Nonlinear Relationships

EC 320: Introduction to Econometrics

Kyle Raze
Fall 2019

Prologue

Housekeeping

Final Exam

Review lecture this Wednesday.

- Come prepared with questions.

Exam: Tuesday, December 10 at 10:15am in Chapman 220.

Office hours TBA for Monday, December 9.

Problem Set 5

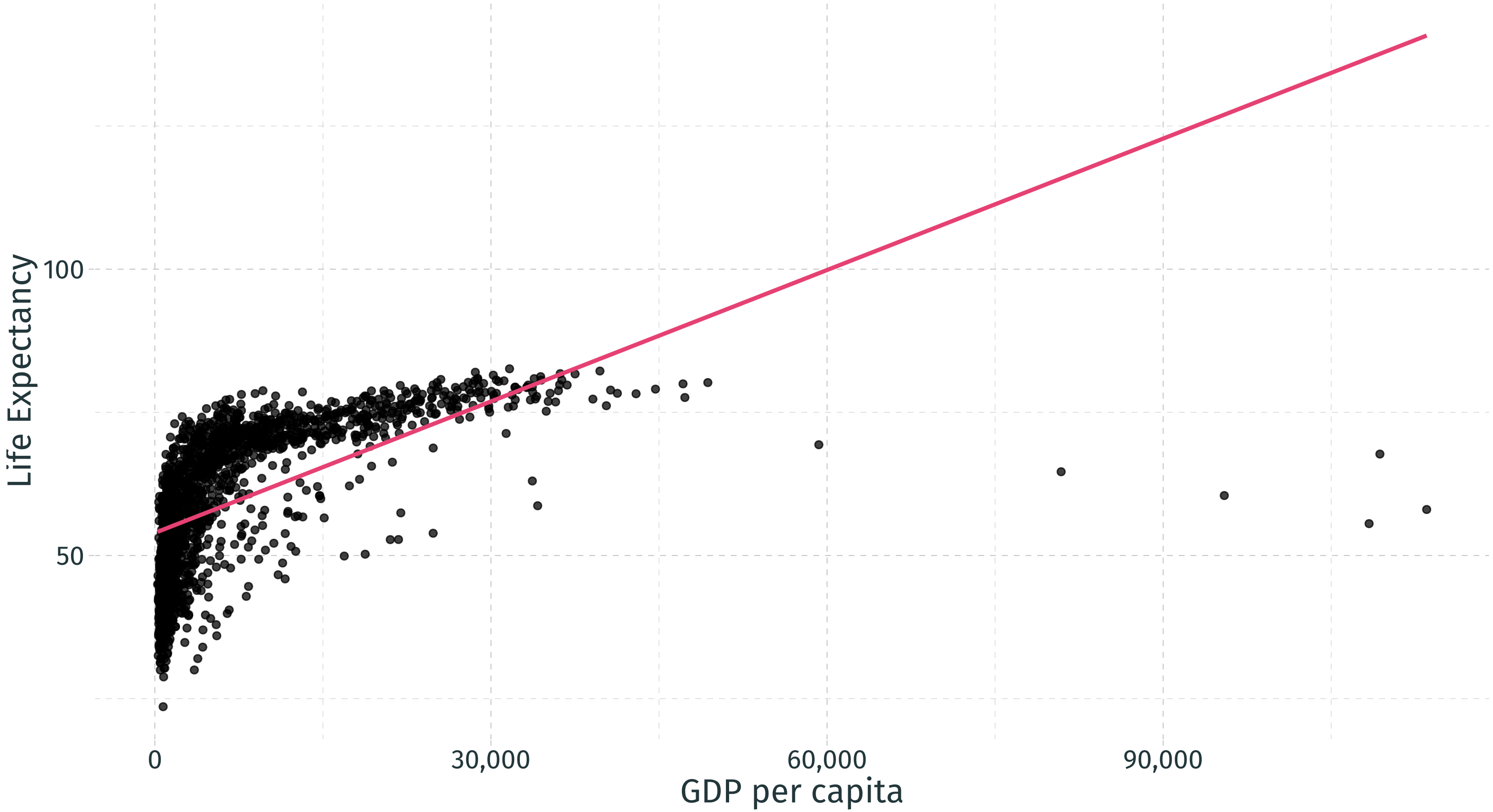
Due Saturday, December 7 by 11:59pm.

- I will post the key immediately after.

Nonlinear Relationships

Can We Do Better?

$$\widehat{(\text{Life Expectancy})}_i = 53.96 + 8 \times 10^{-4} \cdot \text{GDP}_i$$



Nonlinear Relationships

Many economic relationships are **nonlinear**.

- *e.g.*, most production functions, profit, diminishing marginal utility, tax revenue as a function of the tax rate, *etc.*

The flexibility of OLS

OLS can accommodate many, but not all, nonlinear relationships.

- Underlying model must be linear-in-parameters.
- Nonlinear transformations of variables are okay.
- Modeling some nonlinear relationships requires advanced estimation techniques, such as *maximum likelihood*.[†]

† Beyond the scope of this class.

Linearity

Linear-in-parameters: Parameters enter model as a weighted sum, where the weights are functions of the variables.

- One of the assumptions required for the unbiasedness of OLS.

Linear-in-variables: Variables enter the model as a weighted sum, where the weights are functions of the parameters.

- Not required for the unbiasedness of OLS.

The standard linear regression model satisfies both properties:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + u_i$$

Linearity

Which of the following is **linear-in-parameters**, **linear-in-variables**, or **neither**?

1. $Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \cdots + \beta_k X_i^k + u_i$

2. $Y_i = \beta_0 X_i^{\beta_1} v_i$

3. $Y_i = \beta_0 + \beta_1 \beta_2 X_i + u_i$

Model 1 is linear-in-parameters, but not linear-in-variables.

Model 2 is neither.

Model 3 is linear-in-variables, but not linear-in-parameters.

We're Going to Take Logs

The natural log is the inverse function for the exponential function:

$\log(e^x) = x$ for $x > 0$.

(Natural) Log Rules

1. Product rule: $\log(AB) = \log(A) + \log(B)$.
2. Quotient rule: $\log(A/B) = \log(A) - \log(B)$.
3. Power rule: $\log(A^B) = B \cdot \log(A)$.
4. Derivative: $f(x) = \log(x) \Rightarrow f'(x) = \dfrac{1}{x}$.
5. $\log(e) = 1$, $\log(1) = 0$, and $\log(x)$ is undefined for $x \leq 0$.

Log-Linear Model

Nonlinear Model

$$Y_i = \alpha e^{\beta_1 X_i} v_i$$

- $Y > 0$, X is continuous, and v_i is a multiplicative error term.
- Cannot estimate parameters with OLS directly.

Logarithmic Transformation

$$\log(Y_i) = \log(\alpha) + \beta_1 X_i + \log(v_i)$$

- Redefine $\log(\alpha) \equiv \beta_0$ and $\log(v_i) \equiv u_i$.

Transformed (Linear) Model

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

- *Can* estimate with OLS, but coefficient interpretation changes.

Log-Linear Model

Regression Model

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

Interpretation

- A one-unit increase in the explanatory variable increases the outcome variable by approximately $\beta_1 \times 100$ percent, on average.
- *Example:* If $\log(\hat{\text{Pay}}_i) = 2.9 + 0.03 \cdot \text{School}_i$, then an additional year of schooling increases pay by approximately 3 percent, on average.

