Stat 306: Finding Relationships in Data.

Lecture 8

Section 3.4 + Section 3.6

Chapter 3

3.1 Least squares with two or more explanatory variables

3.4 Statistical software output for multiple regression
- R^2 and adjR^2 and 3.4.1 Properties of R^2 and σ^2
- Sum of squares decomposition

3.5 Important explanatory variables

3.6 Interval estimates and standard errors

3.7 Denominator of the residual SD

3.8 Residual plots

3.9 Categorical explanatory variables

3.10 Partial correlation

multiple linear regression

Age vs. Money

PREDICTOR *variables*

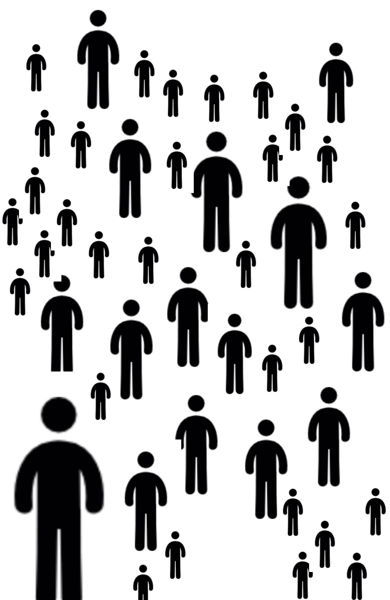
x_1 → Age in Years

x_2 → Income in thousands of $.

RESPONSE *variable*

Y → Cash in pocket dollars ($)

Population



Population parameters

$\beta_0, \beta_1, \beta_2, \sigma^2$

Hypothesis Test

$H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$

Sample statistics

$b_0 = 23.26$

$b_1 = 0.68$

$b_2 = -0.28$

$s = 13.9$

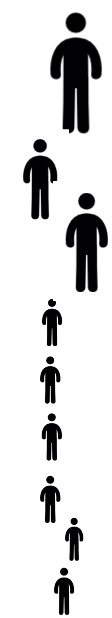
$R^2 = 0.65$

For parameter β_1 :

95% C.I. = [0.18, 1.18]

p-value = 0.016

Sample, n=9



x_1	x_2	y
82	26	71
45	49	54
71	76	43
22	37	45
29	40	21
9	0	11
12	2	30
18	10	45
24	92	10

multiple linear regression

Age vs. Money

PREDICTOR variables

covariates

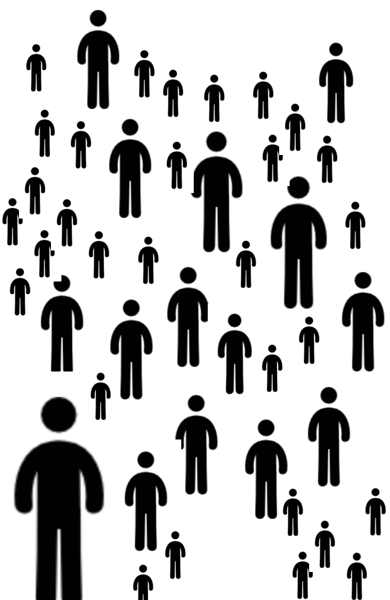
x_1 → Age in Years

x_2 → Income in thousands of $.

RESPONSE variable

Y → Cash in pocket dollars ($)

Population



Population parameters

$$\beta_0, \beta_1, \beta_2, \sigma^2$$

Hypothesis Test

$$H_0 : \beta_1 = 0$$
$$H_1 : \beta_1 \neq 0$$

Sample statistics

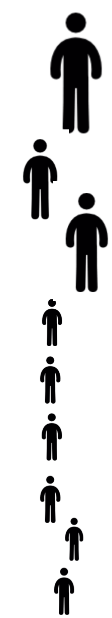
$b_0 = 23.26$
$b_1 = 0.68$
$b_2 = -0.28$
$s = 13.9$
$R^2 = 0.65$

For parameter β_1 :

95% C.I. = [0.18, 1.18]

p-value = 0.016

Sample, n=9



x_1	x_2	y
82	26	71
45	49	54
71	76	43
22	37	45
29	40	21
9	0	11
12	2	30
18	10	45
24	92	10

multiple linear regression

Age vs. Money

PREDICTOR variables

x_1 → Age in Years

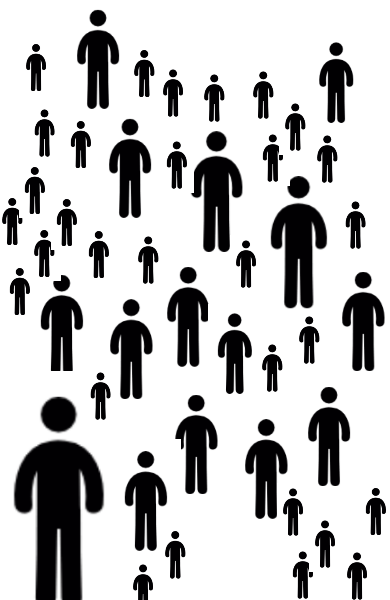
x_2 → Income in thousands of $.

RESPONSE variable

Y → Cash in pocket dollars ($)

p = 2
k = 3

Population



Population parameters

$\beta_0, \beta_1, \beta_2, \sigma^2$

Hypothesis Test

$$H_0 : \beta_1 = 0$$
$$H_1 : \beta_1 \neq 0$$

Sample statistics

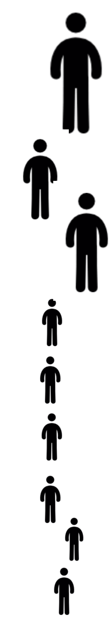
b_0 = 23.26
b_1 = 0.68
b_2 = -0.28
s = 13.9
R^2 = 0.65

For parameter β_1 :

95% C.I. = [0.18, 1.18]

p-value = 0.016

Sample, n=9



x_1	x_2	y
82	26	71
45	49	54
71	76	43
22	37	45
29	40	21
9	0	11
12	2	30
18	10	45
24	92	10

multiple linear regression

Age vs. Money

PREDICTOR *variables*

x_1 → Age in Years

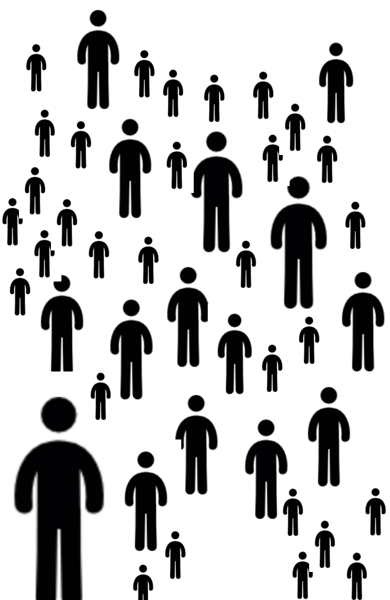
x_2 → Income in thousands of $.

RESPONSE *variable*

Y → Cash in pocket dollars ($)

$b=(X^TX)^{-1}X^Ty$

Population



Population parameters

$\beta_0, \beta_1, \beta_2, \sigma^2$

Hypothesis Test

$H_0 : \beta_1 = 0$

$H_1 : \beta_1 \neq 0$

Sample statistics

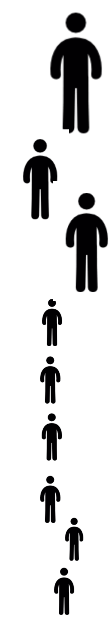
b_0 = 23.26
b_1 = 0.68
b_2 = -0.28
s = 13.9
R^2 = 0.65

For parameter β_1 :

