
      13.84463
```

Marginal Effect Plot, II

```
> psis2<-fit2$coef[3]+(ConsSim*fit2$coef[4])
> psis2.ses<-sqrt(vcov(fit2)[3,3] + (ConsSim)^2*vcov(fit2)[4,4]
  + 2*ConsSim*vcov(fit2)[3,4])

> plot(ConsSim,psis2,t="l",lwd=2,xlab="Respondent's
  Conservatism",ylab="Marginal Effect of Clinton's
  Conservatism",ylim=c(-10,20))
> lines(ConsSim,psis2+(1.96*psis2.ses),lty=2,lwd=2)
> lines(ConsSim,psis2-(1.96*psis2.ses),lty=2,lwd=2)
> abline(h=0,lty=2,lwd=1,col="red")
```

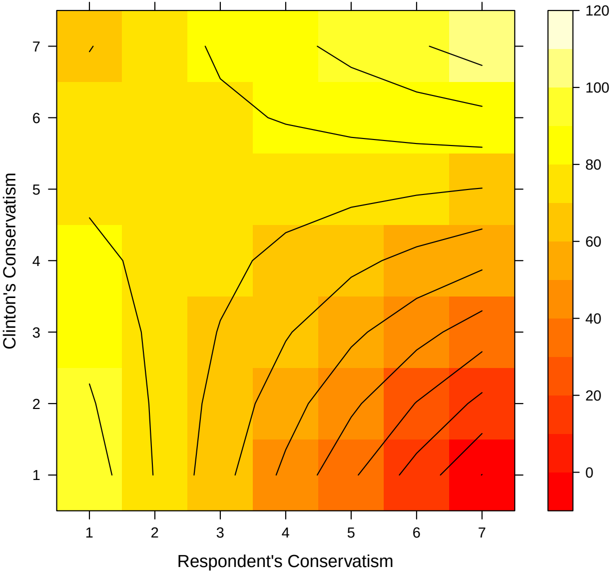

Marginal Effect of Clinton's Conservatism
−10
−5
0
5
10
15
20
1
2
3
4
5
6
7
Respondent's Conservatism

Same, Using `plot_me`



Predicted Values: A Contour Plot

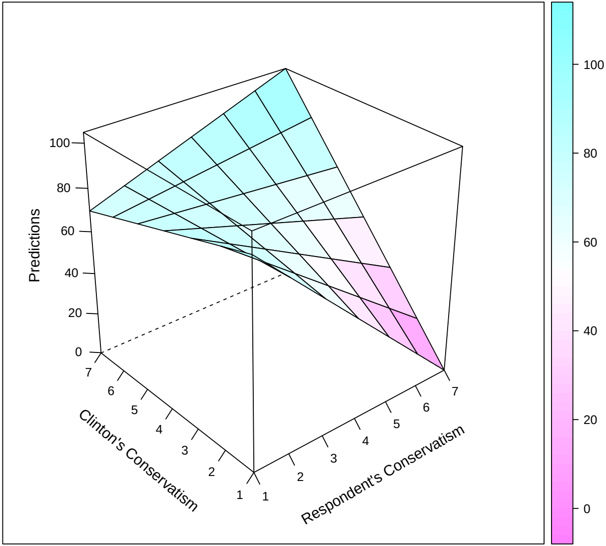
```
> library(lattice)
> grid<-expand.grid(RConserv=seq(1,7,1),
  ClintonConserv=seq(1,7,1))
> hats<-predict(fit2,newdata=grid)

> levelplot(hats~grid$RConserv*grid$ClintonConserv,
  contour=TRUE,
  cuts=12,pretty=TRUE,xlab="Respondent's Conservatism",
  ylab="Clinton's Conservatism",
  col.regions=heat.colors)
```



Clinton's Conservatism
Respondent's Conservatism
1
2
3
4
5
6
7
120
100
80
60
40
20
0

Predicted Values: A Wireframe Plot

```
> trellis.par.set("axis.line",list(col="transparent"))