Log-Linear Model

Derivation

Consider the log-linear model

$$\log(Y) = \beta_0 + \beta_1 \, X + u$$

and differentiate

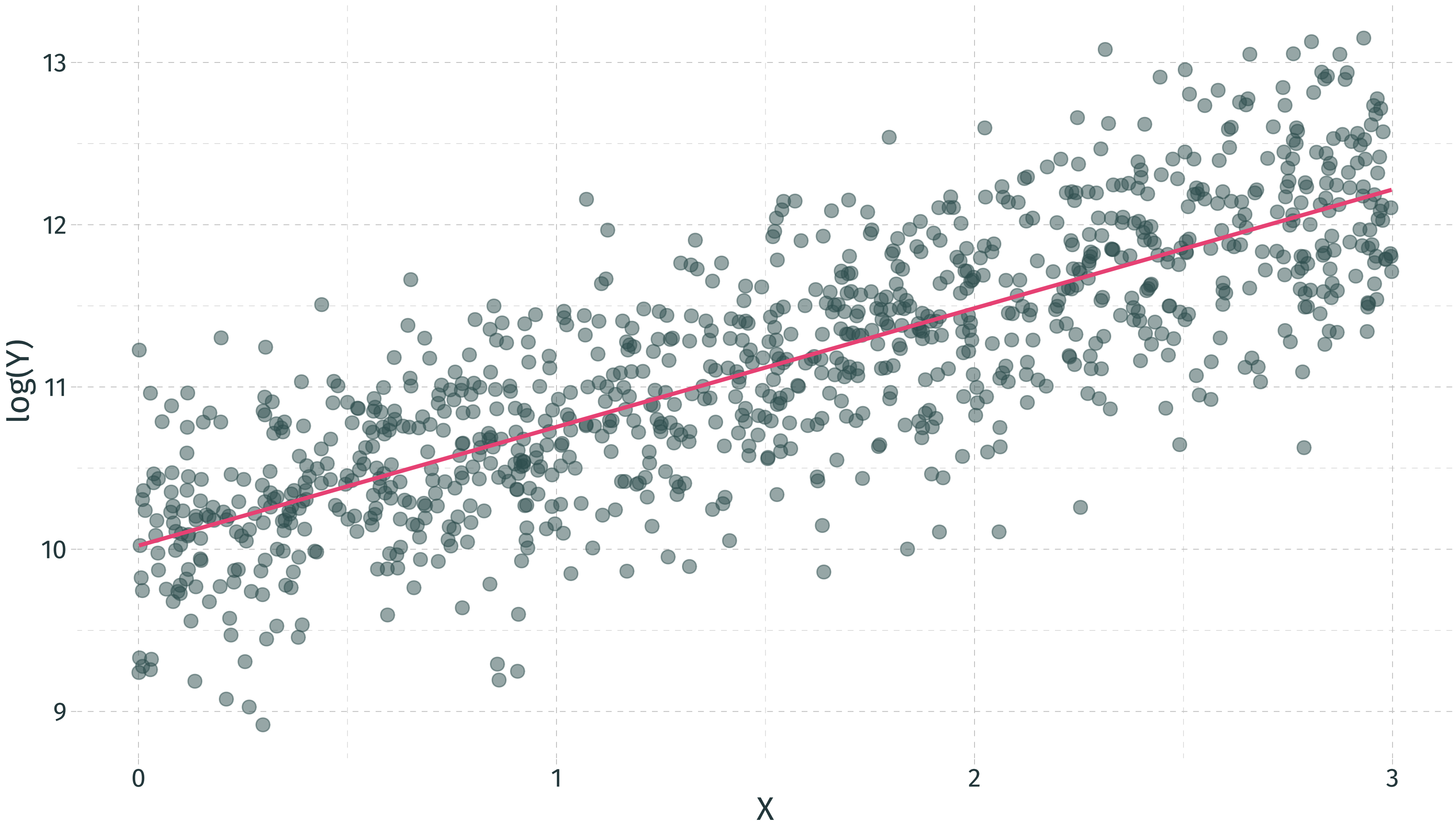
$$\frac{dY}{Y} = \beta_1 dX$$

A marginal (small) change in X (*i.e.*, dX) leads to a $\beta_1 dX$ **proportionate change** in Y.

- Multiply by 100 to get the **percentage change** in Y.

Log-Linear Example

$$\log(\hat{Y}_i) = 10.02 + 0.73 \cdot \mathrm{X}_i$$



Log-Linear Model

Note: If you have a log-linear model with a binary indicator variable, the interpretation of the coefficient on that variable changes.

Consider

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

for binary variable X.

Interpretation of β_1:

- When X changes from 0 to 1, Y will increase by $100 \times \;\; e^{\beta_1} - 1$ percent.
- When X changes from 1 to 0, Y will decrease by $100 \times \;\; e^{-\beta_1} - 1$ percent.

Log-Linear Example

Binary explanatory variable: `trained`

- `trained == 1` if employee received training.
- `trained == 0` if employee did not receive training.

```
lm(log(productivity) ~ trained, data = df2) %>% tidy()
```

```
#> # A tibble: 2 x 5
#>   term        estimate std.error statistic  p.value
#>   <chr>          <dbl>     <dbl>     <dbl>    <dbl>
#> 1 (Intercept)    9.94     0.0446    223.   0.
#> 2 trained        0.557    0.0631      8.83 4.72e-18
```

Q: How do we interpret the coefficient on `trained`?

A_1: Trained workers 64.2 percent more productive than untrained workers.

A_2: Untrained workers 21.08 percent less productive than trained workers.

Log-Log Model

Nonlinear Model

$$Y_i = \alpha X_i^{\beta_1} v_i$$

- $Y > 0$, $X > 0$, and v_i is a multiplicative error term.
- Cannot estimate parameters with OLS directly.

Logarithmic Transformation

$$\log(Y_i) = \log(\alpha) + \beta_1 \log(X_i) + \log(v_i)$$

- Redefine $\log(\alpha) \equiv \beta_0$ and $\log(v_i) \equiv u_i$.

Transformed (Linear) Model

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

- *Can* estimate with OLS, but coefficient interpretation changes.

Log-Log Model

Regression Model

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

Interpretation

- A one-percent increase in the explanatory variable leads to a β_1-percent change in the outcome variable, on average.
- Often interpreted as an elasticity.
- *Example:* If $\widehat{\log(\text{Quantity Demanded}_i)} = 0.45 - 0.31 \cdot \log(\text{Income}_i)$, then each one-percent increase in income decreases quantity demanded by 0.31 percent.