95% C.I. = [0.18, 1.18]

p-value = 0.016

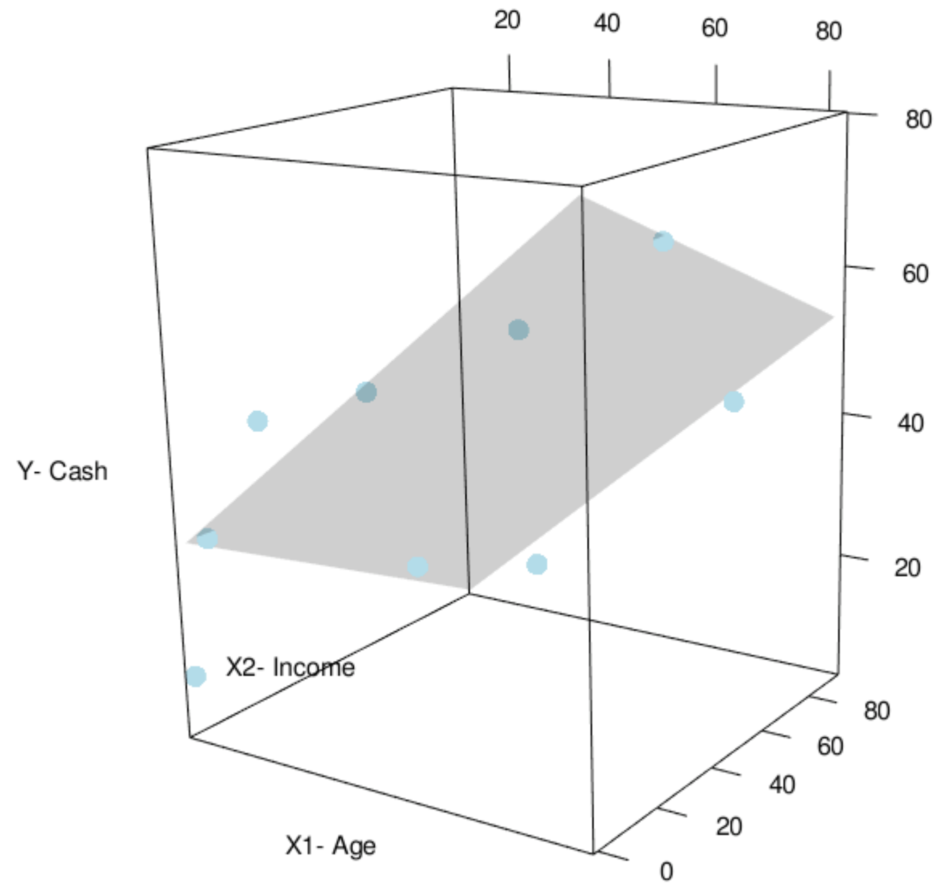
Sample, n=9



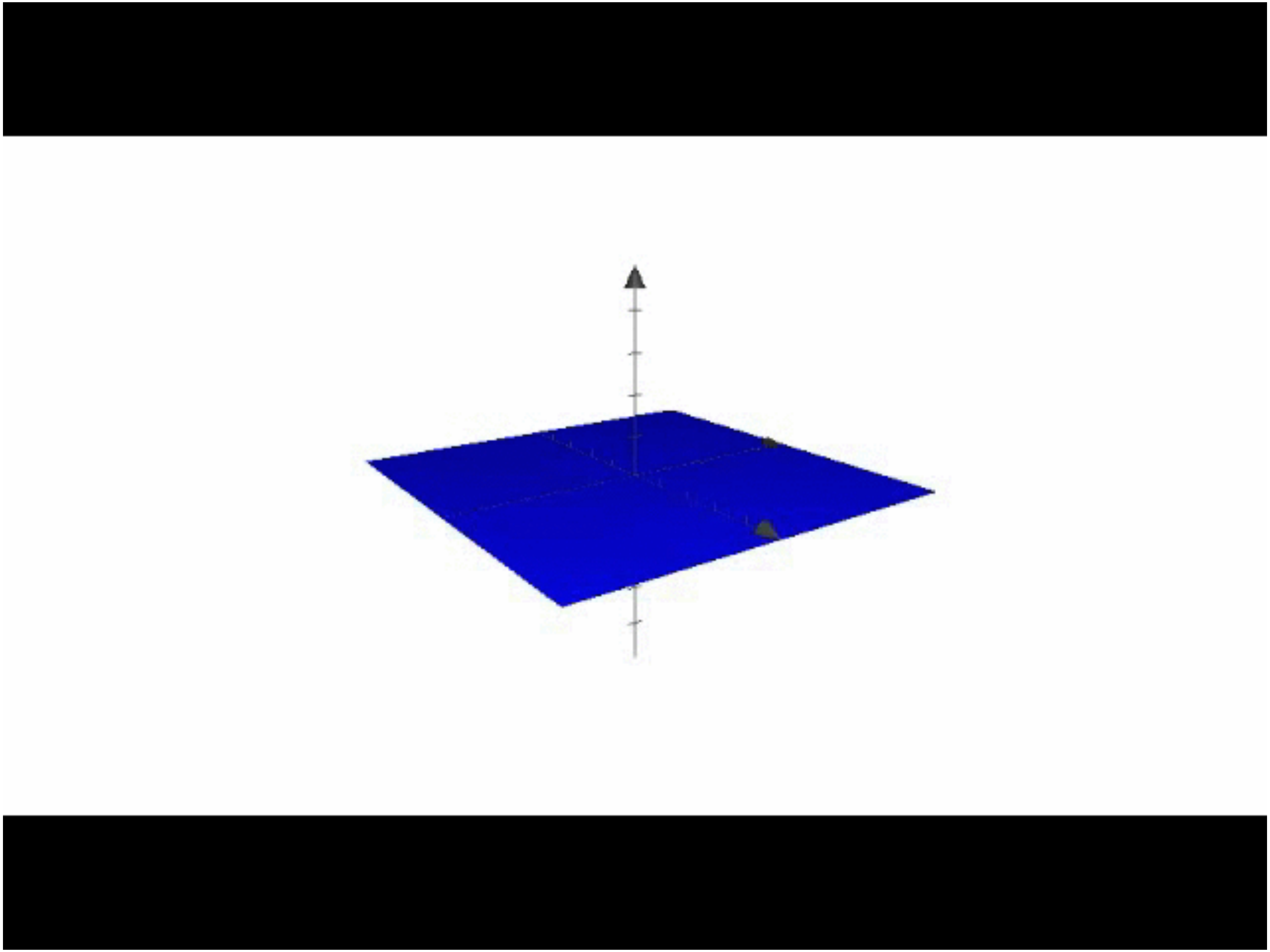
x_1	x_2	y
82	26	71
45	49	54
71	76	43
22	37	45
29	40	21
9	0	11
12	2	30
18	10	45
24	92	10

3.1 Least squares with two or more explanatory variables

"hyperplane equation"



https://commons.wikimedia.org/wiki/File:2d_multiple_linear_regression.gif



```
> x1 <- c(82, 45, 71, 22, 29, 9, 12, 18, 24)
> x2 <- c(26, 49, 76, 37, 40,  0,  2, 10, 92)
> y <- c(71, 54, 43, 45, 21, 11, 30, 45, 10)
>
> X <- matrix(c(1,1,1,1,1,1,1,1,1,x1,x2),nrow=n,ncol=3)
> Xt <- t(X)
> y <- matrix(y, nrow=n, ncol=1)
> dim(X)
[1] 9 3
> dim(Xt)
[1] 3 9
> dim(y)
[1] 9 1


> k <- dim(X)[2]
> k
[1] 3
>
> betahat <- solve(Xt %*% X) %*% Xt %*% y
> c(betahat)
[1] 23.2660767  0.6814606 -0.2771398
>
> yhat <- X%*%betahat
> c(yhat)
[1] 71.94021 40.35196 50.58716 28.00404 31.94284 29.39922 30.88932
[8] 32.76097 14.12427
>
> residuals <- y - yhat
> c(residuals)
[1]  -0.9402139  13.6480442  -7.5871581  16.9959612 -10.9428438
[6] -18.3992224  -0.8893247  12.2390298  -4.1242722
>
```

3.4 Statistical software output for multiple regression

- Sum of squares of residuals

$$SS(Res) = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2. \quad (3.41)$$

- Mean square of residuals or estimated σ^2:

Typo: should be n-2

$$\hat{\sigma}^2 = (n-k)^{-1} \sum_{i=1}^{n} e_i^2 = (n-k)^{-1} \sum_{i=1}^{n} (e_i - \bar{e})^2 = \frac{SS(Res)}{(n-k)} = MS(Res). \quad (3.42)$$

The residual standard deviation (called residual standard error in R output) is the sample standard deviation of the residuals with a denominator of $n-k$ instead of ~~$n-1$.~~ A mathematical explanation of this denominator is given in Section 3.7. A property of the residuals after a least squares fit is that

$$\bar{e} = n^{-1} \sum_{i=1}^{n} e_i = 0 \quad (3.43)$$

3.4 Statistical software output for multiple regression

- Sum of squares of residuals

$$SS(Res) = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2. \qquad (3.41)$$

S

- Mean square of residuals or estimated σ^2:

Typo: should be n-2

$$\hat{\sigma}^2 = (n-k)^{-1} \sum_{i=1}^{n} e_i^2 = (n-k)^{-1} \sum_{i=1}^{n} (e_i - \bar{e})^2 = \frac{SS(Res)}{(n-k)} = MS(Res). \qquad (3.42)$$

