Interpreting (Even Tricky) Regression Coefficients

Scaling of Dependent and Independent Variables

Module 2: Scaling of Dependent and Independent Variables

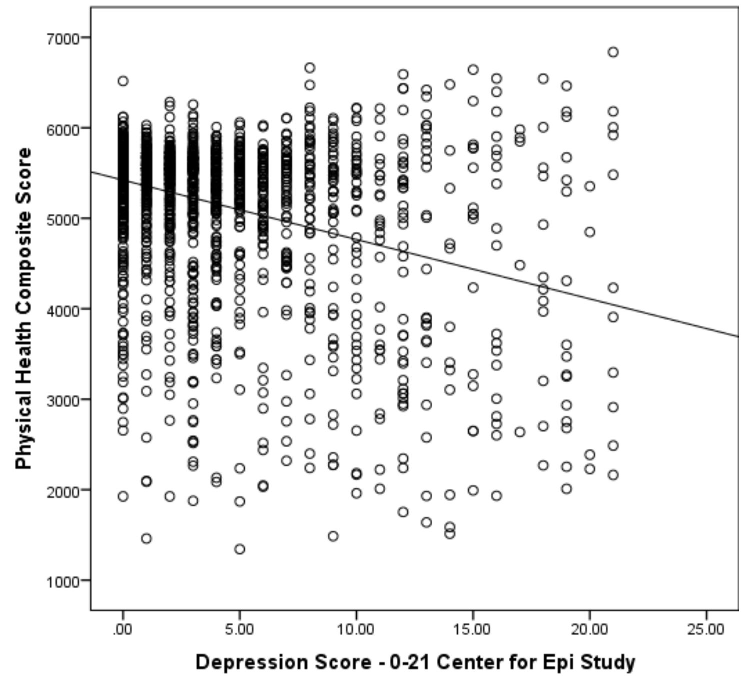
1. Types of Transformations
2. Linear Transformations: Centering
3. Linear Transformations: Rescaling
4. Linear Transformations: Standardized Coefficients
5. Nonlinear Transformations

Types of Transformations

Transformations of X and Y

$$b_0 = E(Y \mid all\ X = 0)$$

$$b_1 = \frac{\Delta \hat{Y}}{\Delta X_1} = \frac{difference\ \ in\ \ E(Y)}{1\ \ unit\ \ difference\ \ in\ \ X_1}$$



Ways to Transform a Numerical Variable

1. Linear Transformations

    Add or subtract a constant

    Multiply or divide by a constant

2. Non-linear Transformations

    exponents, logs, inverse, trig functions, etc.

