

GSERM 2024

Regression for Publishing

June 18, 2024

Link To The “Friday Poll”

<https://bit.ly/RegForPub-2024>

Implicit in

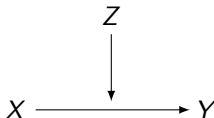
$$Y = \mathbf{X}\beta + \mathbf{u}$$

is that

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

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

Moderating variable Z:

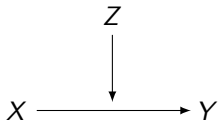


Intuition: The marginal association between X and Y varies with / depends on the value(s) of Z.

Moderating variables imply interactive models.

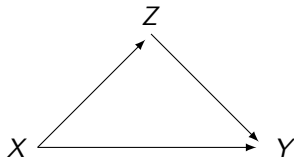
Don't Get Confused

Don't mistake a moderator for a mediator...



Z is a Moderator...

vs.



Z is a Mediator...

Multiplicative interaction:

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

So:

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

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

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

Similarly:

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

which implies:

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

Note that 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 .

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:

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

Means that:

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

and

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

Types of Interactions: Dichotomous X s

For

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

we have:

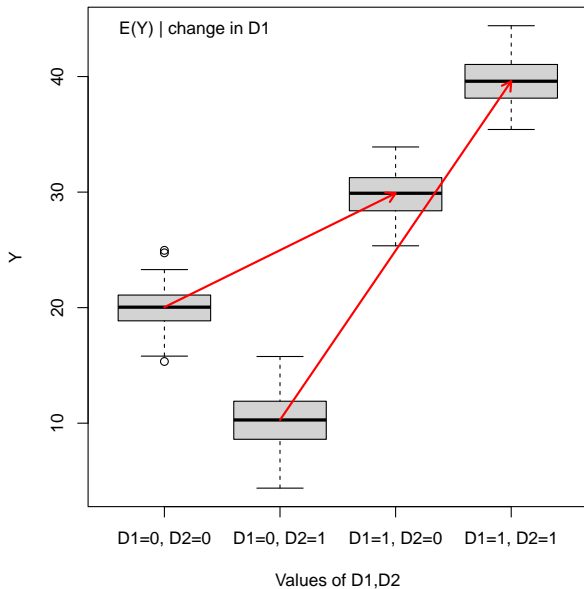
$$E(Y|D_1 = 0, D_2 = 0) = \beta_0$$

$$E(Y|D_1 = 1, D_2 = 0) = \beta_0 + \beta_1$$

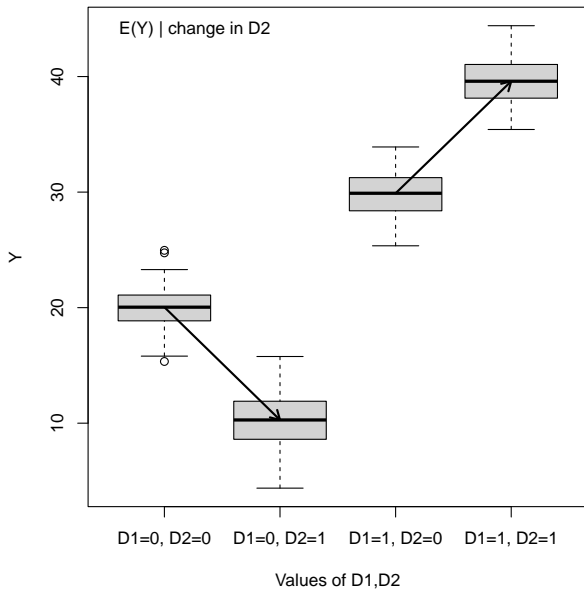
$$E(Y|D_1 = 0, D_2 = 1) = \beta_0 + \beta_2$$

$$E(Y|D_1 = 1, D_2 = 1) = \beta_0 + \beta_1 + \beta_2 + \beta_3$$

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



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



Dichotomous and Continuous X s

The model:

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

gives:

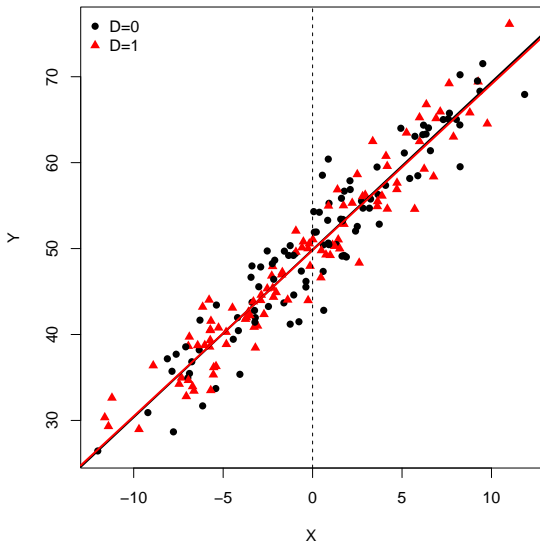
$$E(Y|X, D = 0) = \beta_0 + \beta_1 X$$

$$E(Y|X, D = 1) = (\beta_0 + \beta_2) + (\beta_1 + \beta_3)X$$

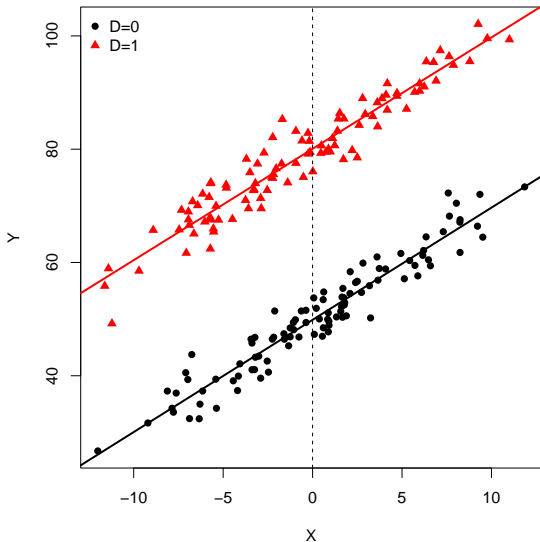
Four possibilities:

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

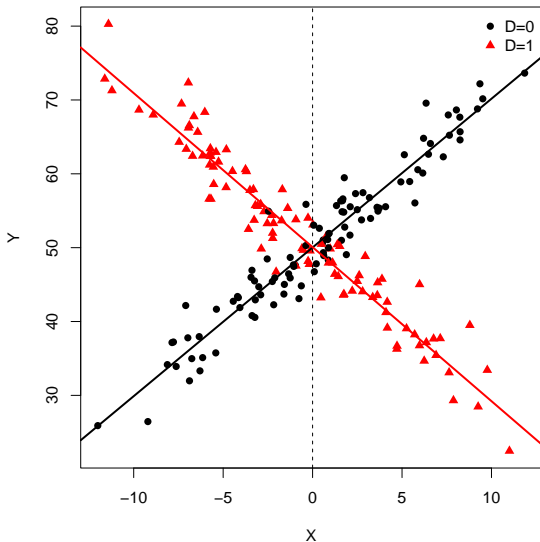
No Slope or Intercept Differences ($\beta_2 = \beta_3 = 0$)