The residual standard deviation (called residual standard error in R output) is the sample standard deviation of the residuals with a denominator of $n-k$ instead of ~~$n-1$~~. A mathematical explanation of this denominator is given in Section 3.7. A property of the residuals after a least squares fit is that

$$\bar{e} = n^{-1} \sum_{i=1}^{n} e_i = 0 \qquad (3.43)$$

```
> SS_res <- sum(residuals^2)
> SS_res
[1] 1159.452
>
> MS_res <- SS_res/(n-k)
> MS_res
[1] 193.2421
```

3.4 Statistical software output for multiple regression

- Total sum of squares for y about its mean, or numerator of sample variance of y:

$$SS(Total) = \sum_{i=1}^{n}(y_i - \overline{y})^2 = (n-1)s_y^2. \tag{3.44}$$

- Multiple correlation coefficient or coefficient of determination :

$$R^2 \stackrel{\text{def}}{=} 1 - \frac{SS(Res)}{SS(Total)}, \tag{3.45}$$

$$\text{adj}R^2 \stackrel{\text{def}}{=} 1 - \frac{SS(Res)/(n-k)}{SS(Total)/(n-1)} = 1 - \frac{\hat{\sigma}^2}{s_y^2}. \tag{3.46}$$

R^2 measures the proportion of total variation in the y-variable about $\overline{y}$ explained by the regression; a better fitting regression model leads to a smaller value of $SS(Res)$ and larger value of R^2. The adjusted R^2 makes an adjustment to R^2 so that it is not always increasing with additional explanatory variables. Note that $R^2 \geq 0$ but adjR^2 could be a little negative when the model is a bad fit.

3.4 Statistical software output for multiple regression

- Total sum of squares for y about its mean, or numerator of sample variance of y:

$$SS(Total) = \sum_{i=1}^{n}(y_i - \overline{y})^2 = (n-1)s_y^2. \tag{3.44}$$

- Multiple correlation coefficient or coefficient of determination :

$$R^2 \overset{\text{def}}{=} 1 - \frac{SS(Res)}{SS(Total)}, \tag{3.45}$$

$$\text{adj}R^2 \overset{\text{def}}{=} 1 - \frac{SS(Res)/(n-k)}{SS(Total)/(n-1)} = 1 - \frac{\hat{\sigma}^2}{s_y^2}. \tag{3.46}$$

R^2 measures the proportion of total variation in the y-variable about $\overline{y}$ explained by the regression; a better fitting regression model leads to a smaller value of $SS(Res)$ and larger value of R^2. The adjusted R^2 makes an adjustment to R^2 so that it is not always ... Note that $R^2 \geq 0$ but adjR^2 could be a little negative

```
> SS_Total <- sum((y-ybar)^2)
> SS_Total
[1] 3318
> (n-1)*sy^2
[1] 3318
>
```

```
> R2 <- 1 - SS_Res/SS_Total
> R2
[1] 0.6505569
>
> adjR2 <- 1 - (SS_Res/(n-k))/(SS_Total/(n-1))
> adjR2
[1] 0.5340758
> 1 - MS_Res/sy^2
[1] 0.5340758
```

3.4 Statistical software output for multiple regression

3.4.1 Properties of R^2 and $\hat{\sigma}^2$

When comparing multiple regression equations with different sets of explanatory variables, larger adjR^2 and smaller $\hat{\sigma}^2$ indicate better prediction equations. Note from (3.46), that as $\hat{\sigma}^2$ decreases, then adjR^2 increases. The results below shows what can happen when additional explanatory variables are included.

1. $0 \le R^2 \le 1$: boundary cases (i) $R^2 = 1$ for perfect fit; (ii) $R^2 = 0$ for a fit that is not useful.

   (i) $e_i = 0 \ \forall i$ and $SS(Res) = 0$.

   (ii) $\hat{\beta}_1 = \cdots = \hat{\beta}_p = 0$ so that $\hat{\beta}_0 = \overline{y} - \hat{\beta}_1\overline{x}_1 - \cdots - \hat{\beta}_p x_p = \overline{y}$, $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_p x_{ip} = \overline{y}$. $e_i = y_i - \hat{y}_i = y_i - \overline{y}$ and $SS(Res) = SS(Total)$.

2. With additional explanatory variables, $SS(Res)$ decreases, R^2 increases, $\hat{\sigma}^2$ need not decrease, adjR^2 need not increase.

3.4 Statistical software output for multiple regression

Although (3.45) is a mathematical definition of R^2, there are alternative forms that give useful interpretations. R^2 is also the square of a correlation coefficient in the following senses.

1. R^2 is the sample squared correlation of $\hat{y}_i$ and y_i, that is,

   (3.58) $$R^2 = \frac{\{\sum_{i=1}^n (y_i - \overline{y})(\hat{y}_i - \overline{\hat{y}})\}^2}{\sum_{i=1}^n (y_i - \overline{y})^2 \cdot \sum_{i=1}^n (\hat{y}_i - \overline{\hat{y}})^2},$$

   where $\overline{\hat{y}} = n^{-1} \sum_{i=1}^n \hat{y}_i$.

2. $R_{y;(x_1,\ldots,x_p)}$ is the maximum correlation between $\{y_i\}$ and $\{b_1 x_{i1} + \cdots + b_p x_{ip}\}$ over choices of $(b_1, \ldots, b_p)$. That is, $\{\hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_p x_{ip}\}$ has maximum correlation with $\{y_i\}$, where $\hat{\beta}_1, \ldots, \hat{\beta}_p$ are the least squares coefficients.

```
> (sum((y-ybar)*(yhat-mean(yhat))))^2 / (sum((y-ybar)^2)*sum((yhat-mean(yhat))^2))
[1] 0.6505569
> cor(y,yhat)^2
          y
y 0.6505569
>
```

3.4.1 Properties of R^2 and $\hat{\sigma}^2$

When comparing multiple regression equations with different sets of explanatory variables, larger adjR^2 and smaller $\hat{\sigma}^2$ indicate better prediction equations. Note from (3.46), that as $\hat{\sigma}^2$ decreases, then adjR^2 increases. The results below shows what can happen when additional explanatory variables are included.

1. $0 \le R^2 \le 1$: boundary cases (i) $R^2 = 1$ for perfect fit; (ii) $R^2 = 0$ for a fit that is not useful.

   (i) $e_i = 0\ \forall i$ and $SS(Res) = 0$.

   (ii) $\hat{\beta}_1 = \cdots = \hat{\beta}_p = 0$ so that $\hat{\beta}_0 = \overline{y} - \hat{\beta}_1\overline{x}_1 - \cdots - \hat{\beta}_p x_p = \overline{y}$, $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_p x_{ip} = \overline{y}$. $e_i = y_i - \hat{y}_i = y_i - \overline{y}$ and $SS(Res) = SS(Total)$.

2. With additional explanatory variables, $SS(Res)$ decreases, R^2 increases, $\hat{\sigma}^2$ need not decrease, adjR^2 need not increase.

(b) $\hat{\sigma}^2(x_1) = SS(Res; x_1)/(n-2)$, $\hat{\sigma}^2(x_1, x_2) = SS(Res; x_1, x_2)/(n-3)$. With additional explanatory variables, the numerator of $\hat{\sigma}^2$ decreases but so does the denominator. If the additional explanatory variables have marginal prediction power, then $SS(Res)$ decreases only marginally but the denominator decreases more and $\hat{\sigma}^2$ increases in this case.

multiple linear regression

Age vs. Money

PREDICTOR *variables*

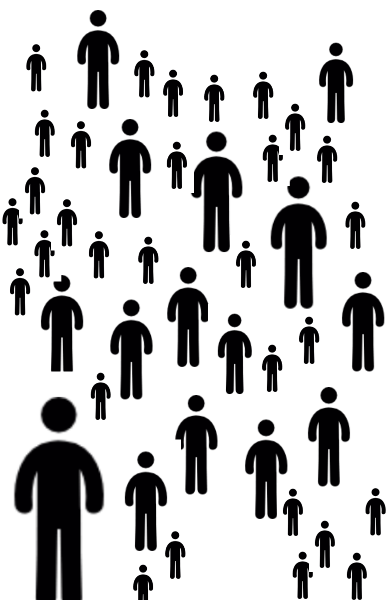
x_1 → Age in Years

x_2 → Income in thousands of $.