Log-Log Model

Derivation

Consider the log-log model

$$\log(Y_i) = \beta_0 + \beta_1 \log(X_i) + u$$

and differentiate

$$\frac{dY}{Y} = \beta_1 \frac{dX}{X}$$

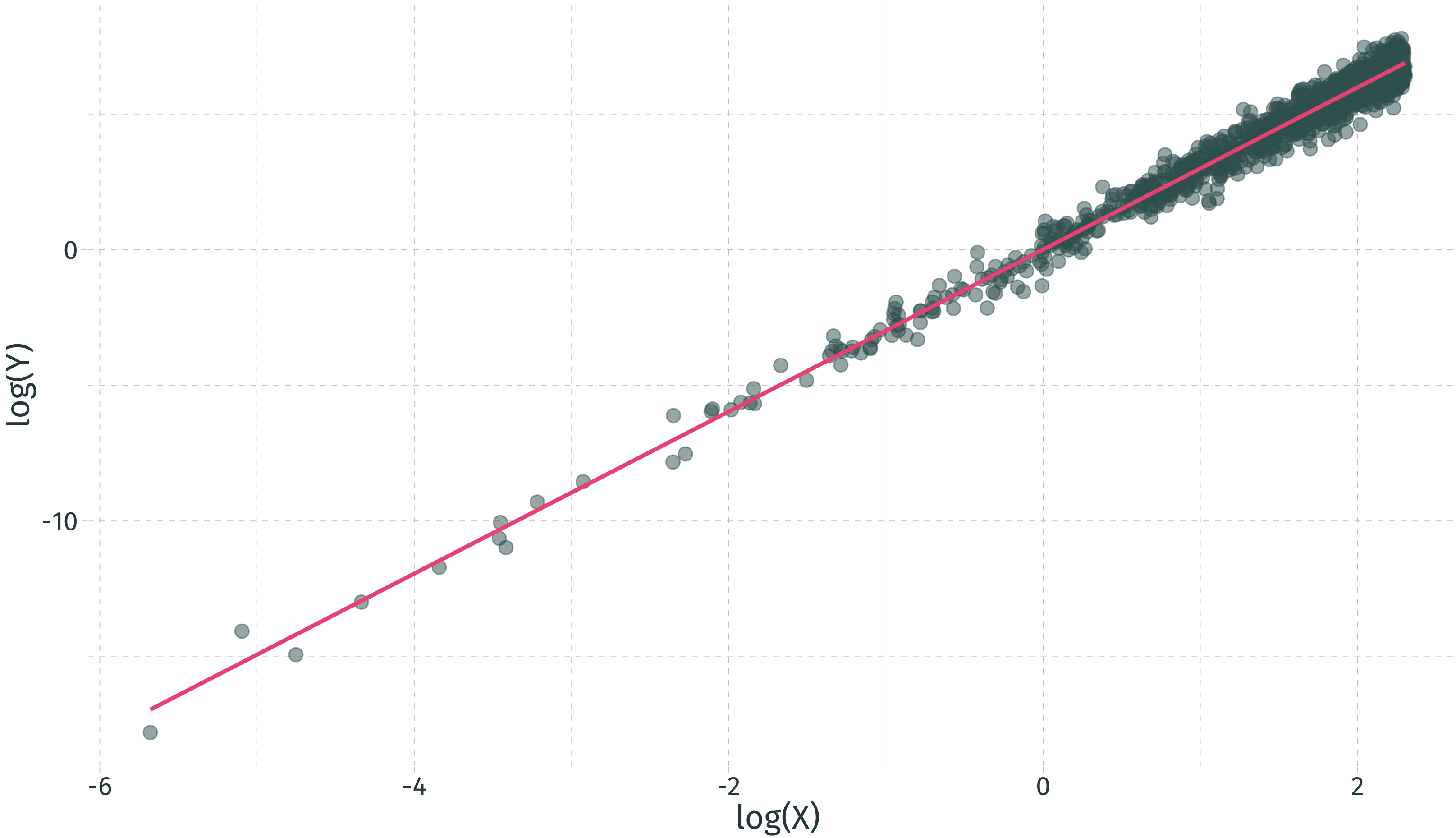
A one-percent increase in X leads to a β_1-percent increase in Y.

- Rearrange to show elasticity interpretation:

$$\frac{dY}{dX}\frac{X}{Y} = \beta_1$$

Log-Log Example

$$\log(\hat{Y}_i) = 0.01 + 2.99 \cdot \log(\mathrm{X}_i)$$



Linear-Log Model

Nonlinear Model

$$e^{Y_i} = \alpha X_i^{\beta_1} v_i$$

- $X > 0$ and v_i is a multiplicative error term.
- Cannot estimate parameters with OLS directly.

Logarithmic Transformation

$$Y_i = \log(\alpha) + \beta_1 \log(X_i) + \log(v_i)$$

- Redefine $\log(\alpha) \equiv \beta_0$ and $\log(v_i) \equiv u_i$.

Transformed (Linear) Model

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

- *Can* estimate with OLS, but coefficient interpretation changes.

Linear-Log Model

Regression Model

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

Interpretation

- A one-percent increase in the explanatory variable increases the outcome variable by approximately $\beta_1 \div 100$, on average.
- *Example:* If $(\hat{\text{Blood Pressure}})_i = 150 - 9.1 \log(\text{Income}_i)$, then a one-percent increase in income decrease blood pressure by 0.091 points.