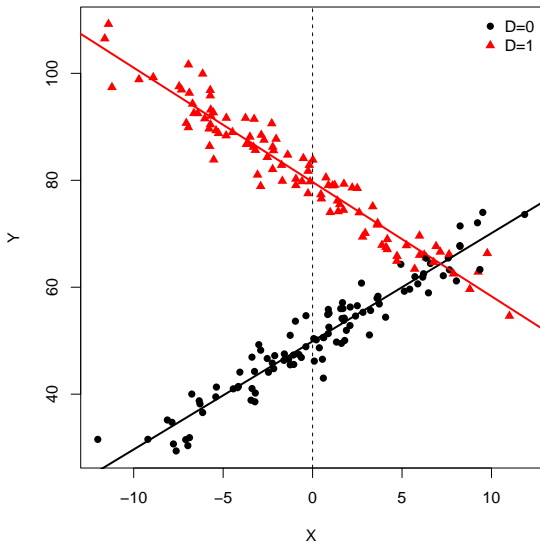
Intercept Shift ($\beta_2 \neq 0, \beta_3 = 0$)



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



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



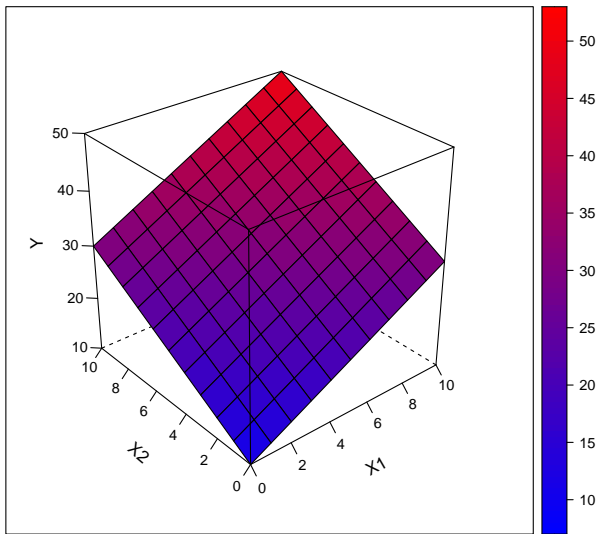
The model:

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

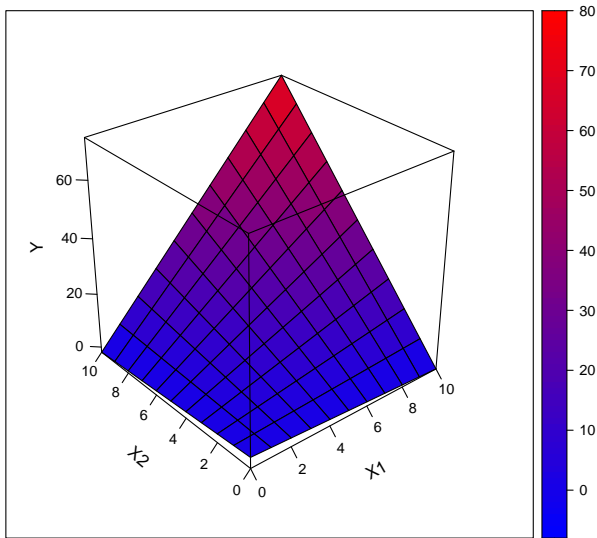
means that:

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

Two Continuous Variables: No Interactive Effects



Two Continuous Variables: Interaction Present



Quadratic, Cubic, and Other Polynomial Effects

For the polynomial-in- X model:

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

in general:

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

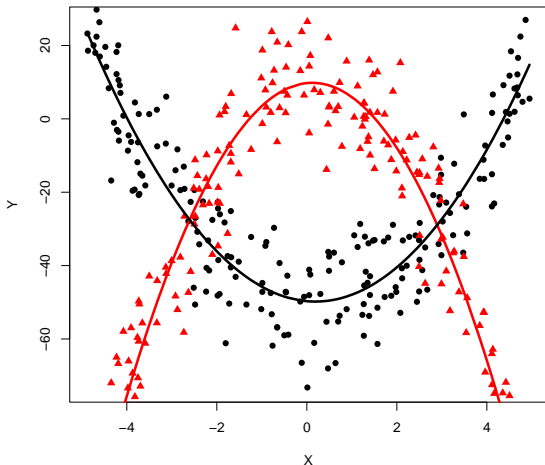
The quadratic case ($j = 2$):

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

implies

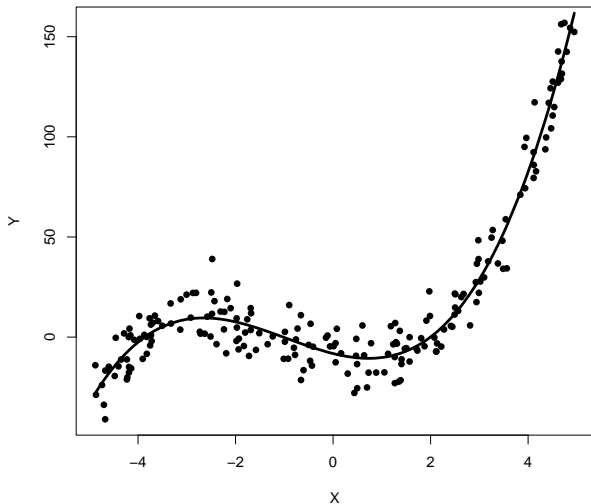
$$\frac{\partial E(Y)}{\partial X} = \beta_1 + 2\beta_2 X.$$

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

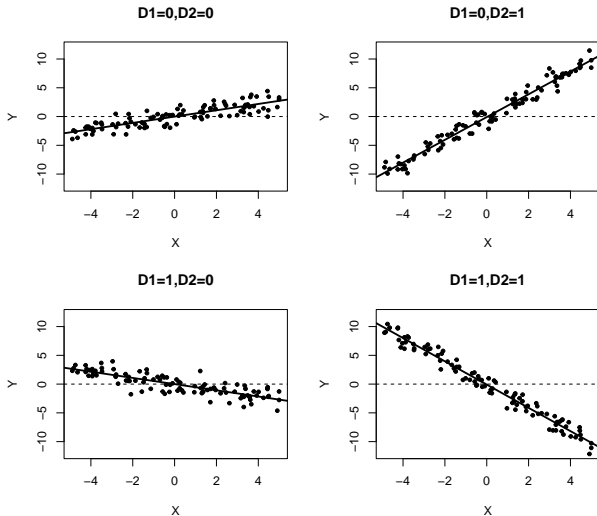
The three-way interaction:

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

...is complex; consider the special case of dichotomous X_1, X_2 :

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

Three-Way Interaction: Two Dummy + One Continuous Covariates



Example: President Clinton's 1996 "Thermometer Score"

Details:

- Source: 1996 [American National Election Study](#)

- $N \approx 1300$

- "Feeling Thermometer":

"In the list that follows, rate that person/party using something we call the feeling thermometer. Ratings between 50 degrees and 100 degrees mean that you feel favorable and warm toward the person/party. Ratings between 0 and 50 degrees mean that you don't feel favorable toward the person/party and that you don't care too much for that person. You would rate the person at the 50-degree mark if you don't feel particularly warm or cold toward the person."