RESPONSE *variable*

Y → Cash in pocket dollars ($)

Population



Population parameters

$$\beta_0, \beta_1, \beta_2, \sigma^2$$

Hypothesis Test

$$H_0 : \beta_1 = 0$$
$$H_1 : \beta_1 \neq 0$$

Sample statistics

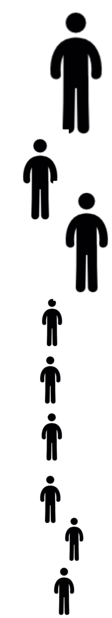
b_0 = 23.26
b_1 = 0.68
b_2 = -0.28
s = 13.9
R^2 = 0.65

For parameter β_1 :

95% C.I. = [0.18, 1.18]

p-value = 0.016

Sample, n=9



x_1	x_2	y
82	26	71
45	49	54
71	76	43
22	37	45
29	40	21
9	0	11
12	2	30
18	10	45
24	92	10

3.4.2 Sum of squares decomposition

In this section, the sum of squares decomposition for multiple regression is introduced. These decompositions are useful for *analysis of variance* associated with multiple regression for experimental data (topic of a follow-up statistics course).

In (3.45), there are two sum of squares terms: $SS(Res)$ and $SS(Total)$. Define $SS(Reg) = SS(Total) - SS(Res)$ as the *regression sum of squares*. Then $R^2 = 1 - SS(Res)/SS(Total) = SS(Reg)/SS(Total)$. It turns out that there is an identity

$$\sum_{i=1}^{n}(y_i - \overline{y})^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2. \tag{3.59}$$

and then

$$SS(Reg) = SS(Total) - SS(Res) = \sum_{i=1}^{n}(y_i - \overline{y})^2 - \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2. \tag{3.60}$$

In order for (3.59) to be valid

$$\sum_{i=1}^{n}(y_i - \overline{y})^2 = \sum_{i=1}^{n}([y_i - \hat{y}_i] + [\hat{y}_i - \overline{y}])^2 \tag{3.61}$$

$$= \sum_{i=1}^{n}[y_i - \hat{y}_i]^2 + \sum_{i=1}^{n}[\hat{y}_i - \overline{y}]^2 + 2\sum_{i=1}^{n}[y_i - \hat{y}_i][\hat{y}_i - \overline{y}]. \tag{3.62}$$

the cross-product term

$$\sum_{i=1}^{n}[y_i - \hat{y}_i][\hat{y}_i - \overline{y}] = 0. \tag{3.63}$$

```
> SS_Reg <- SS_Total - SS_Res
> SS_Reg
[1] 2158.548
> sum((yhat-ybar)^2)
[1] 2158.548
```

Chapter 3

3.1 Least squares with two or more explanatory variables

3.4 Statistical software output for multiple regression
- R^2 and adjR^2 and 3.4.1 Properties of R^2 and σ^2
- Sum of squares decomposition

3.5 Important explanatory variables

3.6 Interval estimates and standard errors

3.7 Denominator of the residual SD

3.8 Residual plots

3.9 Categorical explanatory variables

3.10 Partial correlation

Step 0: From θ, define estimator, $\hat{\theta}$

Step 1: Consider the sample statistic, $\hat{\theta}$, as a random variable $\hat{\Theta}$

Step 2: Determine $E[\hat{\Theta}]$ (to confirm it's unbiased) $Var[\hat{\Theta}]$ (to calculate se)

Step 3: Define se($\hat{\theta}$) = estimate of $\sqrt{\mathrm{Var}(\hat{\Theta})}$

Step 4: Define (1-α)% C.I. = $\hat{\theta} \pm c \times se(\hat{\theta})$

Population parameter or "something we would like to estimate"	**Sample statistic ("estimator")**	**Estimator as a Random Variable**	**Expected Value of the estimator**	**Variance of the estimator**	**Standard Error of estimator**	**Confidence Interval**
β_0	b_0	B_0	$E[B_0]$	$Var[B_0]$	$se(b_0)$	C.I. for β_0
β_1	b_1	B_1	$E[B_1]$	$Var[B_1]$	$se(b_1)$	C.I. for β_1
σ^2	s^2	S^2	$E[S^2]$	$Var[S^2]$	$se(s^2)$	C.I. for σ^2
$\mu_Y(x)$	$(\hat{\mu}_Y(x))$	$(\hat{\mu}_Y(x))$	$E(\hat{\mu}_Y(x))$	$Var(\hat{\mu}_Y(x))$	$se(\hat{\mu}_Y(x))$	C.I. for $\mu_Y(x)$

Step 0: From θ, define estimator, $\hat{\theta}$

Step 1: Consider the sample statistic, $\hat{\theta}$, as a random variable $\hat{\Theta}$

Step 2: Determine
E[$\hat{\Theta}$] (to confirm it's unbiased)
Var[$\hat{\Theta}$] (to calculate se)

Step 3: Define se($\hat{\theta}$) = estimate of $\sqrt{\mathrm{Var}\,(\hat{\Theta})}$

Step 4: Define (1-α)% C.I. = $\hat{\theta} \pm c \times se(\hat{\theta})$

Population parameter or "something we would like to estimate"	**Sample statistic ("estimator")**	**Estimator as a Random Variable**	**Expected Value of the estimator**	**Variance of the estimator**	**Standard Error of estimator**	**Confidence Interval**
$\boldsymbol{\beta}$	**b** $= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$	**B**	E[**B**] = $\boldsymbol{\beta}$	Var[**B**]	se(**b**)	C.I. for $\boldsymbol{\beta}$
σ^2	**s^2 or MS(Res)**	S^2	E[S^2]	Var[S^2]	se(s^2)	C.I. for σ^2
μ_Y(**x**)	$(\hat{\mu}_Y(x))$	$(\hat{\mu}_Y(x))$	E $(\hat{\mu}_Y(x))$	Var$(\hat{\mu}_Y(x))$	$se(\hat{\mu}_Y(x))$	C.I. for $\mu_Y(x)$

3.6 Interval estimates and standard errors

From (3.11), with $\hat{\mathbf{B}} = \hat{\boldsymbol{\beta}}$ as a random vector, and $k = p + 1$ as the dimension of $\hat{\boldsymbol{\beta}}$,

$$\hat{\mathbf{B}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} = \mathbf{A}\mathbf{Y}, \tag{3.66}$$

$$\mathbf{A} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T = \begin{pmatrix} \mathbf{a}_0^T \\ \mathbf{a}_1^T \\ \vdots \\ \mathbf{a}_p^T \end{pmatrix}, \tag{3.67}$$

$$(k \times n) \qquad (k \times k) \quad (k \times n) \tag{3.68}$$

```
> A <- solve(t(X)%*%X)%*%t(X)
> A
           [,1]        [,2]         [,3]          [,4]          [,5]         [,6]
   -0.092635599 0.024117815 -0.184577261  0.1738592755  0.1301994825  0.347163627
x1  0.011078248 0.001406368  0.005187597 -0.0027688020 -0.0014525385 -0.003017751
x2 -0.004887629 0.001036605  0.003140558  0.0009010023  0.0008475803 -0.003563056
           [,7]         [,8]         [,9]
    0.326356335  0.273004422  0.002511903
x1 -0.002503536 -0.001754540 -0.006175045
x2 -0.003482240 -0.002739829  0.008747009
>
```

3.6 Interval estimates and standard errors

From (3.11), with $\hat{\mathbf{B}} = \hat{\boldsymbol{\beta}}$ as a random vector, and $k = p + 1$ as the dimension of $\hat{\boldsymbol{\beta}}$,

$$\hat{\mathbf{B}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} = \mathbf{A}\mathbf{Y}, \tag{3.66}$$

$$\mathbf{A} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T = \begin{pmatrix} \mathbf{a}_0^T \\ \mathbf{a}_1^T \\ \vdots \\ \mathbf{a}_p^T \end{pmatrix}, \tag{3.67}$$

$$(k \times n) \qquad (k \times k) \quad (k \times n) \tag{3.68}$$

- **Thing 1:**
  - Linear combinations of independent normal random variables also have normal distributions! (see Appendix B)