Linear-Log Model

Derivation

Consider the log-linear model

$$Y = \beta_0 + \beta_1 \log(X) + u$$

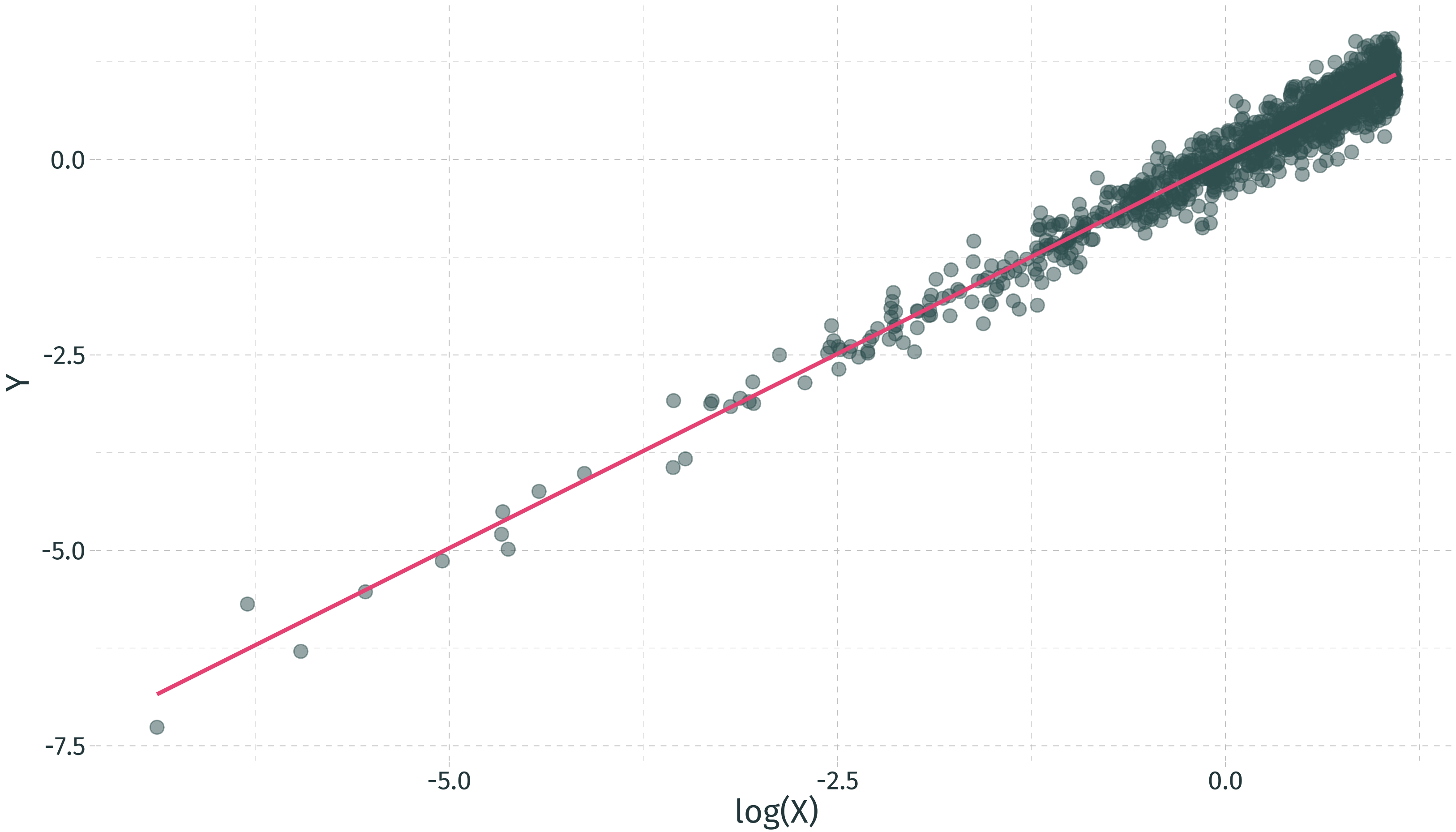
and differentiate

$$dY = \beta_1 \frac{dX}{X}$$

A one-percent increase in X leads to a $\beta_1 \div 100$ **change** in Y.

Linear-Log Example

$$\hat{Y}_i = 0 + 0.99 \cdot \log(\mathrm{X}_i)$$

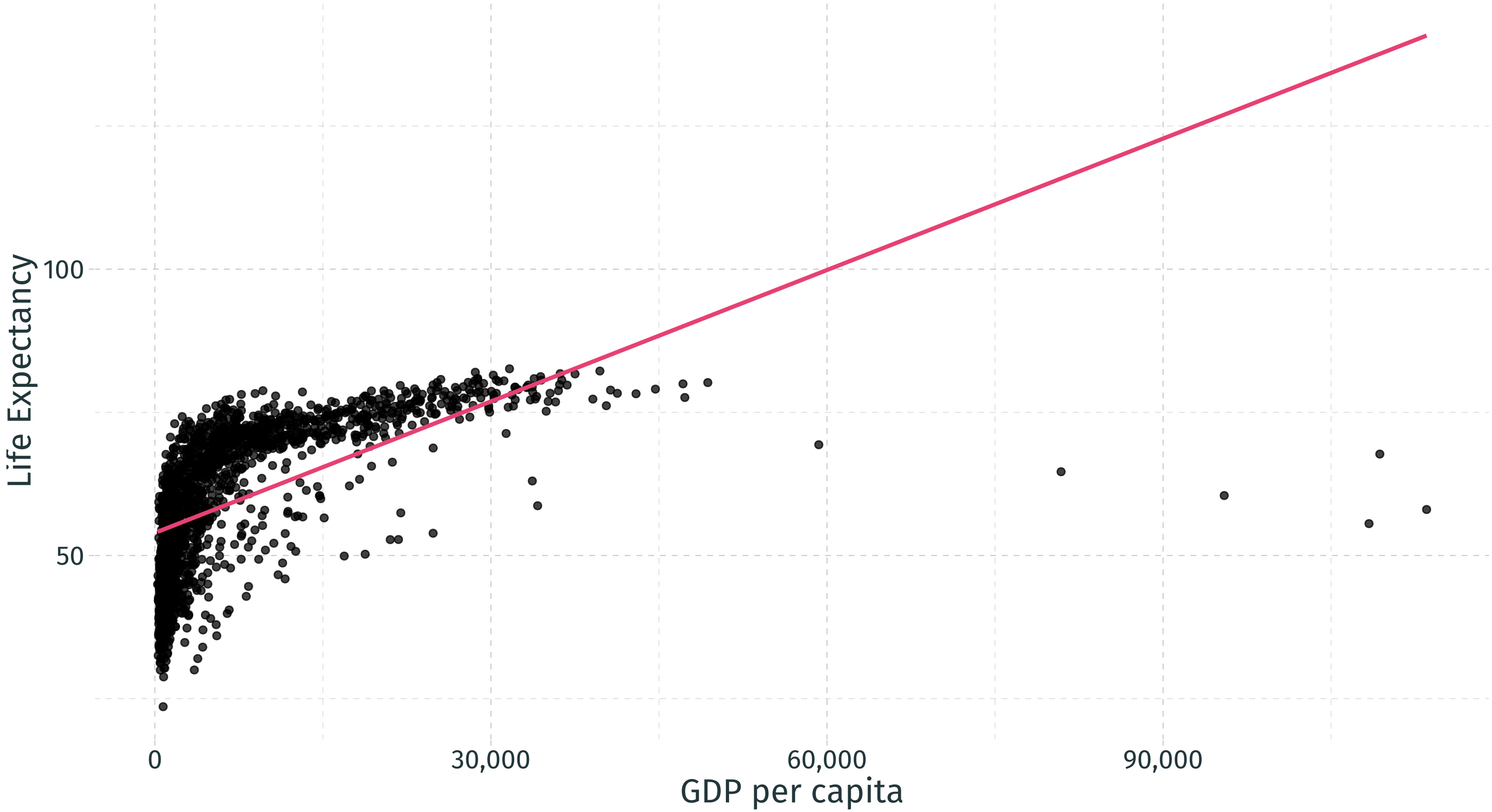


(Approximate) Coefficient Interpretation

Model	β_1 **Interpretation**
Level-level $Y_i = \beta_0 + \beta_1 X_i + u_i$	$\Delta Y = \beta_1 \cdot \Delta X$ *A one-unit increase in X leads to a β_1-unit increase in Y*
Log-level $\log(Y_i) = \beta_0 + \beta_1 X_i + u_i$	$\%\Delta Y = 100 \cdot \beta_1 \cdot \Delta X$ *A one-unit increase in X leads to a $\beta_1 \cdot 100$-percent increase in Y*
Log-log $\log(Y_i) = \beta_0 + \beta_1 \log(X_i) + u_i$	$\%\Delta Y = \beta_1 \cdot \%\Delta X$ *A one-percent increase in X leads to a β_1-percent increase in Y*
Level-log $Y_i = \beta_0 + \beta_1 \log(X_i) + u_i$	$\Delta Y = (\beta_1 \div 100) \cdot \%\Delta X$ *A one-percent increase in X leads to a $\beta_1 \div 100$-unit increase in Y*