3. Make it non-numerical

    Ordered categories

    Unordered categories

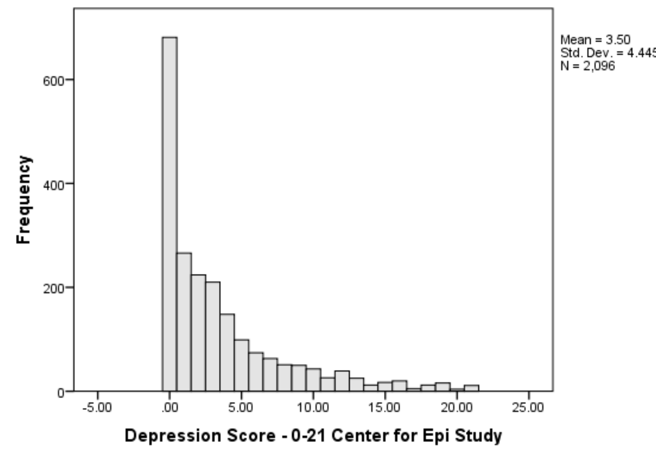
    Ranks

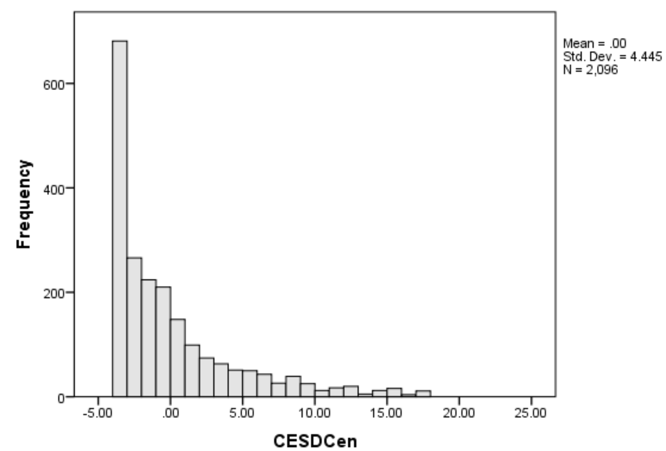
Linear Transformations

Centering

Rescaling

Standardizing

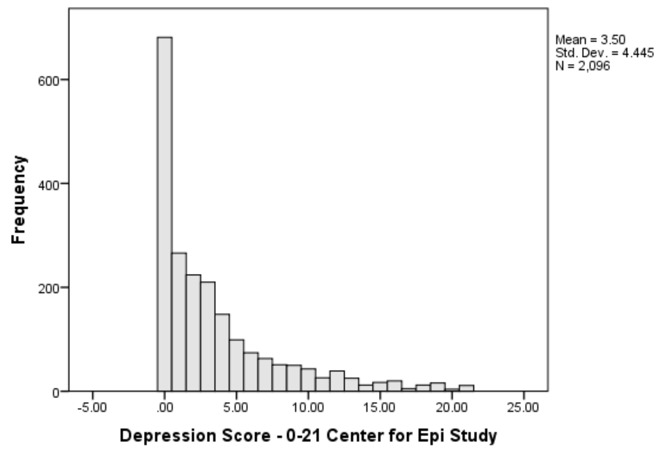


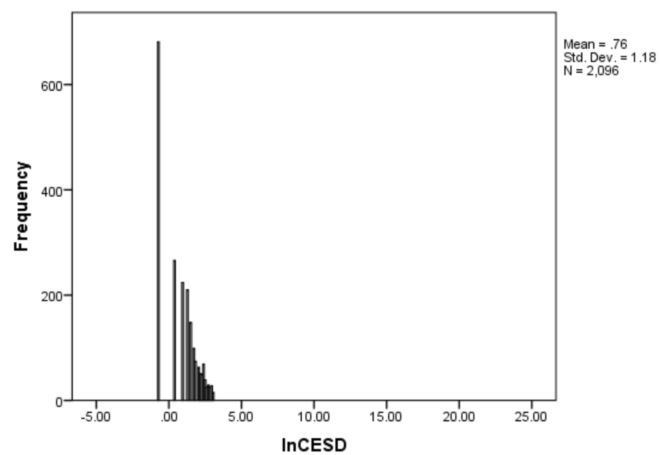


Nonlinear Transformations

Logarithms

Square Roots





Linear Transformations: Centering

Centering

Recoding a predictor variable by subtracting a constant from every observation's value

Typical constants to subtract:

- mean
- other meaningful central value

Center Mental Health Score

How:

Create new variable:
MCSCen = MCS2000 – 5280.05

MCS2000	MCSCen
5860	579.95
5789	508.95
5792	511.95
5501	220.95
6079	798.95
5792	511.95
5545	264.95
6076	795.95
3466	-1814.05
6716	1435.95

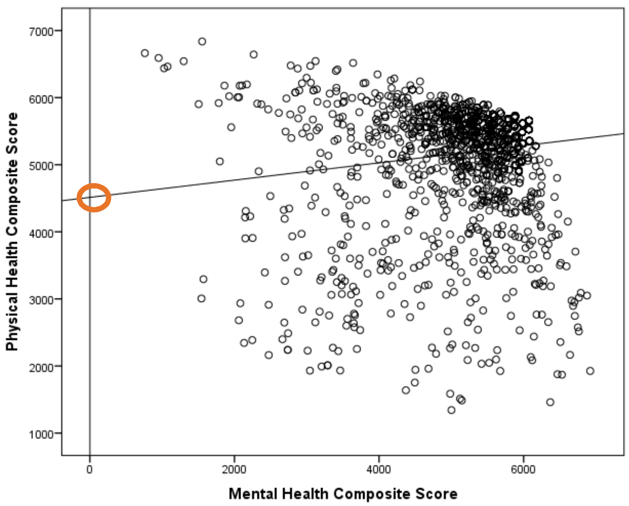
Descriptive Statistics

	N	Min	Max	Mean	Std Dev
MCS2000	2089	760	6921	5280.05	873.510

Descriptive Statistics

	N	Min	Max	Mean	Std Dev
MCSCen	2089	-4520.05	1640.95	.0031	873.510

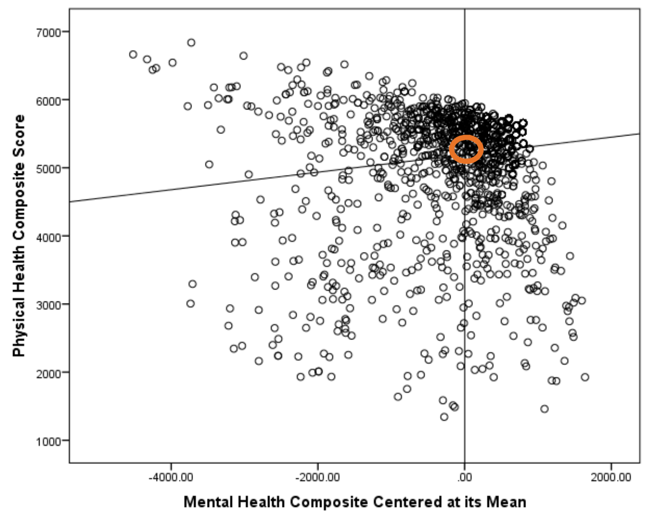
Centering Affects Intercept, not Slope
iff there are no multiplicative terms in the model



