Multivariate: Linear regression

Table of contents

1 Goals

1.1 Goals

1.1.1 Goals of this lecture

- Introduce the concept of **composites** and the **statistical operations** we can perform on them

- Review **linear regression**
- Summarize / review **ordinary least squares estimation**

2 Composites

2.1 Composites or linear combinations

2.1.1 Composites or linear combinations

All multivariate procedures (and most statistical procedures, in general) rely on **composites of variables**, also called **linear combinations** of variables

Statistical procedures **create** these linear combinations and then **do something** with them

- Usually **minimize** or **maximize** some quantity
  - Least squares estimation (*minimize* sum of squared residuals)
  - Maximum likelihoood (*maximize* likelihood function)

2.1.2 Composites

A **composite** or **linear combination** is a way to combine multiple variables into a single variable

To make a composite, you need **variables** and **weights**

Usually:

- One set of **weights** for all subjects (j subscript for variable j)
- Each subject has their own **variable** values (ij subscript for subject i and variable j)

2.1.3 Composites

In general, composites look like:

$$u_i = \Sigma a_j X_{ij} = a_1 X_{i1} + a_2 X_{i2} + \cdots + a_p X_{ip}$$

for subject i across variables $j = 1$ to p

- The a_js are the weights and the X_{ij}s are the variables

Remember:

- One set of **weights**, each subject has value for each **variable**
- One composite score for each subject (subscript i)

2.1.4 Examples of composites

Calculating GPA: Total of 18 units

- 5 unit class with an A: $\frac{5}{18}$ of the grade
- 4 unit class with a B: $\frac{4}{18}$ of the grade
- 4 unit class with a C: $\frac{4}{18}$ of the grade
- 5 unit class with a B: $\frac{5}{18}$ of the grade

Variables: A = 4.0, B = 3.0, C = 2.0

$GPA = \frac{5}{18}(4.0) + \frac{4}{18}(3.0) + \frac{4}{18}(2.0) + \frac{5}{18}(3.0) = 1.11 + 0.67 + 0.44 + 0.83 = 3.05$

2.1.5 Examples of composites

Predicted score for linear regression

- Three predictors (variables): X_1, X_2, and X_3
- Three regression coefficients (weights): b_1, b_2, b_3

$\hat{Y}_i = b_1 X_{i1} + b_2 X_{i2} + b_3 X_{i3}$

- Variables can vary across people (subscript i)
- Weights are the same for everyone (no subscipt i)
- Composite is **predicted value** for each person (subscript i)

2.1.6 Weights

The general strategy in multivariate analysis is to

- **Select a set of weights**
- That form a **composite**
- That leads to a **specific desired outcome**

For example: Least squares criterion for linear regression

- Desired outcome: Minimize the sum of the squared residuals
- **Choose** weights (b_js) that **minimize** $\Sigma(Y - \hat{Y})^2$

2.2 Composites in multivariate analysis

2.2.1 Composites in multivariate analysis

Composites are the basis for all multivariate analyses

Focus on the **relationship** between

- A statistic calculated on a **composite**
- A statistic calculated on the **individual measures** that go into the composite

We will do all of this in matrix algebra

2.2.2 Composites in multivariate analysis

Any statistic on a composite can be written as a composite of the corresponding statistics on the original variables (where the weights are the same)

One common example:

- The mean of a composite = the composite of the means of all the variables that went into the composite

2.3 Forming a composite

2.3.1 Form a composite, algebra-style

Subject 1: $u_1 = a_1X_{11} + a_2X_{12} + a_3X_{13} + \cdots + a_pX_{1p}$

Subject 2: $u_2 = a_1X_{21} + a_2X_{22} + a_3X_{23} + \cdots + a_pX_{2p}$

Subject n: $u_n = a_1X_{n1} + a_2X_{n2} + a_3X_{n3} + \cdots + a_pX_{np}$

- **Same** weights for all subjects
- **Different** variable values for each subject

- **Different** composite values for each subject

2.3.2 Form a composite, matrix-style

- Data matrix $\mathbf{X}$ with n subjects and p variables

Spreadsheet representation

Subject	X_1	$\cdots$	X_j	X_p
1	X_{11}	$\cdots$	X_{1j}	X_{1p}
2	X_{21}	$\cdots$	X_{2j}	X_{2p}
3	X_{31}	$\cdots$	X_{3j}	X_{3p}
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
n	X_{n1}	$\cdots$	X_{nj}	X_{np}

2.3.3 Form a composite, matrix-style

- Data matrix $\mathbf{X}$ is an $n \times p$ matrix

Matrix representation

$$\mathbf{X} = \begin{bmatrix} X_{11} & \cdots & X_{1j} & X_{1p} \\ X_{21} & \cdots & X_{2j} & X_{2p} \\ X_{31} & \cdots & X_{3j} & X_{3p} \\ \vdots & \ddots & \vdots & \vdots \\ X_{n1} & \cdots & X_{nj} & X_{np} \end{bmatrix}$$

2.3.4 Form a composite, matrix-style

Weight vector $\underline{a}$

- $\underline{a}$ is a $p \times 1$ vector
- One element per **variable**

$$\underline{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}$$

2.3.5 Form a composite, matrix-style

Composite vector $\underline{u}$

- $\underline{u}$ is an $n \times 1$ vector
- One element per **subject**

$$\underline{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} = \mathbf{X}\underline{a} = \begin{bmatrix} X_{11} & \cdots & X_{1j} & X_{1p} \\ X_{21} & \cdots & X_{2j} & X_{2p} \\ \vdots & \ddots & \vdots & \vdots \\ X_{n1} & \cdots & X_{nj} & X_{np} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}$$

2.4 Mean, variation, and variance of a composite

2.4.1 Mean of a composite

A **composite** is something like weighted GPA or predicted score in regression

- Calculated from variables (Xs) and weights (as)

If we wanted to get the **mean of a composite**, there are two *equivalent* ways to do that

1. Calculate each person's **composite**, then get the **mean** of those values [Last "Form a composite, matrix-style" slide]
2. Calculate the **mean** of each variable, then calculate the **composite** using those means [Next slide]