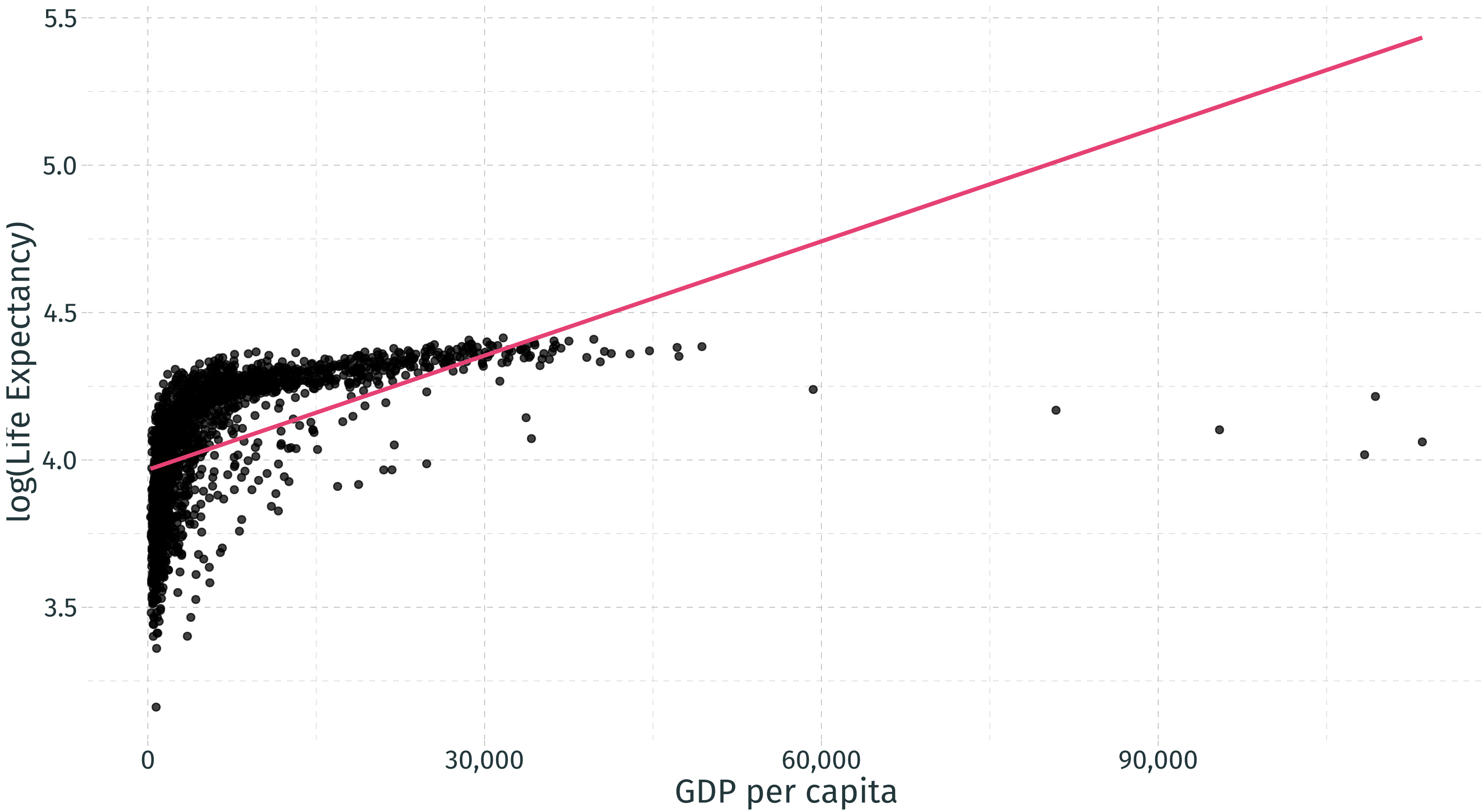
Can We Do Better?

$$\widehat{(\text{Life Expectancy})}_i = 53.96 + 8 \times 10^{-4} \cdot \text{GDP}_i \qquad R^2 = 0.34$$



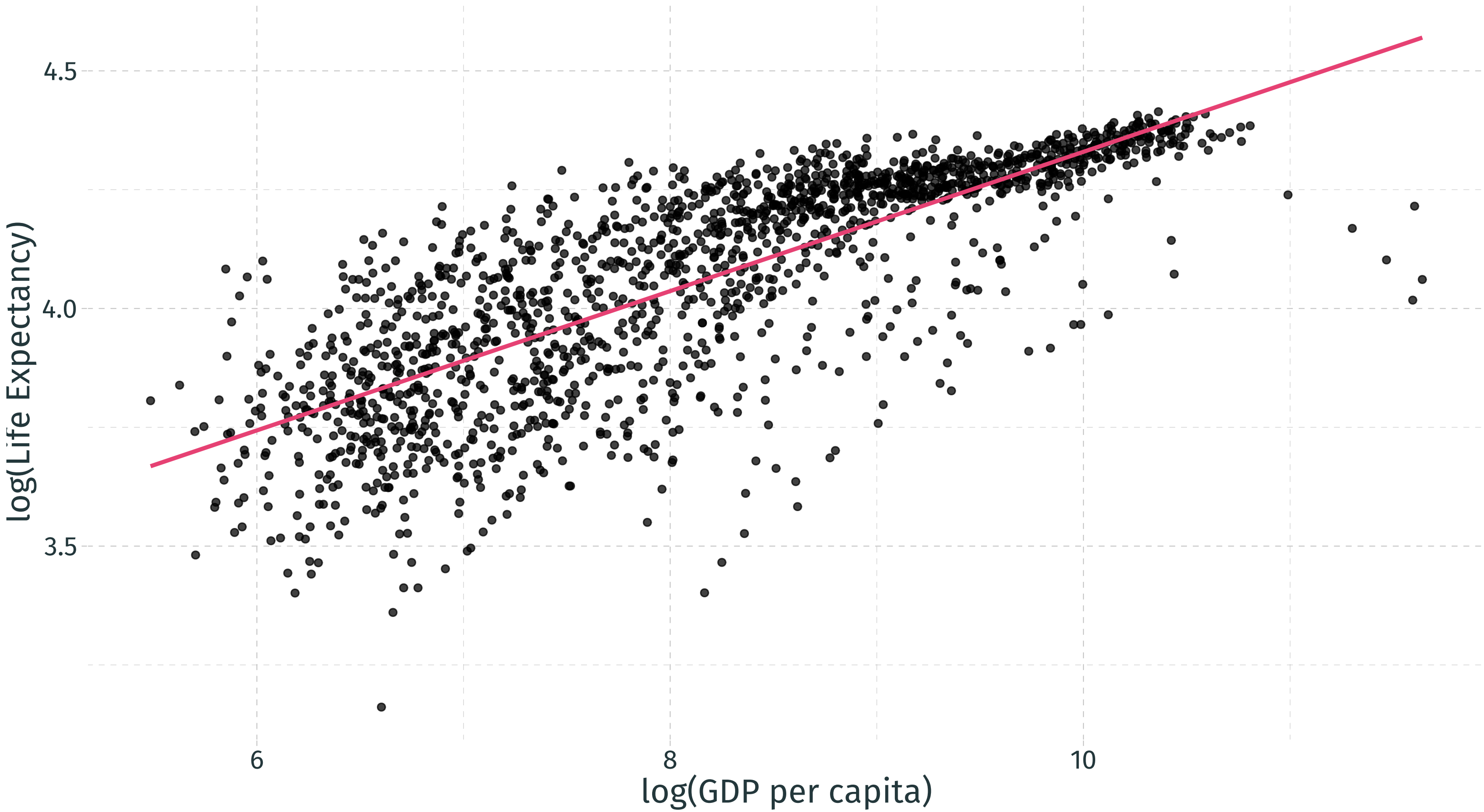
Can We Do Better?

$$\log(\widehat{\text{Life Expectancy}_i}) = 3.97 + 1.3 \times 10^{-5} \cdot \text{GDP}_i \qquad R^2 = 0.3$$