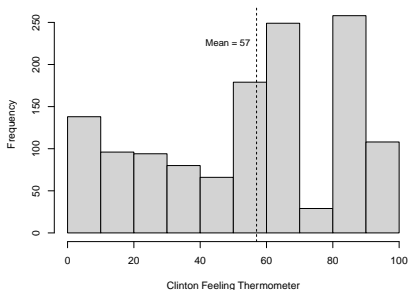
- Ideology measures:

- Seven-point scale (1 = "very left / liberal," 7 = "very right / conservative")
- Each respondent places themselves + President Clinton

- Party ID: Republican (GOP=1) vs. Democrat/Independent (GOP=0)

Clinton “Feeling Thermometer” Data

Distribution:



Summary statistics:

```
> describe(ClintonTherm,skew=FALSE)
```

	vars	n	mean	sd	median	min	max	range	se
caseid	1	1297	2000.70	718.42	1854	1001	3403	2402	19.95
ClintonTherm	2	1297	57.00	30.00	60	0	100	100	0.83
RConserv	3	1297	4.32	1.39	4	1	7	6	0.04
ClintonConserv	4	1297	2.98	1.37	3	1	7	6	0.04
PID	5	1297	2.06	1.05	2	1	5	4	0.03
GOP	6	1297	0.32	0.47	0	0	1	1	0.01

A Basic Regression

```
> fit0<-lm(ClintonTherm~RConserv+GOP,data=CT)
> summary(fit0)
```

Call:

```
lm(formula = ClintonTherm ~ RConserv + GOP, data = CT)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	93.476	2.228	42.0	<2e-16 ***
RConserv	-6.487	0.537	-12.1	<2e-16 ***
GOP	-26.670	1.606	-16.6	<2e-16 ***

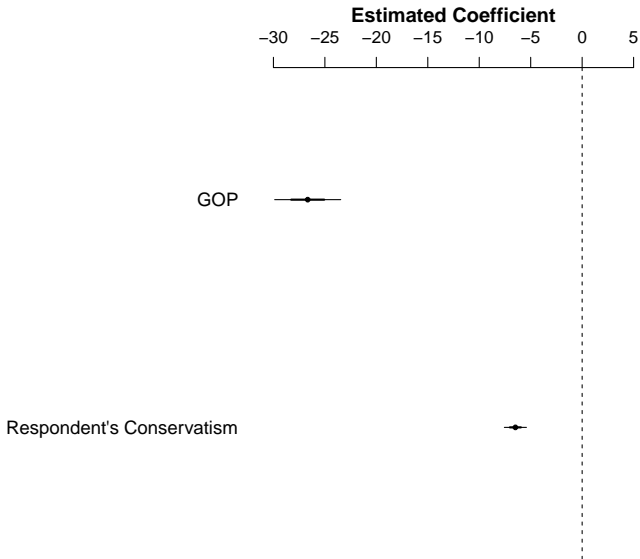
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 23.6 on 1294 degrees of freedom

Multiple R-squared: 0.38, Adjusted R-squared: 0.379

F-statistic: 396 on 2 and 1294 DF, p-value: <2e-16

Coefficient Plot: Non-Interactive Model



An Interactive Model

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

Call:

```
lm(formula = ClintonTherm ~ RConserv + GOP + RConserv * GOP,
    data = CT)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	89.927	2.487	36.16	<2e-16 ***
RConserv	-5.571	0.609	-9.15	<2e-16 ***
GOP	-6.484	6.569	-0.99	0.3238
RConserv:GOP	-4.058	1.281	-3.17	0.0016 **

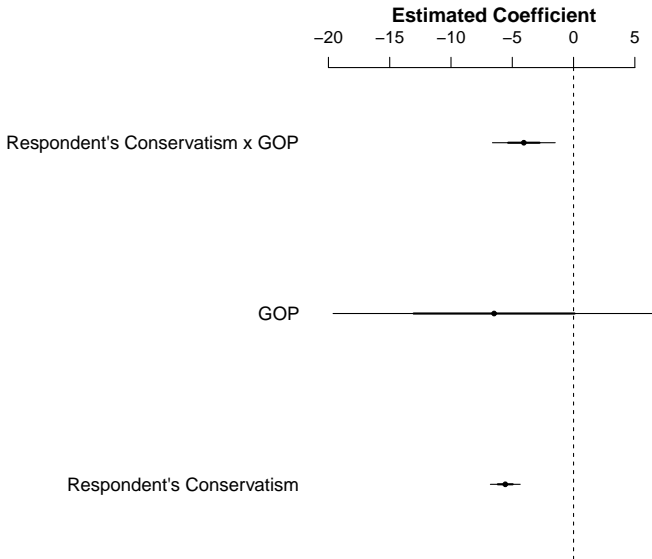
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 23.6 on 1293 degrees of freedom

Multiple R-squared: 0.384, Adjusted R-squared: 0.383

F-statistic: 269 on 3 and 1293 DF, p-value: <2e-16

Coefficient Plot: Interactive Model



Different ways to specify the (mathematically identical, but sometimes computationally different) interactive model in R :