Regression Coefficients

Dependent Variable: PCS2000

Variable	B	se	t	p
Intercept	4513.771	110.530	40.837	.000
MCS2000	.129	.021	6.225	.000



Regression Coefficients

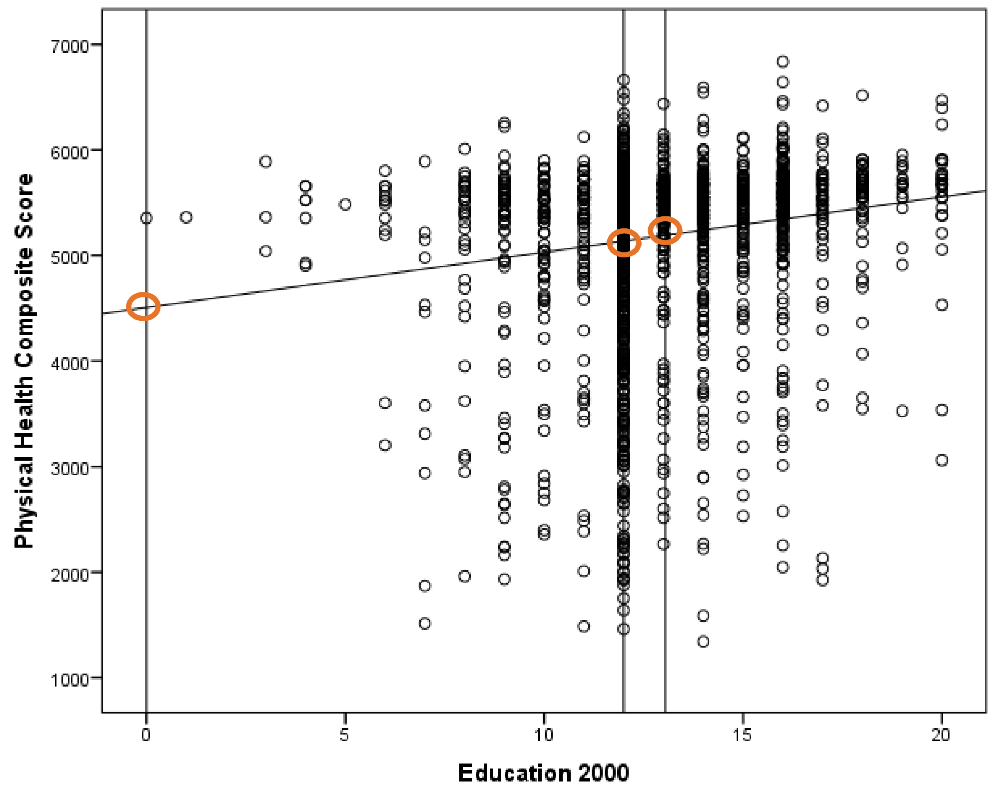
Dependent Variable: PCS2000

Variable	B	se	t	p
Intercept	5192.579	18.036	287.896	.000
MCSCen	.129	.021	6.225	.000