3.6 Interval estimates and standard errors

where $\mathbf{a}_j^T$ is a $1 \times n$ row vector. The covariance matrix of $\mathbf{Y}$ is $\mathbf{\Sigma_Y} = \sigma^2 \mathbf{I}_n$ ($n \times n$ identity matrix because the ϵ_i are independent and identically distributed $N(0, \sigma^2)$ random variables). From the Appendix A for linear combinations,

$$
\begin{aligned}
\text{(3.69)} \quad \operatorname{Var}(\hat{B}_1) &= \operatorname{Var}(\mathbf{a}_1^T \mathbf{Y}) = \mathbf{a}_1^T \mathbf{\Sigma_Y} \mathbf{a}_1 = \mathbf{a}_1^T (\sigma^2 \mathbf{I}_n) \mathbf{a}_1 = \sigma^2 \mathbf{a}_1^T \mathbf{a}_1 \\
\text{(3.70)} \quad \operatorname{Var}(\hat{B}_2) &= \operatorname{Var}(\mathbf{a}_2^T \mathbf{Y}) = \mathbf{a}_2^T \mathbf{\Sigma_Y} \mathbf{a}_2 = \sigma^2 \mathbf{a}_2^T \mathbf{a}_2 \\
\vdots &= \vdots \\
\text{(3.71)} \quad \operatorname{Var}(\hat{B}_p) &= \operatorname{Var}(\mathbf{a}_p^T \mathbf{Y}) = \mathbf{a}_p^T \mathbf{\Sigma_Y} \mathbf{a}_p = \sigma^2 \mathbf{a}_p^T \mathbf{a}_p \\
\text{(3.72)} \quad \operatorname{Cov}(\hat{B}_1, \hat{B}_2) &= \operatorname{Cov}(\mathbf{a}_1^T \mathbf{Y}, \mathbf{a}_2^T \mathbf{Y}) = \mathbf{a}_1^T \mathbf{\Sigma_Y} \mathbf{a}_2 = \sigma^2 \mathbf{a}_1^T \mathbf{a}_2 \\
\vdots &= \vdots
\end{aligned}
$$

$$
\begin{aligned}
\operatorname{Var}(B) &= \operatorname{Var}(AY) \\
\operatorname{Var}(B) &= A \operatorname{Var}(Y) A^T
\end{aligned}
$$

https://en.wikipedia.org/wiki/Covariance_matrix#Generalization_of_the_variance

3.6 Interval estimates and standard errors

From (3.11), with $\hat{\mathbf{B}} = \hat{\boldsymbol{\beta}}$ as a random vector, and $k = p + 1$ as the dimension of $\hat{\boldsymbol{\beta}}$,

(3.66) $$\hat{\mathbf{B}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} = \mathbf{A}\mathbf{Y},$$

(3.67) $$\mathbf{A} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T = \begin{pmatrix} \mathbf{a}_0^T \\ \mathbf{a}_1^T \\ \vdots \\ \mathbf{a}_p^T \end{pmatrix},$$

(3.68) $$(k \times n) \qquad (k \times k) \quad (k \times n)$$

$$\mathbf{Y} \sim Normal(\mu, \sigma^2 I_n)$$

$$\begin{bmatrix} y_1 \\ y_2 \\ .. \\ y_n \end{bmatrix} \sim Normal \begin{bmatrix} \mu_1 & , & \sigma^2 & 0\ldots & 0 \\ \mu_2 & , & 0 & \sigma^2\ldots & 0 \\ \vdots & & \vdots & \ddots & \vdots \\ \mu_n & & 0 & \ldots 0 & \sigma^2 \end{bmatrix}$$

3.6 Interval estimates and standard errors

From (3.11), with $\hat{\mathbf{B}} = \hat{\boldsymbol{\beta}}$ as a random vector, and $k = p + 1$ as the dimension of $\hat{\boldsymbol{\beta}}$,

(3.66) $$\hat{\mathbf{B}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} = \mathbf{A}\mathbf{Y},$$

(3.67) $$\mathbf{A} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T = \begin{pmatrix} \mathbf{a}_0^T \\ \mathbf{a}_1^T \\ \vdots \\ \mathbf{a}_p^T \end{pmatrix},$$

(3.68) $$(k \times n) \qquad (k \times k) \quad (k \times n)$$

Var(B) = Var(AY)

Var (B) = A Var(Y) A^T

$$\mathbf{Y} \sim Normal(\mu, \sigma^2 I_n)$$

$$\begin{bmatrix} y_1 \\ y_2 \\ .. \\ y_n \end{bmatrix} \sim Normal \begin{bmatrix} \mu_1 & \sigma^2 & 0 \ldots & 0 \\ \mu_2 \quad , & 0 & \sigma^2 \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \mu_n & 0 & \ldots 0 & \sigma^2 \end{bmatrix}$$

3.6 Interval estimates and standard errors

From (3.11), with $\hat{\mathbf{B}} = \hat{\boldsymbol{\beta}}$ as a random vector, and $k = p + 1$ as the dimension of $\hat{\boldsymbol{\beta}}$,

$$\hat{\mathbf{B}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} = \mathbf{A}\mathbf{Y}, \tag{3.66}$$

$$\mathbf{A} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T = \begin{pmatrix} \mathbf{a}_0^T \\ \mathbf{a}_1^T \\ \vdots \\ \mathbf{a}_p^T \end{pmatrix}, \tag{3.67}$$

$$(k \times n) \qquad (k \times k) \quad (k \times n) \tag{3.68}$$

Variance – Covariance Matrix of **Y**

Var(B) = Var(AY)

Var (B) = A Var(Y) A^T

$$\mathbf{Y} \sim Normal(\mu, \sigma^2 I_n)$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \cdot\cdot \\ y_n \end{bmatrix} \sim Normal \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}, \begin{pmatrix} \sigma^2 & 0 \ldots & 0 \\ 0 & \sigma^2 \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots 0 & \sigma^2 \end{pmatrix}$$

3.6 Interval estimates and standard errors

$$
\begin{aligned}
&(3.69) \quad & \mathrm{Var}(\hat{B}_1) &= \mathrm{Var}(\mathbf{a}_1^T\mathbf{Y}) = \mathbf{a}_1^T\boldsymbol{\Sigma}_{\mathbf{Y}}\mathbf{a}_1 = \mathbf{a}_1^T(\sigma^2\mathbf{I}_n)\mathbf{a}_1 = \sigma^2\mathbf{a}_1^T\mathbf{a}_1 \\
&(3.70) \quad & \mathrm{Var}(\hat{B}_2) &= \mathrm{Var}(\mathbf{a}_2^T\mathbf{Y}) = \mathbf{a}_2^T\boldsymbol{\Sigma}_{\mathbf{Y}}\mathbf{a}_2 = \sigma^2\mathbf{a}_2^T\mathbf{a}_2 \\
& & \vdots &= \vdots \\
&(3.71) \quad & \mathrm{Var}(\hat{B}_p) &= \mathrm{Var}(\mathbf{a}_p^T\mathbf{Y}) = \mathbf{a}_p^T\boldsymbol{\Sigma}_{\mathbf{Y}}\mathbf{a}_p = \sigma^2\mathbf{a}_p^T\mathbf{a}_p \\
&(3.72) \quad & \mathrm{Cov}(\hat{B}_1, \hat{B}_2) &= \mathrm{Cov}(\mathbf{a}_1^T\mathbf{Y}, \mathbf{a}_2^T\mathbf{Y}) = \mathbf{a}_1^T\boldsymbol{\Sigma}_{\mathbf{Y}}\mathbf{a}_2 = \sigma^2\mathbf{a}_1^T\mathbf{a}_2 \\
& & \vdots &= \vdots
\end{aligned}
$$

Putting everything together, one gets:

$$
\begin{aligned}
&(3.73) \quad & \begin{pmatrix} \mathrm{Var}(\hat{B}_0) & \mathrm{Cov}(\hat{B}_0, \hat{B}_1) & \cdots & \mathrm{Cov}(\hat{B}_0, \hat{B}_p) \\ \mathrm{Cov}(\hat{B}_1, \hat{B}_0) & \mathrm{Var}(\hat{B}_1) & \cdots & \mathrm{Cov}(\hat{B}_1, \hat{B}_p) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(\hat{B}_p, \hat{B}_0) & \mathrm{Cov}(\hat{B}_p, \hat{B}_1) & \cdots & \mathrm{Var}(\hat{B}_p) \end{pmatrix} &= \sigma^2 \begin{pmatrix} \mathbf{a}_0^T\mathbf{a}_0 & \mathbf{a}_0^T\mathbf{a}_1 & \cdots & \mathbf{a}_0^T\mathbf{a}_p \\ \mathbf{a}_1^T\mathbf{a}_0 & \mathbf{a}_1^T\mathbf{a}_1 & \cdots & \mathbf{a}_1^T\mathbf{a}_p \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{a}_p^T\mathbf{a}_0 & \mathbf{a}_p^T\mathbf{a}_1 & \cdots & \mathbf{a}_p^T\mathbf{a}_p \end{pmatrix} \\
&(3.74) \quad & &= \sigma^2 \begin{pmatrix} \mathbf{a}_0^T \\ \mathbf{a}_1^T \\ \cdots \\ \mathbf{a}_p^T \end{pmatrix} \begin{pmatrix} \mathbf{a}_0 & \cdots & \mathbf{a}_p \end{pmatrix} = \sigma^2\mathbf{A}\mathbf{A}^T \\
&(3.75) \quad & &= \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} \\
&(3.76) \quad & &= \sigma^2(\mathbf{X}^T\mathbf{X})^{-1} \stackrel{\text{def}}{=} \boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}}.
\end{aligned}
$$