> wireframe(hats~grid$RConserv*grid$ClintonConserv,
  drape=TRUE,
  xlab=list("Respondent's Conservatism",rot=30),
  ylab=list("Clinton's Conservatism",
  rot=-40),zlab=list("Predictions",rot=90),
  scales=list(arrows=FALSE,col="black"),
  zoom=0.85,pretty=TRUE),
  col.regions=colorRampPalette(c("blue","red"))(100))
```



Predictions
100
80
60
40
20
0
Clinton's Conservatism
7
6
5
4
3
2
1
Respondent's Conservatism
1
2
3
4
5
6
7
100
80
60
40
20
0

Variable Transformations

Why Transform?

- Induce / conform to linearity
- Induce / conform to additivity
- Induce normality in the u_is
- Facilitate interpretation
- **Make the model fit the theory**

This:

$$Y_i = \beta_0 X_i^{\beta_1} u_i$$

becomes this:

$$\ln(Y_i) = \ln(\beta_0) + \beta_1 X_i + \ln(u_i)$$

And this:

$$\exp(Y_i) = \beta_0 + \beta_1 X_i + u_i$$

becomes this:

$$Y_i = \ln(\beta_0) + \beta_1 \ln(X_i) + \ln(u_i)$$

Monotonic Transformations

The "Ladder of Powers":

Transformation	p	$f(X)$	Fox's $f(X)$
Cube	3	X^3	$\frac{X^3-1}{3}$
Square	2	X^2	$\frac{X^2-1}{2}$
(None/Identity)	(1)	(X)	(X)
Square Root	$\frac{1}{2}$	$\sqrt{X}$	$2(\sqrt{X}-1)$
Cube Root	$\frac{1}{3}$	$\sqrt[3]{X}$	$3(\sqrt[3]{X}-1)$
Log	0 (sort of)	$\ln(X)$	$\ln(X)$
Inverse Cube Root	-$\frac{1}{3}$	$\frac{1}{\sqrt[3]{X}}$	$\frac{\left(\frac{1}{\sqrt[3]{X}}-1\right)}{-\frac{1}{3}}$
Inverse Square Root	-$\frac{1}{2}$	$\frac{1}{\sqrt{X}}$	$\frac{\left(\frac{1}{\sqrt{X}}-1\right)}{-\frac{1}{2}}$
Inverse	-1	$\frac{1}{X}$	$\frac{\left(\frac{1}{X}-1\right)}{-1}$
Inverse Square	-2	$\frac{1}{X^2}$	$\frac{\left(\frac{1}{X^2}-1\right)}{-2}$
Inverse Cube	-3	$\frac{1}{X^3}$	$\frac{\left(\frac{1}{X^3}-1\right)}{-3}$

Using higher-order power transformations (e.g., squares, cubes, etc.) "inflates" large values and "compresses" small ones; conversely, using lower-order power transformations (logs, etc.) "compresses" large values and "inflates" (or "expands") smaller ones.

Power Transformations: Two Issues

1. X must be *positive*; so:

$$X^* = X + (|X_\ell| + \epsilon)$$

with (CZ's Rule of Thumb):

$$\epsilon = \frac{X_{\ell+1} - X_\ell}{2}$$

2. Power transformations generally require that:

$$\frac{X_h}{X_l} > 5 \text{ (or so)}$$

A Note On Logarithms

Note that:

$$\ln(X|X \leq 0) \text{ is undefined.}$$

For $X = 0$, we might:

1. exclude observations,
2. add some arbitrary amount (perhaps 1.0) to *all observations*
3. add some arbitrary amount (perhaps 1.0) to *observations where* $X = 0$
4. add some arbitrary amount (perhaps 1.0) to *observations where* $X = 0$, and include a variable D_i in your regression, where:

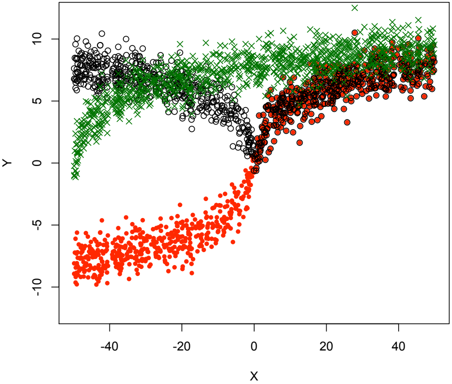
$$D_i = \begin{cases} 1 \text{ if } X_i = 0 \\ 0 \text{ otherwise} \end{cases}$$

The short answer: **Do #4.** Find out more at this poster.

