Univariate Regression

Econ 2560, Fall 2023

Prof. Josh Abel

(Chapters 4 and 5)

Introduction

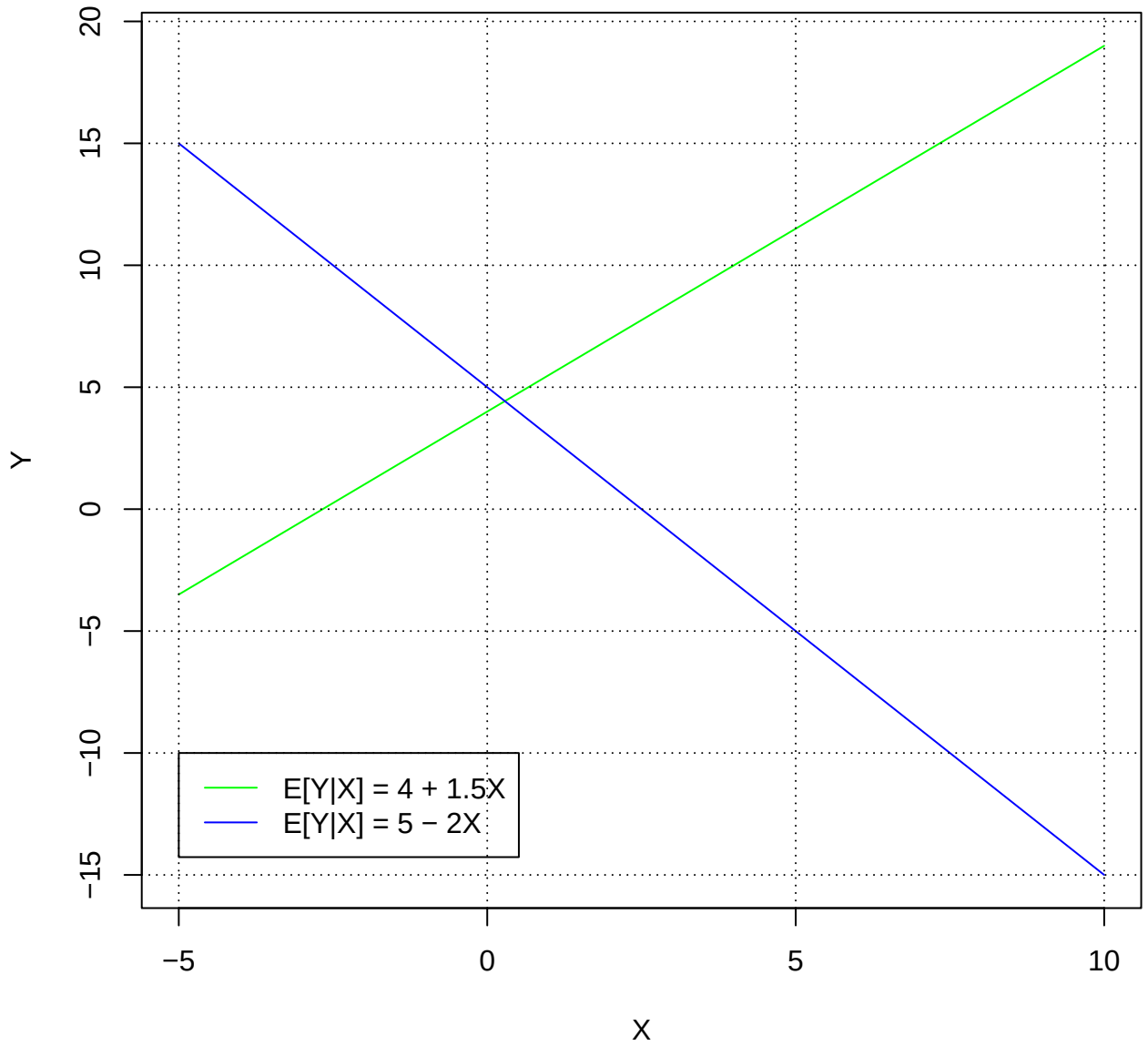
- How to estimate a conditional mean ($E[Y|X]$)?
  - E.g. Average hourly wage (Y) as a function of years of education (X)
- We previously discussed pros and cons of assuming a linear relationship
  - i.e. Assume $E[Y|X] = \beta_0 + \beta_1 \cdot X$
- Will now cover how to estimate β_0 and β_1

The linear regression function

$$E[Y_i|X_i] = \beta_0 + \beta_1 \cdot X_i$$

- Coefficients:
  - β_1 ("Beta one"): Slope, effect of 1-unit change in X
  - β_0 ("Beta zero"): Intercept, mean when $X = 0$

The linear regression function

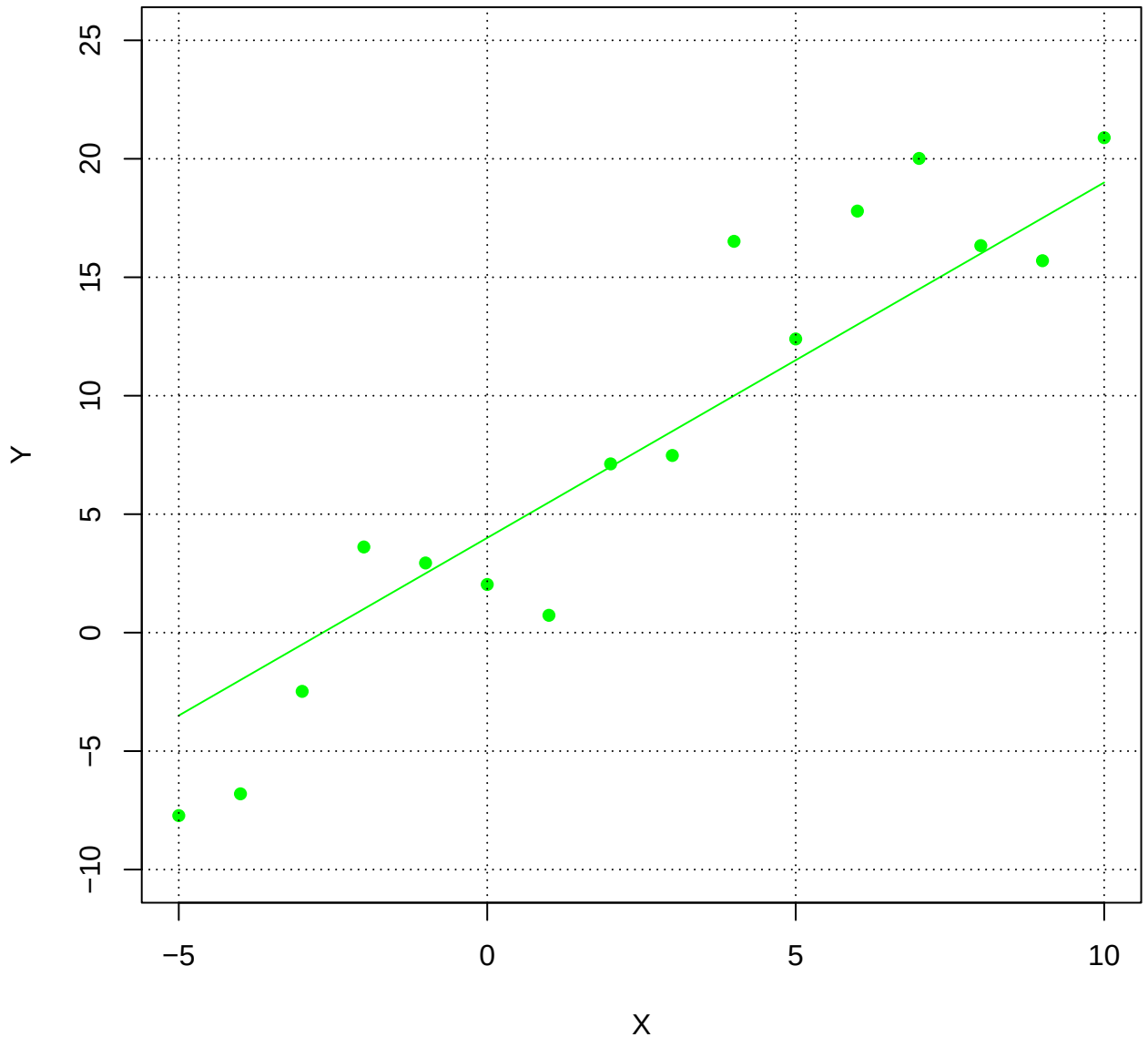


The linear regression function

$$E[Y_i|X_i] = \beta_0 + \beta_1 \cdot X_i$$

- Coefficients:
  - β_1 ("Beta one"): Slope, effect of 1-unit change in X
  - β_0 ("Beta zero"): Intercept, mean when $X = 0$
- Not all data lies on the regression line, even if truth is linear
  - Randomness/other factors also drive Y

The linear regression function



The linear regression function

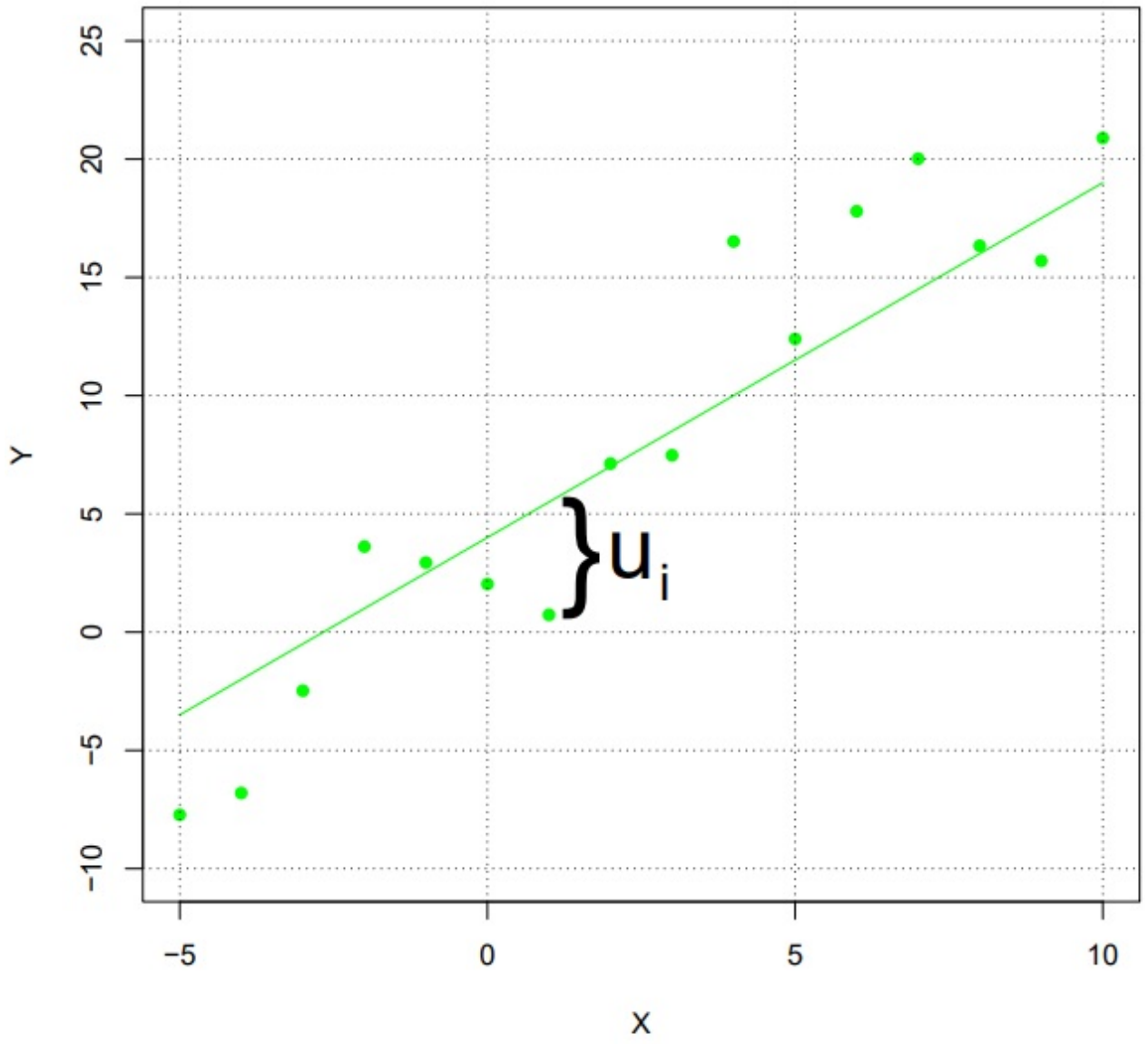
$$E[Y_i|X_i] = \beta_0 + \beta_1 \cdot X_i$$

- Coefficients:
  - β_1 ("Beta one"): Slope, effect of 1-unit change in X
  - β_0 ("Beta zero"): Intercept, mean when $X = 0$
- Not all data lies on the regression line, even if truth is linear
  - Randomness/other factors also drive Y

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

  - u_i ("u-i"): error term

The linear regression function



The linear regression function

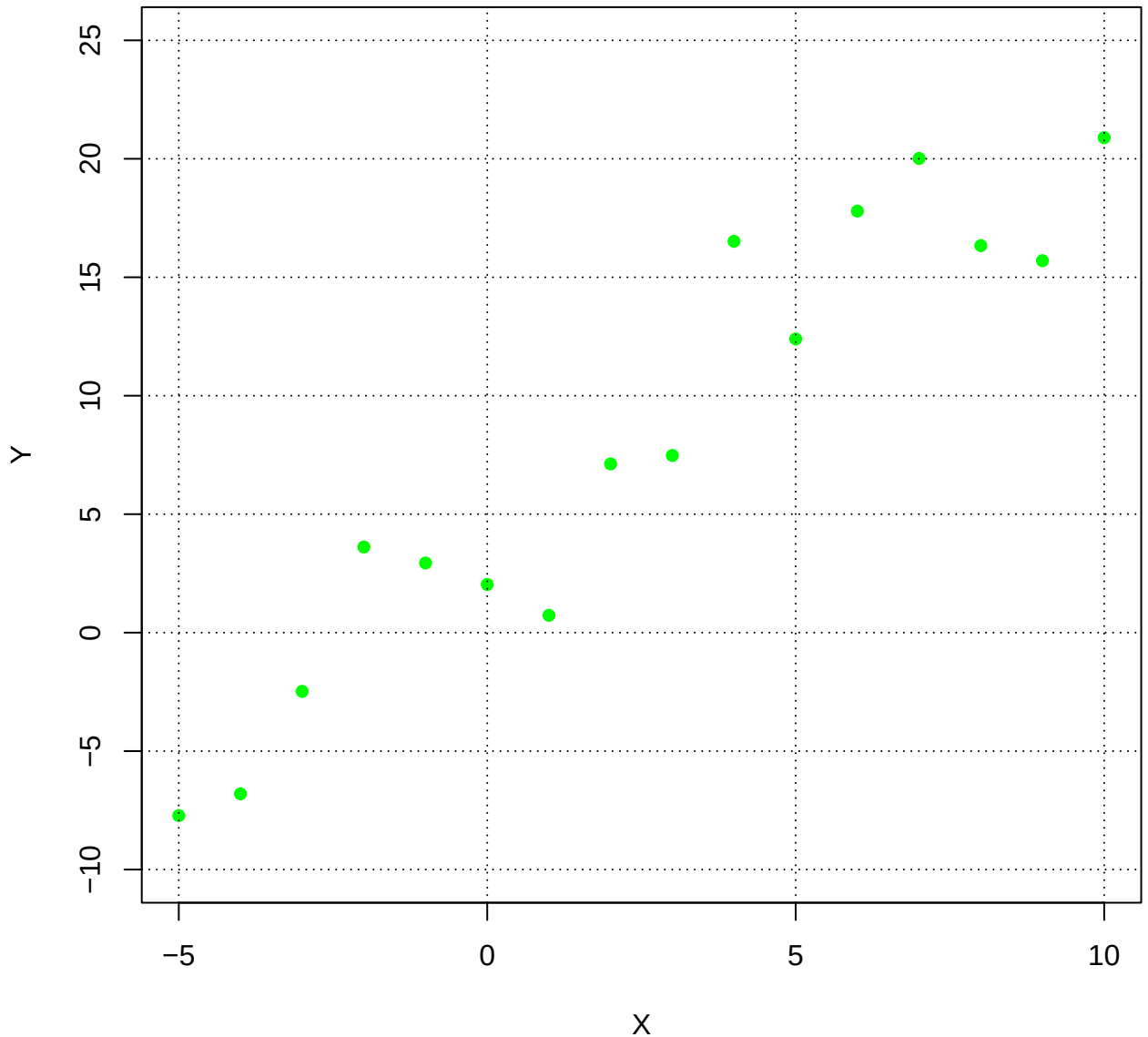
$$E[Y_i|X_i] = \beta_0 + \beta_1 \cdot X_i$$

- Coefficients:
  - β_1 ("Beta one"): Slope, effect of 1-unit change in X
  - β_0 ("Beta zero"): Intercept, mean when $X = 0$
- Not all data lies on the regression line, even if truth is linear
  - Randomness/other factors also drive Y

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

  - u_i ("u-i"): error term
- In reality, we only see the data points – need to infer the underlying relationship