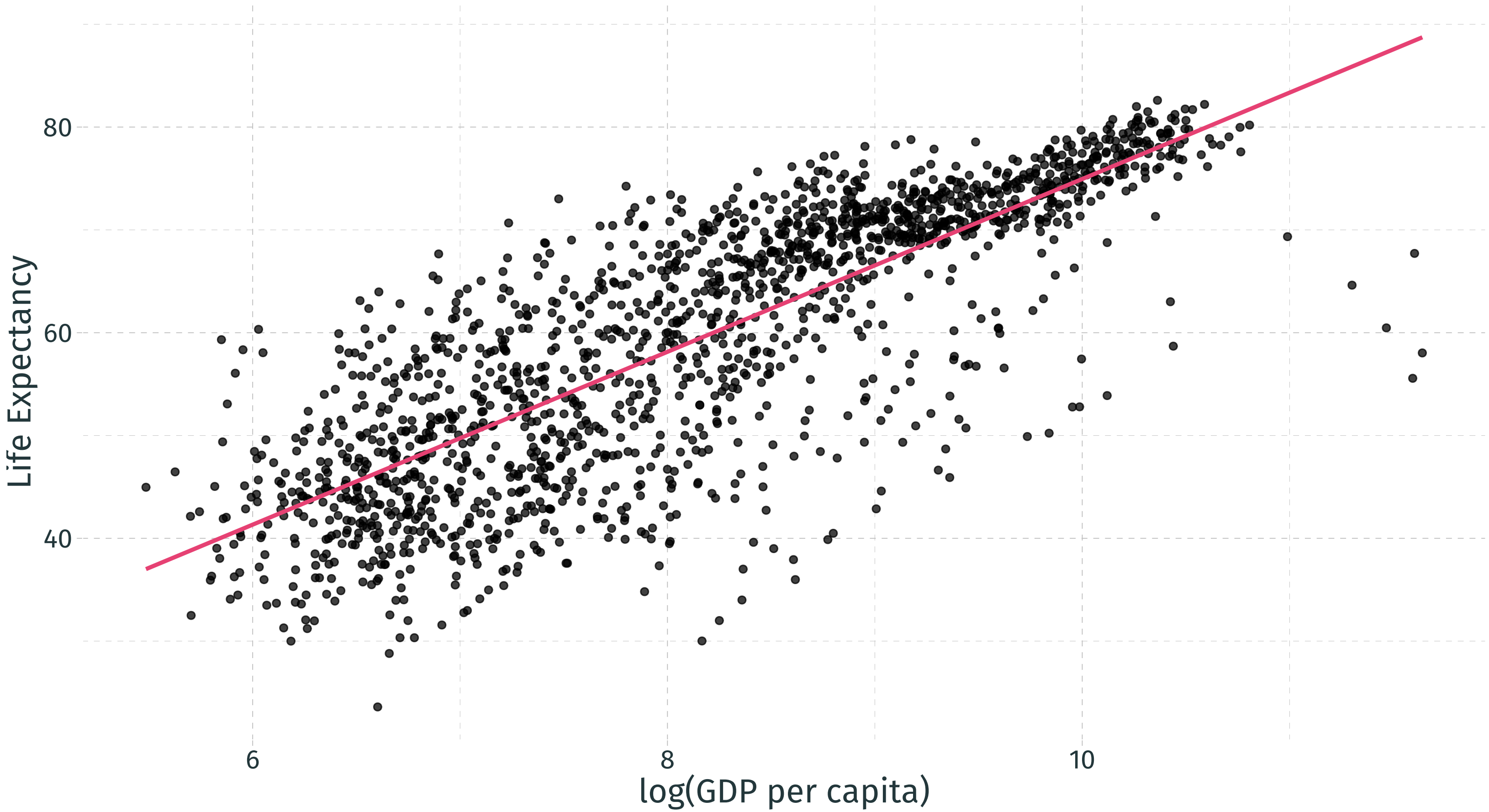
Can We Do Better?

$$\log(\widehat{\text{Life Expectancy}}_i) = 2.86 + 0.15 \cdot \log(\text{GDP}_i) \qquad R^2 = 0.61$$



Can We Do Better?

$$\widehat{(\text{Life Expectancy})}_i = -9.1 + 8.41 \cdot \log(\text{GDP}_i) \qquad R^2 = 0.65$$



Practical Considerations

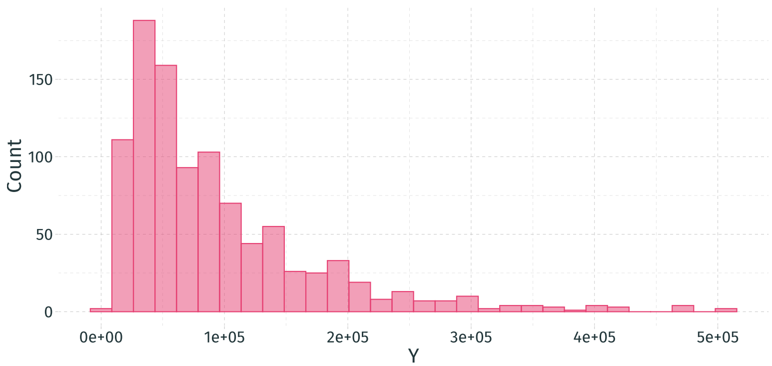
Consideration 1: Do your data take negative numbers or zeros as values?

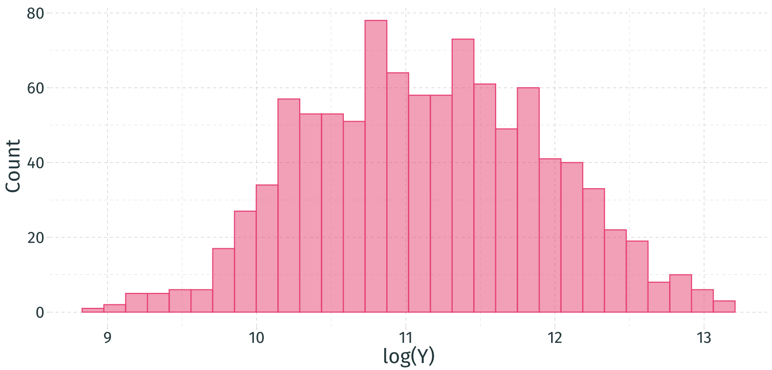
```
log(0)
```

```
#> [1] -Inf
```

Consideration 2: What coefficient interpretation do you want? Unit change? Unit-free percent change?