2.4.2 Mean of a composite

The mean of a composite is the composite of the means (of the variables that went into the composite)

$$\overline{\mathbf{U}} = \underline{\overline{X}}\ \underline{a}$$

Three Xs:

$$\overline{\mathbf{U}} = \begin{bmatrix} \overline{X}_1 & \overline{X}_2 & \overline{X}_3 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = a_1\,\overline{X}_1 + a_2\,\overline{X}_2 + a_3\,\overline{X}_3$$

2.4.3 Mean of a composite: Example

$$\mathbf{X} = \begin{bmatrix} 5 & 1 & 2 \\ 9 & 2 & 5 \\ 4 & 6 & 3 \\ 2 & 3 & 6 \end{bmatrix} \qquad \underline{a} = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}$$

2.4.4 Mean of a composite V1: Composite first, then mean

Step 1: Get the vector of composites $\underline{u} = \mathbf{X}\underline{a}$

$$\underline{u} = \mathbf{X}\underline{a} = \begin{bmatrix} 5 & 1 & 2 \\ 9 & 2 & 5 \\ 4 & 6 & 3 \\ 2 & 3 & 6 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix} =$$

$$\begin{bmatrix} (5 \times 2) + (1 \times 3) + (2 \times 1) \\ (9 \times 2) + (2 \times 3) + (5 \times 1) \\ (4 \times 2) + (6 \times 3) + (3 \times 1) \\ (2 \times 2) + (3 \times 3) + (6 \times 1) \end{bmatrix} = \begin{bmatrix} 15 \\ 29 \\ 29 \\ 19 \end{bmatrix}$$

2.4.5 Mean of a composite V1: Composite first, then mean

Step 2: Calculate the mean composite $\overline{\mathbf{U}}$ from $\underline{u}$

$$\overline{\mathbf{U}} = \tfrac{1}{n}\,\underline{1}'\,\underline{u} = \tfrac{1}{4}\begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 15 \\ 29 \\ 29 \\ 19 \end{bmatrix} =$$

$$\tfrac{1}{4}\,\big[(1 \times 15) + (1 \times 29) + (1 \times 29) + (1 \times 19)\big] =$$

$$\tfrac{1}{4}\,(92) = 23$$

2.4.6 Mean of a composite V2: Mean first, then composite

Step 1: Get the mean vector the variables $\underline{\overline{x}} = \frac{1}{n}\,\underline{1}'\,\mathbf{X}$

$$\underline{\overline{x}} = \tfrac{1}{n}\,\underline{1}'\,\mathbf{X} = \tfrac{1}{4}\begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 5 & 1 & 2 \\ 9 & 2 & 5 \\ 4 & 6 & 3 \\ 2 & 3 & 6 \end{bmatrix} =$$

$$\tfrac{1}{4}\,\big[(1 \times 5) + (1 \times 9) + (1 \times 4) + (1 \times 2) \quad (1 \times 1) + (1 \times 2) + (1 \times 6) + (1 \times 3) \quad (1 \times 2) + (1 \times 5) + (1 \times 3) + (1 \times$$

$\frac{1}{4}\ \begin{bmatrix} 5+9+4+2 & 1+2+6+3 & 2+5+3+6 \end{bmatrix} =$

$\frac{1}{4}\ \begin{bmatrix} 20 & 12 & 16 \end{bmatrix} = \begin{bmatrix} 5 & 3 & 4 \end{bmatrix}$

2.4.7 Mean of a composite V2: Mean first, then composite

Step 2: Calculate the mean composite $\overline{\mathbf{U}}$ from $\underline{\overline{x}}$

$\overline{\mathbf{U}} = \underline{\overline{x}}\ \underline{a} = \begin{bmatrix} 5 & 3 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix} =$

$\left[(5 \times 2) + (3 \times 3) + (4 \times 1)\right] = \left[10 + 9 + 4\right] = 23$

2.4.8 Variation of a composite 1

Variation of a single variable X:

$$SS_X = \underline{x}'\ \underline{x} - \frac{1}{n}\ \underline{x}'\ \mathbf{E}\ \underline{x}$$

Variation of a composite:

$$SS_u = \underline{u}'\ \underline{u} - \frac{1}{n}\ \underline{u}'\ \mathbf{E}\ \underline{u}$$

2.4.9 Variation of a composite 2

Substitute in the expression for a composite ($\underline{u} = \mathbf{X}\underline{a}$ or $\underline{u}' = \underline{a}'\mathbf{X}'$):

$$SS_u = \underline{a}'\ \mathbf{X}'\ \mathbf{X}\ \underline{a} - \frac{1}{n}\ \underline{a}'\ \mathbf{X}\ \mathbf{E}\ \mathbf{X}'\ \underline{a}$$

Factor out terms: **pre**-multipliers get **pre**-factored, **post**-multipliers get **post**-factored:

$$SS_u = \underline{a}' \left(\mathbf{X}'\ \mathbf{X} - \frac{1}{n}\ \mathbf{X}'\ \mathbf{E}\ \mathbf{X} \right)\ \underline{a}$$

2.4.10 Variation of a composite 3

Remember the *variation covariation matrix* $\mathbf{P}$:

$$\mathbf{P} = \mathbf{X}' \, \mathbf{X} - \frac{1}{n} \, \mathbf{X}' \, \mathbf{E} \, \mathbf{X}$$

Substitute P into the expression for variation of a composite:

$$SS_u = \underline{a}' \, \mathbf{P} \, \underline{a}$$

2.4.11 Variation of a composite 4

Variation of a composite $\underline{u}$: $SS_u = \underline{a}' \, \mathbf{P} \, \underline{a}$

Two important points:

1. We can calculate a statistic (mean, variation, variance) about a composite **without ever having to compute the composite $\underline{u}$ itself**
2. $\underline{a}' \, \mathbf{P} \, \underline{a}$ is called a **quadratic form**

- weight vector $\times$ matrix $\times$ weight vector
- quadratic = squared (e.g., $(X - \overline{X})^2$)

2.4.12 Variance of a composite

Variance of a composite $\underline{u}$:

$$s_u^2 = \underline{a}' \, \mathbf{S} \, \underline{a}$$

where $\mathbf{S}$ is the **variance covariance matrix**:

$$\mathbf{S} = \frac{1}{n-1} \, \left(\mathbf{X}' \, \mathbf{X} - \frac{1}{n} \, \mathbf{X}' \, \mathbf{E} \, \mathbf{X} \right) = \frac{1}{n-1} \, \mathbf{P}$$

2.4.13 So…

Why do we care about the mean and variance of composites?