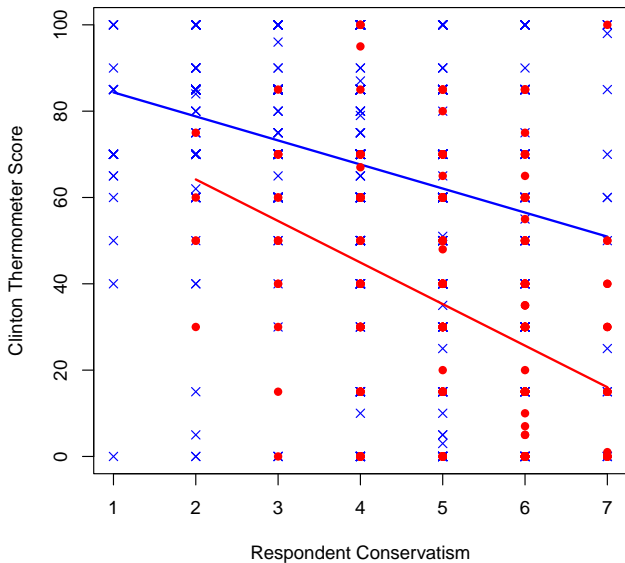
```
> summary(lm(ClintonTherm~RConserv+GOP+RConserv*GOP,data=CT))
> summary(lm(ClintonTherm~RConserv+GOP+RConserv:GOP,data=CT))
> summary(lm(ClintonTherm~RConserv*GOP,data=CT))
> summary(lm(ClintonTherm~(RConserv+GOP)^2,data=CT))
>
> CT$GOPxRC<-CT$GOP * CT$RConserv
> summary(lm(ClintonTherm~RConserv+GOP+GOPxRC,data=CT))
```

Interpretation: Two Regressions, Sort Of

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

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

Thermometer Scores by Conservatism, GOP and Non-GOP



Interactive Results \approx Separate Regressions

```
> NonReps<-subset(CT,GOP==0)
> Reps<-subset(CT,GOP==1)
> split<-list("Interactive" = lm(ClintonTherm~RConserv*GOP,data=CT),
+           "Non-GOP" = lm(ClintonTherm~RConserv,data=NonReps),
+           "GOP" = lm(ClintonTherm~RConserv,data=Reps))
> modelsummary(split,output="SplitModels-24.tex",
+           gof_omit="DF|Deviance|LogLik|AIC|BIC")
```

Feeling Thermometer Models

	Interactive	Non-GOP	GOP
(Intercept)	89.927 (2.487)	89.927 (2.469)	83.443 (6.170)
RConserv	-5.571 (0.609)	-5.571 (0.604)	-9.629 (1.144)
GOP	-6.484 (6.569)		
RConserv \times GOP	-4.058 (1.281)		
Num.Obs.	1297	887	410
R ²	0.384	0.088	0.148
R ² Adj.	0.383	0.087	0.146
F	269.011	84.961	70.878
RMSE	23.53	23.38	23.86

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	73	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

>

> # Note: Same as t-test:

> sqrt(73)

[1] 8.544

An Easier Way: linearHypothesis()

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

Hypothesis:

$\text{GOP} + 7 \text{ RConserv:GOP} = 0$

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 ***

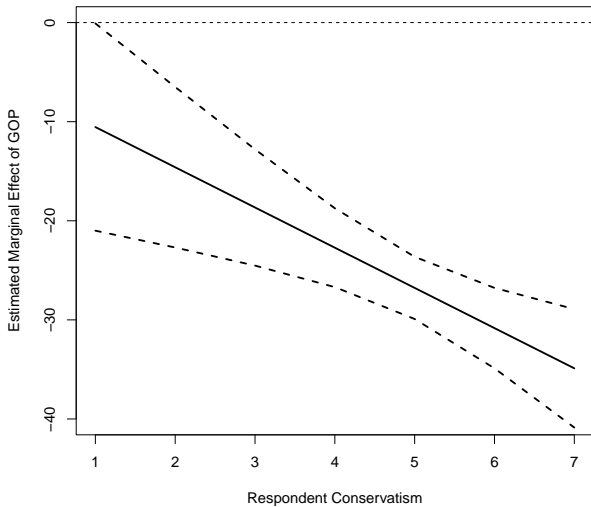
Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Marginal Effects Plots, I

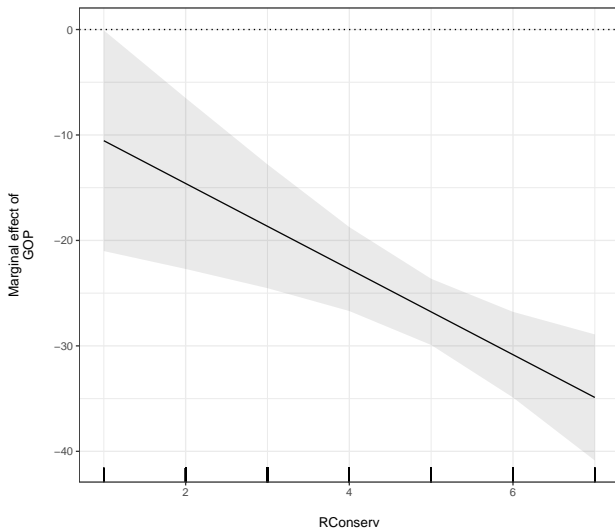
```
> ConsSim<-seq(1,7,1)
> psis<-fit1$coeff[3]+(ConsSim * fit1$coeff[4])
> psis.ses<-sqrt(vcov(fit1)[3,3] +
  (ConsSim)^2*vcov(fit1)[4,4] + 2*ConsSim*vcov(fit1)[3,4])

> 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)
```

The Plot...



Same, Using `plot_me` (`plotMElm` package)



Interacting Two Continuous Covariates

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

Call:

```
lm(formula = ClintonTherm ~ RConserv + ClintonConserv + RConserv *
    ClintonConserv, data = CT)
```

Residuals:

Min	1Q	Median	3Q	Max
-95.48	-13.48	0.48	15.48	100.15

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	119.351	5.163	23.11	< 2e-16 ***
RConserv	-19.567	1.036	-18.88	< 2e-16 ***
ClintonConserv	-7.931	1.648	-4.81	0.0000017 ***
RConserv:ClintonConserv	3.629	0.339	10.69	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 22 on 1293 degrees of freedom

Multiple R-squared: 0.462, Adjusted R-squared: 0.461

F-statistic: 370 on 3 and 1293 DF, p-value: <2e-16

Hypothesis Tests (“by hand”)

```
> psi<-fit2$coef[2]+(1*fit2$coef[4])
> psi
RConserv
-15.94

> varpsi<-sqrt(vcov(fit2)[2,2] + (1)^2*vcov(fit2)[4,4] + 2*1*vcov(fit2)[2,4])
> varpsi
[1] 0.744

> psi/varpsi # t-test
RConserv
-21.42

> 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

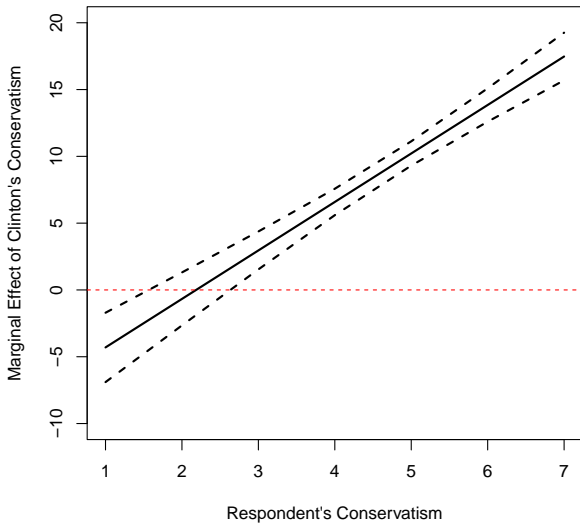
   Res.Df    RSS Df Sum of Sq   F Pr(>F)
1    1294 850442
2    1293 627658   1    222784 459 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> (psi/varpsi)^2
RConserv
458.9
```