Data by itself



Estimator for the linear regression function

$$\hat{E}[Y_i|X_i] = \hat{\beta}_0 + \hat{\beta}_1 \cdot X_i$$

- Coefficients:
  - $\hat{\beta}_0$ ("Beta zero hat"): Intercept estimator
  - $\hat{\beta}_1$ ("Beta one hat"): Slope estimator
- $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 \cdot X_i$ ("y-i hat"): Fitted value

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 \cdot X_i + \hat{u}_i = \hat{Y}_i + \hat{u}_i$$

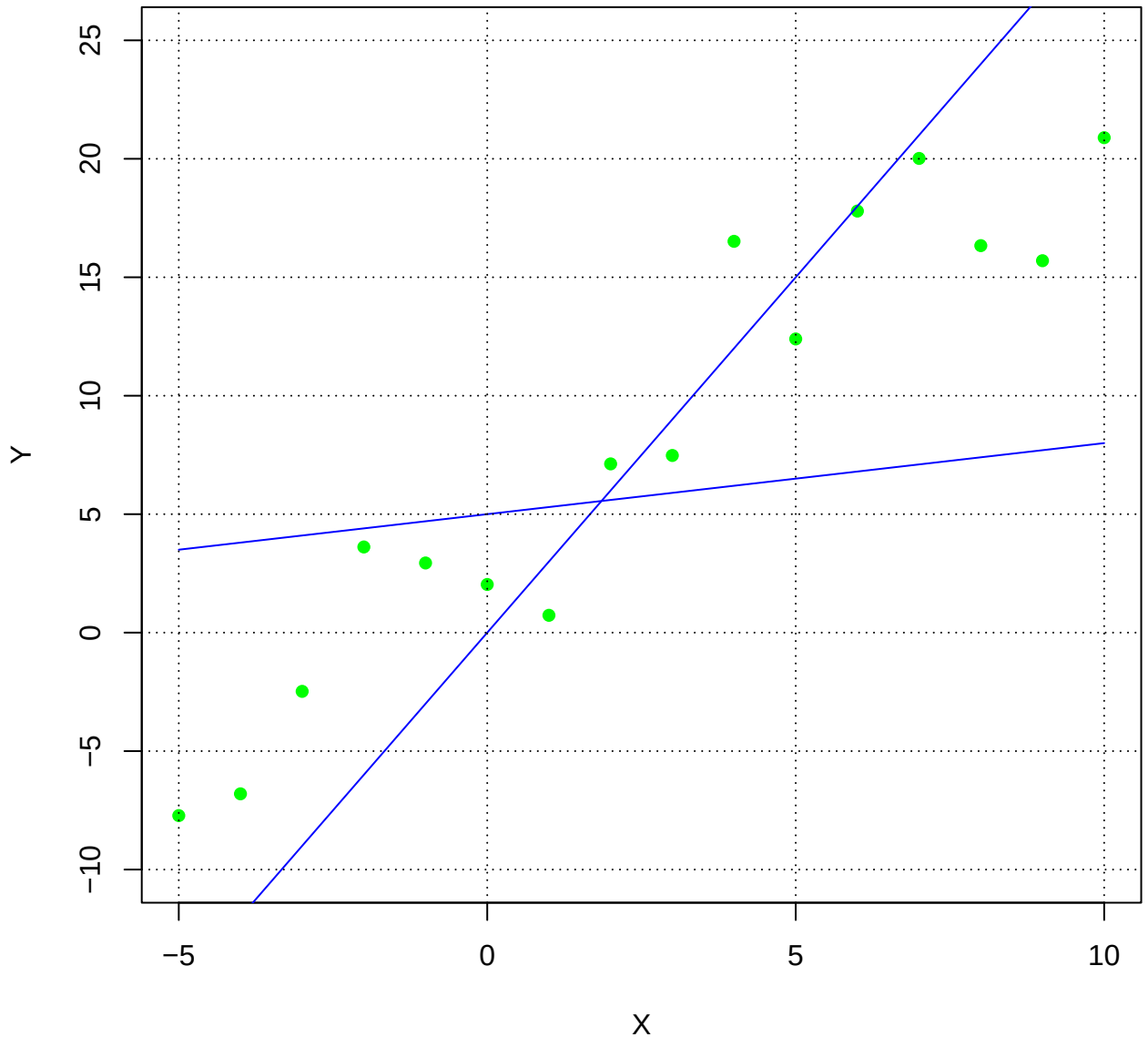
- $\hat{u}_i$ ("u-i hat"): residual

Key insight: the $\hat{\beta}$s and the $\hat{u}$s are two sides of the same coin – $\hat{\beta}$s determine $\hat{u}$s.

How to choose an estimator?

- We know our estimator will be a line
  - Our assumed $\hat{E}$ function has slope-intercept form
- But we have ∞ potential lines to choose from

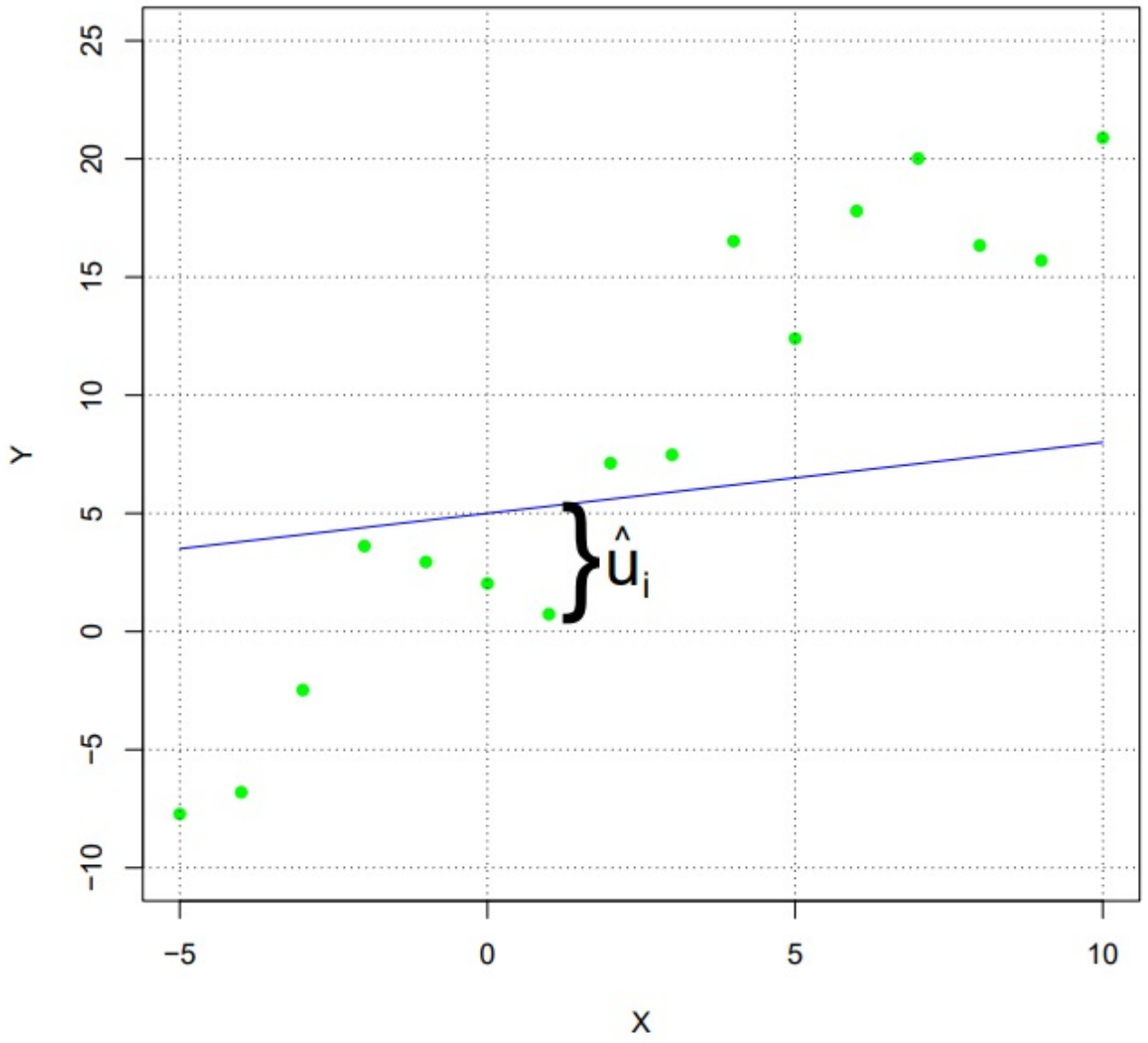
How to choose an estimator?



How to choose an estimator?

- We know our estimator will be a line
  - Our assumed $\hat{E}$ function has slope-intercept form
- But we have ∞ potential lines to choose from
- The key will be to "minimize residuals" in some sense

How to choose an estimator?



Ordinary Least Squares (OLS)

- We will choose the line that minimizes the sum of squared residuals:

$$\sum_{i=1}^{n} \hat{u}_i^2$$

- Goal is to get small residuals and, especially, avoid really large residuals as much as possible
- This approach is known as "**Ordinary Least Squares" (OLS)**
- If $E[Y|X]$ is truly linear, OLS is unbiased and consistent
- If $E[Y|X]$ is not linear, OLS is unbiased and consistent *for the best linear approximation* to the true function

Deriving the OLS estimators

- OLS chooses $\hat{\beta}_0, \hat{\beta}_1$ to minimize:

$$\sum_{i=1}^{n} \left(Y_i - [\hat{\beta}_0 + \hat{\beta}_1 \cdot X_i] \right)^2$$

- We know from calculus that we can find the minimum by taking derivatives with respect to our choice variables and setting them equal to 0

First Order Conditions (FOCs)

$$SSR = \min_{\hat{\beta}_0, \hat{\beta}_1} \sum_{i=1}^{n} \left(Y_i - [\hat{\beta}_0 + \hat{\beta}_1 \cdot X_i] \right)^2$$

First Order Conditions (FOCs)

$$SSR = \min_{\hat{\beta}_0, \hat{\beta}_1} \sum_{i=1}^{n} \left(Y_i - [\hat{\beta}_0 + \hat{\beta}_1 \cdot X_i] \right)^2$$

$$\frac{\partial SSR}{\partial \hat{\beta}_0} = -2 \cdot \sum_{i=1}^{n} \left(Y_i - [\hat{\beta}_0 + \hat{\beta}_1 \cdot X_i] \right) = 0$$

$$\frac{1}{n} \sum_{i=1}^{n} \left(Y_i - [\hat{\beta}_0 + \hat{\beta}_1 \cdot X_i] \right) = 0$$

First Order Conditions (FOCs)

$$SSR = \min_{\hat{\beta}_0, \hat{\beta}_1} \sum_{i=1}^{n} \left(Y_i - [\hat{\beta}_0 + \hat{\beta}_1 \cdot X_i] \right)^2$$

$$\frac{\partial SSR}{\partial \hat{\beta}_0} = -2 \cdot \sum_{i=1}^{n} \left(Y_i - [\hat{\beta}_0 + \hat{\beta}_1 \cdot X_i] \right) = 0$$

$$\frac{1}{n} \sum_{i=1}^{n} \left(Y_i - [\hat{\beta}_0 + \hat{\beta}_1 \cdot X_i] \right) = 0$$

$$\frac{\partial SSR}{\partial \hat{\beta}_1} = -2 \cdot \sum_{i=1}^{n} X_i \left(Y_i - [\hat{\beta}_0 + \hat{\beta}_1 \cdot X_i] \right) = 0$$

$$\frac{1}{n} \sum_{i=1}^{n} X_i \left(Y_i - [\hat{\beta}_0 + \hat{\beta}_1 \cdot X_i] \right) = 0$$

Making sense of FOCs

- There are 2 directions we can go once we have the FOCs
  - Explicit solution
  - Implicit characterization
- Both are useful; implicit is ultimately much more powerful

Explicit solution for OLS

$$\hat{\beta}_1 = \frac{\frac{1}{n}\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\frac{1}{n}\sum(X_i - \bar{X})^2} = \frac{\sim \text{Sample cov of X and Y}}{\sim \text{Sample var of X}}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \cdot \bar{X}$$

- The slope coefficient ($\hat{\beta}_1$) measures how strongly X and Y co-vary, normalized by the variance of X to measure the impact of a 1-unit change in X
- The intercept ensures that the model is right "on average": average X is associated with average Y

[Proof of previous result]

$$\text{FOC 1: } \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - [\hat{\beta}_0 + \hat{\beta}_1 \cdot X_i]\right) = 0$$

$$\frac{1}{n}\sum_{i=1}^{n} Y_i - \hat{\beta}_0 - \hat{\beta}_1 \cdot \frac{1}{n}\sum_{i=1}^{n} X_i =$$

$$\bar{Y} - \hat{\beta}_1 \cdot \bar{X} = \hat{\beta}_0$$

[Proof of previous result (2)]

$$\text{FOC 2: } \frac{1}{n}\sum_{i=1}^{n} X_i\left(Y_i - [\hat{\beta}_0 + \hat{\beta}_1 \cdot X_i]\right) = 0$$

$$\frac{1}{n}\sum_{i=1}^{n}\left((X_i \cdot (Y_i - \hat{\beta}_0)\right) - \hat{\beta}_1 \cdot \frac{1}{n}\sum_{i=1}^{n} X_i^2 =$$

$$\frac{1}{n}\sum_{i=1}^{n}\left((X_i \cdot (Y_i - \bar{Y})\right) - \hat{\beta}_1 \cdot \frac{1}{n}\sum_{i=1}^{n}\left(X_i \cdot (X_i - \bar{X})\right) =$$

$$\frac{\frac{1}{n}\sum_{i=1}^{n}\left((X_i \cdot (Y_i - \bar{Y})\right)}{\frac{1}{n}\sum_{i=1}^{n}\left(X_i \cdot (X_i - \bar{X})\right)} = \hat{\beta}_1$$

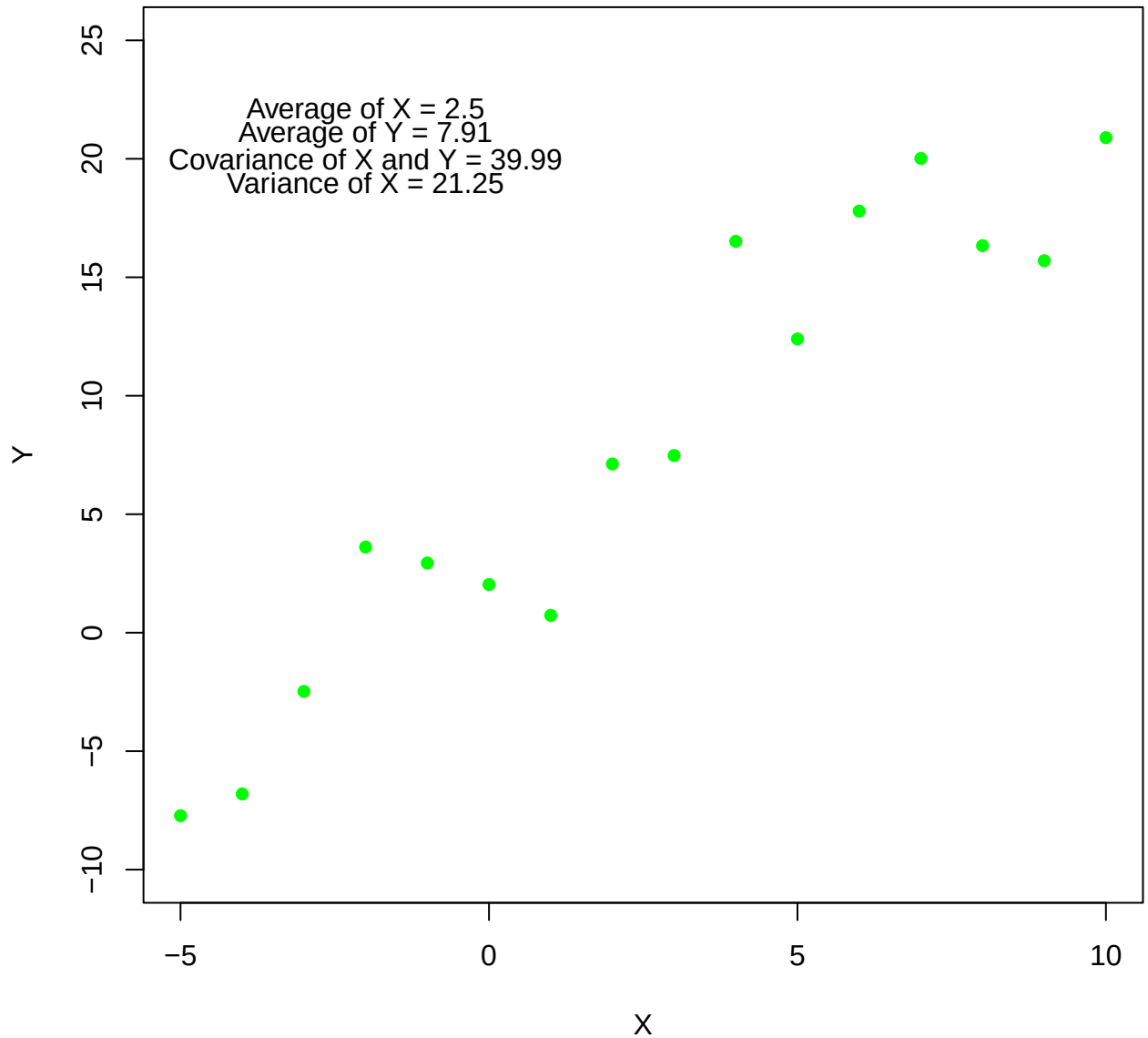
[Proof of previous result (3)]

$$\hat{\beta}_1 = \frac{\frac{1}{n}\sum_{i=1}^{n}((X_i \cdot (Y_i - \bar{Y}))}{\frac{1}{n}\sum_{i=1}^{n}(X_i \cdot (X_i - \bar{X}))}$$