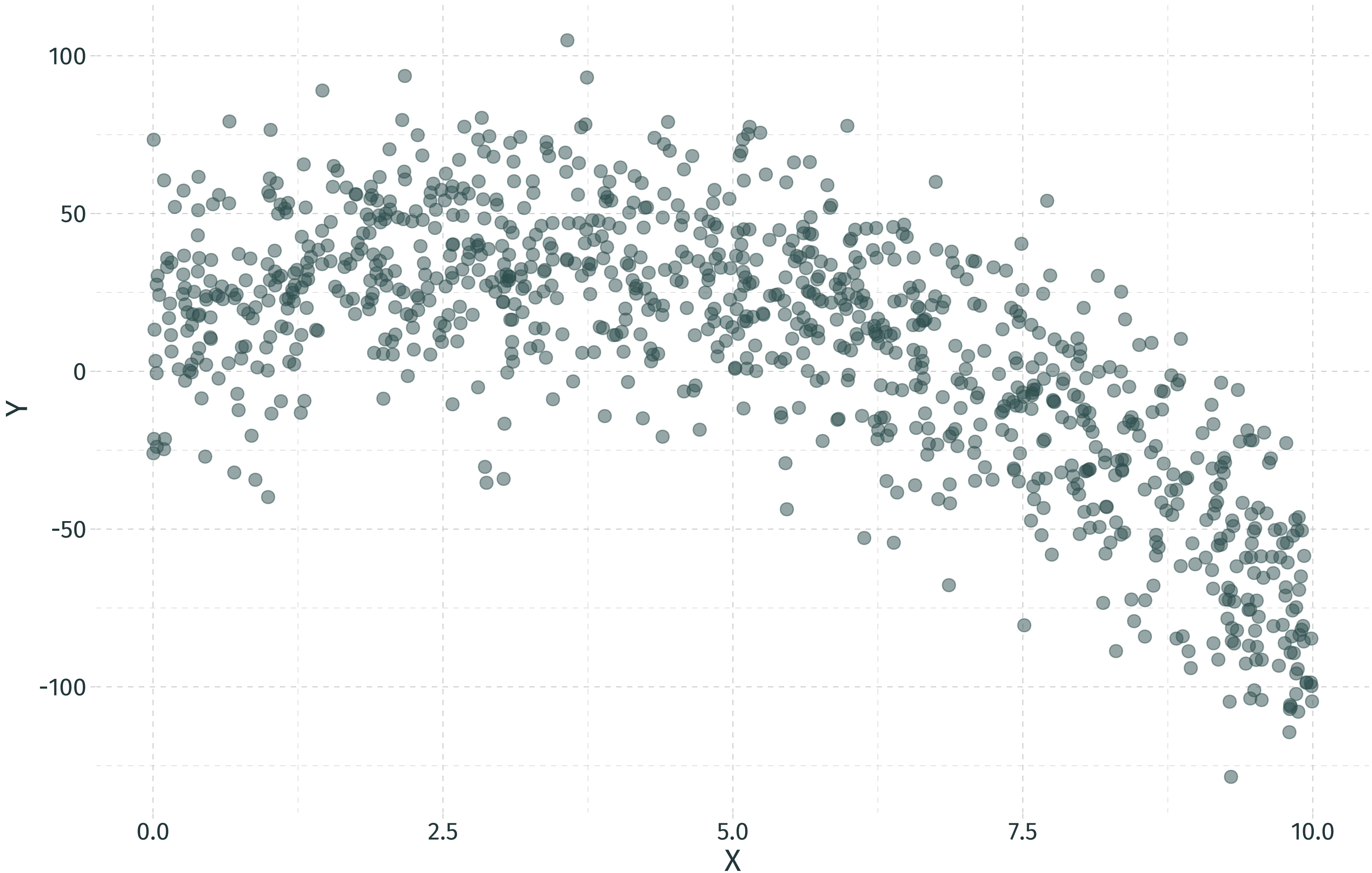
Consideration 3: Are your data skewed?





Quadratic Regression

Quadratic Data



Quadratic Regression

Regression Model

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

Interpretation

Sign of β_2 indicates whether the relationship is convex (+) or concave (-)

Sign of β_1?🤷‍♀️

Partial derivative of Y with respect to X is the **marginal effect** of X on Y:

$$\frac{\partial Y}{\partial X} = \beta_1 + 2\beta_2 X$$

- Effect of X depends on the level of X

Quadratic Regression

```
lm(y ~ x + I(x^2), data = quad_df) %>% tidy()
```

```
#> # A tibble: 3 x 5
#>   term        estimate std.error statistic   p.value
#>   <chr>          <dbl>     <dbl>     <dbl>     <dbl>
#> 1 (Intercept)    13.2     2.26        5.81 8.30e-  9
#> 2 x              15.7     1.03       15.3  1.99e- 47
#> 3 I(x^2)         -2.50    0.0982    -25.4  2.46e-110
```

What is the marginal effect of X on Y?

$$\frac{\partial \mathrm{Y}}{\partial \mathrm{X}} = \hat{\beta}_1 + 2\hat{\beta}_2 X = 15.69 + -4.99X$$

Quadratic Regression

```
lm(y ~ x + I(x^2), data = quad_df) %>% tidy()
```

```
#> # A tibble: 3 x 5
#>   term        estimate std.error statistic   p.value
#>   <chr>          <dbl>     <dbl>     <dbl>     <dbl>
#> 1 (Intercept)    13.2     2.26        5.81 8.30e-  9
#> 2 x              15.7     1.03       15.3  1.99e- 47
#> 3 I(x^2)         -2.50    0.0982    -25.4  2.46e-110
```

What is the marginal effect of X on Y when $X = 0$?

$$\left.\frac{\partial \mathrm{Y}}{\partial \mathrm{X}}\right|_{\mathrm{X}=0} = \hat{\beta}_1 = \mathbf{15.69}$$

Quadratic Regression

```
lm(y ~ x + I(x^2), data = quad_df) %>% tidy()
```

```
#> # A tibble: 3 x 5
#>   term        estimate std.error statistic   p.value
#>   <chr>          <dbl>     <dbl>     <dbl>     <dbl>
#> 1 (Intercept)    13.2      2.26         5.81 8.30e-  9
#> 2 x              15.7      1.03        15.3  1.99e- 47
#> 3 I(x^2)         -2.50     0.0982     -25.4  2.46e-110
```

What is the marginal effect of X on Y when $X = 2$?

$$\left.\frac{\partial \mathrm{Y}}{\partial \mathrm{X}}\right|_{\mathrm{X}=2} = \hat{\beta}_1 + 2\hat{\beta}_2 \cdot (2) = 15.69 - 9.99 = 5.71$$