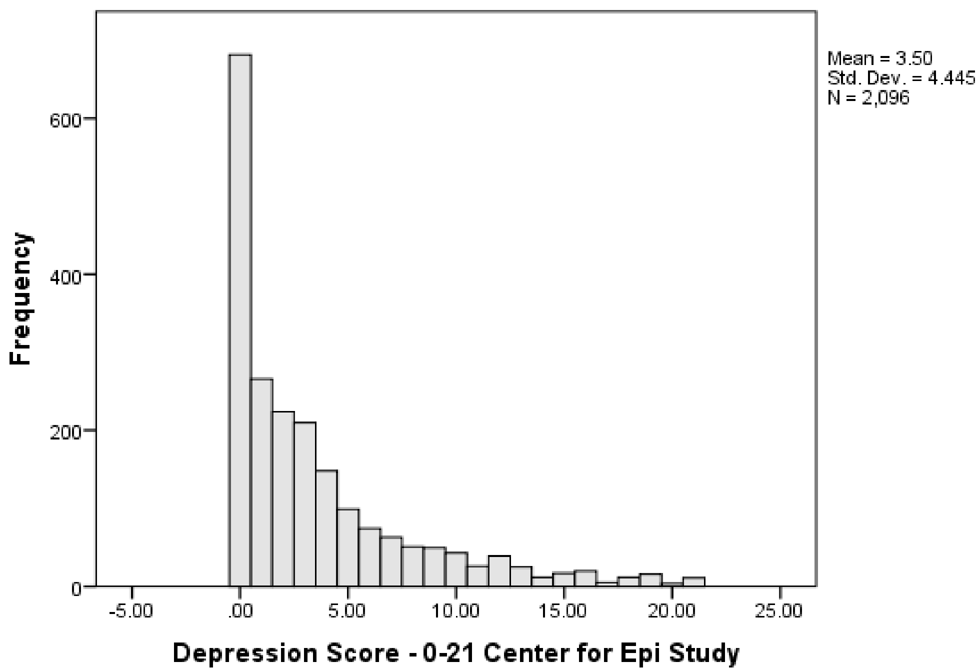
Where to Center? Education

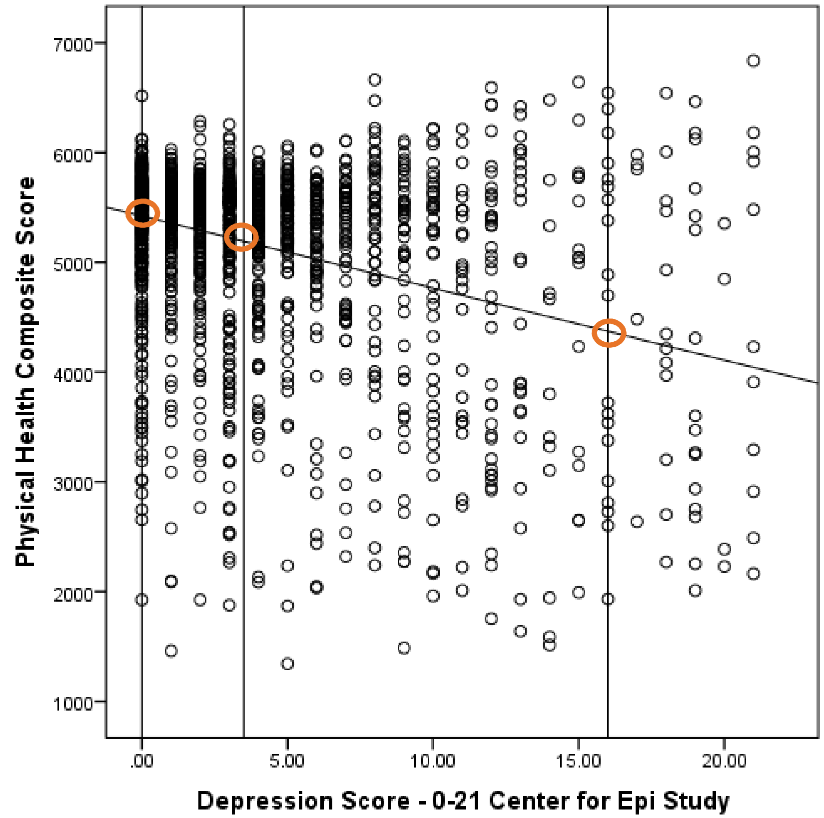
Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Education2000	2102	0	20	13.05	2.537



Where to Center? Depression Score





Center Mental Health Score and Education

Regression Coefficients

Dependent Variable: PCS2000

Variable	B	Se	t	p
Intercept	5931.907	182.535	32.497	.000
Education2000	30.748	6.753	4.553	.000
NumberBioStepAdoptChildHH2000	34.881	12.772	2.731	.006
CESD2000Total	-84.633	5.217	-16.222	.000
MCS2000	-.169	.026	-6.489	.000

Regression Coefficients

Dependent Variable: PCS2000

Variable	B	Se	t	p
Intercept	5109.760	25.707	198.772	.000
EducationCen	30.748	6.753	4.553	.000
NumberBioStepAdoptChildHH2000	34.881	12.772	2.731	.006
CESDCen	-84.633	5.217	-16.222	.000
MCSCen	-.169	.026	-6.489	.000

Linear Transformations: Rescaling

Rescaling

Recoding a variable by multiplying or dividing every observation's value by a constant

Typical constants to multiply or divide:

- Meaningful change in units
- Making variances more comparable

Rescaling

How: Create new variables

gestation_days = gestation * 7

bwt_pnds = bwt / 16

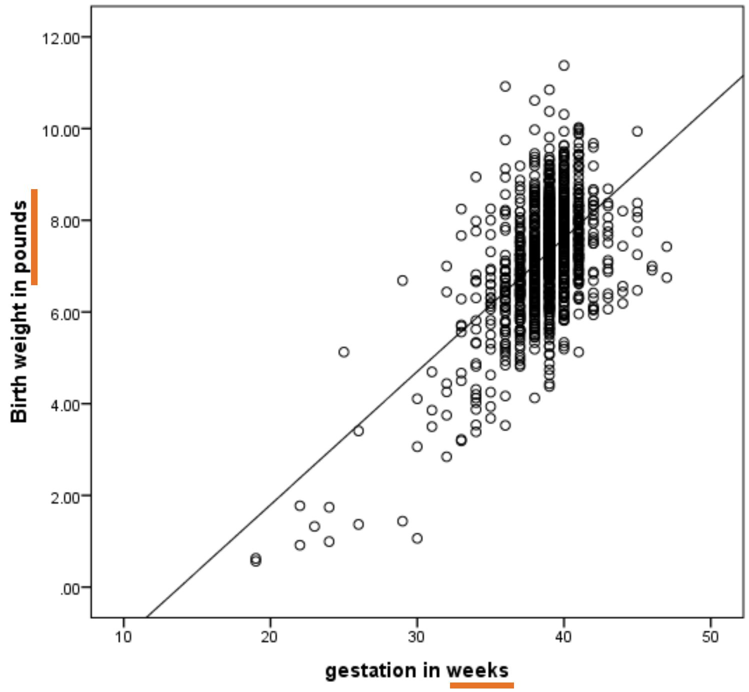
bwt	bwt_pnds	gestation	gestation_days
100.99	6.31	37	259
93.02	5.81	38	266
132.00	8.25	38	266
105.01	6.56	39	273
102.33	6.40	43	301
119.05	7.44	39	273
106.99	6.69	37	259
117.99	7.37	39	273
123.99	7.75	39	273
110.58	6.91	46	322