Statistical procedures create composites and then

- **Do something** with them: usually *minimize* or *maximize*
- **Minimize** sum of squared residuals in least squares
- **Maximize** variance explained by a factor or component

Calculating the variance of the composite **directly** is computationally easier

Also, quadratic form will be helpful later

2.5 Multiple composites

2.5.1 Two composites on the same variables

.	Composite 1	Composite 2
Variables	$\mathbf{X}$	$\mathbf{X}$
Weights	$\underline{a}$	$\underline{c}$
Composite	$\underline{u} = \mathbf{X}\ \underline{a}$	$\underline{w} = \mathbf{X}\ \underline{c}$
Mean of composite	$\overline{U} = \overline{\underline{X}}\ \underline{a}$	$\overline{W} = \overline{\underline{X}}\ \underline{c}$
Variation of composite	$\underline{a}'\ \mathbf{P}_{XX}\ \underline{a}$	$\underline{c}'\ \mathbf{P}_{XX}\ \underline{c}$
Variance of composite	$\underline{a}'\ \mathbf{S}_{XX}\ \underline{a}$	$\underline{c}'\ \mathbf{S}_{XX}\ \underline{c}$
Covariation bet composites	$SP_{UW} =$	$\underline{a}'\ \mathbf{P}_{XX}\ \underline{c}$
Covariance bet composites	$s_{UW} =$	$\underline{a}'\ \mathbf{S}_{XX}\ \underline{c}$

2.5.2 Two composites on two sets of variables

.	Comp 1 on Xs	Comp 2 on Ys
Variables	$\mathbf{X}$	$\mathbf{Y}$
Weights	$\underline{a}$	$\underline{d}$
Composite	$\underline{u} = \mathbf{X}\ \underline{a}$	$\underline{z} = \mathbf{Y}\ \underline{d}$
Mean of composite	$\overline{U} = \overline{\underline{X}}\ \underline{a}$	$\overline{Z} = \overline{\underline{Y}}\ \underline{d}$
Variation of comp	$SS_U = \underline{a}'\ \mathbf{P}_{XX}\ \underline{a}$	$SS_Z = \underline{d}'\ \mathbf{P}_{YY}\ \underline{d}$
Variance of comp	$s^2_U = \underline{a}'\ \mathbf{S}_{XX}\ \underline{a}$	$s^2_Z = \underline{d}'\ \mathbf{S}_{YY}\ \underline{d}$
Covariation bet comp	$SP_{UZ} =$	$\underline{a}'\ \mathbf{P}_{XY}\ \underline{d}$
Covariance bet comp	$s_{UZ} =$	$\underline{a}'\ \mathbf{S}_{XY}\ \underline{d}$

3 Partitioned Matrices

3.1 Partitioned data matrix

3.1.1 Partitioned data matrix

$$\mathbf{M} = \begin{bmatrix} \mathbf{X} & \mathbf{Y} \end{bmatrix}$$

Order $(n, p+q)$: there are p X variables and q Y variables

Subjects	*Predictors*			*Outcomes*		
	X_1	...	X_p	Y_1	...	Y_q
1	X_{11}	...	X_{1p}	Y_{11}	...	Y_{1q}
...	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
n	X_{n1}	...	X_{np}	Y_{n1}	...	Y_{nq}

3.2 Partitioned covariation matrix

3.2.1 Partitioned covariation matrix

$$\mathbf{P}_{XX,YY} = \mathbf{M}'\,\mathbf{M} - \frac{1}{n}\mathbf{M}'\,\mathbf{E}\,\mathbf{M} = \left[\begin{array}{c|c} \mathbf{P}_{XX} & \mathbf{P}_{XY} \\ \hline \mathbf{P}_{YX} & \mathbf{P}_{YY} \end{array}\right]$$

$$= \left[\begin{array}{c|c} \mathbf{X}'\,\mathbf{X} - \frac{1}{n}\mathbf{X}'\,\mathbf{E}\,\mathbf{X} & \mathbf{X}'\,\mathbf{Y} - \frac{1}{n}\mathbf{X}'\,\mathbf{E}\,\mathbf{Y} \\ \hline \mathbf{Y}'\,\mathbf{X} - \frac{1}{n}\mathbf{Y}'\,\mathbf{E}\,\mathbf{X} & \mathbf{Y}'\,\mathbf{Y} - \frac{1}{n}\mathbf{Y}'\,\mathbf{E}\,\mathbf{Y} \end{array}\right]$$

3.2.2 Partitioned covariation matrix

$$\mathbf{P}_{XX,YY} =$$

$$\left[\begin{array}{cccc|cccc} SS_{x1} & SP_{x1,x2} & \cdots & SP_{x1,xp} & SP_{x1,y1} & SP_{x1,y2} & \cdots & SP_{x1,yq} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ SP_{xp,x1} & SP_{xp,x2} & \cdots & SS_{xp} & SP_{xp,y1} & SP_{xp,y2} & \cdots & SP_{xp,yq} \\ \hline SP_{y1,x1} & SP_{y1,x2} & \cdots & SP_{y1,xp} & SS_{y1} & SP_{y1,y2} & \cdots & SP_{y1,yq} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ SP_{yq,x1} & SP_{yq,x2} & \cdots & SP_{yq,xp} & SP_{yq,y1} & SP_{yq,y2} & \cdots & SS_{yq} \end{array}\right]$$