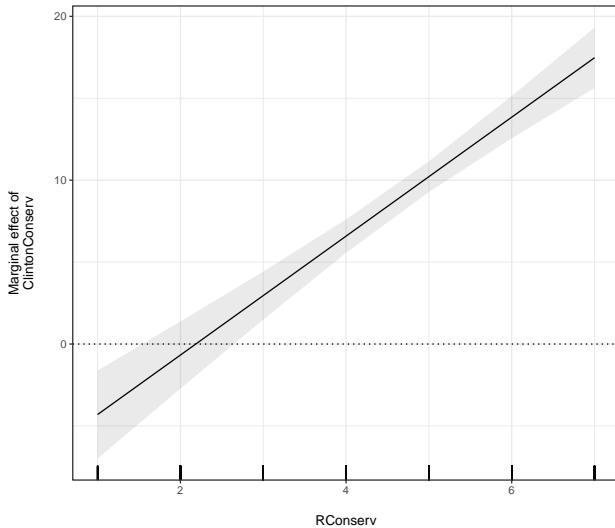

Marginal Effect Plot, II

```
> 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")
```



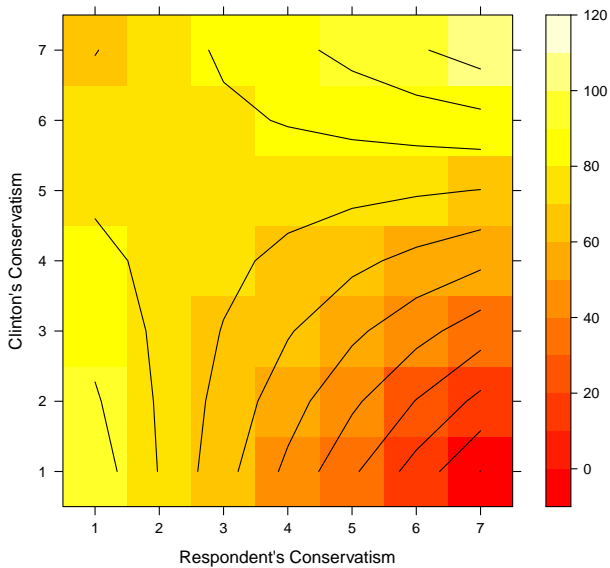
Same, Using plot_me



Predicted Values: A Contour Plot

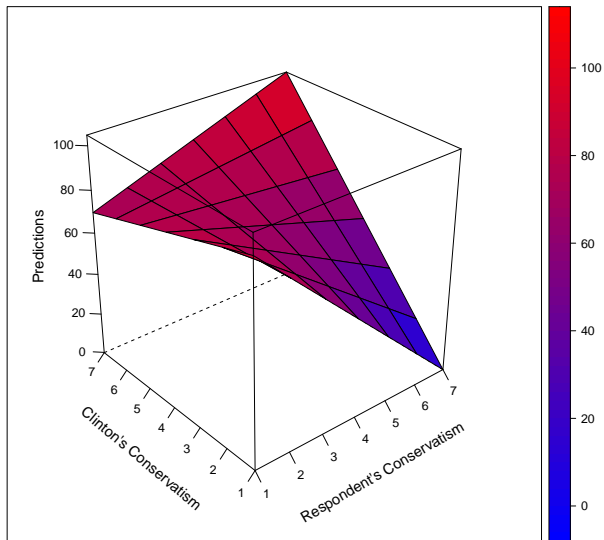
```
> library(lattice)
> grid<-expand.grid(RConserv=seq(1,7,1),
+   ClintonConserv=seq(1,7,1))
> hats<-predict(fit2,newdata=grid)

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



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))
```



Interpreting Interactions: Other Tools

A partial list of packages and things:

- `interactions` ([vignette](#))
- `emmeans` (see, e.g., [this vignette](#))
- `plot_model` in `sjPlot` ([tutorial](#))
- `interactionR`
- `interplot`
- `InteractionPowerR` (power analyses for interactive models)

Variable Transformations

Why Transform?

- Induce / conform to linearity
- Induce / conform to additivity
- Induce normality in the u_i s
- 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	$-\frac{1}{3}$	$\frac{1}{\sqrt[3]{X}}$	$\frac{(\frac{1}{\sqrt[3]{X}} - 1)}{-\frac{1}{3}}$
Inverse Square Root	$-\frac{1}{2}$	$\frac{1}{\sqrt{X}}$	$\frac{(\frac{1}{\sqrt{X}} - 1)}{-\frac{1}{2}}$
Inverse	-1	$\frac{1}{X}$	$\frac{(\frac{1}{X} - 1)}{-1}$
Inverse Square	-2	$\frac{1}{X^2}$	$\frac{(\frac{1}{X^2} - 1)}{-2}$
Inverse Cube	-3	$\frac{1}{X^3}$	$\frac{(\frac{1}{X^3} - 1)}{-3}$

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:

$\ln(X|X \leq 0)$ 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:

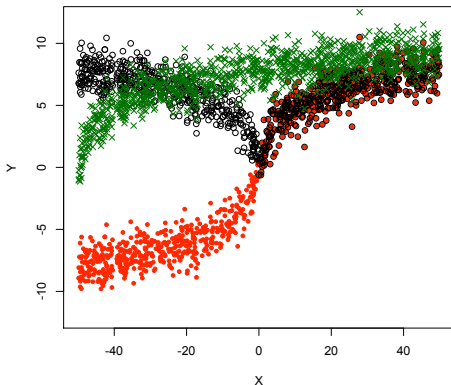
$$D_i = \begin{cases} 1 & \text{if } X_i = 0 \\ 0 & \text{otherwise} \end{cases}$$

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 $X < 0$:

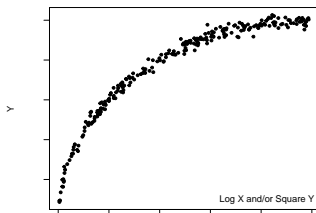
1. a “shift”, where the logarithmic form starts at values of X less than zero,
2. a “V-curve,” where $E(Y|X = k) = E(Y|X = -k)$, or
3. an “S-curve,” where the $X - Y$ relationship for $X < 0$ “mirrors” that for $X > 0$ [so $E(Y|X = k) = -E(Y|X = -k)$]



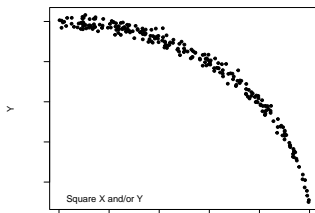
Which is correct? **It depends on your theory.** Again: find out more at [this poster](#).