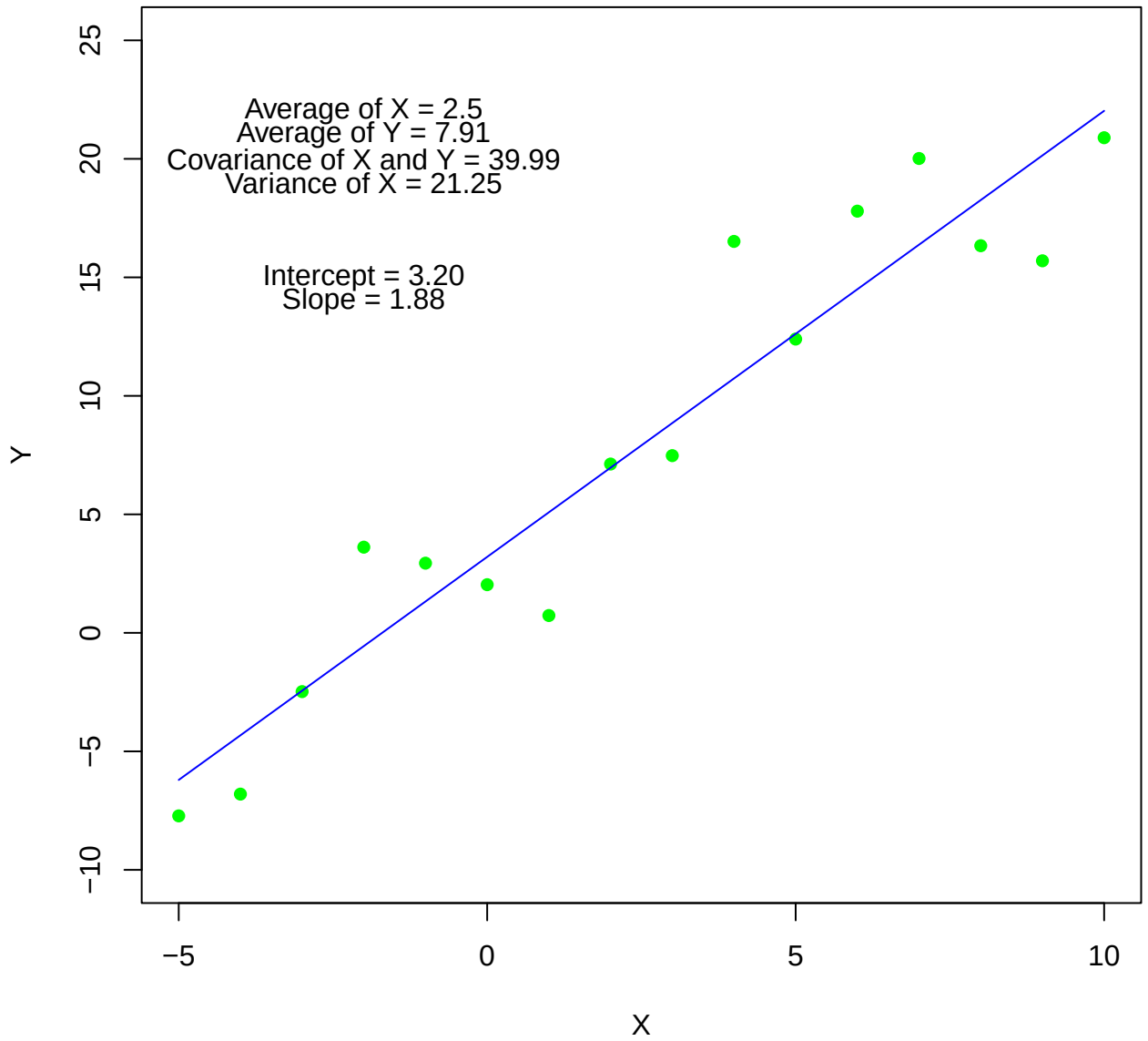
$$= \frac{\frac{1}{n}\sum_{i=1}^{n}((X_i \cdot (Y_i - \bar{Y})) - \frac{1}{n}\bar{X}\sum_{i=1}^{n}(Y_i - \bar{Y})}{\frac{1}{n}\sum_{i=1}^{n}(X_i \cdot (X_i - \bar{X})) - \frac{1}{n}\bar{X}\sum_{i=1}^{n}(X_i - \bar{X})}$$

$$= \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}) \cdot (Y_i - \bar{Y})}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{cov(X, Y)}{var(X)}$$

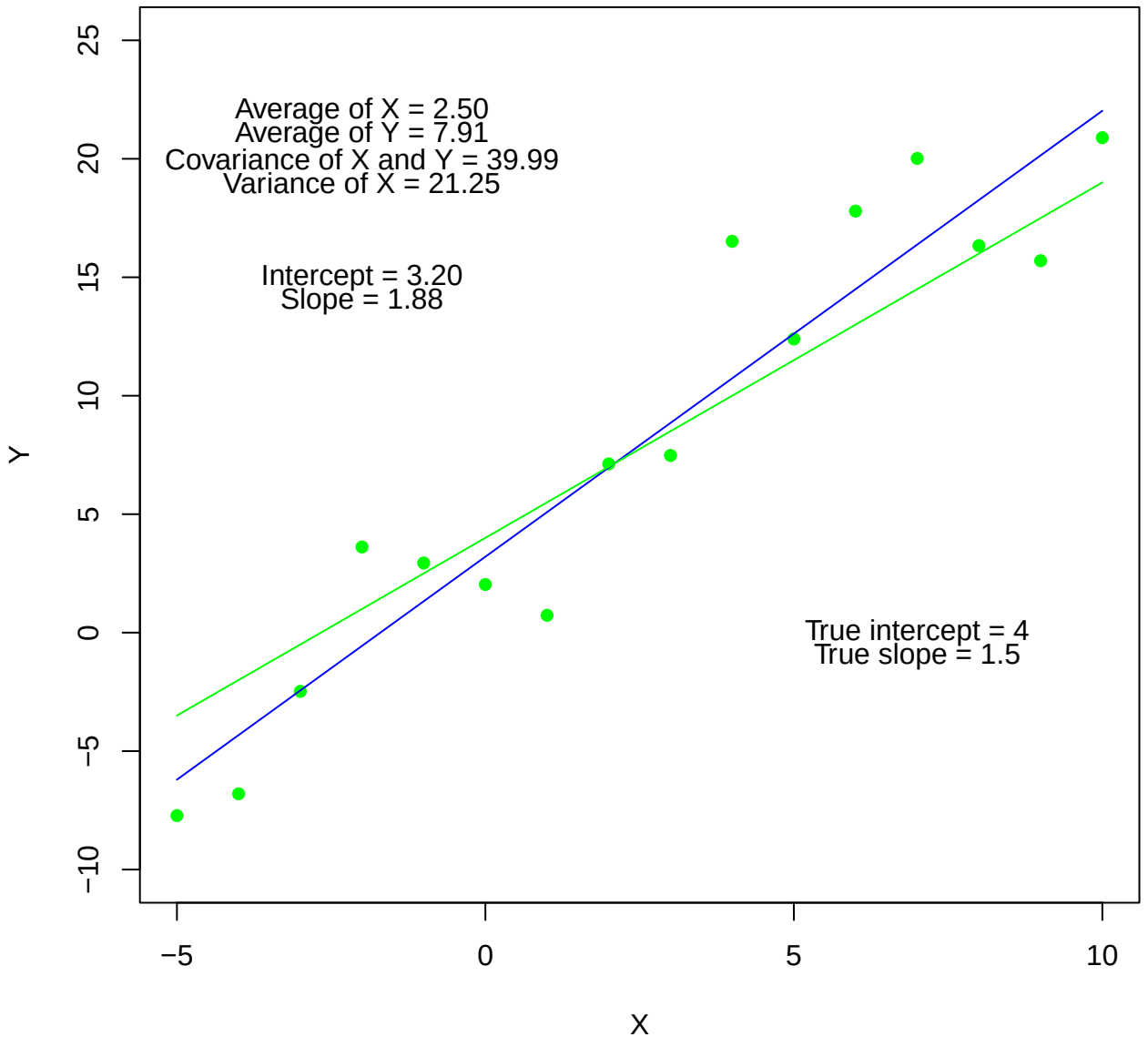
Computing best fit line



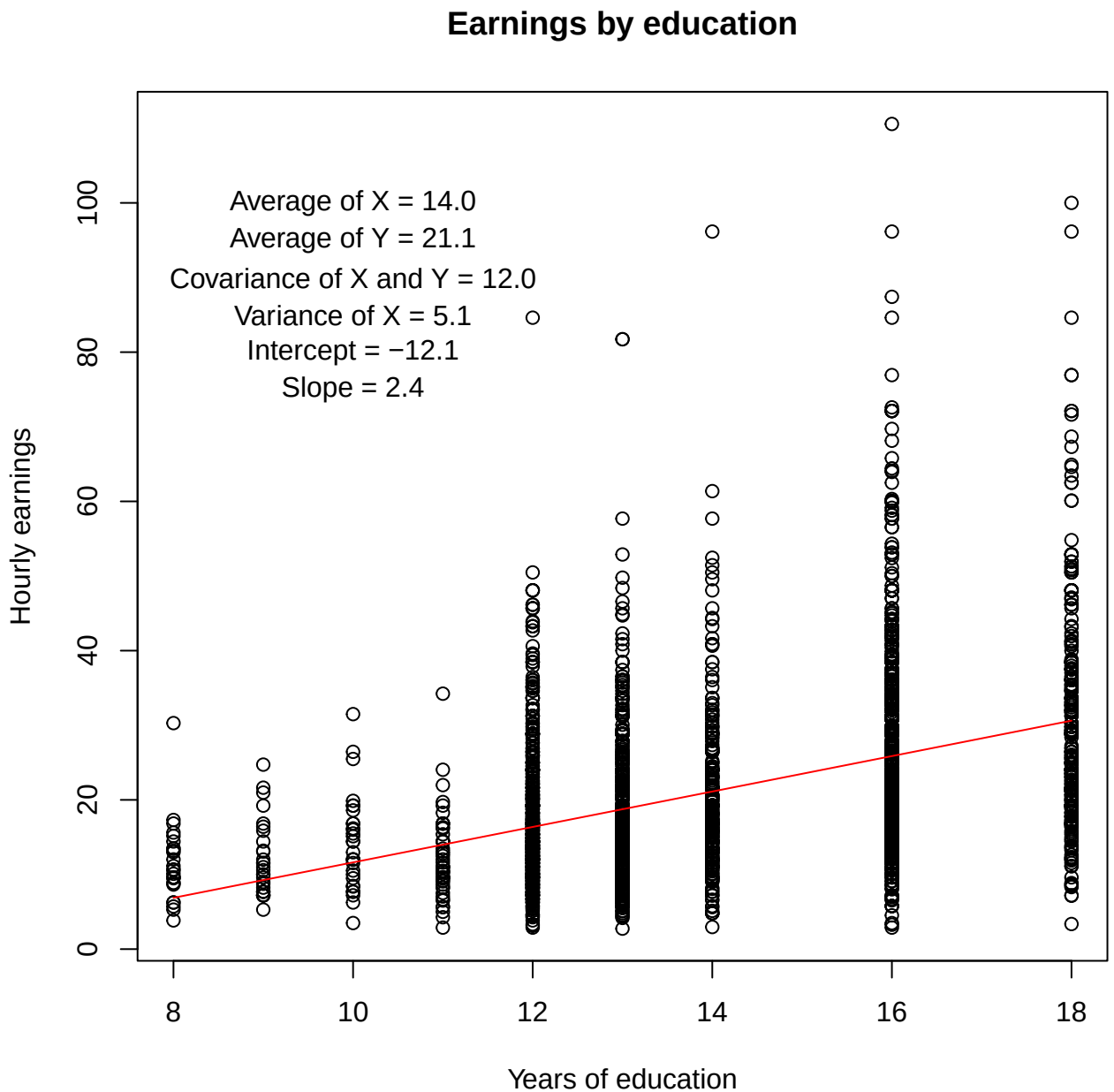
Computing best fit line



Computing best fit line



Computing best fit line



Implicit characterization of OLS solution

$$\frac{1}{n}\sum_{i=1}^{n}\hat{u}_i = 0$$

$$\frac{1}{n}\sum_{i=1}^{n}X_i \cdot \hat{u}_i = 0$$

Implicit characterization of OLS solution

$$\frac{1}{n}\sum_{i=1}^{n}\hat{u}_i = 0$$

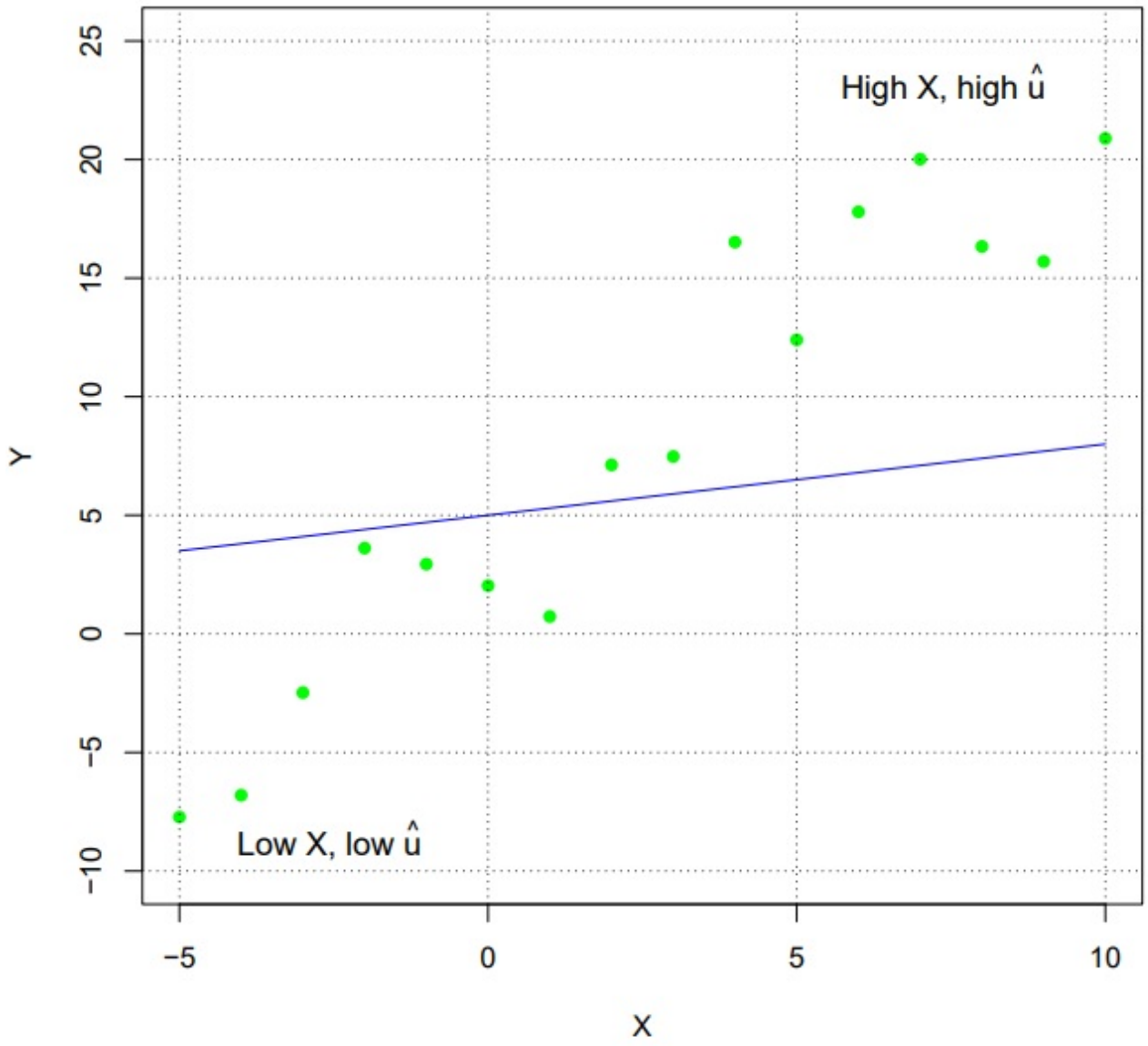
$$\frac{1}{n}\sum_{i=1}^{n}X_i \cdot \hat{u}_i = 0$$

- Model is correct on average (similar to what we've already seen)
- Residual is uncorrelated with X
  - Will show this formally on your problem set
  - X and $\hat{u}$ are **"orthogonal"**

Orthogonality

- Why does the solution require X and $\hat{u}$ to be uncorrelated?
- Suppose X and $\hat{u}$ were positively correlated