A Note On Logarithms (continued)

For $X < 0$, we should think about how we expect X and Y to covary when $X < 0$:

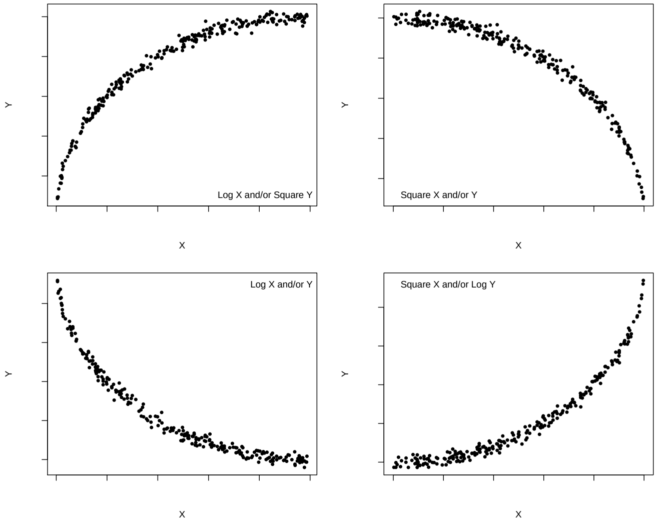
1. a "shift", where the logarithmic form starts at values of X less than zero,
2. a "V-curve," where $E(Y|X = k) = E(Y|X = -k)$, or
3. an "S-curve," where the $X - Y$ relationship for $X < 0$ "mirrors" that for $X > 0$ [so $E(Y|X = k) = -E(Y|X = -k)$]



Which is correct? **It depends on your theory.** Again: find out more at this poster.

Which Transformation?

Mosteller and Tukey's "Bulging Rule":



Transformed Xs: Interpretation

For:

$$\ln(Y_i) = \beta_0 + \beta_1 X_i + u_i,$$

then:

$$\mathrm{E}(Y) = \exp(\beta_0 + \beta_1 X_i)$$

and so:

$$\frac{\partial \mathrm{E}(Y)}{\partial X} = \exp(\beta_1).$$

Transformed Xs: Interpretation

Similarly, for:

$$Y_i = \beta_0 + \beta_1 \ln(X_i) + u_i$$

we have:

$$\frac{\partial \mathrm{E}(Y)}{\partial \ln(X)} = \beta_1.$$

So doubling X (say, from X_ℓ to $2X_\ell$):

$$\begin{aligned} \Delta \mathrm{E}(Y) &= \mathrm{E}(Y|X = 2X_\ell) - \mathrm{E}(Y|X = X_\ell) \\ &= [\beta_0 + \beta_1 \ln(2X_\ell)] - [\beta_0 + \beta_1 \ln(X_\ell)] \\ &= \beta_1[\ln(2X_\ell) - \ln(X_\ell)] \\ &= \beta_1 \ln(2) \end{aligned}$$

Specifying:

$$\ln(Y_i) = \beta_0 + \beta_1 \ln(X_i) + ... + u_i$$

means:

$$\text{Elasticity}_{YX} \equiv \frac{\%\Delta Y}{\%\Delta X} = \beta_1.$$

IOW, a one-percent change in X leads to a $\hat{\beta}_1$-percent change in Y.

Simple solution: Polynomials...

- Second-order / quadratic:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + u_i$$

- Third-order / cubic:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + u_i$$

- *p*th-order:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + ... + \beta_p X_i^p + u_i$$

Read coefficients "left to right." So, for:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + u_i$$

then:

	$\hat{\beta}_2$		
$\hat{\beta}_1$	< 0	$= 0$	> 0
< 0	E(Y) decreases in X at an increasing rate	E(Y) decreases linearly in X	E(Y) decreases in X at low values of X, but increases in X at high values of X
$= 0$	E(Y) decreases in X^2	E(Y) is (quadratically) unrelated to X	E(Y) increases in X^2
> 0	E(Y) increases in X at low values of X, but decreases in X at high values of X	E(Y) increases linearly in X	E(Y) increases in X at an increasing rate

An Example: Military Spending and GDP

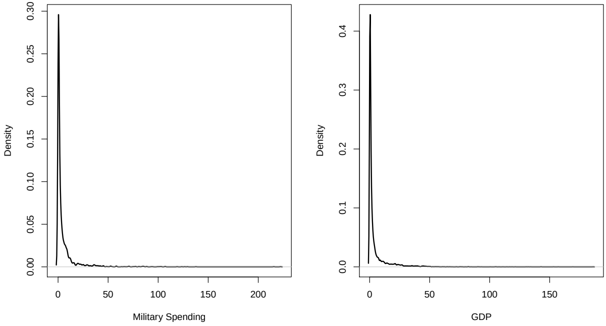
Q: Are Militaries Luxury Goods?

Data are from the CNTS Data archive...

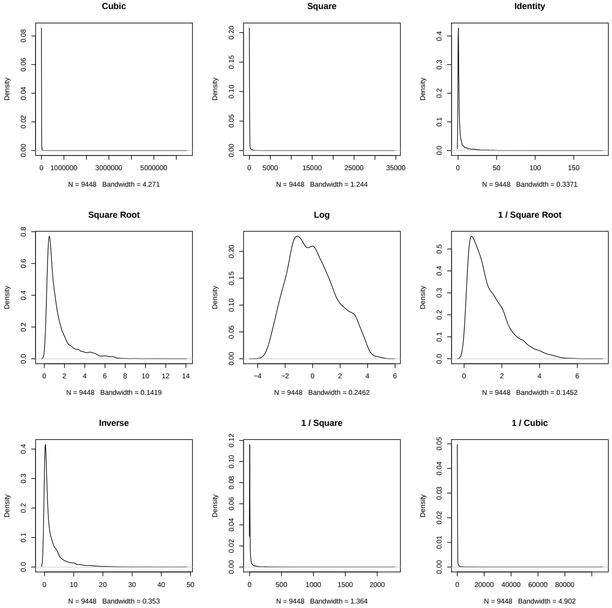
- Panel-type data on $\approx$ 200 countries over $\approx$ 70 years (1946-2014)
- Two variables:
  - Y = National Defense Expenditures Per Capita (in thousands of constant $U.S.)
  - X = GDP Per Capita (at factor cost, in thousands of constant $U.S.)