Which Transformation?

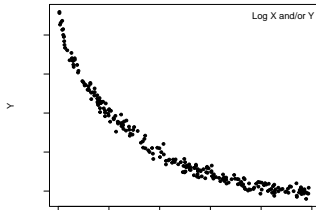
Mosteller and Tukey's "Bulging Rule":



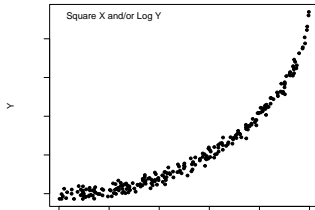
X



X



X



X

Transformed X s: Interpretation

For:

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

then:

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

and so:

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

Transformed X s: Interpretation

Similarly, for:

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

we have:

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

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

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

Specifying:

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

means:

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

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

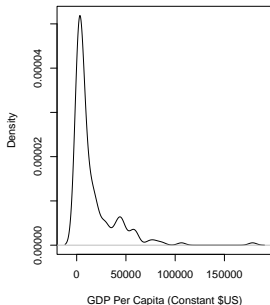
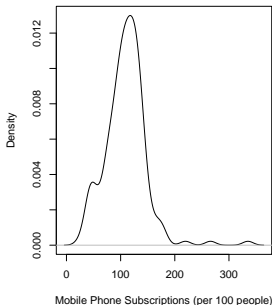
An Example: Cell Phones and Wealth

Data are a tiny subset of the **World Development Indicators** (2018 *only*)...

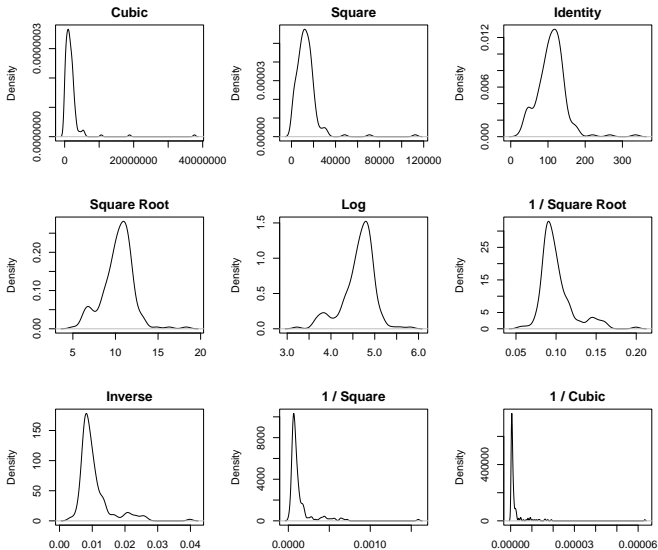
- ISO3 - The country's ISO3 (alphanumeric) code.
- country - The name of the country.
- GDPPerCapita - GDP per capita (constant 2010 \$US)
- MobileCellSubscriptions - Mobile / cellular subscriptions per 100 people

```
> describe(WDI)
```

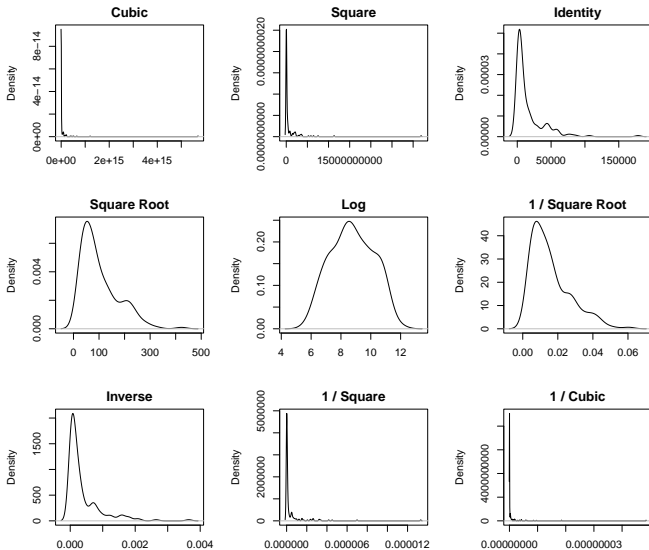
	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
ISO3*	1	186	93.5	53.84	93.5	93.5	68.94	1.00	186.0	185	0.00	-1.22	3.95
Country*	2	186	93.5	53.84	93.5	93.5	68.94	1.00	186.0	185	0.00	-1.22	3.95
GDPPerCapita	3	186	15769.2	22904.65	6257.3	11146.0	7358.36	274.13	178571.4	178297	3.05	14.09	1679.45
MobileCellSubscriptions	4	186	109.0	38.01	110.1	108.4	30.00	25.11	335.1	310	1.32	7.02	2.79



“Ladder of Powers”: Mobile Subscriptions

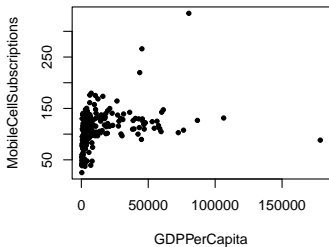


“Ladder of Powers”: Wealth / GDP

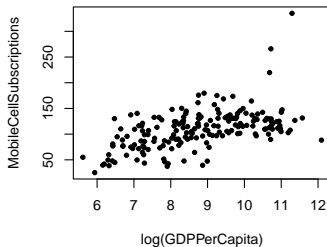


Scatterplots

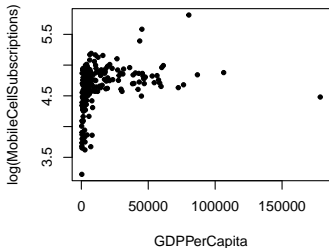
Linear-Linear



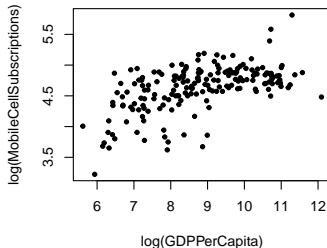
Linear-Log



Log-Linear



Log-Log

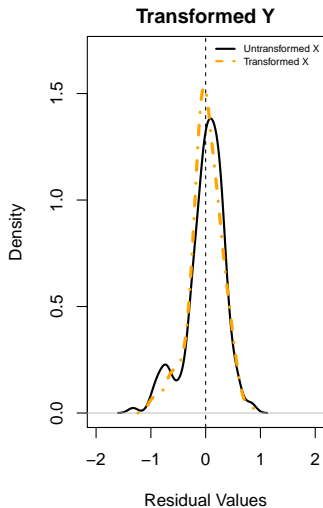
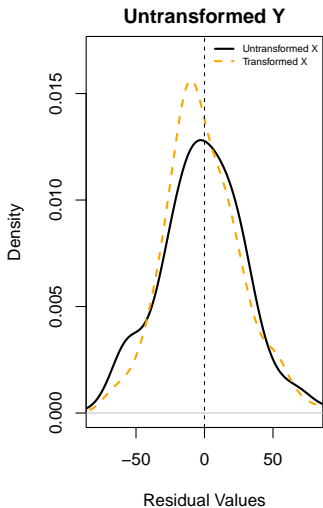