Positive correlation between X and $\hat{u}$



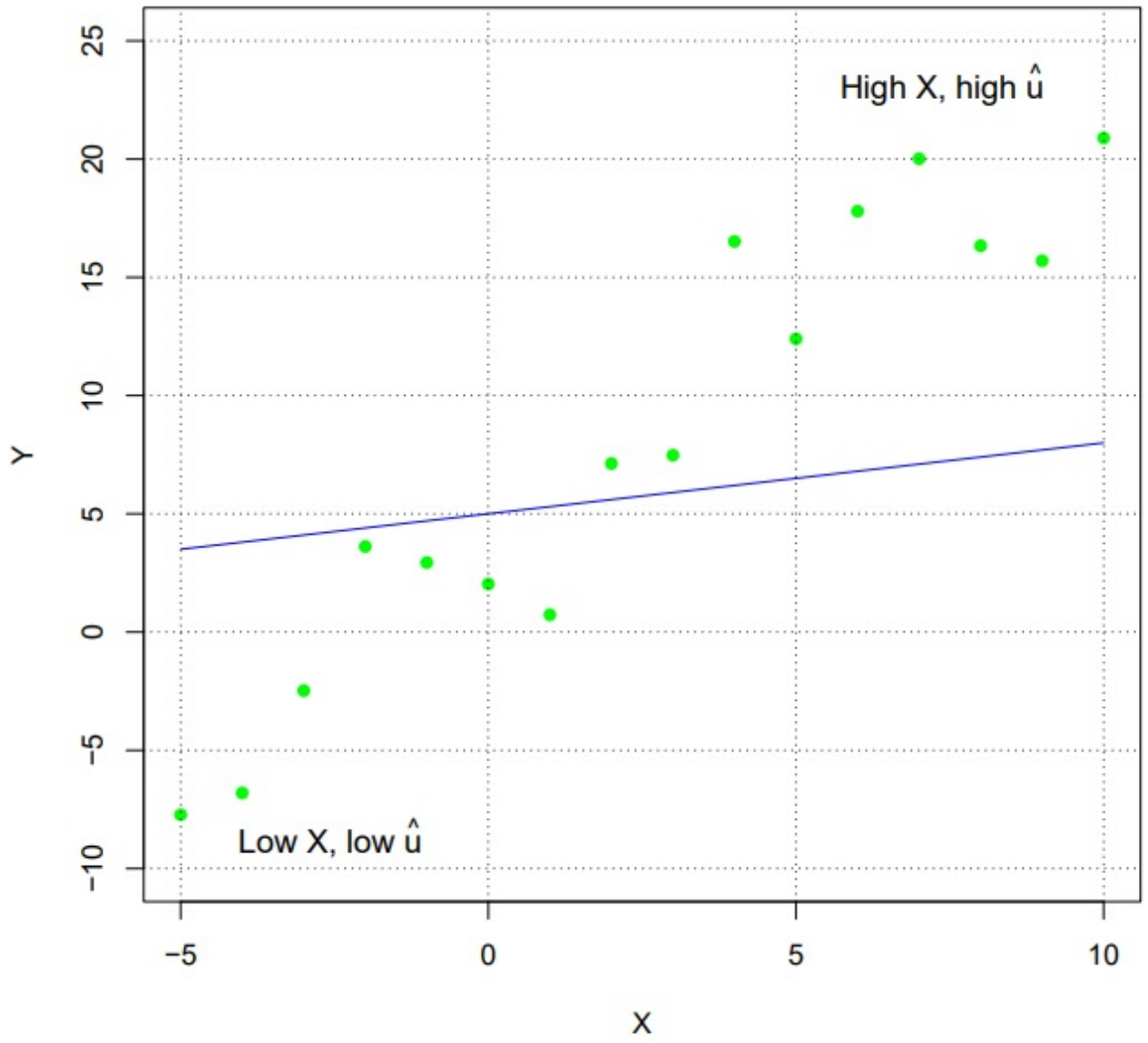
Orthogonality

- Why does the solution require X and $\hat{u}$ to be uncorrelated?
- Suppose X and $\hat{u}$ were positively correlated
- That means Y tends to be higher than you predicted when X is high and lower when X is low
- How can you improve your predictions, then?

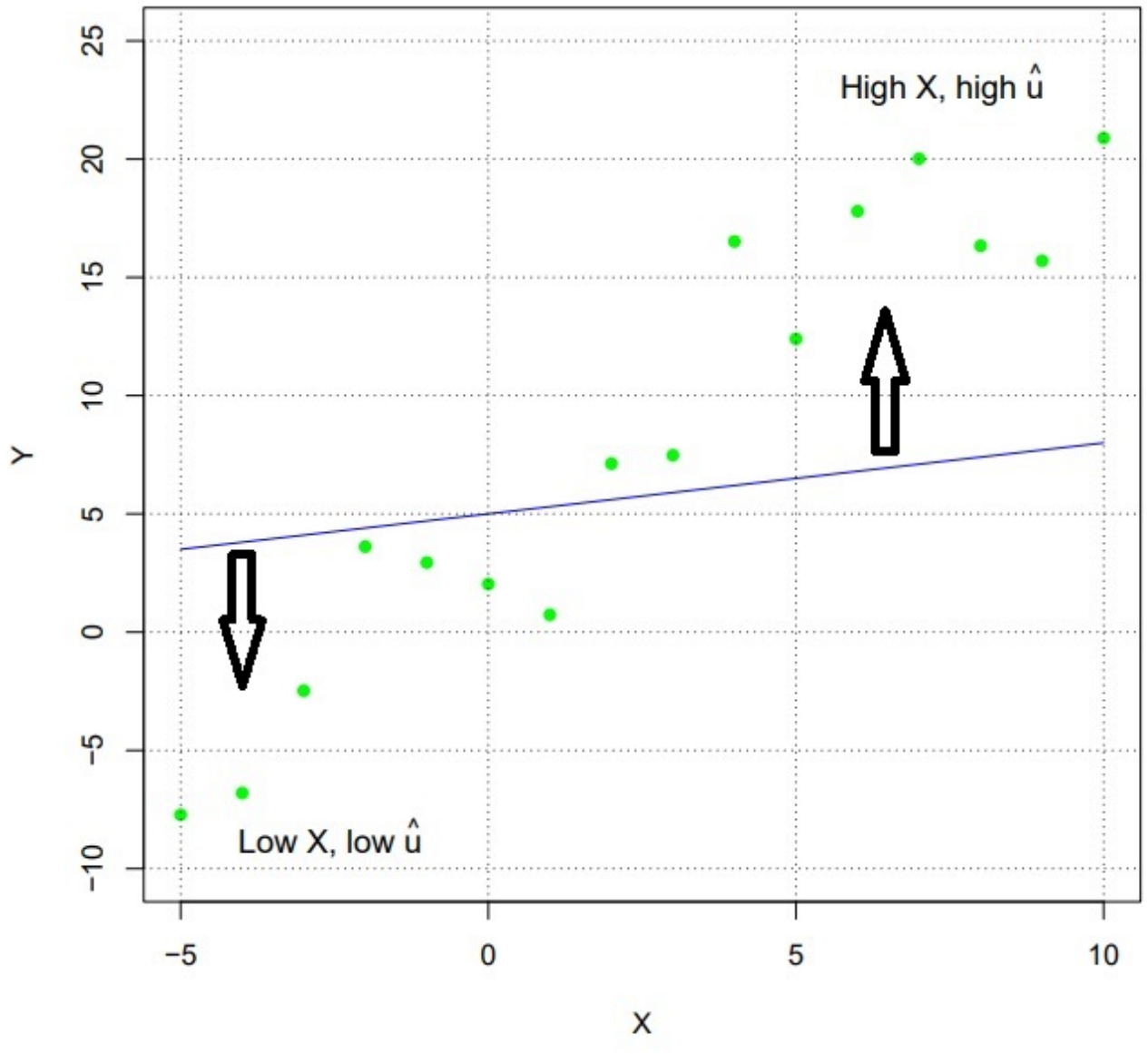
Orthogonality

- Why does the solution require X and $\hat{u}$ to be uncorrelated?
- Suppose X and $\hat{u}$ were positively correlated
- That means Y tends to be higher than you predicted when X is high and lower when X is low
- How can you improve your predictions, then?
- Increase $\hat{\beta}_1$!
  - That will increase your predictions when X is high...
  - ...and decrease your predictions when X is low
    - (It may also require adjusting $\hat{\beta}_0$ to make sure you're right on average)

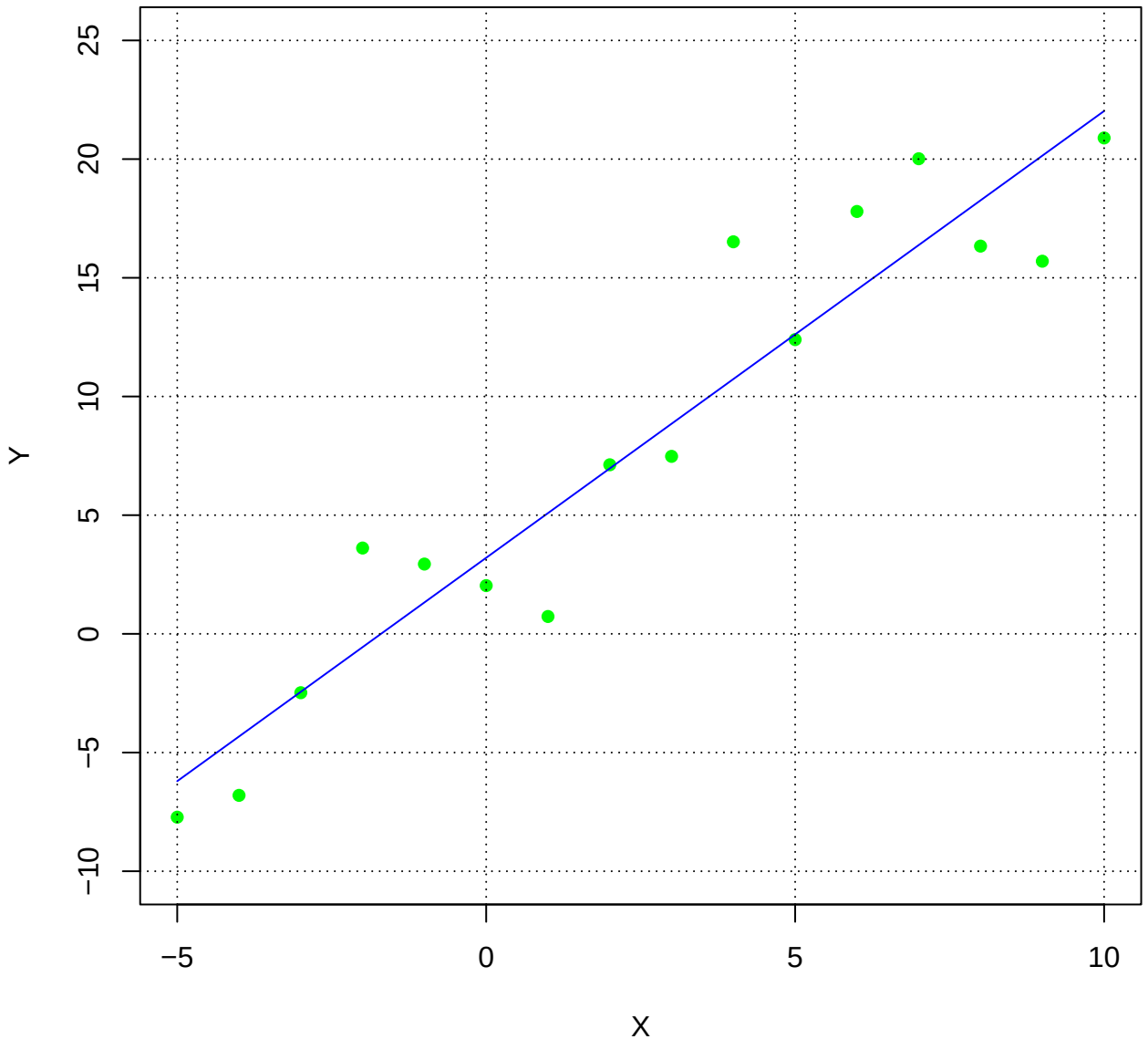
Positive correlation between X and $\hat{u}$



Positive correlation between X and $\hat{u}$



Positive correlation between X and $\hat{u}$



Orthogonality

- Why does the solution require X and $\hat{u}$ to be uncorrelated?
- Suppose X and $\hat{u}$ were positively correlated
- That means Y tends to be higher than you predicted when X is high and lower when X is low
- How can you improve your predictions, then?
- Increase $\hat{\beta}_1$!
  - That will increase your predictions when X is high...
  - ...and decrease your predictions when X is low
    - (It may also require adjusting $\hat{\beta}_0$ to make sure you're right on average)
- Until X and $\hat{u}$ are orthogonal, such improvements will always be possible

Inference for OLS

- Let's now talk about quantifying the uncertainty around $\hat{\beta}_1$
- Because the coefficient estimators are basically fancy sample means, LLN and CLT give us large-sample Normality
- So, hypothesis testing and confidence intervals for β_1 (or β_0) work just like they did for a sample means

$$t = \frac{\bar{Y} - \mu_0}{\hat{\sigma}_{\bar{Y}}} \rightarrow t = \frac{\hat{\beta}_1 - \beta_{1,H_0}}{\hat{\sigma}_{\hat{\beta}_1}}$$

$$95\%CI = [\bar{Y} \pm 1.96 \cdot \hat{\sigma}_{\bar{Y}}] \rightarrow 95\%CI = [\hat{\beta}_1 \pm 1.96 \cdot \hat{\sigma}_{\hat{\beta}_1}]$$

- The only new thing is, what is $\hat{\sigma}_{\hat{\beta}_1}$ the **"standard error (SE)"** of $\hat{\beta}_1$?

Standard error of $\hat{\beta}_1$

- It turns out:

$$\sigma^2_{\hat{\beta}_1} = \frac{1}{n} \cdot \frac{var\left[(X_i - \mu_X) \cdot u_i\right]}{var^2(X_i)}$$

and

$$\hat{\sigma}^2_{\hat{\beta}_1} = \frac{1}{n} \cdot \frac{\frac{1}{n-2}\sum_{i=1}^{n}\left[(X_i - \bar{X})^2 \cdot \hat{u}_i^2\right]}{\left(\frac{1}{n}\sum_{i=1}^{n}\left[(X_i - \bar{X})^2\right]\right)^2}$$

- Yuck. What to make of this?

Homoskedasticity

- Consider special case of **"homoskedasticity"**:

$$var(u_i|X) = \sigma_u^2,$$

  i.e. the variance of the error is constant across all X values.

- Expression simplifies to:

$$\hat{\sigma}^2_{\hat{\beta}_1} = \frac{1}{n} \cdot \frac{\frac{1}{n-2}\sum_{i=1}^{n} \hat{u}_i^2}{\frac{1}{n}\sum_{i=1}^{n}\left[(X_i - \bar{X})^2\right]}$$

Interpreting the standard error of $\hat{\beta}_1$

$$\hat{\sigma}^2_{\hat{\beta}_1} = \frac{1}{n} \cdot \frac{\frac{1}{n-2}\sum_{i=1}^{n} \hat{u}_i^2}{\frac{1}{n}\sum_{i=1}^{n}\left[(X_i - \bar{X})^2\right]}$$

- Your standard error is driven by 3 things

Interpreting the standard error of $\hat{\beta}_1$

$$\hat{\sigma}^2_{\hat{\beta}_1} = \frac{1}{n} \cdot \frac{\frac{1}{n-2}\sum_{i=1}^{n} \hat{u}_i^2}{\frac{1}{n}\sum_{i=1}^{n}\left[(X_i - \bar{X})^2\right]}$$

- Your standard error is driven by 3 things
  - Sample size (n)
    - Larger samples will have smaller standard errors – more information reduces uncertainty

Interpreting the standard error of $\hat{\beta}_1$

$$\hat{\sigma}^2_{\hat{\beta}_1} = \frac{1}{n} \cdot \frac{\frac{1}{n-2}\sum_{i=1}^{n} \hat{u}_i^2}{\frac{1}{n}\sum_{i=1}^{n}\left[(X_i - \bar{X})^2\right]}$$