3.3 Partitioned covariance matrix

3.3.1 Partitioned covariance matrix

$$\mathbf{S}_{XX,YY} = \frac{1}{(n-1)}\left(\mathbf{M}'\,\mathbf{M} - \frac{1}{n}\mathbf{M}'\,\mathbf{E}\,\mathbf{M}\right) = \left[\begin{array}{c|c} \mathbf{S}_{XX} & \mathbf{S}_{XY} \\ \hline \mathbf{S}_{YX} & \mathbf{S}_{YY} \end{array}\right]$$

$$= \left[\begin{array}{c|c} \frac{1}{(n-1)}(\mathbf{X}'\,\mathbf{X} - \frac{1}{n}\mathbf{X}'\,\mathbf{E}\,\mathbf{X}) & \frac{1}{(n-1)}(\mathbf{X}'\,\mathbf{Y} - \frac{1}{n}\mathbf{X}'\,\mathbf{E}\,\mathbf{Y}) \\ \hline \frac{1}{(n-1)}(\mathbf{Y}'\,\mathbf{X} - \frac{1}{n}\mathbf{Y}'\,\mathbf{E}\,\mathbf{X}) & \frac{1}{(n-1)}(\mathbf{Y}'\,\mathbf{Y} - \frac{1}{n}\mathbf{Y}'\,\mathbf{E}\,\mathbf{Y}) \end{array}\right]$$

3.3.2 Partitioned covariance matrix

$$\mathbf{S}_{XX,YY} =$$

$$\left[\begin{array}{cccc|cccc} s_{x1}^2 & s_{x1,x2} & \cdots & s_{x1,xp} & s_{x1,y1} & s_{x1,y2} & \cdots & s_{x1,yq} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ s_{xp,x1} & s_{xp,x2} & \cdots & s_{xp}^2 & s_{xp,y1} & s_{xp,y2} & \cdots & s_{xp,yq} \\ \hline s_{y1,x1} & s_{y1,x2} & \cdots & s_{y1,xp} & s_{y1}^2 & s_{y1,y2} & \cdots & s_{y1,yq} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ s_{yq,x1} & s_{yq,x2} & \cdots & s_{yq,xp} & s_{yq,y1} & s_{yq,y2} & \cdots & s_{yq}^2 \end{array}\right]$$

3.4 Partitioned correlation matrix

3.4.1 Partitioned correlation matrix

$$\mathbf{R}_{XX,YY} = \left[\begin{array}{c|c} \mathbf{R}_{XX} & \mathbf{R}_{XY} \\ \hline \mathbf{R}_{YX} & \mathbf{R}_{YY} \end{array}\right] =$$

$$\left[\begin{array}{cccc|cccc} 1 & r_{x1,x2} & \cdots & r_{x1,xp} & r_{x1,y1} & r_{x1,y2} & \cdots & r_{x1,yq} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{xp,x1} & r_{xp,x2} & \cdots & 1 & r_{xp,y1} & r_{xp,y2} & \cdots & r_{xp,yq} \\ \hline r_{y1,x1} & r_{y1,x2} & \cdots & r_{y1,xp} & 1 & r_{y1,y2} & \cdots & r_{y1,yq} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{yq,x1} & r_{yq,x2} & \cdots & r_{yq,xp} & r_{yq,y1} & r_{yq,y2} & \cdots & 1 \end{array}\right]$$

4 Linear Regression

4.1 Regression review

4.1.1 Linear regression

Also called OLS (ordinary least squares) regression, normal regression, just "regression"

Data:

- 1 predictor variable, X
- 1 outcome variable, Y
- Measured on n subjects

Problem:

Find an equation that "best" summarizes the relationship between X and Y

4.1.2 Linear regression

4.1.3 Linear regression

4.1.4 Linear regression: $\hat{Y} = b_0 + b_1 X$

$$\widehat{weight} = -253.94 + 5.8 height$$

- b_0 is the predicted value of *weight* when $height = 0$
  - Predicted *weight* for a 0 inch tall person = -253.94
- For a 1-unit difference in X, we expect Y to differ by b_1 units
  - Expect 5.8 lb diff in *weight* for 1 inch diff in *height*

Each obs has one outcome value (Y_i), one predicted value ($\hat{Y}_i$), and one residual ($Y_i - \hat{Y}_i$)