Various Regressions

Models with and Without Transformations

	Linear Y & X	Linear Y, Log X	Log Y, Linear X	Log Y and X
(Intercept)	100.235 (3.199)	82.944 (3.904)	4.546 (0.032)	4.357 (0.037)
GDP Per Capita (1000s)	0.557 (0.115)		0.005 (0.001)	
Logged GDP Per Capita (1000s)		14.047 (1.669)		0.146 (0.016)
Num.Obs.	186	186	186	186
R2	0.113	0.278	0.103	0.314
R2 Adj.	0.108	0.274	0.098	0.310
F	23.370	70.830	21.142	84.305
RMSE	35.71	32.22	0.35	0.31

Density Plots of \hat{u}_i s



(One) simple solution: Polynomials...

- First-order / linear ($P = 1$):

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

- Second-order / quadratic ($P = 2$):

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

- Third-order / cubic ($P = 3$):

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

- ... p th-order ($P = p$):

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

Understanding Polynomials

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

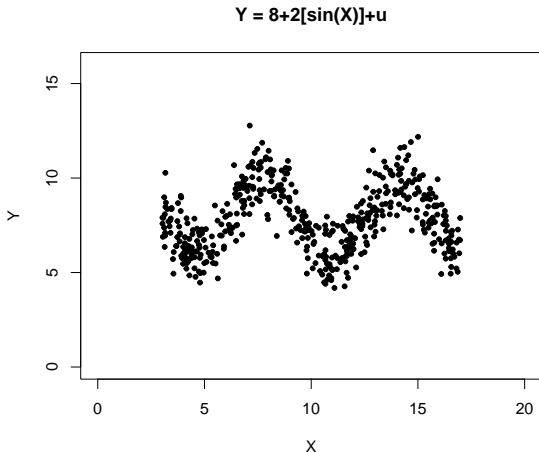
$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + u_i$$

then:

$\hat{\beta}_1$	$\hat{\beta}_2$		
	< 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

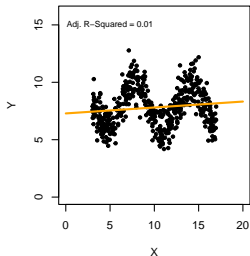
Polynomials: Simulated Example

```
> N<-500  
> set.seed(7222009)  
> X<-runif(N,3,17)  
> Y<-8+2*sin(X)+rnorm(N)
```

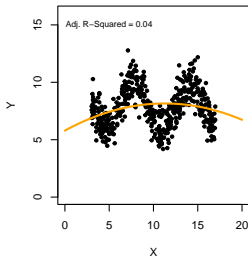


Some Polynomial Regressions

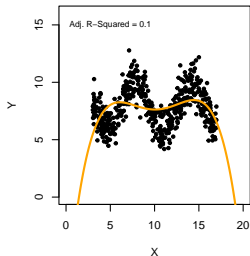
Linear ($P=1$)



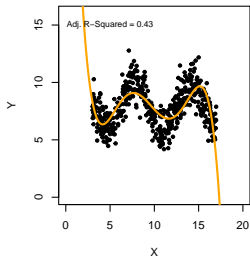
Quadratic ($P=2$)



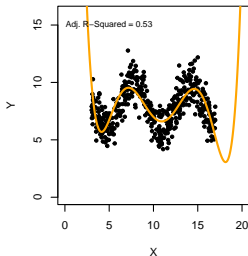
Fourth-Degree ($P=4$)



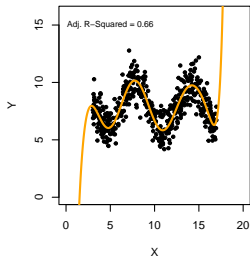
Fifth-Degree ($P=5$)



Sixth-Degree ($P=6$)



Twelfth-Degree ($P=12$)



"Raw" vs. Orthogonal Polynomials

Check out the $P = 12$ regression:

```
> summary(R.12)
```

Call:

```
lm(formula = Y ~ X + I(X^2) + I(X^3) + I(X^4) + I(X^5) + I(X^6) +  
    I(X^7) + I(X^8) + I(X^9) + I(X^10) + I(X^11) + I(X^12))
```

Coefficients: (1 not defined because of singularities)

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-92.9216900503	816.2489413571	-0.11	0.91
X	149.4055086103	1212.1158531764	0.12	0.90
I(X^2)	-104.6855472388	788.3527710495	-0.13	0.89
I(X^3)	46.9192208616	296.7267961681	0.16	0.87
I(X^4)	-14.5570311719	71.9111574249	-0.20	0.84
I(X^5)	3.1175787785	11.8013895994	0.26	0.79
I(X^6)	-0.4537511003	1.3406553442	-0.34	0.74
I(X^7)	0.0442854569	0.1056218156	0.42	0.68
I(X^8)	-0.0028398562	0.0056659448	-0.50	0.62
I(X^9)	0.0001145568	0.0001974535	0.58	0.56
I(X^10)	-0.0000026340	0.0000040301	-0.65	0.51
I(X^11)	0.0000000263	0.0000000366	0.72	0.47
I(X^12)	NA	NA	NA	NA

Residual standard error: 0.986 on 488 degrees of freedom

Multiple R-squared: 0.669, Adjusted R-squared: 0.662

F-statistic: 89.7 on 11 and 488 DF, p-value: <2e-16

“Raw” vs. Orthogonal Polynomials (continued)

What's going on?

- The “raw” polynomial terms are (often strongly) correlated with each other...

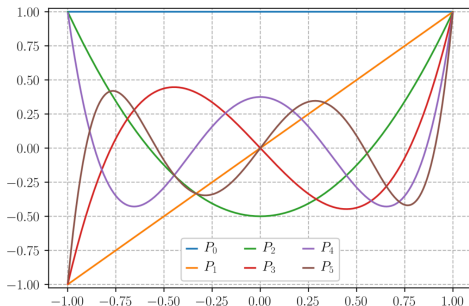
```
> cor(X,I(X^2))  
[1] 0.984
```

- → large standard errors / imprecision in the estimates
- Can also lead to numerical instability in estimation...