3.6 Interval estimates and standard errors

- **Step 0:** From θ, define estimator, $\hat{\theta}$
- **Step 1:** Consider the sample statistic, $\hat{\theta}$, as a random variable $\hat{\Theta}$
- **Step 2:** Determine $E[\hat{\Theta}]$ (to confirm it's unbiased) $Var[\hat{\Theta}]$ (to calculate se)
- **Step 3:** Define $se(\hat{\theta})$ = estimate of $\sqrt{\mathrm{Var}(\hat{\Theta})}$
- **Step 4:** Define (1-α)% C.I. = $\hat{\theta} \pm c \times se(\hat{\theta})$

Population parameter or "something we would like to estimate"	**Sample statistic ("estimator")**	**Estimator as a Random Variable**	**Expected Value of the estimator**	**Variance of the estimator**	**Standard Error of estimator**	**Confidence Interval**
β	**b**	**B**	E[**B**]	Var[**B**]	se(**b**)	C.I. for **β**
σ^2	s^2 or MS(Res)	S^2	$E[S^2]$	$Var[S^2]$	$se(s^2)$	C.I. for σ^2
μ_Y(**x**)	$(\hat{\mu}_Y(x))$	$(\hat{\mu}_Y(x))$	E $(\hat{\mu}_Y(x))$	Var$(\hat{\mu}_Y(x))$	$se(\hat{\mu}_Y(x))$	C.I. for $\mu_Y(x)$

3.6 Interval estimates and standard errors

Step 0: From θ, define estimator, $\hat{\theta}$

Step 1: Consider the sample statistic, $\hat{\theta}$, as a random variable $\hat{\Theta}$

Step 2: Determine
E[$\hat{\Theta}$] (to confirm it's unbiased)
Var[$\hat{\Theta}$] (to calculate se)

Step 3: Define se($\hat{\theta}$) = estimate of $\sqrt{\mathrm{Var}(\hat{\Theta})}$

Step 4: Define (1-α)% C.I. = $\hat{\theta} \pm c \times se(\hat{\theta})$

Population parameter or "something we would like to estimate"	**Sample statistic ("estimator")**	**Estimator as a Random Variable**	**Expected Value of the estimator**	**Variance of the estimator**	**Standard Error of estimator**	**Confidence Interval**
β	**b** $= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$	**B**	E[**B**] = **β**	Var[**B**] $= \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}$	se(**b**)	C.I. for **β**
σ^2	**s^2 or MS(Res)**	S^2	E[S^2]	Var[S^2]	se(s^2)	C.I. for σ^2
μ_Y(**x**)	$(\hat{\mu}_Y(x))$	$(\hat{\mu}_Y(x))$	E$(\hat{\mu}_Y(x))$	Var$(\hat{\mu}_Y(x))$	$se(\hat{\mu}_Y(x))$	C.I. for $\mu_Y(x)$

3.6 Interval estimates and standard errors

Variance – Covariance Matrix of **β**

Putting everything together, one gets:

$$\begin{pmatrix} \mathrm{Var}\,(\hat{B}_0) & \mathrm{Cov}\,(\hat{B}_0,\hat{B}_1) & \cdots & \mathrm{Cov}\,(\hat{B}_0,\hat{B}_p) \\ \mathrm{Cov}\,(\hat{B}_1,\hat{B}_0) & \mathrm{Var}\,(\hat{B}_1) & \cdots & \mathrm{Cov}\,(\hat{B}_1,\hat{B}_p) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}\,(\hat{B}_p,\hat{B}_0) & \mathrm{Cov}\,(\hat{B}_p,\hat{B}_1) & \cdots & \mathrm{Var}\,(\hat{B}_p) \end{pmatrix} = \sigma^2 \begin{pmatrix} \mathbf{a}_0^T\mathbf{a}_0 & \mathbf{a}_0^T\mathbf{a}_1 & \cdots & \mathbf{a}_0^T\mathbf{a}_p \\ \mathbf{a}_1^T\mathbf{a}_0 & \mathbf{a}_1^T\mathbf{a}_1 & \cdots & \mathbf{a}_1^T\mathbf{a}_p \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{a}_p^T\mathbf{a}_0 & \mathbf{a}_p^T\mathbf{a}_1 & \cdots & \mathbf{a}_p^T\mathbf{a}_p \end{pmatrix} \tag{3.73}$$

$$= \sigma^2 \begin{pmatrix} \mathbf{a}_0^T \\ \mathbf{a}_1^T \\ \cdots \\ \mathbf{a}_p^T \end{pmatrix} (\mathbf{a}_0 \quad \cdots \quad \mathbf{a}_p) = \sigma^2\mathbf{A}\mathbf{A}^T \tag{3.74}$$

$$= \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} \tag{3.75}$$

$$= \sigma^2(\mathbf{X}^T\mathbf{X})^{-1} \stackrel{\text{def}}{=} \mathbf{\Sigma}_{\hat{\beta}}. \tag{3.76}$$

The above implies that $\mathrm{Var}\,(\hat{B}_j)$ is the jth diagonal element of $\mathbf{\Sigma}_{\hat{\beta}}$. with row/column indexing $j = 0, 1, \ldots, p$. Since a standard error is defined as an estimated square root of the variance of an estimator,

$$se(\hat{\beta}_j) = \hat{\sigma}\sqrt{[(\mathbf{X}^T\mathbf{X})^{-1}]_{jj}}\,, \quad j = 0, 1, \ldots, p. \tag{3.77}$$

3.6 Interval estimates and standard errors

Step 0:
From θ, define estimator, $\hat{\theta}$

Step 1:
Consider the sample statistic, $\hat{\theta}$, as a random variable $\hat{\Theta}$

Step 2:
Determine
E[$\hat{\Theta}$] (to confirm it's unbiased)
Var[$\hat{\Theta}$] (to calculate se)

Step 3:
Define
se($\hat{\theta}$) =
estimate of $\sqrt{\mathrm{Var}\,(\hat{\Theta})}$

Step 4:
Define
(1-α)% C.I. =
$\hat{\theta} \pm c \times se(\hat{\theta})$

Population parameter **or "something we would like to estimate"**	**Sample statistic ("estimator")**	**Estimator as a Random Variable**	**Expected Value of the estimator**	**Variance of the estimator**	**Standard Error of estimator**	**Confidence Interval**
β	**b** $= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$	**B**	E[**B**] = **β**	Var[**B**] $= \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}$	se(**b**) $= \hat{\sigma}\sqrt{[(\mathbf{X}^T\mathbf{X})^{-1}]_{jj}}$	C.I. for **β**
σ^2	**s^2 or MS(Res)**	S^2	E[S^2]	Var[S^2]	se(s^2)	C.I. for σ^2
μ_Y**(x)**	$(\hat{\mu}_Y(x))$	$(\hat{\mu}_Y(x))$	E $(\hat{\mu}_Y(x))$	Var$(\hat{\mu}_Y(x))$	$se(\hat{\mu}_Y(x))$	C.I. for $\mu_Y(x)$

3.6 Interval estimates and standard errors

Step 0:
From θ, define estimator, $\hat{\theta}$

Step 1:
Consider the sample statistic, $\hat{\theta}$, as a random variable $\hat{\Theta}$

Step 2:
Determine
E[$\hat{\Theta}$] (to confirm it's unbiased)
Var[$\hat{\Theta}$] (to calculate se)

Step 3:
Define
se($\hat{\theta}$) =
estimate of $\sqrt{\mathrm{Var}\,(\hat{\Theta})}$

Step 4:
Define
(1-α)% C.I. =
$\hat{\theta} \pm c \times se(\hat{\theta})$

Population parameter or "something we would like to estimate"	**Sample statistic ("estimator")**	**Estimator as a Random Variable**	**Expected Value of the estimator**	**Variance of the estimator**	**Standard Error of estimator**	**Confidence Interval**
β	**b** $= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$	**B**	E[**B**] = **β**	Var[**B**] $= \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}$	se(**b**) $= \hat{\sigma}\sqrt{[(\mathbf{X}^T\mathbf{X})^{-1}]_{jj}}$	C.I. for **β**
σ^2	**s^2 or MS(Res)**	S^2	E[S^2]	Var[S^2]	se(s^2)	C.I. for σ^2
μ_Y(**x**)	$(\hat{\mu}_Y(x))$	$(\hat{\mu}_Y(x))$	E $(\hat{\mu}_Y(x))$	Var$(\hat{\mu}_Y(x))$	$se(\hat{\mu}_Y(x))$	C.I. for $\mu_Y(x)$

For 95% confidence intervals for β's or subpopulation means, or for 95% prediction intervals, the appropriate SE is multiplied by $t_{n-k,0.975}$ to get the margin of error to add/substract from the point estimate.