```
> summary(Data)
   BanksCode         WBCode           Year          BanksCountry
 Min.   :  10             : 722   Min.   :1946   Afghanistan:   69
 1st Qu.: 310    AFG     :  69   1st Qu.:1972   Albania    :   69
 Median : 680    ALB     :  69   Median :1987   Argentina  :   69
 Mean   : 661    ARG     :  69   Mean   :1986   Australia  :   69
 3rd Qu.:1000    AUS     :  69   3rd Qu.:2001   Belgium    :   69
 Max.   :1300    BEL     :  69   Max.   :2014   Bolivia    :   69
                 (Other):9589                   (Other)    :10242
    AreaKM2            Population            GDP           MilitarySpending
 Min.   :       0   Min.   :      1   Min.   :  0.021   Min.   :  0.00
 1st Qu.:   30000   1st Qu.:   1596   1st Qu.:  0.280   1st Qu.:  0.38
 Median :  164000   Median :   6130   Median :  0.884   Median :  1.17
 Mean   :  818610   Mean   :  28738   Mean   :  4.563   Mean   :  5.38
 3rd Qu.:  600000   3rd Qu.:  18148   3rd Qu.:  3.411   3rd Qu.:  4.68
 Max.   :22402000   Max.   :1354503   Max.   :186.243   Max.   :222.22
                    NA's   :6         NA's   :1208      NA's   :7790
```

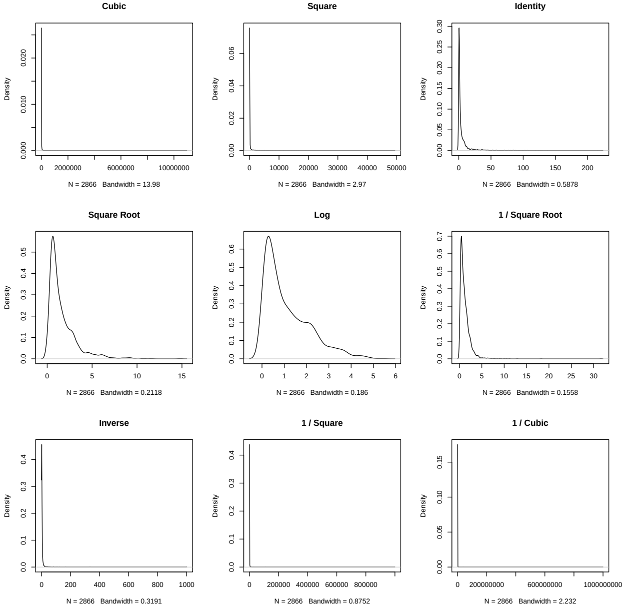
```
> with(Data, describe(MilitarySpending))
   vars    n mean    sd median trimmed  mad min    max  range skew kurtosis   se
X1    1 2866 5.38 13.44   1.17    2.48 1.48   0 222.22 222.22 6.79    68.39 0.25