Orthogonal Polynomials

An alternative is to use *orthogonal polynomials*...

- Think of these as orthogonal (uncorrelated) versions of the polynomials above
- There are many of them; probably the most commonly-used are the **Legendre polynomials**:



- The math is a bit complex; the R command is `poly()`

"Raw" polynomials using poly():

```
> P.12R<-lm(Y~poly(X,degree=12,raw=TRUE))
> summary(P.12R)
```

Call:

```
lm(formula = Y ~ poly(X, degree = 12, raw = TRUE))
```

Residuals:

Min	1Q	Median	3Q	Max
-2.8041	-0.6935	-0.0245	0.6881	3.0645

Coefficients: (1 not defined because of singularities)

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-92.9216900503	816.2489413571	-0.11	0.91
poly(X, degree = 12, raw = TRUE)1	149.4055086103	1212.1158531764	0.12	0.90
poly(X, degree = 12, raw = TRUE)2	-104.6855472388	788.3527710495	-0.13	0.89
poly(X, degree = 12, raw = TRUE)3	46.9192208616	296.7267961681	0.16	0.87
poly(X, degree = 12, raw = TRUE)4	-14.5570311719	71.9111574249	-0.20	0.84
poly(X, degree = 12, raw = TRUE)5	3.1175787785	11.8013895994	0.26	0.79
poly(X, degree = 12, raw = TRUE)6	-0.4537511003	1.3406553442	-0.34	0.74
poly(X, degree = 12, raw = TRUE)7	0.0442854569	0.1056218156	0.42	0.68
poly(X, degree = 12, raw = TRUE)8	-0.0028398562	0.0056659448	-0.50	0.62
poly(X, degree = 12, raw = TRUE)9	0.0001145568	0.0001974535	0.58	0.56
poly(X, degree = 12, raw = TRUE)10	-0.0000026340	0.0000040301	-0.65	0.51
poly(X, degree = 12, raw = TRUE)11	0.0000000263	0.0000000366	0.72	0.47
poly(X, degree = 12, raw = TRUE)12	NA	NA	NA	NA

Residual standard error: 0.986 on 488 degrees of freedom

Multiple R-squared: 0.669, Adjusted R-squared: 0.662

F-statistic: 89.7 on 11 and 488 DF, p-value: <2e-16

Our Example (continued)

Orthogonal polynomials:

```
> P.12<-lm(Y~poly(X,degree=12))
> summary(P.12)
```

Call:

```
lm(formula = Y ~ poly(X, degree = 12))
```

Residuals:

Min	1Q	Median	3Q	Max
-2.8255	-0.6953	-0.0286	0.6894	3.0733

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	7.7989	0.0441	176.80	< 2e-16 ***
poly(X, degree = 12)1	4.7901	0.9863	4.86	0.00000161352 ***
poly(X, degree = 12)2	-6.2379	0.9863	-6.32	0.00000000058 ***
poly(X, degree = 12)3	2.5039	0.9863	2.54	0.011 *
poly(X, degree = 12)4	-9.5937	0.9863	-9.73	< 2e-16 ***
poly(X, degree = 12)5	-21.5763	0.9863	-21.87	< 2e-16 ***
poly(X, degree = 12)6	12.0295	0.9863	12.20	< 2e-16 ***
poly(X, degree = 12)7	12.4067	0.9863	12.58	< 2e-16 ***
poly(X, degree = 12)8	-5.2176	0.9863	-5.29	0.00000018541 ***
poly(X, degree = 12)9	-1.4389	0.9863	-1.46	0.145
poly(X, degree = 12)10	2.0529	0.9863	2.08	0.038 *
poly(X, degree = 12)11	0.7097	0.9863	0.72	0.472
poly(X, degree = 12)12	-0.3987	0.9863	-0.40	0.686

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.986 on 487 degrees of freedom

Multiple R-squared: 0.669, Adjusted R-squared: 0.661

F-statistic: 82.1 on 12 and 487 DF, p-value: <2e-16

What Degree Polynomial?

```
> for(degree in 1:12) {  
+   fit <- lm(Y~poly(X,degree))  
+   assign(paste("P", degree, sep = "."), fit)  
+ }  
> anova(P.1,P.2,P.3,P.4,P.5,P.6,P.7,P.8,P.9,P.10,P.11,P.12)  
Analysis of Variance Table
```

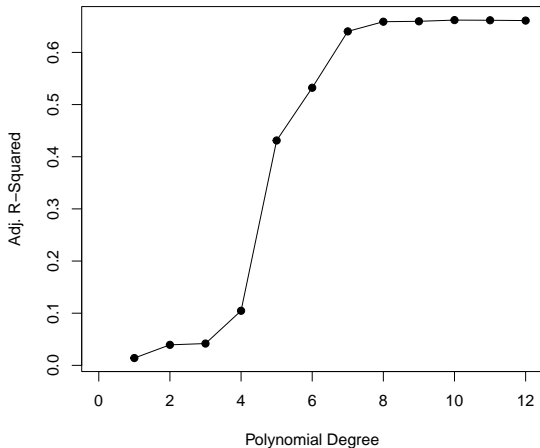
```
Model 1: Y ~ poly(X, degree)  
Model 2: Y ~ poly(X, degree)  
Model 3: Y ~ poly(X, degree)  
Model 4: Y ~ poly(X, degree)  
Model 5: Y ~ poly(X, degree)  
Model 6: Y ~ poly(X, degree)  
Model 7: Y ~ poly(X, degree)  
Model 8: Y ~ poly(X, degree)  
Model 9: Y ~ poly(X, degree)  
Model 10: Y ~ poly(X, degree)  
Model 11: Y ~ poly(X, degree)  
Model 12: Y ~ poly(X, degree)
```

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	498	1409				
2	497	1370	1	39	40.00	0.00000000058 ***
3	496	1364	1	6	6.44	0.011 *
4	495	1272	1	92	94.60	< 2e-16 ***
5	494	807	1	466	478.51	< 2e-16 ***
6	493	662	1	145	148.74	< 2e-16 ***
7	492	508	1	154	158.22	< 2e-16 ***
8	491	481	1	27	27.98	0.00000018541 ***
9	490	479	1	2	2.13	0.145
10	489	474	1	4	4.33	0.038 *
11	488	474	1	1	0.52	0.472
12	487	474	1	0	0.16	0.686

```
----  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

What Degree Polynomial?

Plotting R^2_{adj} for different polynomial degrees...



Good things...

- Polynomials are flexible functional forms for nonlinear marginal associations
- They are also easy to fit, and easily interpretable

Cautions...

- Polynomials can be prone to overfitting, which...
- ...can lead to poor out-of-sample generalizability / predictive power
- This is especially true outside the observed values of the data (extrapolation)

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