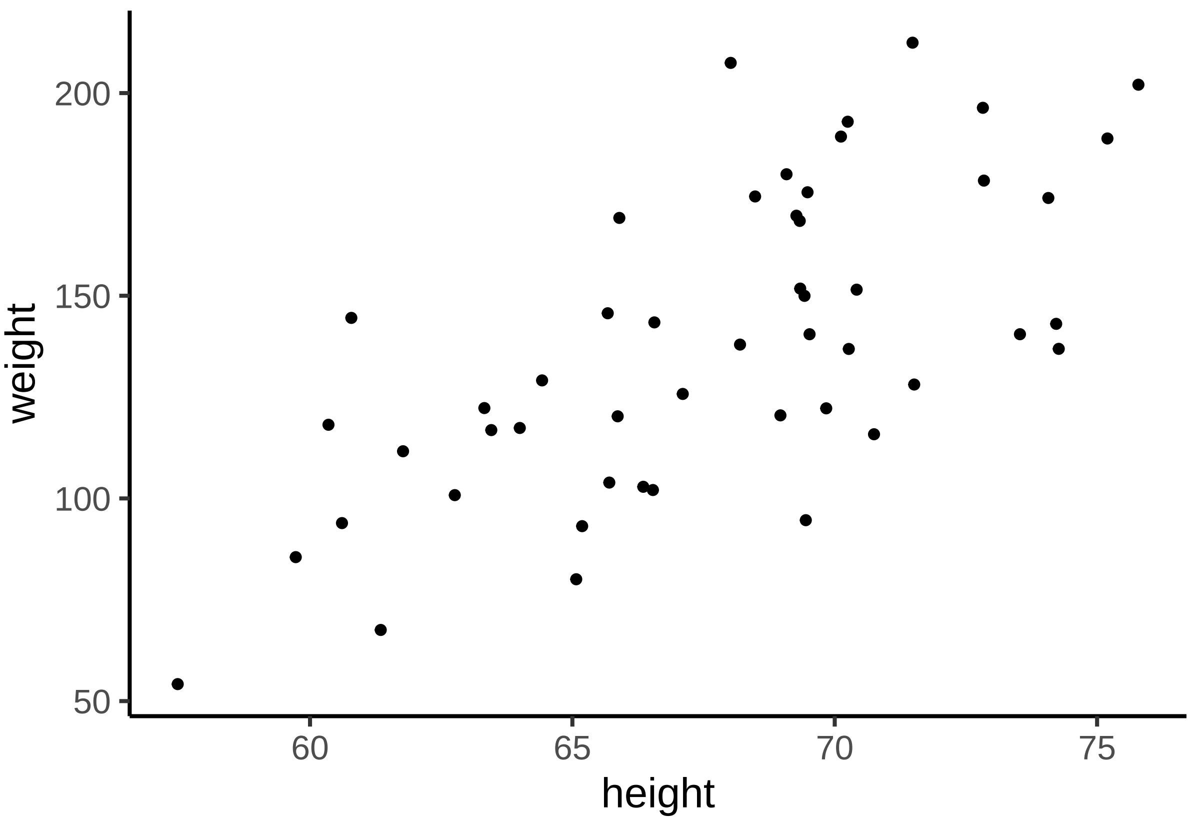


Figure 1: Relationship between height and weight

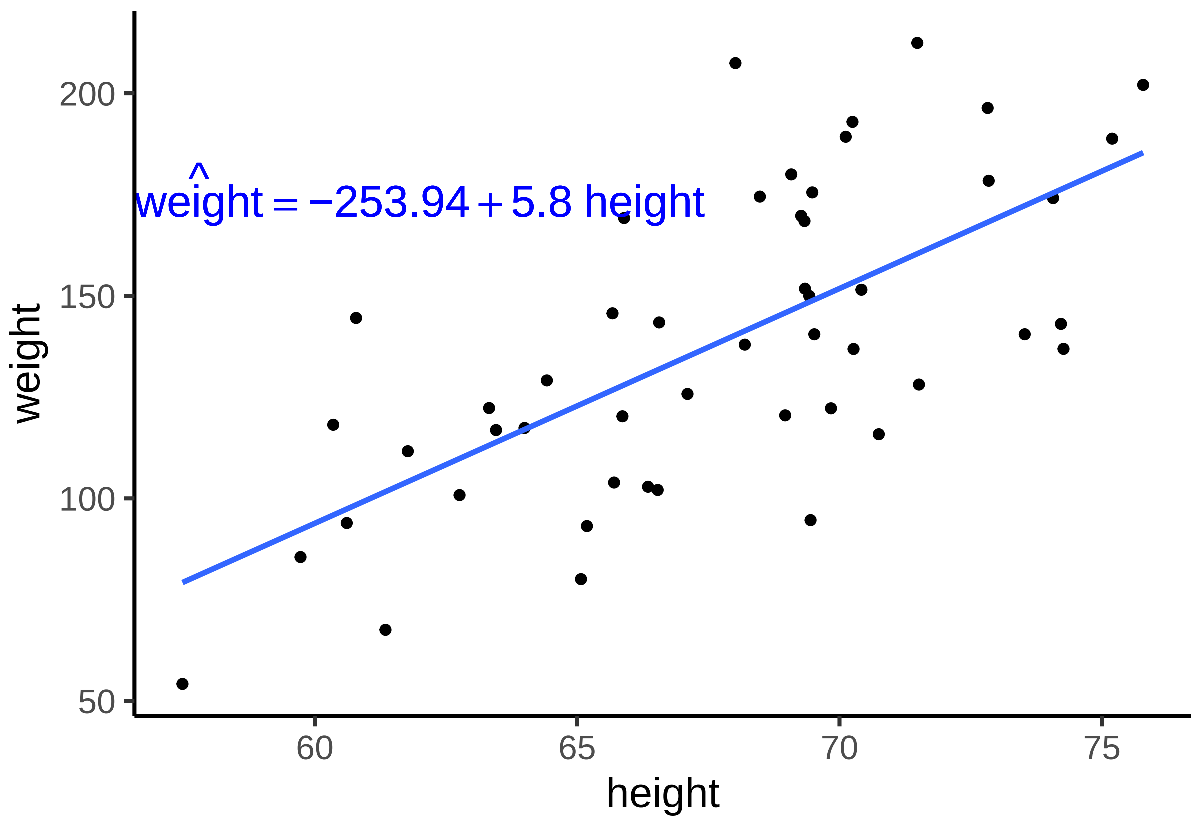


Figure 2: Relationship between height and weight with linear fit

4.2 Least squares estimation

4.2.1 Least squares estimation

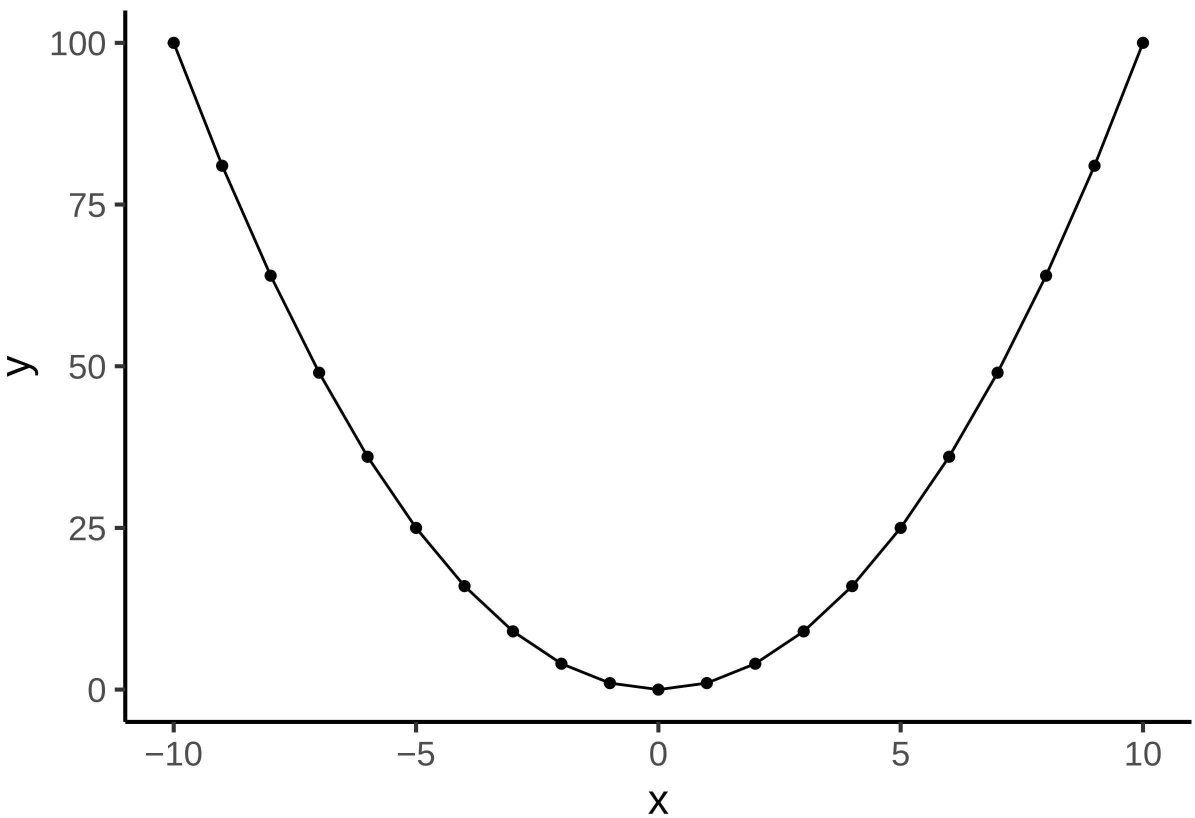
Least squares criterion:

- How we estimate the regression coefficients, b_0 and b_1
- Find b_0 and b_1 that give the smallest $\Sigma\left((Y_i - \hat{Y}_i)^2\right)$
- This is our "best fit" line

For linear regression, there is **one** value of b_0 and **one** value of b_1 that minimize the residuals

- This is not true for other methods of estimation that we'll look at later in this course]

4.2.2 $Y = X^2$

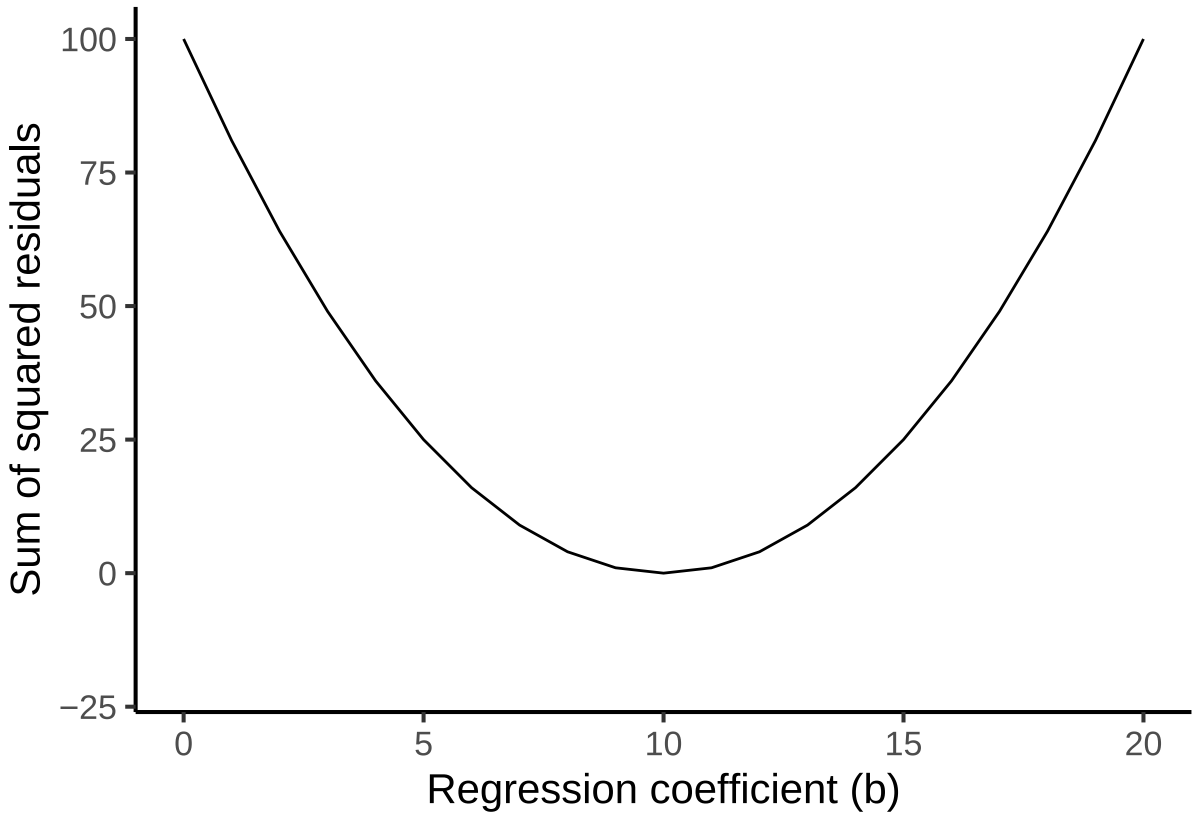


4.2.3 Functions involving squares

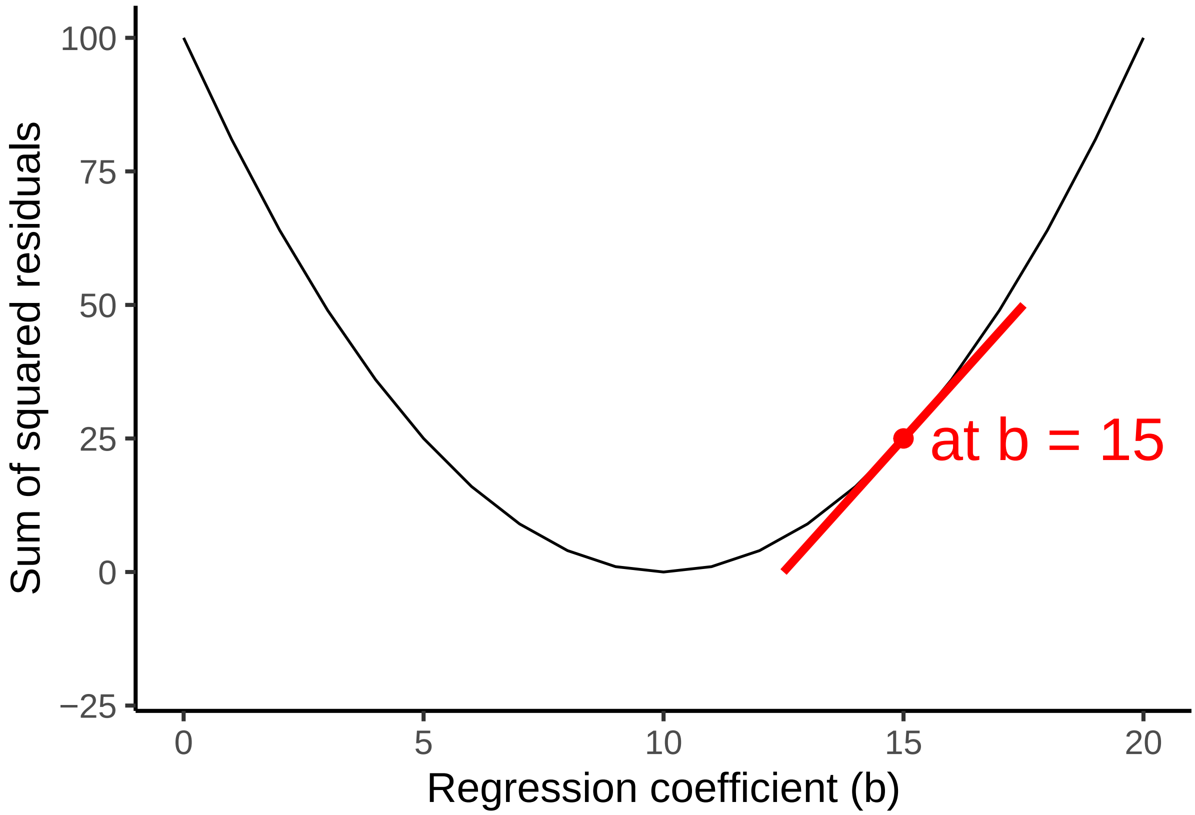
- Functions that have squares in them (like the sum of squared residuals) look like a "U"
  - To find the **minimum** of this function, we need to find the **bottom** of the "U"
- That happens using **calculus** (*which you don't need to know*)
- But you need to understand what is going on in the process