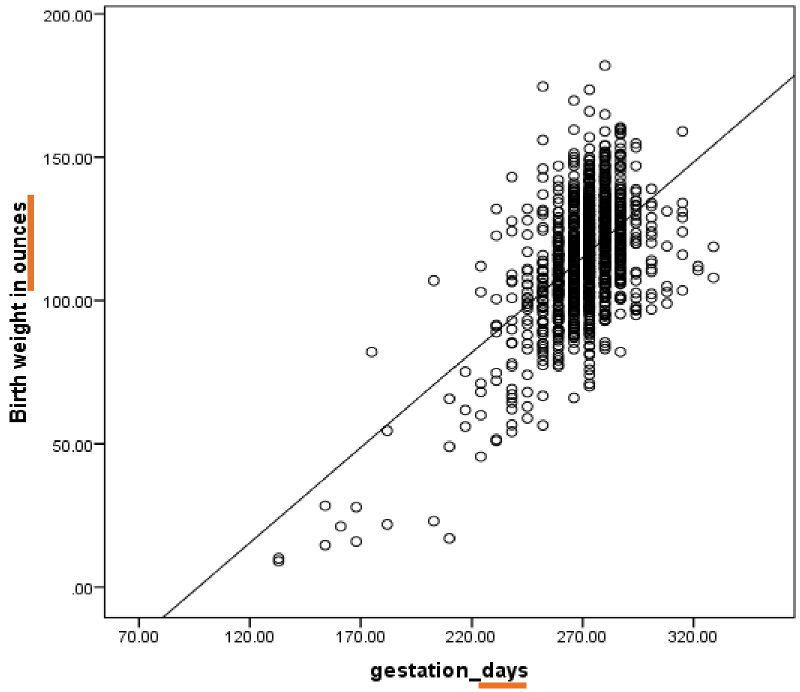
Rescaling

$$b_0 = E(Y \mid all\ X = 0)$$

$$b_1 = \frac{\Delta \hat{Y}}{\Delta X_1} = \frac{difference\ \ in\ \ E(Y)}{1\ \ unit\ \ difference\ \ in\ \ X_1}$$

Rescaling affects intercept and slope through units, not relationships





Rescaling affects intercept and slope through units, not relationships

Regression Coefficients

Dependent Variable: Birth weight in pounds

Variable	B	se	t	p
Intercept	-4.015	.475	-8.445	.000
gestation	.291	.012	23.681	.000

Regression Coefficients

Dependent Variable: Birth weight in ounces

Variable	B	se	t	p
Intercept	-64.246	7.608	-8.445	.000
gestation_days	.664	.028	23.681	.000

-4.015 * 16 = -64.24

(.291 * 16)/7 = .665

Linear Transformations: Standardized Coefficients

Standardized Coefficients

$$b_1 = \frac{\Delta \hat{Y}}{\Delta X_1} = \frac{difference \cdot in \cdot E(Y)}{1 \cdot unit \cdot difference \cdot in \cdot X_1}$$

Standardized Coefficients

Both Response and all Predictor Variables are standardized:

$$Y_i' = \left(\frac{Y_i - \overline{Y}}{s_Y} \right) \qquad X_i' = \left(\frac{X_i - \overline{X}}{s_X} \right)$$

Standardized Coefficients

Standardized Regression Model:

$$Y_i^{'} = \beta_1^{'} X_{i1}^{'} + \beta_2^{'} X_{i2}^{'} + \beta_3^{'} X_{i3}^{'} + \varepsilon_i^{'}$$

Standardized Coefficients

Standardized Regression Model:

$$Y_i' = \beta_1' X_{i1}' + \beta_2' X_{i2}' + \beta_3' X_{i3}' + \varepsilon_i'$$

No intercept

$$\beta_0 = \overline{Y} - \beta_1 \overline{X_1} - \beta_2 \overline{X_2} - \beta_3 \overline{X_3}$$

Standardized Coefficients

Standardized Regression Model:

$$Y_i^{'} = \beta_1^{'} X_{i1}^{'} + \beta_2^{'} X_{i2}^{'} + \beta_3^{'} X_{i3}^{'} + \varepsilon_i^{'}$$