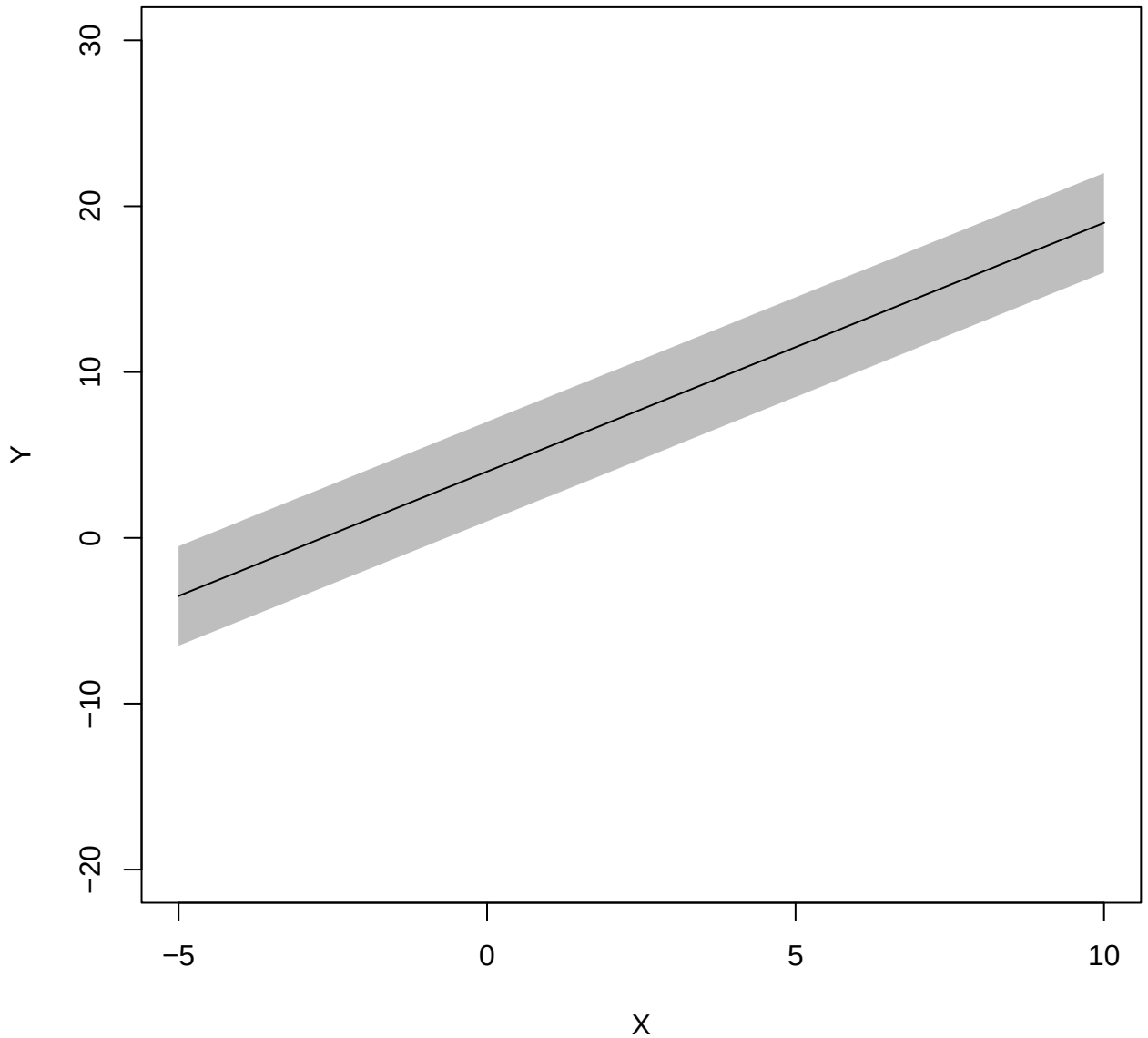
- Your standard error is driven by 3 things
  - Sample size (n)
    - Larger samples will have smaller standard errors – more information reduces uncertainty
  - Variance of residuals (numerator of expression)
    - The noisier the data (higher variance), the greater the uncertainty

Interpreting the standard error of $\hat{\beta}_1$

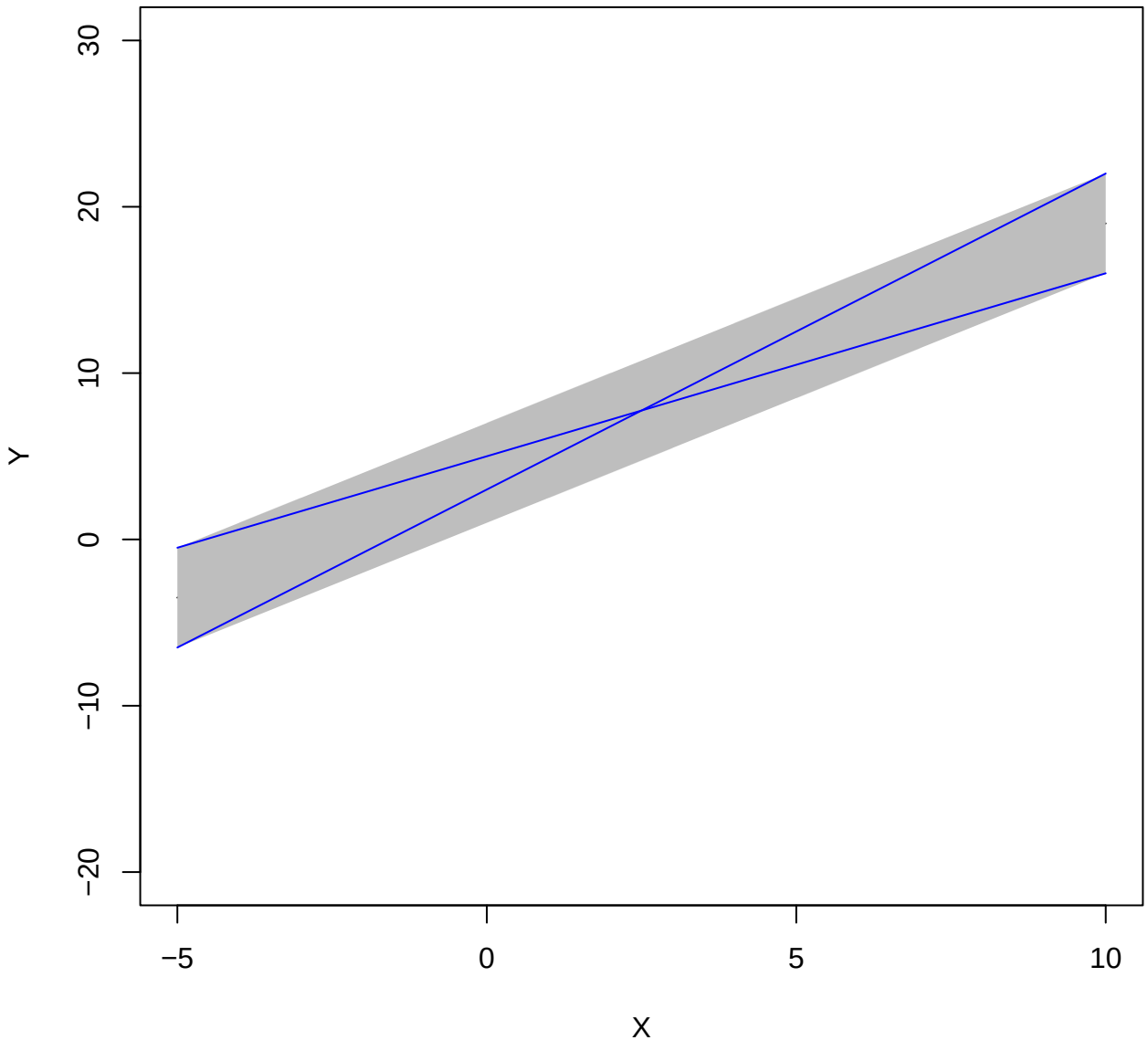
$$\hat{\sigma}^2_{\hat{\beta}_1} = \frac{1}{n} \cdot \frac{\frac{1}{n-2}\sum_{i=1}^{n} \hat{u}_i^2}{\frac{1}{n}\sum_{i=1}^{n}\left[(X_i - \bar{X})^2\right]}$$

- Your standard error is driven by 3 things
  - Sample size (n)
    - Larger samples will have smaller standard errors – more information reduces uncertainty
  - Variance of residuals (numerator of expression)
    - The noisier the data (higher variance), the greater the uncertainty
  - Variance of explanatory variable (denominator of expression)
    - To see how Y varies when X varies, we need X to vary! The more it does, the less uncertainty we have about the relationship.
    - Imagine if all observations had the same X. You literally could not estimate a relationship between X and Y (variance of ∞).