3.6 Interval estimates and standard errors

The above implies that Var $(\hat{B}_j)$ is the jth diagonal element of $\mathbf{\Sigma}_{\hat{\boldsymbol{\beta}}}$. with row/column indexing $j = 0, 1, \ldots, p$. Since a standard error is defined as an estimated square root of the variance of an estimator,

$$se(\hat{\beta}_j) = \hat{\sigma}\sqrt{[(\mathbf{X}^T\mathbf{X})^{-1}]_{jj}}\,, \quad j = 0, 1, \ldots, p. \tag{3.77}$$

For 95% confidence intervals for β's or subpopulation means, or for 95% prediction intervals, the appropriate SE is multiplied by $t_{n-k,0.975}$ to get the margin of error to add/substract from the point estimate.

```
> betahat[1] - qt(0.975, n-k)*sqrt(MS_Res)*sqrt(solve(t(X)%*%X)[1,1])
[1] 1.96994
> betahat[1] + qt(0.975, n-k)*sqrt(MS_Res)*sqrt(solve(t(X)%*%X)[1,1])
[1] 44.56221
>
> betahat[2] - qt(0.975, n-k)*sqrt(MS_Res)*sqrt(solve(t(X)%*%X)[2,2])
[1] 0.1792634
> betahat[2] + qt(0.975, n-k)*sqrt(MS_Res)*sqrt(solve(t(X)%*%X)[2,2])
[1] 1.183658
>
> betahat[3] - qt(0.975, n-k)*sqrt(MS_Res)*sqrt(solve(t(X)%*%X)[3,3])
[1] -0.6870086
> betahat[3] + qt(0.975, n-k)*sqrt(MS_Res)*sqrt(solve(t(X)%*%X)[3,3])
[1] 0.1327291
>
> confint(lm(y~x1+x2))
                 2.5 %     97.5 %
(Intercept)  1.9699396 44.5622139
x1           0.1792634  1.1836579
x2          -0.6870086  0.1327291
```

3.6 Interval estimates and standard errors

Next consider the subpopulation mean $\mu_Y(\mathbf{x}^*)$, where $\mathbf{x}^* = (1, x_1^*, \ldots, x_p^*)^T$. The point estimate is

$$\hat{\mu}_Y(\mathbf{x}^*) = \hat{\beta}_0 + \hat{\beta}_1 x_i^* + \cdots + \hat{\beta}_p x_p^* = \mathbf{x}^{*T}\hat{\boldsymbol{\beta}}. \tag{3.78}$$

Step 0:
From θ, define estimator,

3.6 Interval estimates and standard errors

Next consider the subpopulation mean $\mu_Y(\mathbf{x}^*)$, where $\mathbf{x}^* = (1, x_1^*, \ldots, x_p^*)^T$. The point estimate is

$$\hat{\mu}_Y(\mathbf{x}^*) = \hat{\beta}_0 + \hat{\beta}_1 x_i^* + \cdots + \hat{\beta}_p x_p^* = \mathbf{x}^{*T}\hat{\boldsymbol{\beta}}. \tag{3.78}$$

Step 0:
From θ, define estimator,

As a random variable, the variance is

$$\mathrm{Var}\,[\hat{\mu}_Y(\mathbf{x}^*)] = \mathrm{Var}\,[\mathbf{x}^{*T}\hat{\mathbf{B}}] = \mathbf{x}^{*T}\boldsymbol{\Sigma}_{\hat{\beta}}\mathbf{x}^* \tag{3.79}$$

$$= \sigma^2\mathbf{x}^{*T}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}^* \tag{3.80}$$

Step 2:
Determine
$E[\hat{\Theta}]$ (to confirm it's unbiased)
$\mathrm{Var}[\hat{\Theta}]$ (to calculate se)

Step 1:
Consider the sample statistic, , as a random variable

3.6 Interval estimates and standard errors

Next consider the subpopulation mean $\mu_Y(\mathbf{x}^*)$, where $\mathbf{x}^* = (1, x_1^*, \ldots, x_p^*)^T$. The point estimate is

Step 0:
From θ, define estimator,

$$\hat{\mu}_Y(\mathbf{x}^*) = \hat{\beta}_0 + \hat{\beta}_1 x_i^* + \cdots + \hat{\beta}_p x_p^* = \mathbf{x}^{*T}\hat{\boldsymbol{\beta}}. \tag{3.78}$$

As a random variable, the variance is

Step 2:
Determine
$E[\hat{\Theta}]$ (to confirm it's unbiased)
$Var[\hat{\Theta}]$ (to calculate se)

$$\mathrm{Var}\,[\hat{\mu}_Y(\mathbf{x}^*)] = \mathrm{Var}\,[\mathbf{x}^{*T}\hat{\mathbf{B}}] = \mathbf{x}^{*T}\boldsymbol{\Sigma}_{\hat{\beta}}\mathbf{x}^* \tag{3.79}$$

$$= \sigma^2\mathbf{x}^{*T}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}^* \tag{3.80}$$

Step 1:
Consider the sample statistic, , as a random variable

For the special case of $p = 1$, check that this is the same as (2.66). With the definition of the standard error,

$$se[\hat{\mu}_Y(\mathbf{x}^*)] = \hat{\sigma}\sqrt{\mathbf{x}^{*T}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}^*}\,. \tag{3.81}$$

Step 3:
Define
se($\hat{\theta}$) =
estimate of $\sqrt{\mathrm{Var}\,(\hat{\Theta})}$

3.6 Interval estimates and standard errors

Next consider the subpopulation mean $\mu_Y(\mathbf{x}^*)$, where $\mathbf{x}^* = (1, x_1^*, \ldots, x_p^*)^T$. The point estimate is

Step 0:
From θ, define estimator,

$$\hat{\mu}_Y(\mathbf{x}^*) = \hat{\beta}_0 + \hat{\beta}_1 x_i^* + \cdots + \hat{\beta}_p x_p^* = \mathbf{x}^{*T} \hat{\boldsymbol{\beta}}. \tag{3.78}$$

As a random variable, the variance is

Step 2:
Determine
$E[\hat{\Theta}]$ (to confirm it's unbiased)
$Var[\hat{\Theta}]$ (to calculate se)

$$\mathrm{Var}\left[\hat{\mu}_Y(\mathbf{x}^*)\right] = \mathrm{Var}\left[\mathbf{x}^{*T}\hat{\mathbf{B}}\right] = \mathbf{x}^{*T}\boldsymbol{\Sigma}_{\hat{\beta}}\mathbf{x}^* \tag{3.79}$$

$$= \sigma^2 \mathbf{x}^{*T}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}^* \tag{3.80}$$

Step 1:
Consider the sample statistic, , as a random variable

For the special case of $p = 1$, check that this is the same as (2.66). With the definition of the standard error,

$$se[\hat{\mu}_Y(\mathbf{x}^*)] = \hat{\sigma}\sqrt{\mathbf{x}^{*T}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}^*}\,. \tag{3.81}$$

Step 3:
Define
se($\hat{\theta}$) =
estimate of $\sqrt{\mathrm{Var}(\hat{\Theta})}$

95% Confidence Interval:

$$\hat{\mu}_Y(\mathbf{x}^*) \ \text{+/-}\ t_{n-k,0.975}\ se[\hat{\mu}_Y(\mathbf{x}^*)]$$

Step 4:
Define
(1-α)% C.I. =
$\hat{\theta} \pm c \times se(\hat{\theta})$

3.6 Interval estimates and standard errors

95% **Confidence Interval** for the **subpopulation mean**:

$$\hat{\mu}_Y(\mathbf{x}^*) \quad +/- \quad t_{n-k,0.975}\ se[\hat{\mu}_Y(\mathbf{x}^*)]$$

where:

$$se[\hat{\mu}_Y(\mathbf{x}^*)] = \hat{\sigma}\sqrt{\mathbf{x}^{*T}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}^*}\ .$$

```
> # 50 year old who make 30K$ in income
> xstar <- c(1,50,30)
> xstar%*%betahat
            y
[1,] 49.02491
>
> mu_xstar <- xstar%*%betahat
> mu_xstar
            y
[1,] 49.02491
>
> se_mu_xstar <- sqrt(MS_Res)*sqrt(t(xstar)%*%solve(t(X)%*%X)%*%xstar)
>
> mu_xstar -qt(0.975, n-k)*se_mu_xstar
            y
[1,] 34.43381
> mu_xstar +qt(0.975, n-k)*se_mu_xstar
            y
[1,] 63.61602
>
> predict(lm(y~x1+x2), newdata=data.frame(1,x1=50,x2=30), se.fit=TRUE,interval="confidence")
```