Standardized Coefficients are related to Unstandardized Coefficients

$$\beta_k^{'} = \left(\frac{s_k}{s_Y} \right) \beta_k$$

where k = 1, 2,…number of Xs

Standardized Coefficients

Standardized Regression Model:

$$Y_i' = \beta_1' X_{i1}' + \beta_2' X_{i2}' + \beta_3' X_{i3}' + \varepsilon_i'$$

Interpretation:

β_k' = Number of standard deviations difference in Y for each one standard deviation difference in X_k

Standardized Coefficients

Regression Coefficients
Dependent Variable: PCS2000

Variable	Unstandardized Coefficients B	Standardized Coefficients β	t	p
Intercept	5931.907		32.497	.000
Depression	-84.633	-.451	-16.222	.000
Education	30.748	.094	4.553	.000
NumberChildren	34.881	.055	2.731	.006
MentalHealth	-.169	-.178	-6.489	.000

E(Physical Health) = – 0.45(Depression) + 0.09(Education)
+.06(Number of Children) - .18(Mental Health)

Nonlinear Transformations

Linear Transformations

Effects:

$[(Y-k)*c]$

- Change units
- Change where we evaluate X=0
- Do not change relationships
- Do not change the spacing of the values of a variable

Nonlinear Transformations

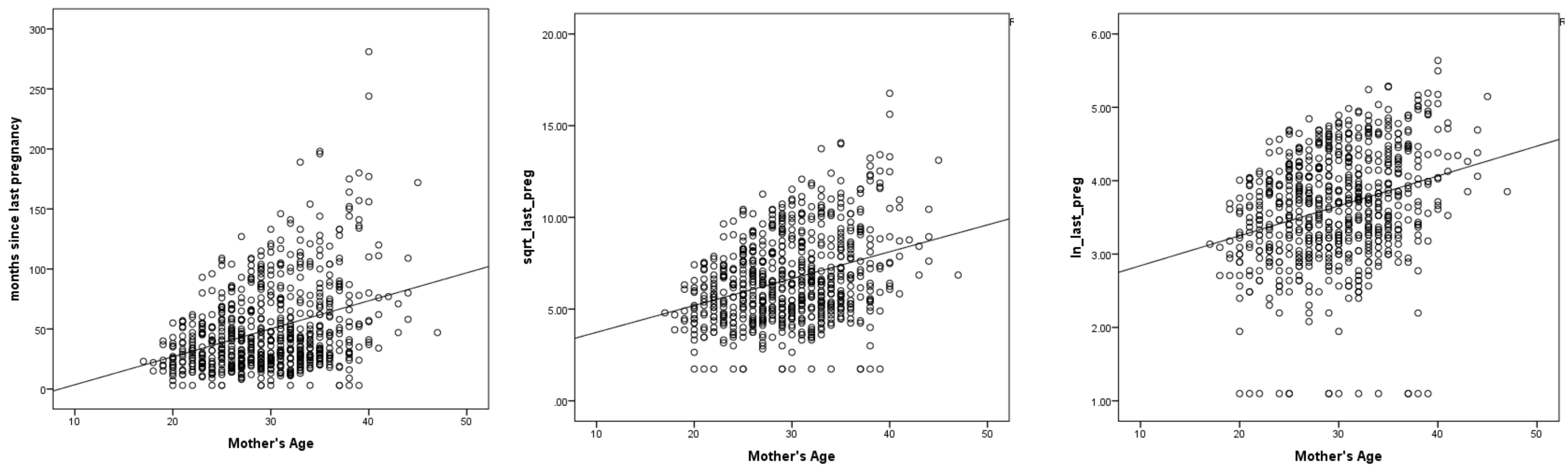
Effects:

- Change units
- Change where we evaluate X=0
- Change relationships
- Change the spacing of the values of a variable

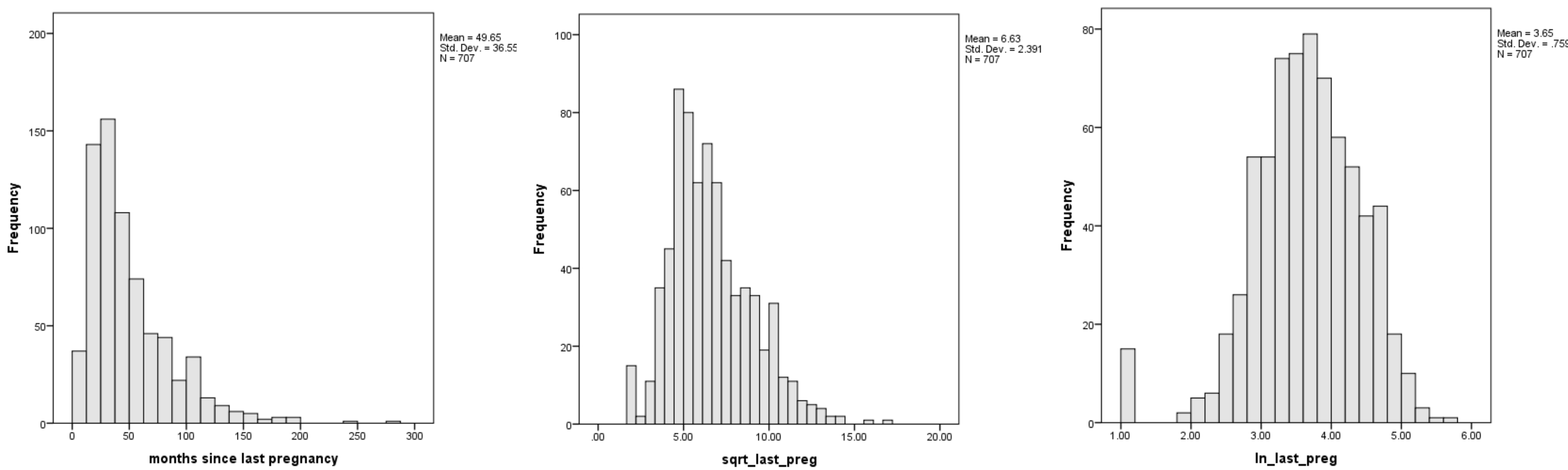
Examples of Common Nonlinear Transformations:

- Logarithms
- Square, cubic, and Square root
- Inverse
- Trigonometric functions, like sine and cosine

Nonlinear Transformations of Y:



Nonlinear Transformations



How Nonlinear Transformations work on X or Y

$W_i = \text{Log}(Y_i)$ $\quad Q_i = \text{Log}(X_i)$ $\quad Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$

$W_i = \text{sqrt}(Y_i)$ $\quad Q_i = e^{Xi}$ $\quad W_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$

$W_i = 1/Y_i$ $\quad Q_i = X_i^2$ $\quad Y_i = \beta_0 + \beta_1 Q_i + \beta_2 X_{2i} + \varepsilon_i$

Nonlinear Transformations are NOT Link Functions

X or Y

Transformation:

$W_i = \text{Log}(Y_i)$

$\text{Log}(Y_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$

μ_Y only

Link Function:

$f(\mu) = \text{Log}(\mu_{Yi})$

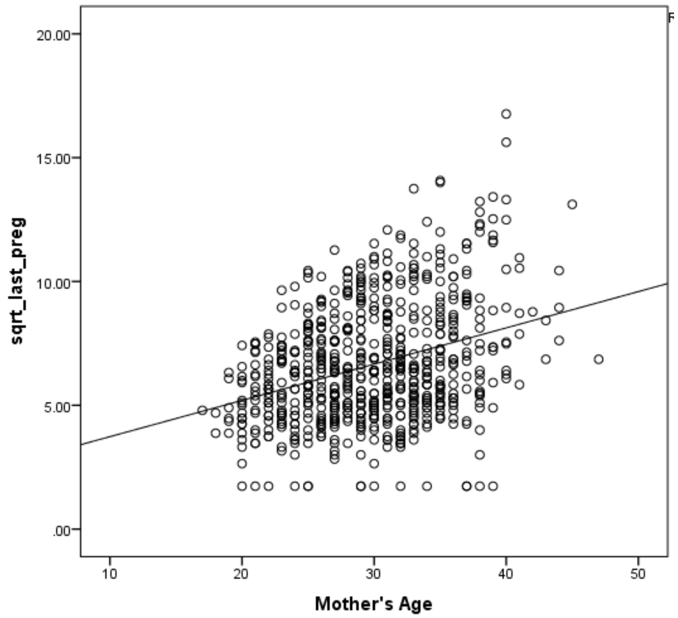
$\text{Log}(\mu_{Yi}) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}$

Interpreting a Coefficient for a Square Rooted Y

Regression Coefficients

Dependent Variable: sqrt_last_preg

Parameter	B	se	t	p
Intercept	2.004	.562	3.566	.000
mage	.145	.016	9.265	.000
pre_wgt	.002	.002	.877	.381

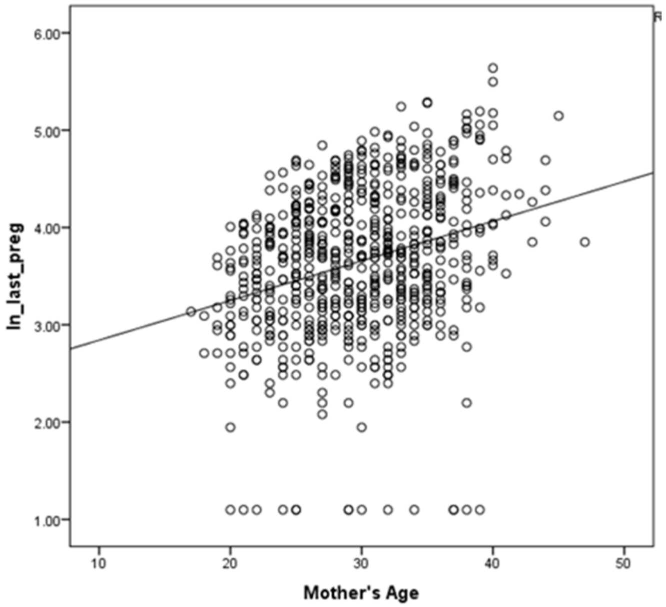


Interpreting a Coefficient for a Log Transformed Y

Regression Coefficients

Dependent Variable: ln_last_preg

Parameter	B	se	t	p
Intercept	2.383	.181	13.164	.000
mage	.041	.005	8.026	.000
pre_wgt	.000	.001	.530	.596