The **tangent line** is a line that touches a curve at a single point

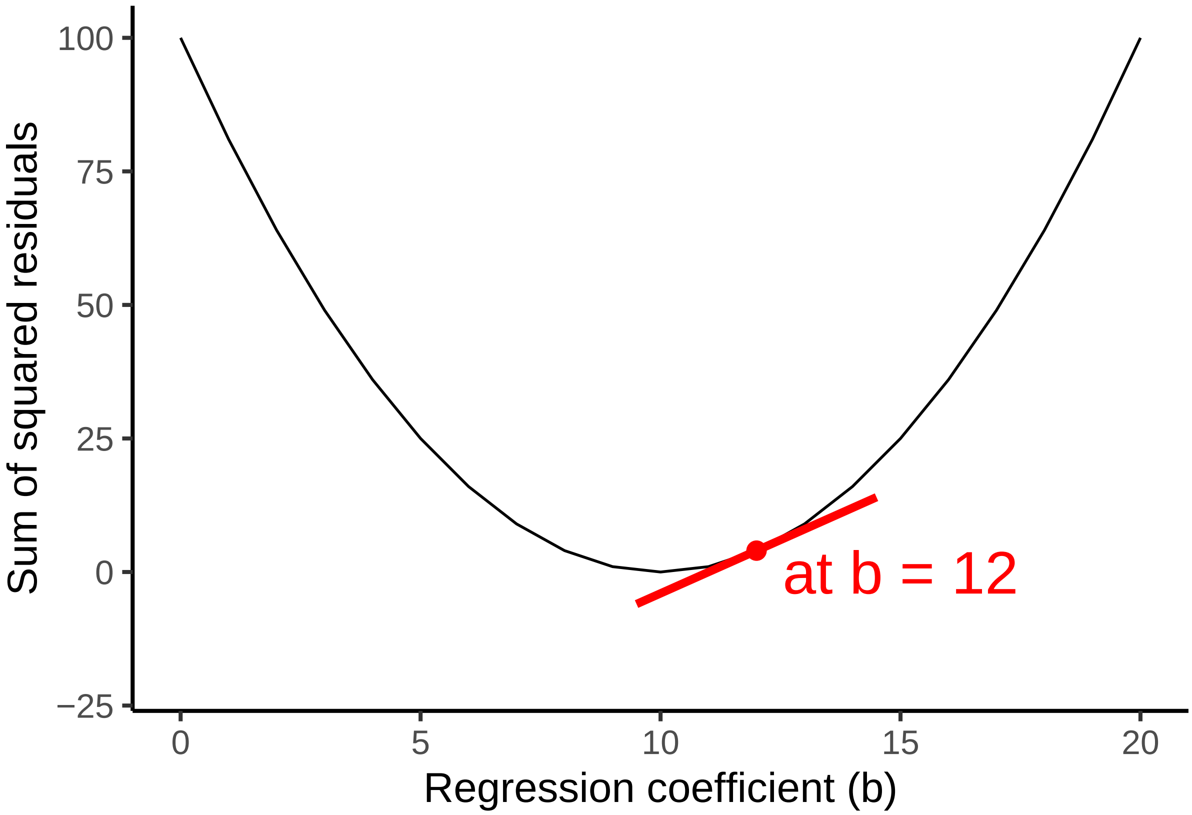
4.2.4 Calculus and tangents



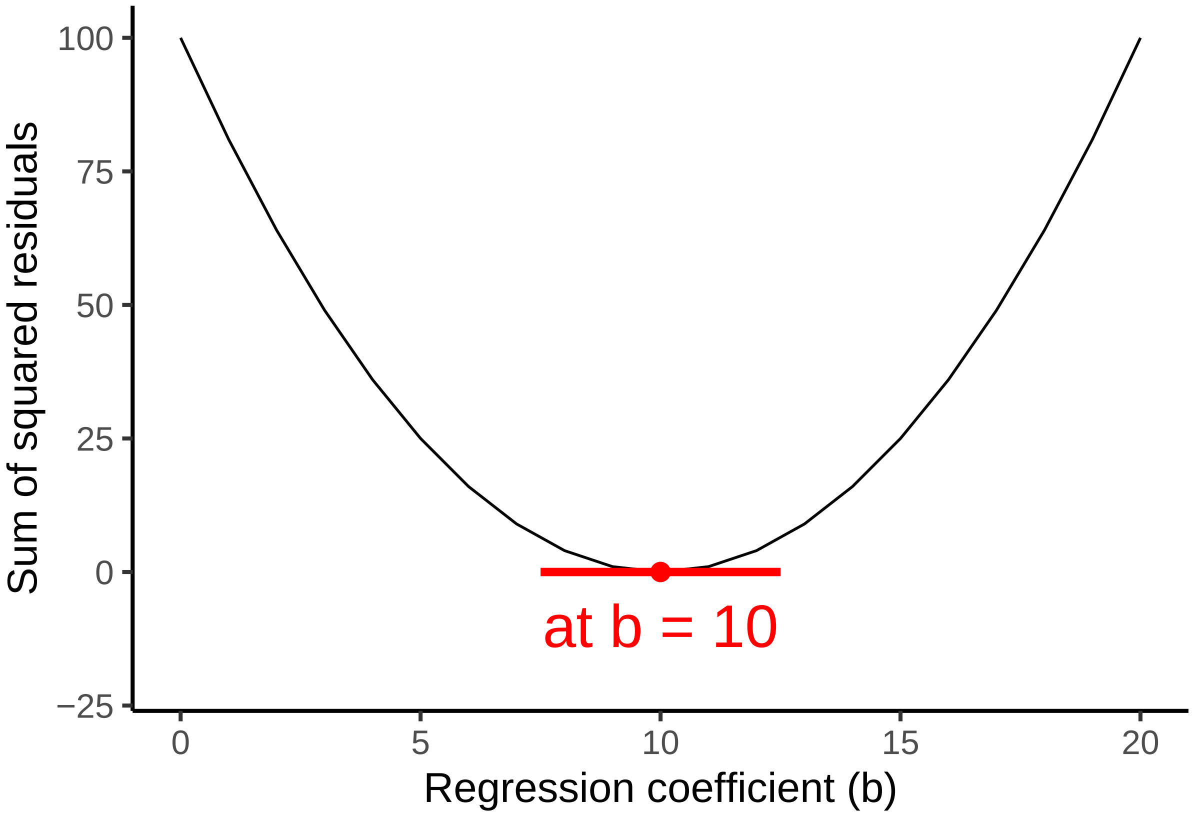
4.2.5 Calculus and tangents



4.2.6 Calculus and tangents



4.2.7 Calculus and tangents



4.2.8 Tangents and minimums

The tangent line is horizontal ($slope = 0$) at the minimum

We want to find the minimum of the **sum of squared residuals**

- We want to find where that tangent line is flat
- Where the tangent line is *flat* is the value of **regression coefficient** that meets the *least squares criterion*

We find the tangent line by using **calculus**

- The **derivative of a function** produces the **tangent line**

4.2.9 Least squares solution

1. State the **function** to be minimized
   - Here, it is the **sum of squared residuals**: $\Sigma(Y_i - \hat{Y}_i)^2$
2. **Differentiate** (take the derivative of) the function, with respect to the constants of interest
   - The constants of interest are b_0 and b_1 here
3. Set those derivatives **equal to 0**
   - These are called the "**normal equations**"
4. **Solve the normal equations** for the constants of interest

4.2.10 Step 1. Function to be minimized

$$\Sigma(Y_i - \hat{Y}_i)^2 =$$

$$\Sigma(Y - (b_1X + b_0))^2 =$$
$$\Sigma(Y - b_1X - b_0)^2 =$$

$$\Sigma(Y^2 + {b_0}^2 + {b_1}^2X^2 - 2b_0Y - 2b_1XY + 2b_0b_1X) =$$
$$\Sigma Y^2 + \Sigma {b_0}^2 + \Sigma {b_1}^2X^2 - \Sigma 2b_0Y - \Sigma 2b_1XY + \Sigma 2b_0b_1X =$$

$$\Sigma Y^2 + n{b_0}^2 + {b_1}^2\Sigma X^2 - 2b_0\Sigma Y - 2b_1\Sigma XY + 2b_0b_1\Sigma X$$

4.2.11 Step 2. Differentiate the functions

For b_1:

$$\frac{\partial \Sigma(Y - \hat{Y})^2}{\partial b_1} = 2b_1\Sigma X^2 - 2\Sigma XY + 2b_0\Sigma X$$

For b_0:

$$\frac{\partial \Sigma(Y - \hat{Y})^2}{\partial b_0} = 2nb_0 - 2\Sigma Y + 2b_1\Sigma X$$

4.2.12 Steps 3. and 4. Solve normal equations

For b_1:

$$2b_1\Sigma X^2 - 2\Sigma XY + 2b_0\Sigma X = 0$$
$$\vdots$$
$$b_1 = \frac{n\Sigma XY - (\Sigma X)(\Sigma Y)}{n\Sigma X^2 - (\Sigma X)^2} = \frac{SP_{XY}}{SS_X} = \frac{s_{XY}}{{s_X}^2}$$

4.2.13 Steps 3. and 4. Solve normal equations

For b_0:

$$2nb_0 - 2\Sigma Y + 2b_1\Sigma X = 0$$
$$\vdots$$
$$b_0 = \overline{Y} - b_1\overline{X}$$

4.3 Multiple regression

4.3.1 Multiple regression

The least squares solution gets more complex with more predictors (and thus more regression coefficients to solve for)

- But similar

Two predictor regression:

- Move from a **regression line** to a **regression plane**
- This requires some geometric thinking

4.3.2 Multiple correlation

- The multiple correlation is the correlation between Y and $\hat{Y}$
- If you used least squares estimation, the multiple correlation is the **maximum** possible correlation between Y and $\hat{Y}$
- The **square** of the multiple correlation (${R^2}_{multiple}$) tells you the **proportion of variation in Y that is accounted for by the set of predictors**
- ${R^2}_{multiple} = {r^2}_{Y\hat{Y}} = \frac{SS_{regression}}{SS_Y} = \frac{predictable\ variation}{total\ variation}$

4.3.3 Multiple regression and composites

Next week:

- The predicted score in multiple regression is a **composite** or **linear combination**
  - $\hat{Y} = b_0 + b_1X_1 + b_2X_2 + b_3X_3$
- From this **scalar** version of regression to the **matrix** version