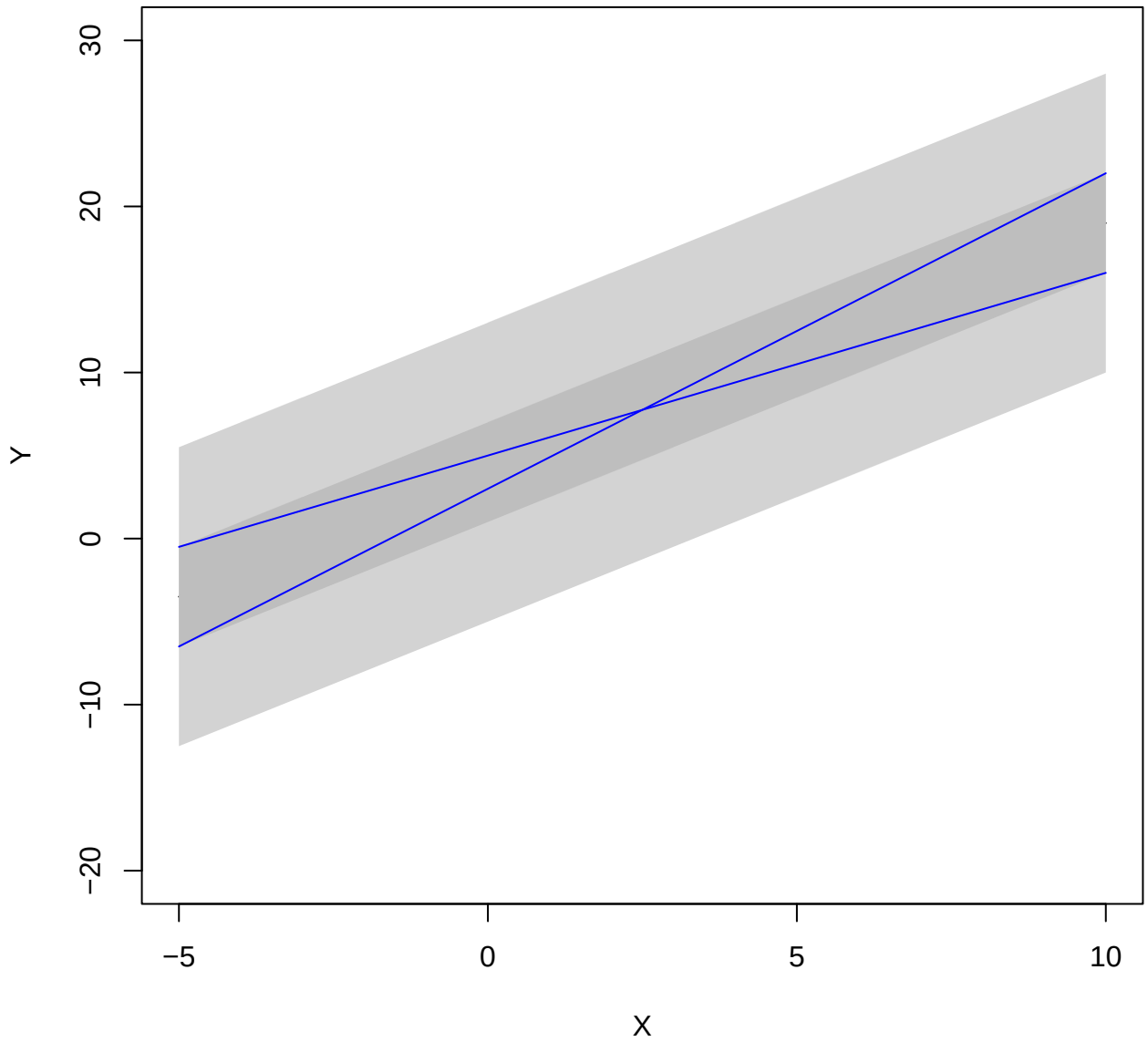
Effect of larger error variance



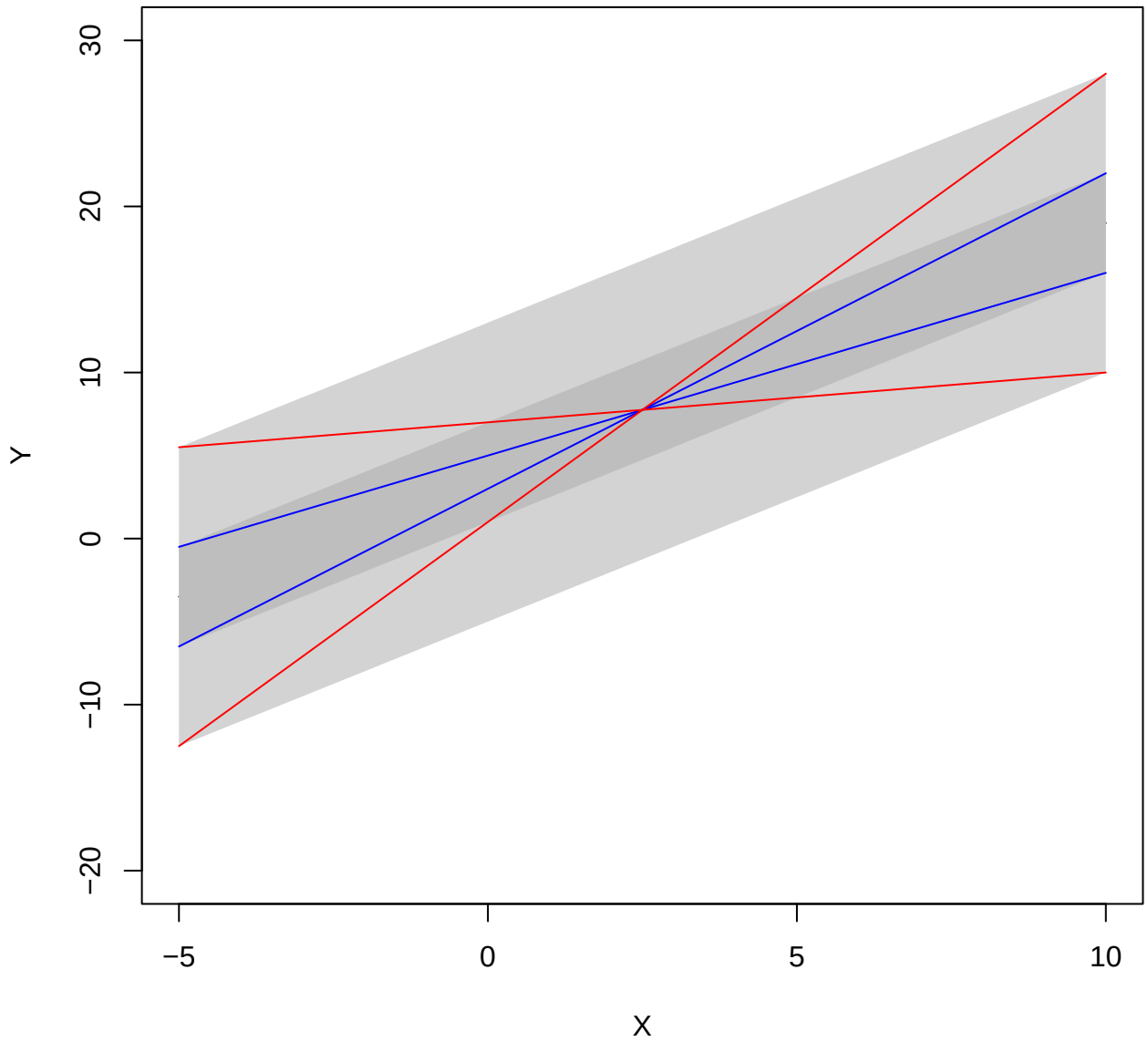
Effect of larger error variance



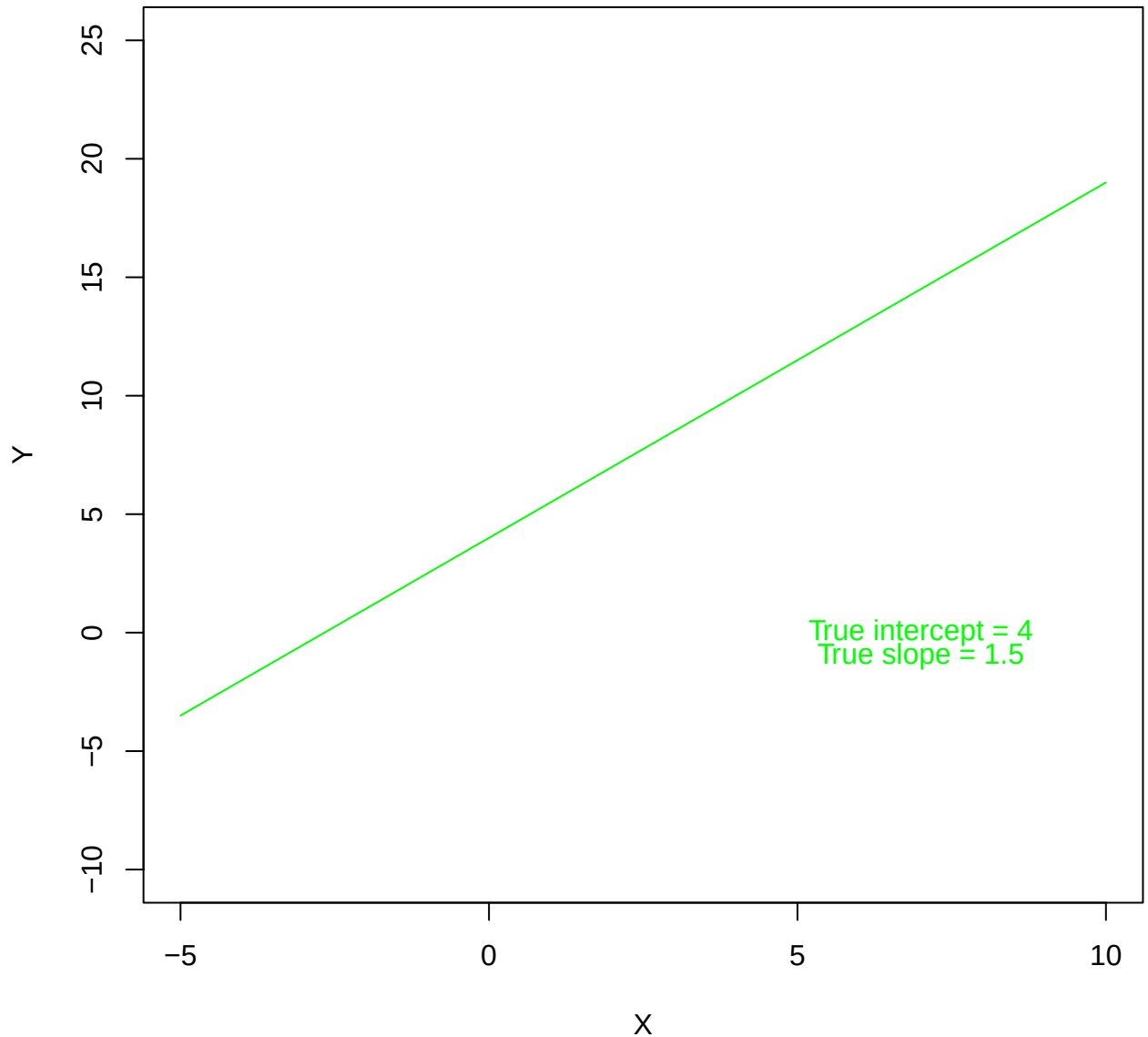
Effect of larger error variance



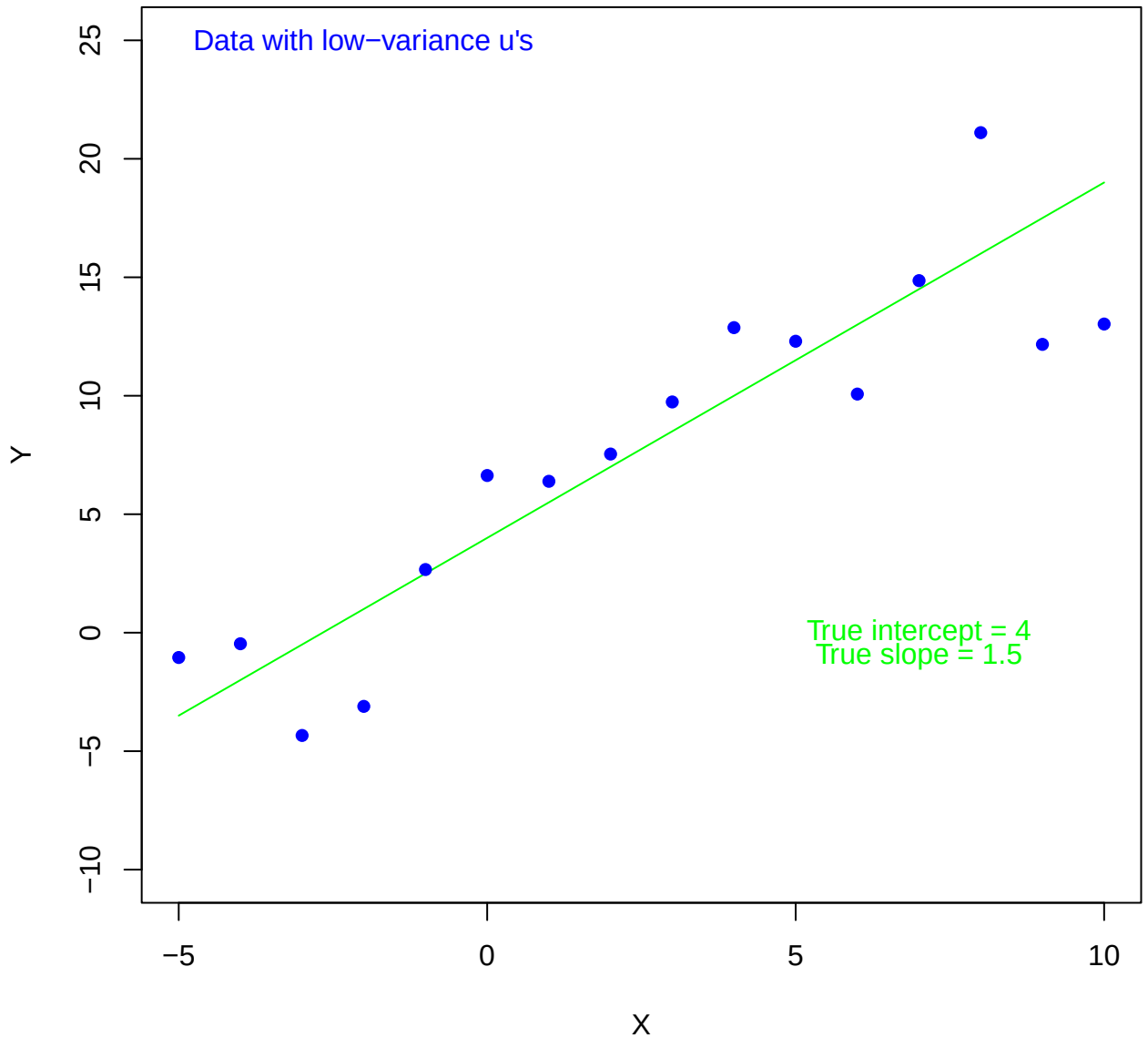
Effect of larger error variance



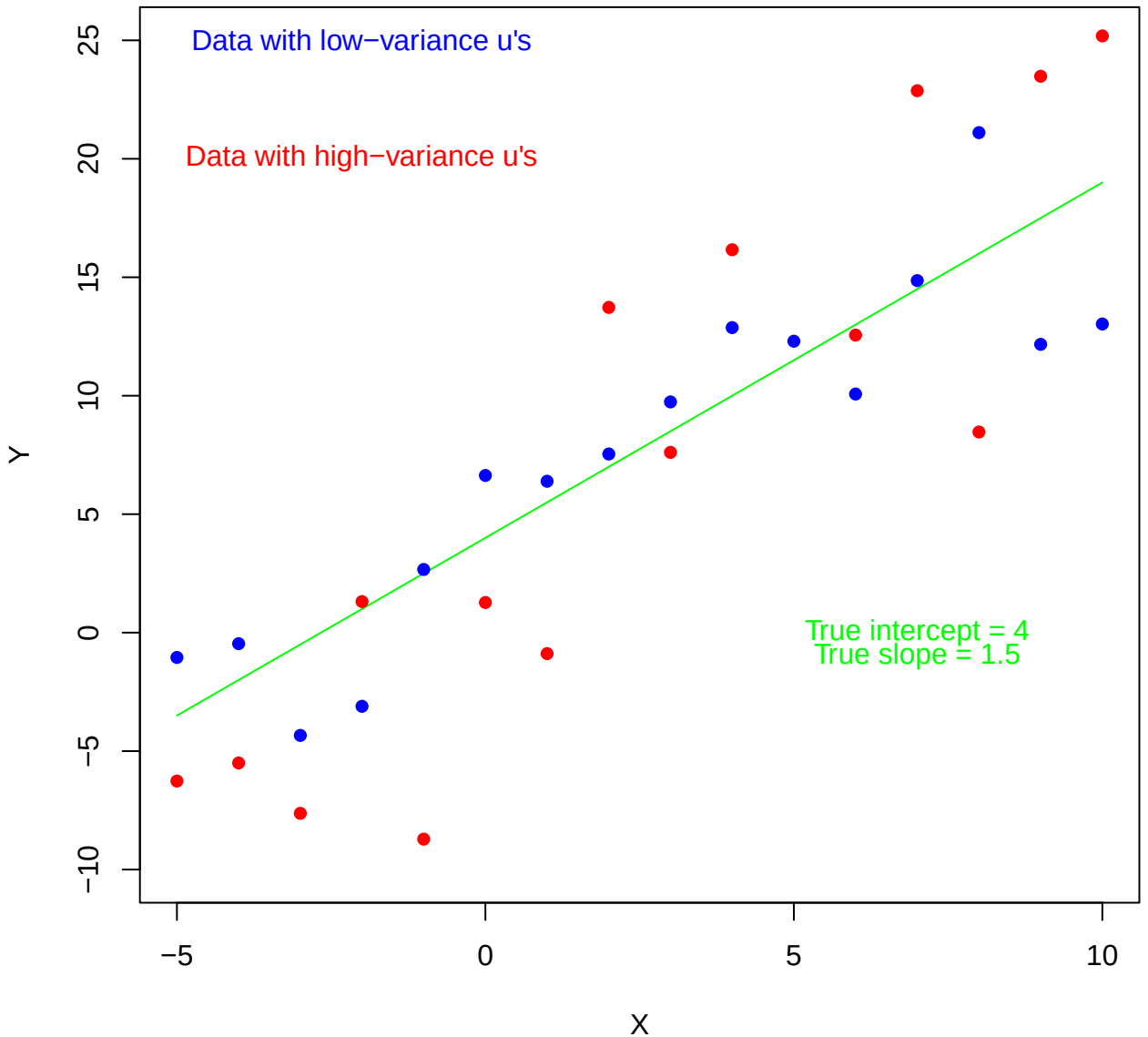
Effect of larger error variance



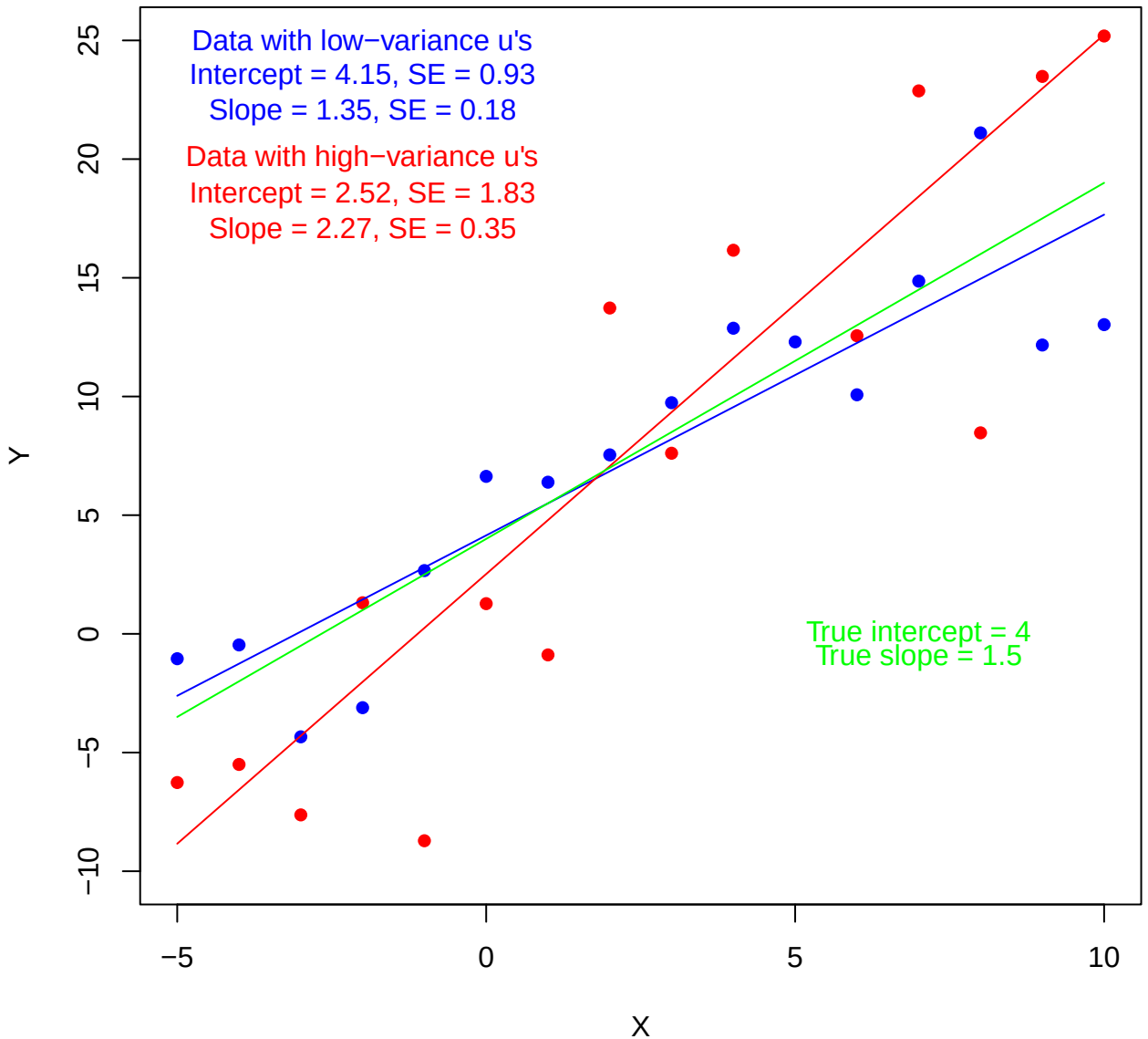
Effect of larger error variance



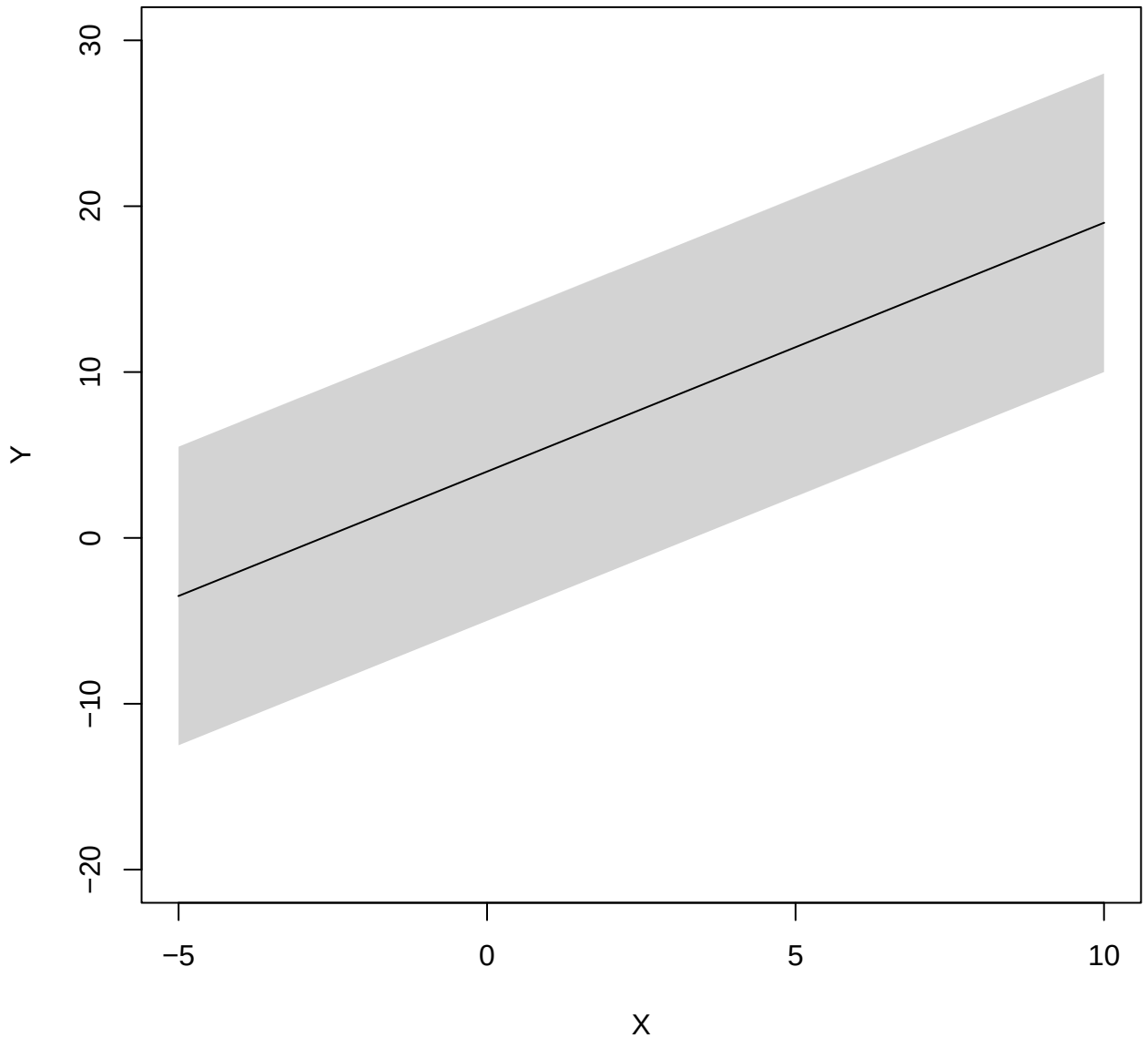
Effect of larger error variance



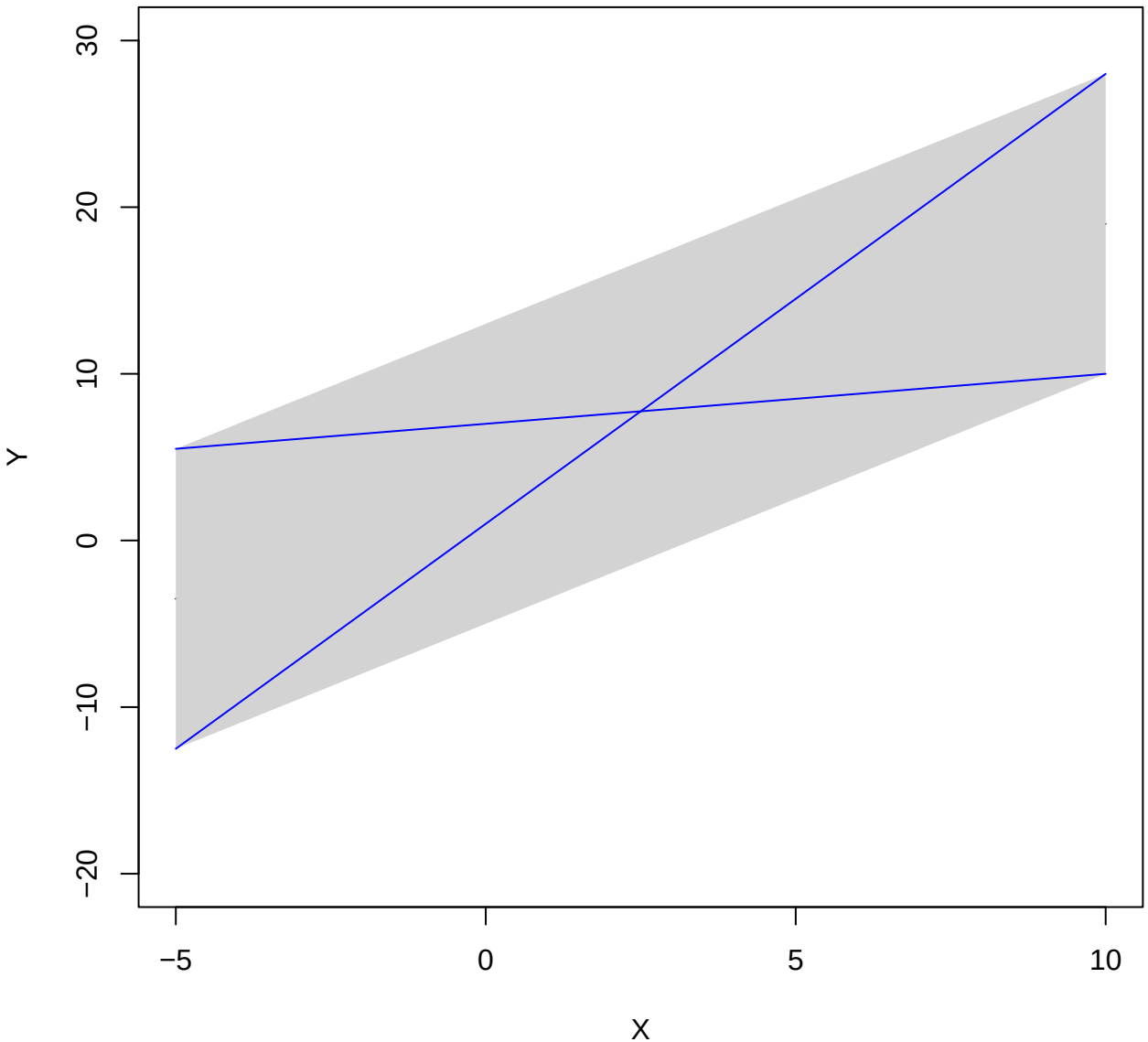
Effect of larger error variance



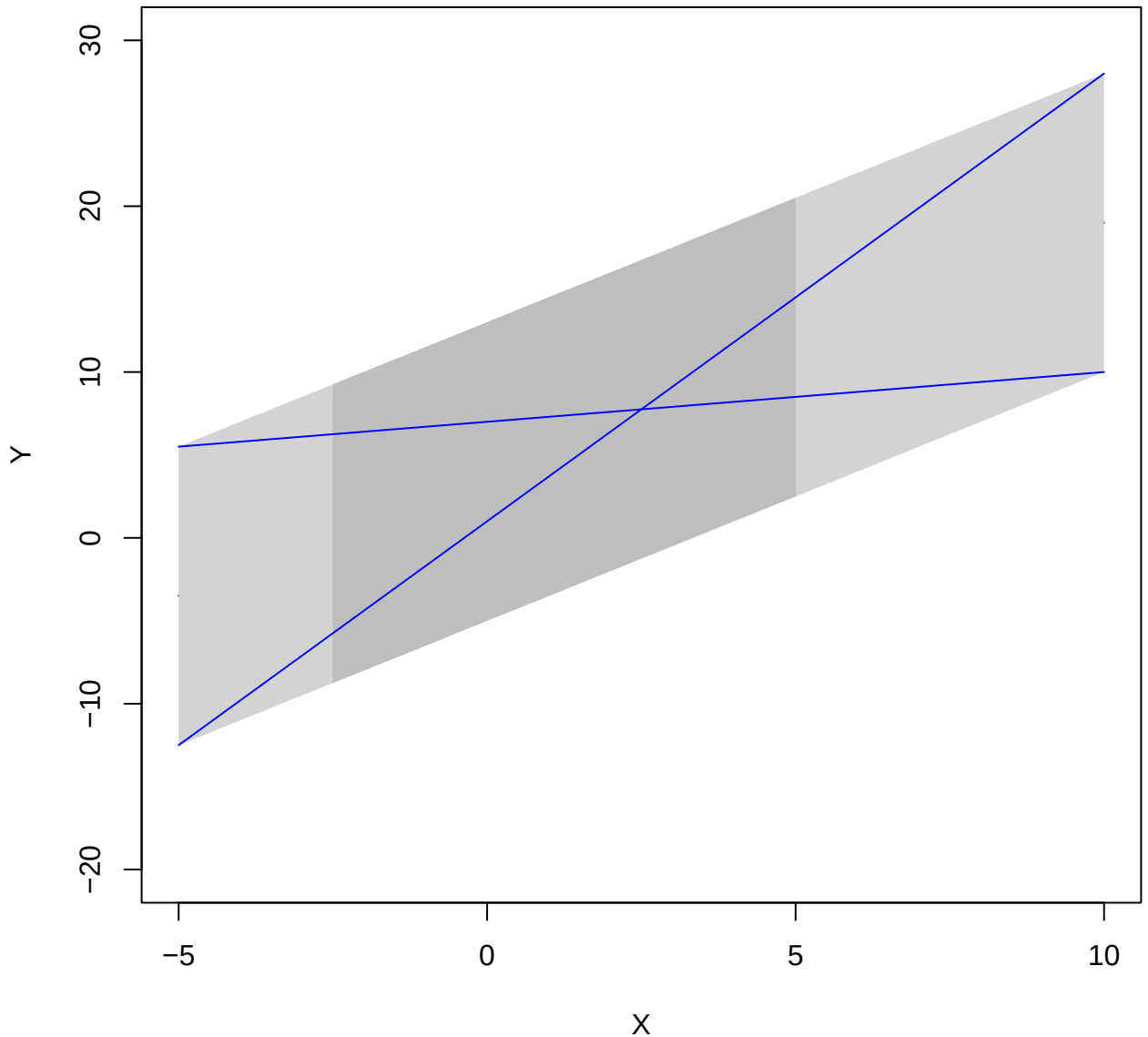