Converting Back to the Raw Scale

Square Roots

$$b_1 = \frac{E(\sqrt{Y})_{x+1} - E(\sqrt{Y})_x}{1}$$

Logarithms

$$b_1 = \frac{E(log\ Y)_{x+1} - E(\log Y)_x}{1}$$

How Logarithms Work

$10^1=10$ $\quad$ $\log_{10}(10)=1$ $\quad$ $e^1=2.72$ $\quad$ $\log_e(2.72)=1$

$10^2=100$ $\quad$ $\log_{10}(100)=2$ $\quad$ $e^2=7.40$ $\quad$ $\log_e(7.40)=2$

$10^3=1,000$ $\quad$ $\log_{10}(1000)=3$ $\quad$ $e^3=20.12$ $\quad$ $\log_e(20.12)=3$

$10^4=10,000$ $\quad$ $\log_{10}(10,000)=4$ $\quad$ $e^4=54.74$ $\quad$ $\log_e(54.74)=4$

$\log(A) + \log(B) = \log(AB)$

$\log(A) - \log(B) = \log(A/B)$

Multiplicative relationship on raw scale

Additive relationship on log scale

When Y is logged

$$b_1 = \frac{E\Delta(log\ Y)}{1\ unit\ \Delta\ in\ X}$$

Difference on log Y scale
Ratio on Y scale

Difference on X scale

When Y is logged

Regression Coefficients

Dependent Variable: ln_last_preg

Parameter	B	se	t	p
Intercept	2.383	.181	13.164	.000
mage	.041	.005	8.026	.000
pre_wgt	.000	.001	.530	.596

Each one-unit difference in X multiplies the expected value of Y by e^b

$$e^{.041} = 1.042$$

When Y is logged: Interpret as a % change

Regression Coefficients

Dependent Variable: ln_last_preg

Parameter	B	se	t	p
Intercept	2.383	.181	13.164	.000
mage	.041	.005	8.026	.000
pre_wgt	.000	.001	.530	.596

$$e^{.041} = 1.042$$

$100(e^b – 1)$ = % change in Y for 1 unit difference in X

= $100*(e^{.041} – 1)$
= 100*(1.042 – 1)
= 4.2%

When X is logged

$$b_1 = \frac{E\Delta(Y)}{1\ unit\ \Delta\ in\ logX}$$

Difference on Y scale

Difference on log X scale
Ratio on X scale

When X is logged: Interpret as % change

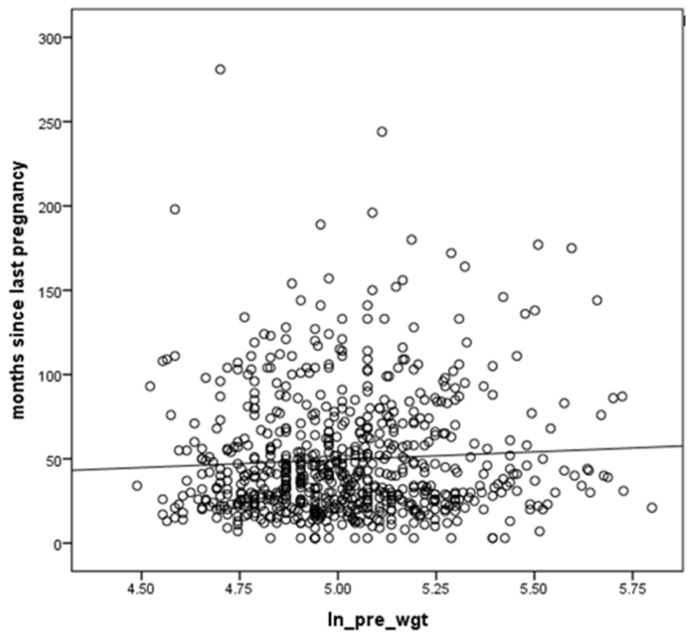
E(ΔY) for a p% increase in X = b*log[(100 + p)/100]

When the % difference in X is small:
b/100 = E(ΔY) for each 1% difference in X

Regression Coefficients

Dependent Variable: last_preg

Variable	B	se	t	p
Intercept	-41.621	28.305	-1.470	.142
mage	2.314	.239	9.696	.000
ln_pre_wgt	4.461	5.580	.799	.424



When both Y and X are logged

b expresses % change in Y for 1% increase in X

Regression Coefficients

Dependent Variable: ln_last_preg

Variable	B	se	t	p
Intercept	2.240	.600	3.736	.000
mage	.041	.005	8.032	.000
ln_pre_wgt	.040	.118	.335	.738