3.6 Interval estimates and standard errors

Next consider the prediction $\hat{Y}(\mathbf{x}^*)$ for a future value at $\mathbf{x}^*$. Then

$$\hat{Y}(\mathbf{x}^*) = \hat{\beta}_0 + \hat{\beta}_1 x_i^* + \cdots + \hat{\beta}_p x_p^* = \mathbf{x}^{*T} \hat{\boldsymbol{\beta}} \tag{3.82}$$

3.6 Interval estimates and standard errors

Next consider the prediction $\hat{Y}(\mathbf{x}^*)$ for a future value at $\mathbf{x}^*$. Then

$$\hat{Y}(\mathbf{x}^*) = \hat{\beta}_0 + \hat{\beta}_1 x_i^* + \cdots + \hat{\beta}_p x_p^* = \mathbf{x}^{*T}\hat{\boldsymbol{\beta}} \tag{3.82}$$

is the same as the subpopulation mean in (3.78), which could be considered as the average of many observations at $\mathbf{x}^*$. Assuming the model is correct, the prediction error is

$$E = \hat{Y}(\mathbf{x}^*) - [\beta_0 + \beta_1 \mathbf{x}_1^* + \cdots + \beta_p x_p^* + \epsilon(\mathbf{x}^*)], \tag{3.83}$$

3.6 Interval estimates and standard errors

Next consider the prediction $\hat{Y}(\mathbf{x}^*)$ for a future value at $\mathbf{x}^*$. Then

$$\hat{Y}(\mathbf{x}^*) = \hat{\beta}_0 + \hat{\beta}_1 x_i^* + \cdots + \hat{\beta}_p x_p^* = \mathbf{x}^{*T}\hat{\boldsymbol{\beta}} \tag{3.82}$$

is the same as the subpopulation mean in (3.78), which could be considered as the average of many observations at $\mathbf{x}^*$. Assuming the model is correct, the prediction error is

$$E = \hat{Y}(\mathbf{x}^*) - [\beta_0 + \beta_1 \mathbf{x}_1^* + \cdots + \beta_p x_p^* + \epsilon(\mathbf{x}^*)], \tag{3.83}$$

where $\epsilon(\mathbf{x}^*) \sim N(0, \sigma^2)$ independent of the previous data. Hence

$$\text{Var}(E) = \text{Var}[\mathbf{x}^{*T}\hat{\mathbf{B}}] + \text{Var}[\epsilon(\mathbf{x}^*)] \tag{3.84}$$

$$= \sigma^2 \mathbf{x}^{*T}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}^* + \sigma^2 \tag{3.85}$$

3.6 Interval estimates and standard errors

Next consider the prediction $\hat{Y}(\mathbf{x}^*)$ for a future value at $\mathbf{x}^*$. Then

$$\hat{Y}(\mathbf{x}^*) = \hat{\beta}_0 + \hat{\beta}_1 x_i^* + \cdots + \hat{\beta}_p x_p^* = \mathbf{x}^{*T}\hat{\boldsymbol{\beta}} \tag{3.82}$$

is the same as the subpopulation mean in (3.78), which could be considered as the average of many observations at $\mathbf{x}^*$. Assuming the model is correct, the prediction error is

$$E = \hat{Y}(\mathbf{x}^*) - [\beta_0 + \beta_1 \mathbf{x}_1^* + \cdots + \beta_p x_p^* + \epsilon(\mathbf{x}^*)], \tag{3.83}$$

where $\epsilon(\mathbf{x}^*) \sim N(0, \sigma^2)$ independent of the previous data. Hence

$$\begin{aligned} \operatorname{Var}(E) &= \operatorname{Var}[\mathbf{x}^{*T}\hat{\mathbf{B}}] + \operatorname{Var}[\epsilon(\mathbf{x}^*)] && (3.84) \\ &= \sigma^2 \mathbf{x}^{*T}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}^* + \sigma^2 && (3.85) \end{aligned}$$

and for the prediction error E,

$$se[E] = \hat{\sigma}\sqrt{1 + \mathbf{x}^{*T}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}^*}\,. \tag{3.86}$$

3.6 Interval estimates and standard errors

Next consider the prediction $\hat{Y}(\mathbf{x}^*)$ for a future value at $\mathbf{x}^*$. Then

$$\hat{Y}(\mathbf{x}^*) = \hat{\beta}_0 + \hat{\beta}_1 x_i^* + \cdots + \hat{\beta}_p x_p^* = \mathbf{x}^{*T}\hat{\boldsymbol{\beta}} \tag{3.82}$$

is the same as the subpopulation mean in (3.78), which could be considered as the average of many observations at $\mathbf{x}^*$. Assuming the model is correct, the prediction error is

$$E = \hat{Y}(\mathbf{x}^*) - [\beta_0 + \beta_1 \mathbf{x}_1^* + \cdots + \beta_p x_p^* + \epsilon(\mathbf{x}^*)], \tag{3.83}$$

where $\epsilon(\mathbf{x}^*) \sim N(0, \sigma^2)$ independent of the previous data. Hence

$$\begin{aligned} \operatorname{Var}(E) &= \operatorname{Var}[\mathbf{x}^{*T}\hat{\mathbf{B}}] + \operatorname{Var}[\epsilon(\mathbf{x}^*)] && (3.84) \\ &= \sigma^2 \mathbf{x}^{*T}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}^* + \sigma^2 && (3.85) \end{aligned}$$

and for the prediction error E,

$$se[E] = \hat{\sigma}\sqrt{1 + \mathbf{x}^{*T}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}^*}\,. \tag{3.86}$$

95% Prediction Interval:

$$\hat{Y}(\mathbf{x}^*) \text{ +/- } t_{n-k,0.975}\; se[E]$$

3.6 Interval estimates and standard errors

95% Prediction Interval:

$$\hat{Y}(\mathbf{x}^*) \text{ +/- } t_{n-k,0.975} \; se[E]$$

$$se[E] = \hat{\sigma}\sqrt{1 + \mathbf{x}^{*T}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}^*}\,.$$

```
> # 50 year old who make 30K$ in income
> xstar <- c(1,50,30)
> xstar%*%betahat
            y
[1,] 49.02491
>
> se_E <- sqrt(MS_Res)*sqrt(1 + t(xstar)%*%solve(t(X)%*%X)%*%xstar)
> se_E
         [,1]
[1,] 15.12614
> mu_xstar -qt(0.975, n-k)*se_E
            y
[1,] 12.01257
> mu_xstar +qt(0.975, n-k)*se_E
            y
[1,] 86.03726
>
> predict(lm(y~x1+x2), newdata=data.frame(1,x1=50,x2=30), se.fit=TRUE,interval="prediction")
$fit
       fit      lwr      upr
1 49.02491 12.01257 86.03726
```

multiple linear regression

Age vs. Money

PREDICTOR *variables*

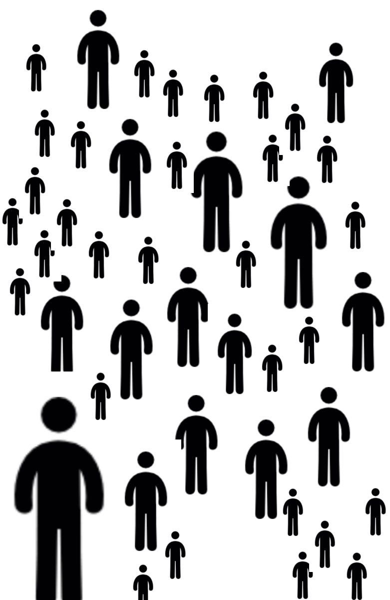
$x_1 \longrightarrow$ Age in Years

$x_2 \longrightarrow$ Income in thousands of $.

RESPONSE *variable*

$Y \longrightarrow$ Cash in pocket dollars ($)

Population



Population parameters

$$\beta_0, \beta_1, \beta_2, \sigma^2$$

Hypothesis Test

$$H_0 : \beta_1 = 0$$
$$H_1 : \beta_1 \neq 0$$

Sample statistics

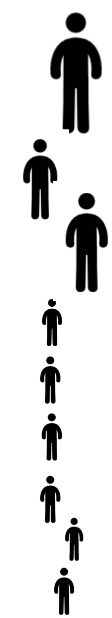
b_0 = 23.26
b_1 = 0.68
b_2 = -0.28
s = 13.9
R^2 = 0.65

For parameter β_1 :

95% C.I. = [0.18, 1.18]

p-value = 0.016

Sample, n=9



x_1	x_2	y
82	26	71
45	49	54
71	76	43
22	37	45
29	40	21
9	0	11
12	2	30
18	10	45
24	92	10