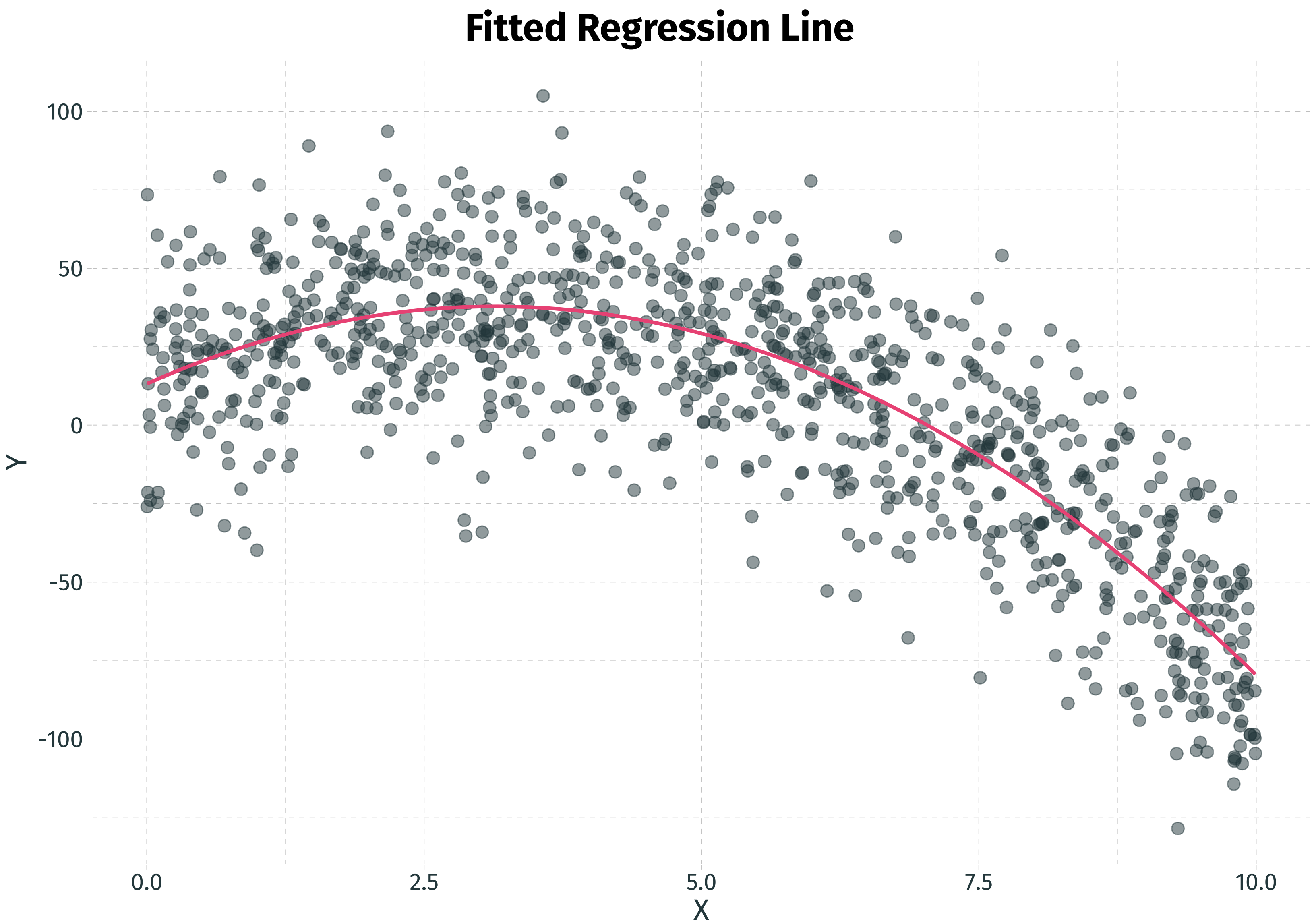
Quadratic Regression

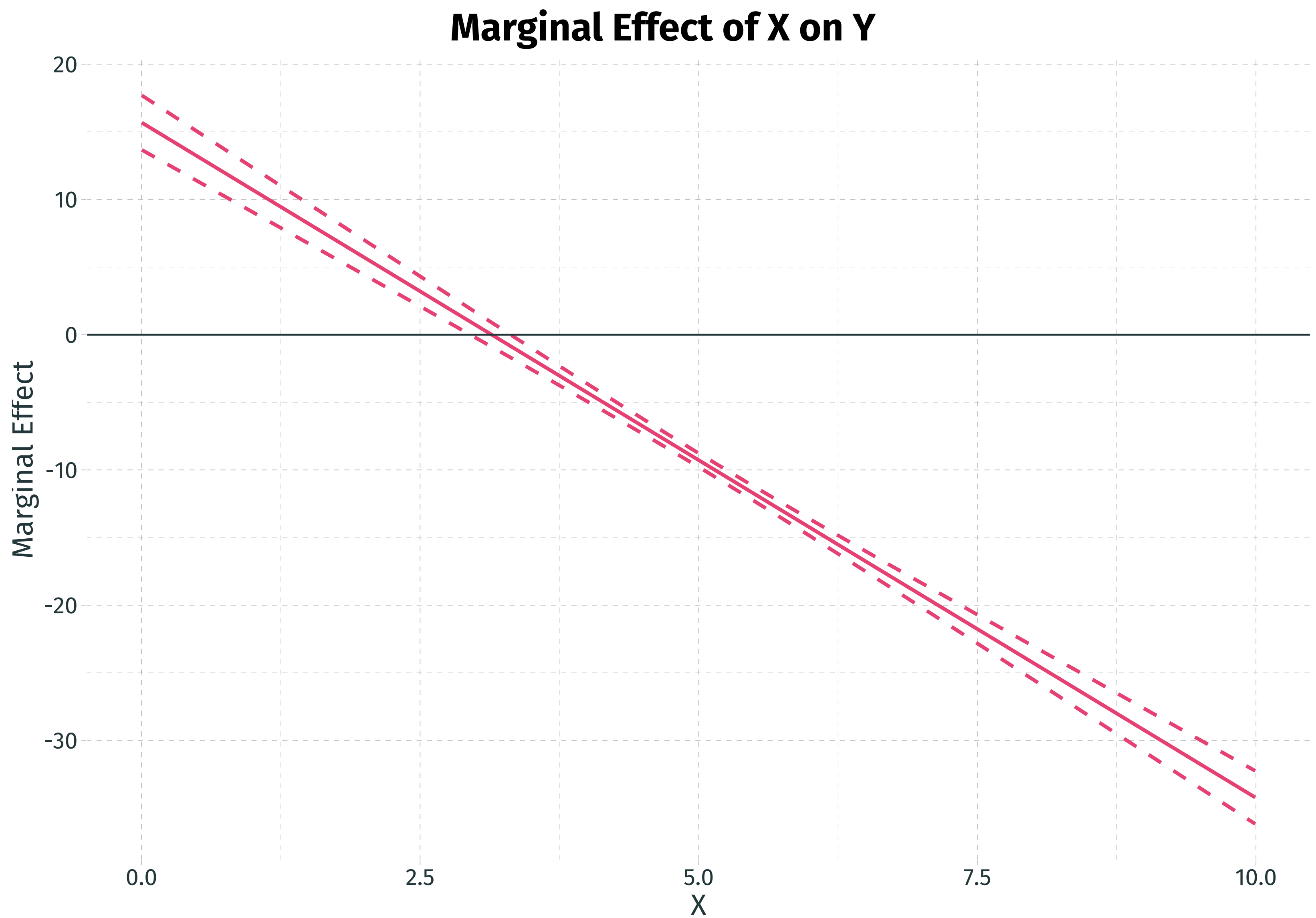
```
lm(y ~ x + I(x^2), data = quad_df) %>% tidy()
```

```
#> # A tibble: 3 x 5
#>   term        estimate std.error statistic   p.value
#>   <chr>          <dbl>     <dbl>     <dbl>     <dbl>
#> 1 (Intercept)    13.2      2.26        5.81 8.30e-  9
#> 2 x              15.7      1.03       15.3  1.99e- 47
#> 3 I(x^2)         -2.50     0.0982    -25.4  2.46e-110
```

What is the marginal effect of X on Y when $X = 7$?

$$\frac{\partial \mathrm{Y}}{\partial \mathrm{X}}\bigg|_{\mathrm{X}=7} = \hat{\beta}_1 + 2\hat{\beta}_2 \cdot (7) = 15.69 - 34.96 = -19.27$$


Fitted Regression Line
100
50
0
-50
-100
Y
0.0
2.5
5.0
7.5
10.0
X


Marginal Effect of X on Y
Marginal Effect
X
20
10
0
-10
-20
-30
0.0
2.5
5.0
7.5
10.0

Quadratic Regression

Where does the regression $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\beta}_2 X_i^2$ *turn*?

- In other words, where is the peak (valley) of the fitted relationship?

Step 1: Take the derivative and set equal to zero.

$$\frac{\partial \mathrm{Y}}{\partial \mathrm{X}} = \hat{\beta}_1 + 2\hat{\beta}_2 X = 0$$

Step 2: Solve for X.

$$X = -\frac{\hat{\beta}_1}{2\hat{\beta}_2}$$

Example: Peak of previous regression occurs at $X = 3.14$.

Anscombe's Quartet

Four *"identical"* regressions: Intercept = 3, Slope = 0.5, R^2 = 0.67