Effect of larger variance in X



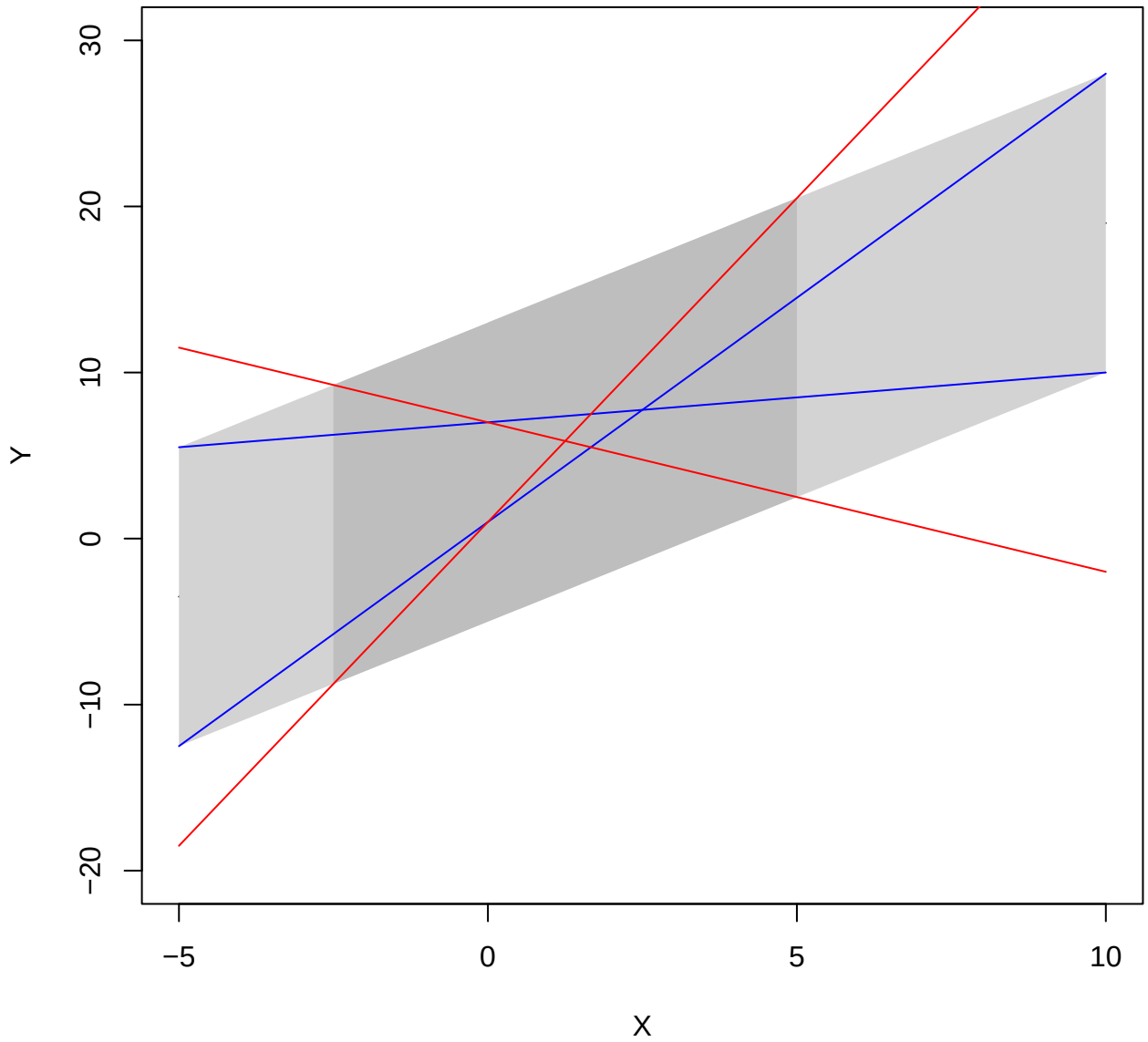
Effect of larger variance in X



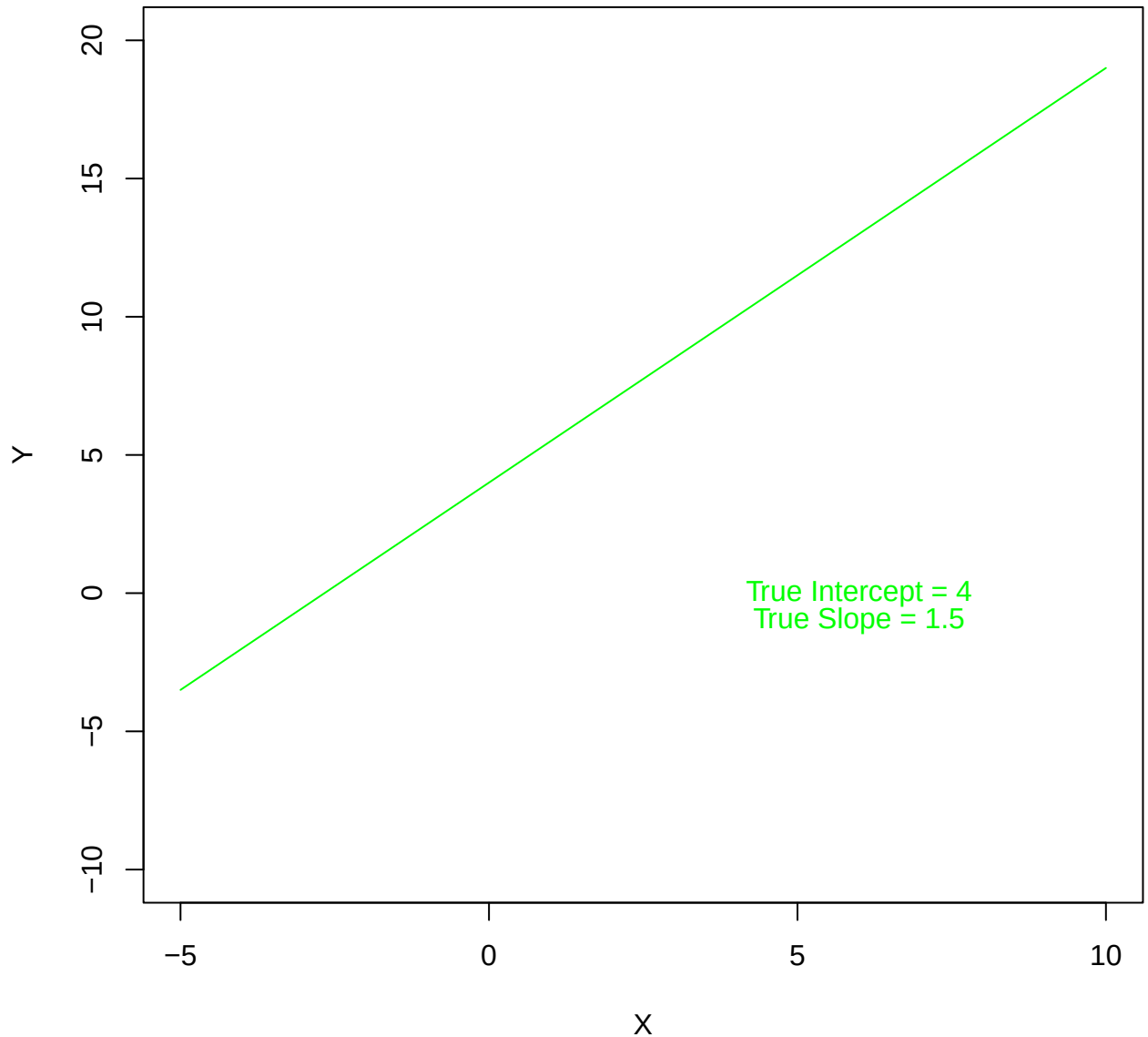
Effect of larger variance in X



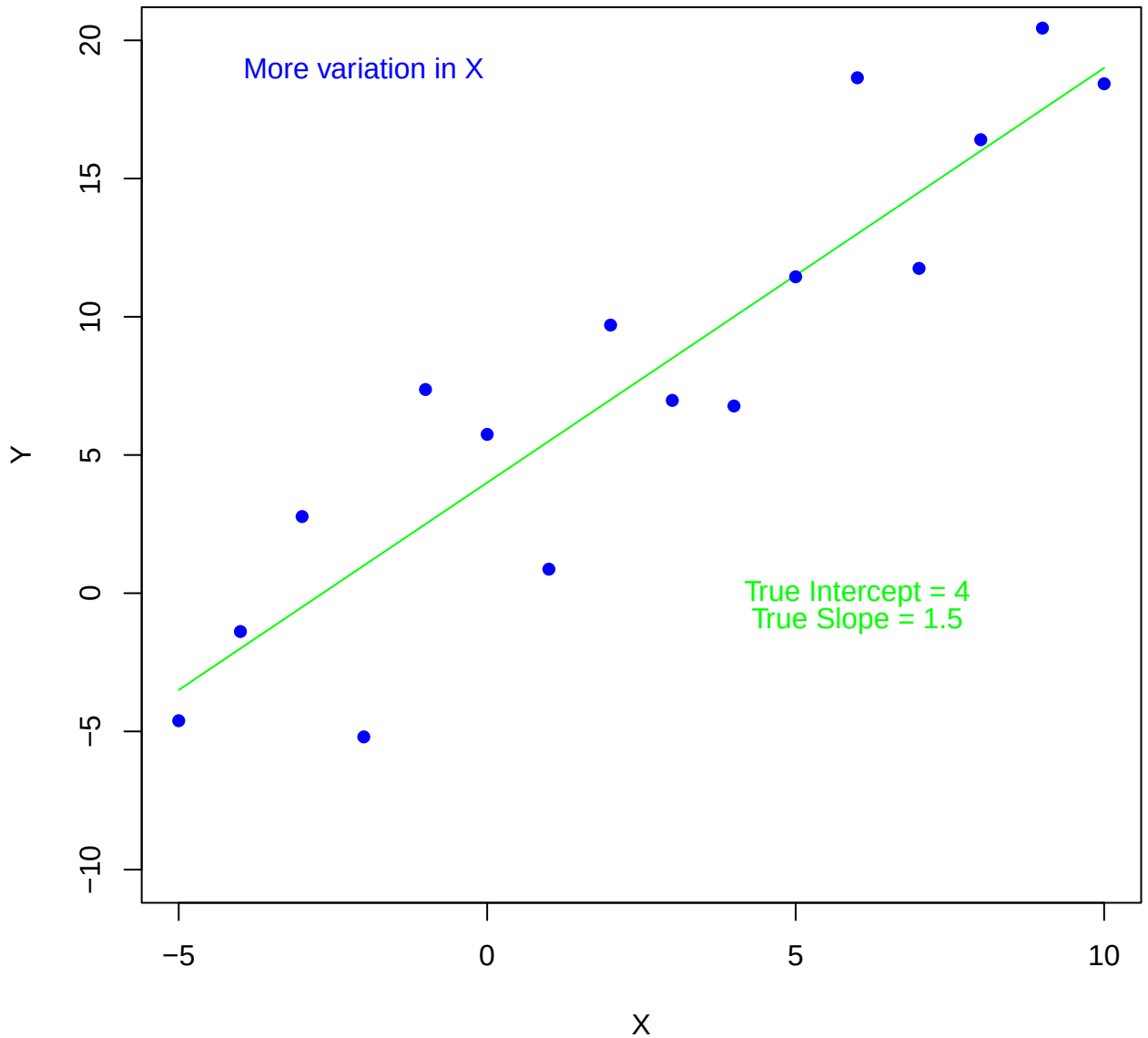
Effect of larger variance in X



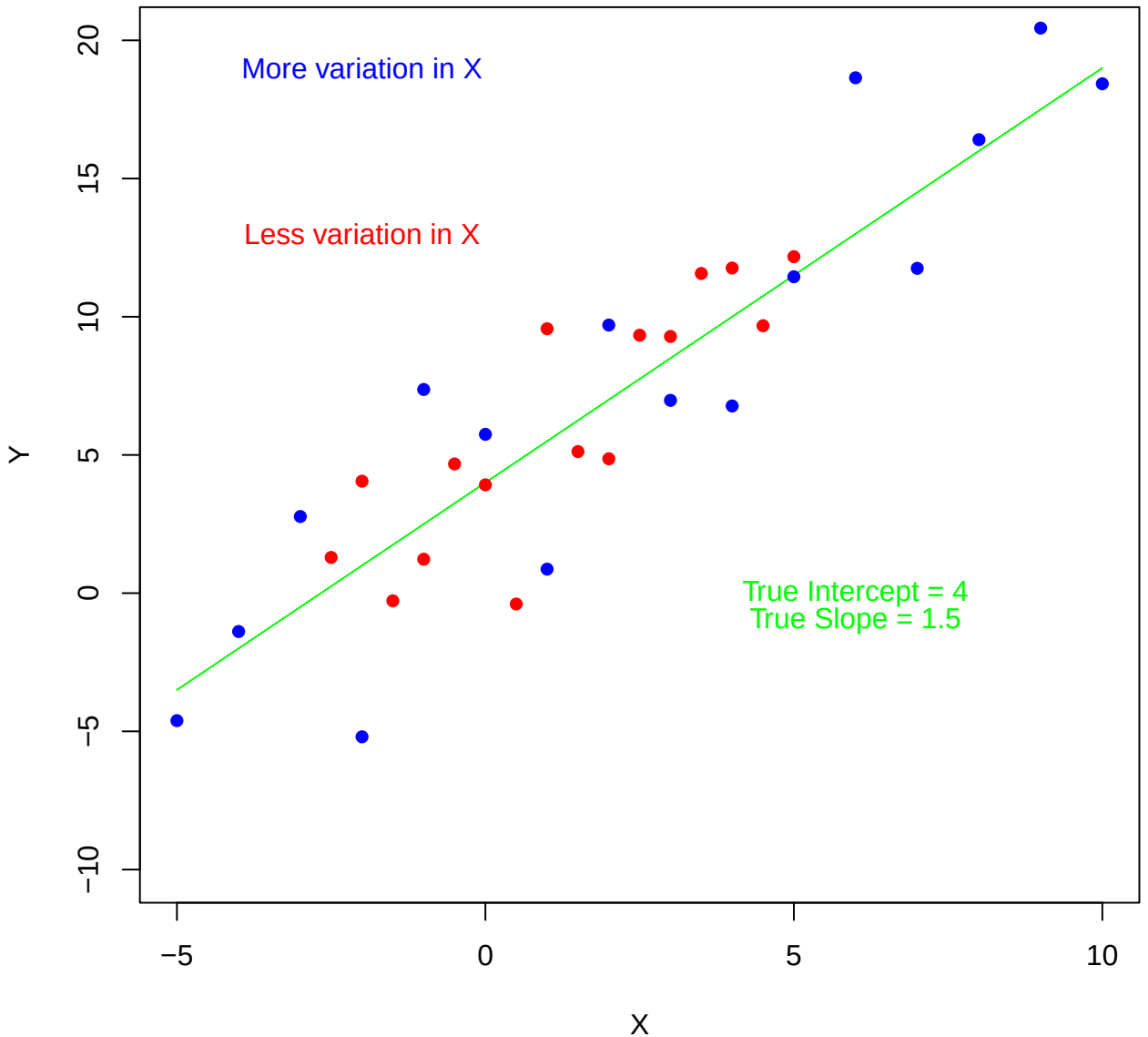
Effect of larger variance in X



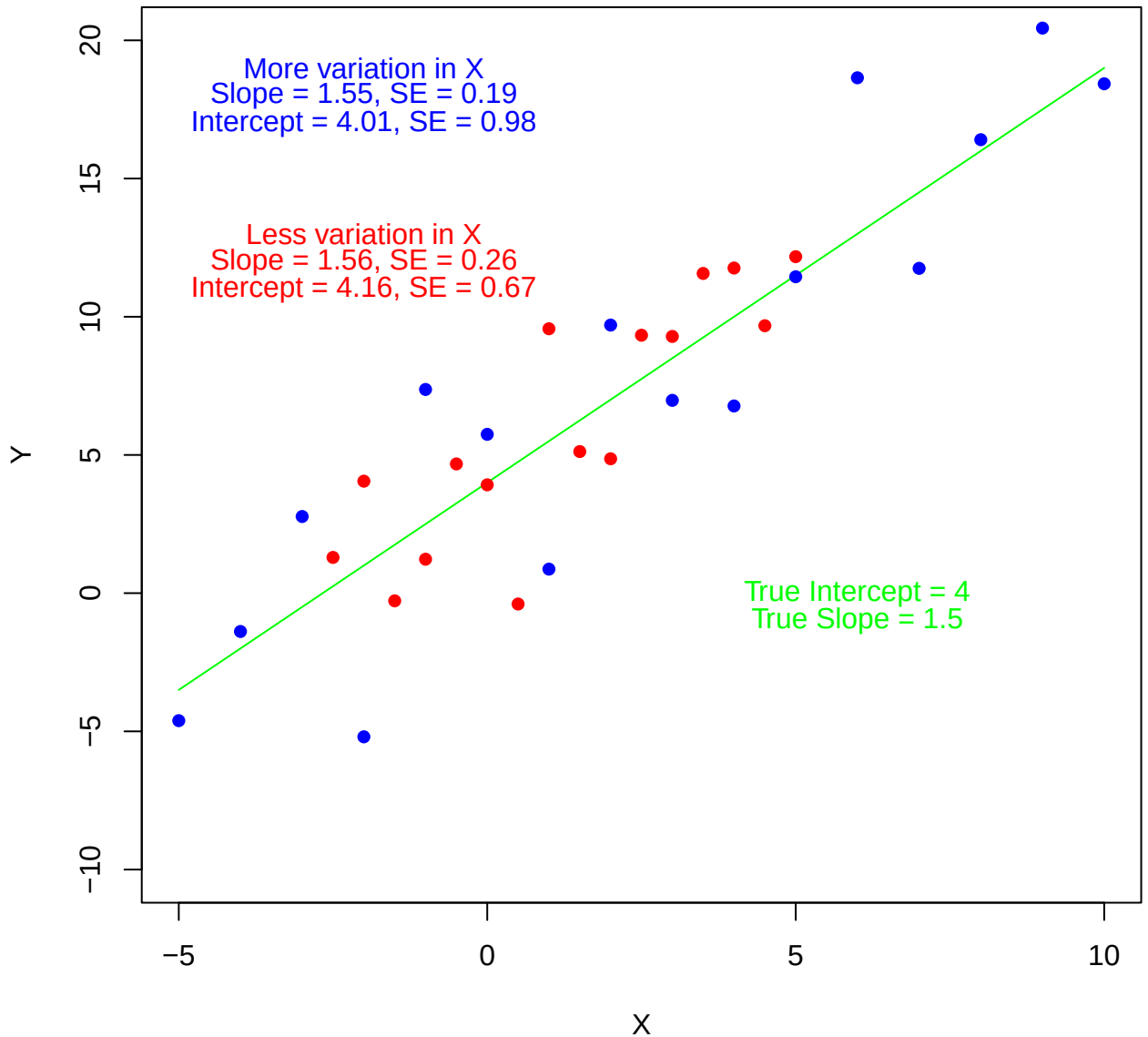
Effect of larger variance in X



Effect of larger variance in X



Effect of larger variance in X

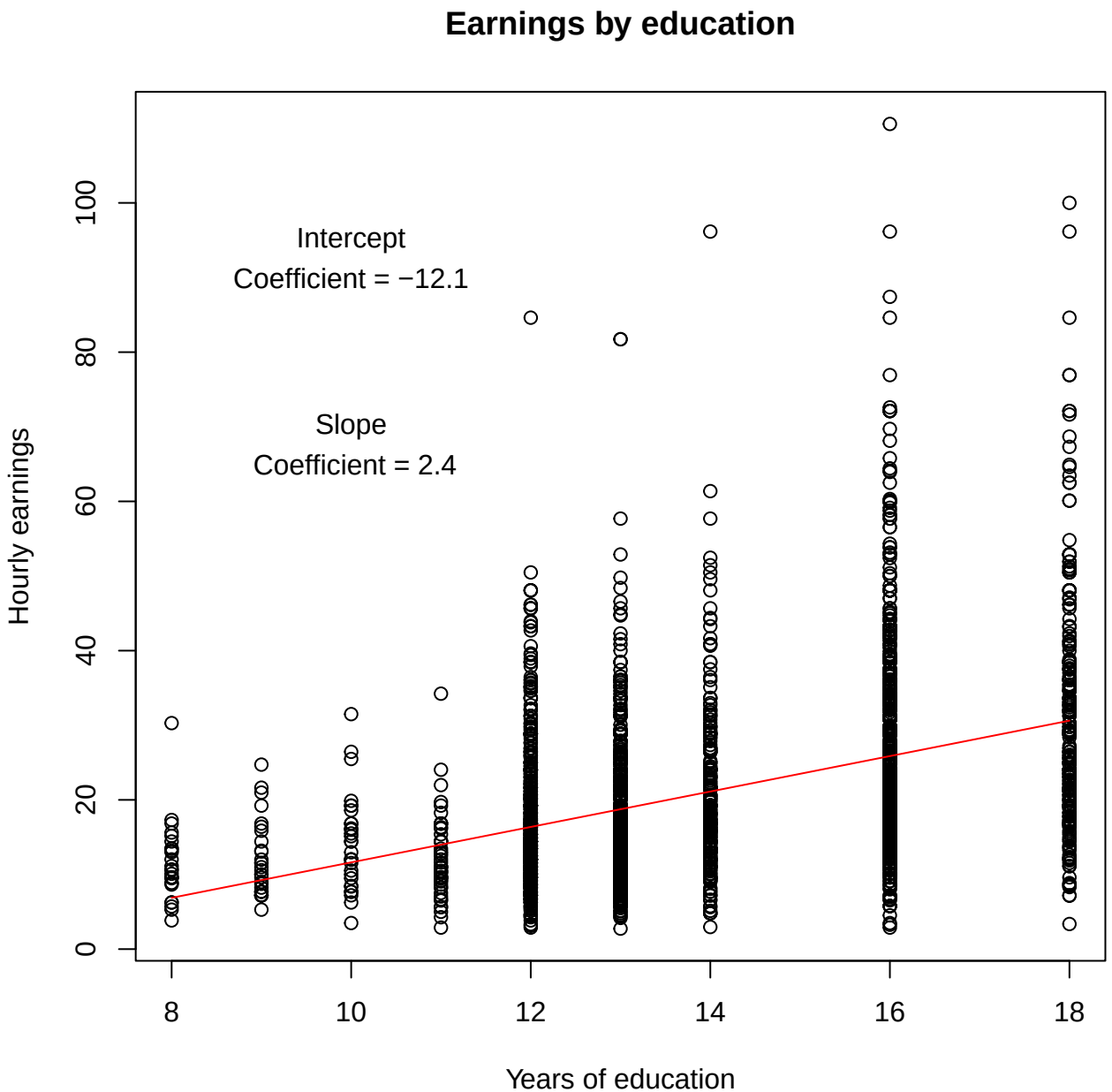


Don't assume homoskedasticity

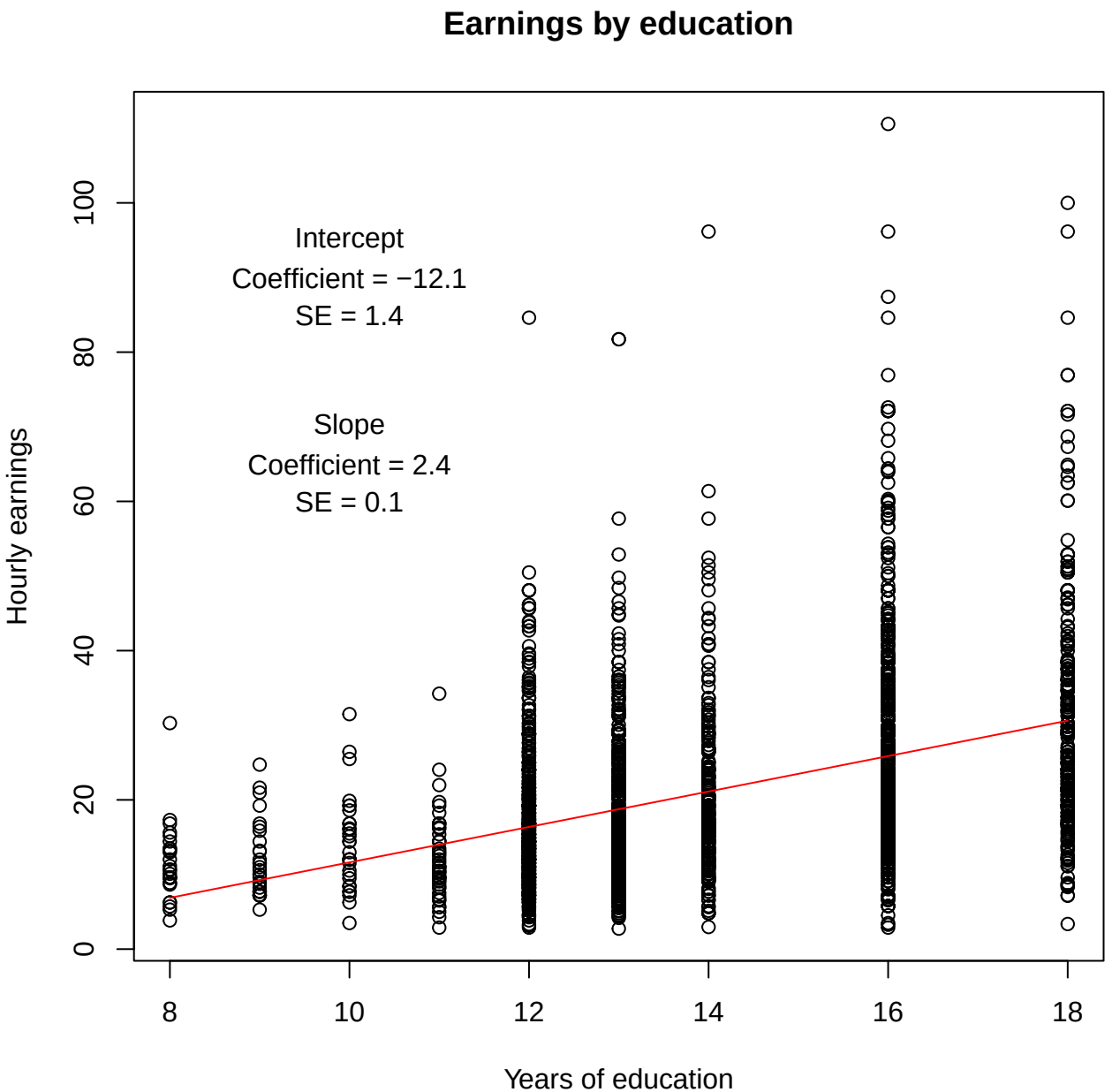
- It almost never makes sense to assume homoskedasticity in practice
- The formula below can handle both the hetero- and homoskedastic cases
  - It is more general than the homoskedastic formula
- As data typically are not homoskedastic, it's in appropriate to use homoskedastic-specific SEs
- We just discussed it for pedagogical purposes: simpler formula!

$$\hat{\sigma}^2_{\hat{\beta}_1} = \frac{1}{n} \cdot \frac{\frac{1}{n-2}\sum_{i=1}^{n}\left[(X_i - \bar{X})^2 \cdot \hat{u}_i^2\right]}{\left(\frac{1}{n}\sum_{i=1}^{n}\left[(X_i - \bar{X})^2\right]\right)^2}$$

Inference on coefficients in real data



Inference on coefficients in real data



Inference on coefficients in real data