> with(Data, describe(GDP))
   vars    n mean    sd median trimmed  mad  min    max  range skew kurtosis   se
X1    1 9448 4.56 10.65   0.88    2.03 1.09 0.02 186.24 186.22 5.85    54.74 0.11
```



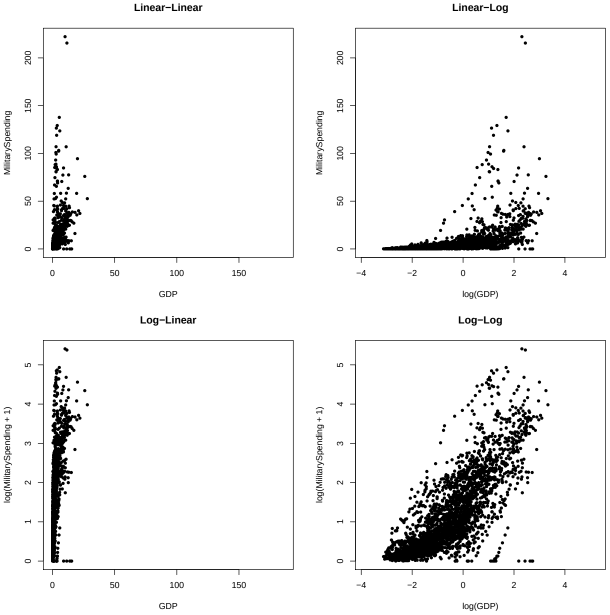
"Ladder of Powers": GDP



“Ladder of Powers”: Military Spending



Scatterplots



Linear-Linear (Untransformed)

Untransformed:

```
> linlin <- with(Data, lm(MilitarySpending~GDP))
> summary(linlin)

Call:
lm(formula = MilitarySpending ~ GDP)

Residuals:
   Min     1Q Median     3Q    Max
-46.08  -1.88  -1.41  -0.46 191.61

Coefficients:
            Estimate Std. Error t value   Pr(>|t|)
(Intercept)   1.2936     0.2430    5.32 0.00000011 ***
GDP           2.9123     0.0823   35.39    < 2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 11.3 on 2810 degrees of freedom
  (7844 observations deleted due to missingness)
Multiple R-squared:  0.308,Adjusted R-squared:  0.308
F-statistic: 1.25e+03 on 1 and 2810 DF,  p-value: <2e-16
```

Linear-Log (Transforming X)

Logging X:

```
> linlog <- with(Data, lm(MilitarySpending~log(GDP)))
> summary(linlog)

Call:
lm(formula = MilitarySpending ~ log(GDP))

Residuals:
   Min     1Q Median     3Q    Max
-22.55  -4.63  -1.25   2.23 201.86

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    8.458      0.241    35.1   <2e-16 ***
log(GDP)       5.155      0.166    31.0   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 11.7 on 2810 degrees of freedom
  (7844 observations deleted due to missingness)
Multiple R-squared:  0.255,Adjusted R-squared:  0.255
F-statistic:  963 on 1 and 2810 DF,  p-value: <2e-16
```

Log-Linear (Transforming Y)

Logging Y:

```
> loglin <- with(Data, lm(log(MilitarySpending+1)~GDP))
> summary(loglin)

Call:
lm(formula = log(MilitarySpending + 1) ~ GDP)

Residuals:
   Min     1Q Median     3Q    Max
-4.751 -0.538 -0.203  0.451  3.284

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.77309    0.01650    46.9   <2e-16 ***
GDP          0.25864    0.00559    46.3   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.766 on 2810 degrees of freedom
  (7844 observations deleted due to missingness)
Multiple R-squared:  0.433,Adjusted R-squared:  0.433
F-statistic: 2.14e+03 on 1 and 2810 DF,  p-value: <2e-16
```

Log-Log (Transforming X and Y)

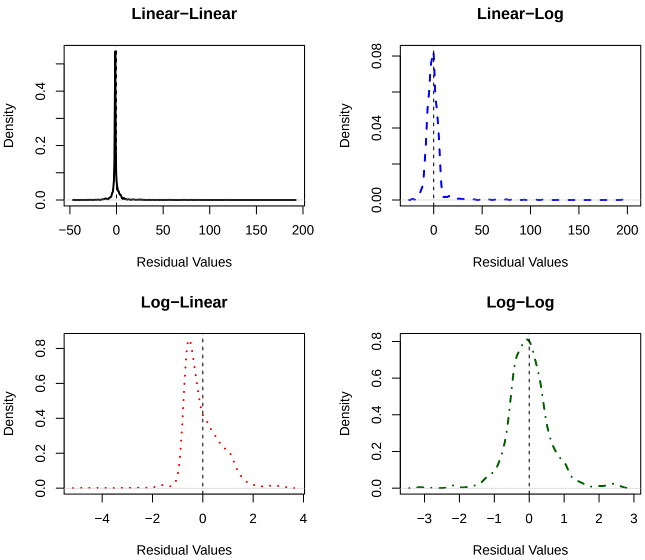
Logging X and Y:

```
> loglog <- with(Data, lm(log(MilitarySpending+1)~log(GDP)))
> summary(loglog)

Call:
lm(formula = log(MilitarySpending + 1) ~ log(GDP))

Residuals:
   Min     1Q Median     3Q    Max
-3.165 -0.359 -0.039  0.302  2.668

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.4980     0.0126   118.7   <2e-16 ***
log(GDP)      0.6101     0.0087    70.1   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.613 on 2810 degrees of freedom
  (7844 observations deleted due to missingness)
Multiple R-squared:  0.636,Adjusted R-squared:  0.636
F-statistic: 4.91e+03 on 1 and 2810 DF,  p-value: <2e-16
```

(Slightly) Prettier Table...

	Linear Y		Logged Y	
	Linear-Linear	Linear-Log	Log-Linear	Log-Log
(Constant)	1.29***	8.46***	0.77***	1.50***
	(0.24)	(0.24)	(0.02)	(0.01)
GDP	2.91***		0.26***	
	(0.08)		(0.01)	
log(GDP)		5.15***		0.61***
		(0.17)		(0.01)
Observations	2,812	2,812	2,812	2,812
R^2	0.31	0.26	0.43	0.64
Adjusted R^2	0.31	0.25	0.43	0.64
Residual Std. Error (df = 2810)	11.28	11.70	0.77	0.61
F Statistic (df = 1; 2810)	1,252.62***	963.03***	2,143.49***	4,912.36***
Note:			*p<0.1; **p<0.05; ***p<0.01	

Density Plots of $\hat{u}_i$s



- **Theory is valuable**.
- **Try different things**.
- **Look at plots**.
- **It takes practice**.