2.8 Exercises

2.1 This question concerns a galaxy data set on Hubble's law from R.J. Larsen and M.L. Marx (1981), *An Introduction to Mathematical Statistics and its Applications*, Prentice-Hall. Since the 1920's it has been known that the universe is expanding and that more distant galaxies are receding the fastest. Hubble's law states that there is a linear relationship between $x =$ the distance of a galaxy from Earth (in millions of light years) and $y =$ the velocity at which the galaxy is moving away from the Earth (say, in thousands of km per second). The constant of proportionality is Hubble's constant, which has to be estimated from empirical data. Table 2.5 contains data for 11 galaxies.

Table 2.5: Data for Hubble's law

cluster	distance (millions-of-light-years)	velocity (thousands-of-km/sec)
Virgo	22	1.21
Pegasus	68	3.86
Perseus	108	5.15
Coma Berenices	137	7.56
Ursa Major No.1	255	15.0
Leo	315	19.3
Corona Borealis	390	21.6
Gemini	405	23.2
Bootes	685	39.4
Ursa Major No.2	700	41.8
Hydra	1100	61.2

2.5.2 Derivations

$$b_1 = r_{xy}\frac{s_y}{s_x}$$

Recall: $r_{xy} = \frac{s_{xy}}{s_x s_y}$

$$= \frac{s_{xy}}{s_x s_y}\frac{s_y}{s_x}$$

Recall: $s_{xy} = (n-1)^{-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}).$

$$= \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x^2}$$

$$= \frac{\sum_{i=1}^{n}(y_i)(x_i - \bar{x})}{(n-1)s_x^2} - \frac{\sum_{i=1}^{n}(\bar{y})(x_i - \bar{x})}{(n-1)s_x^2}$$

$$= \frac{\sum_{i=1}^{n}(y_i)(x_i - \bar{x})}{(n-1)s_x^2} - \bar{y}\frac{\sum_{i=1}^{n}(x_i - \bar{x})}{(n-1)s_x^2}$$

0

Useful identity number 1.

$\sum_{i=1}^{n}(w_i - \bar{w}) = 0$

$\sum_{i=1}^{n}(w_i - \bar{w}) = -n\bar{w} + \sum_{i=1}^{n}(w_i)$

$= -n\bar{w} + n\frac{1}{n}\sum_{i=1}^{n}(w_i)$

$= -n\bar{w} + n\bar{w} = 0$

Useful identity number 2.

$$\sum_{i=1}^{n}(w_i-\bar{w})^2=\sum_{i=1}^{n}(w_i^2)-n\bar{w}^2$$

$$\sum_{i=1}^{n}(w_i-\bar{w})^2=\sum_{i=1}^{n}(w_i^2-2w_i\bar{w}+\bar{w}^2)$$

$$=\sum_{i=1}^{n}(w_i^2)-2\bar{w}\sum_{i=1}^{n}(w_i)+n\bar{w}^2$$

$$=\sum_{i=1}^{n}(w_i^2)-2\bar{w}n\frac{1}{n}\sum_{i=1}^{n}(w_i)+n\bar{w}^2$$

$$=\sum_{i=1}^{n}(w_i^2)-2\bar{w}n\bar{w}+n\bar{w}^2$$

$$=\sum_{i=1}^{n}(w_i^2)-2n\bar{w}^2+n\bar{w}^2$$

$$=\sum_{i=1}^{n}(w_i^2)-n\bar{w}^2$$

Useful identity number 3.

$$\sum_{i=1}^{n}(w_i - \bar{w})(z_i - \bar{z}) = \sum_{i=1}^{n}(w_i - \bar{w})z_i$$

$$\sum_{i=1}^{n}(w_i - \bar{w})(z_i - \bar{z}) = \sum_{i=1}^{n}(w_i z_i - \bar{w}z_i - \bar{z}w_i \bar{w}\bar{z})$$

$$= \sum_{i=1}^{n}(x_i - \bar{x})z_i + \sum_{i=1}^{n}(\bar{w}\bar{z}) - \sum_{i=1}^{n} \bar{z}w_i$$

$$= \sum_{i=1}^{n}(x_i - \bar{x})z_i + n\bar{w}\bar{z} - \bar{z}\frac{n}{n}\sum_{i=1}^{n} w_i$$

$$= \sum_{i=1}^{n}(x_i - \bar{x})z_i + n\bar{w}\bar{z} - n\bar{w}\bar{z}$$

$$= \sum_{i=1}^{n}(w_i - \bar{w})z_i$$

simple linear regression

Age vs. Money

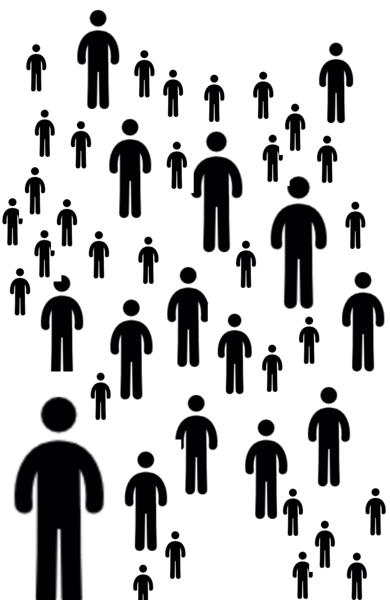
PREDICTOR *variable*

x → Age in Years

RESPONSE *variable*

Y → Cash in pocket dollars ($)

Population



Population parameters

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

Hypothesis Test

$H_0 : \beta_1 = 0$

$H_1 : \beta_1 \neq 0$

Sample statistics

$b_0 = 17.7$

$b_1 = 0.55$

$s = 15.5$

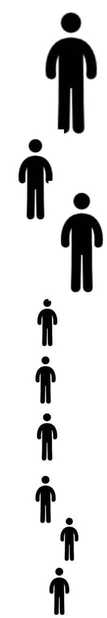
$R^2 = 0.49$

For parameter β_1 :

95% C.I. = [0.05, 1.05]

p-value = 0.036

Sample, n=9



x	y
82	71
45	54
71	43
22	45
29	21
9	11
12	30
18	45
24	10

Age vs. Money

Sample statistics

$b_0 = 17.7$
$b_1 = 0.55$
$s = 15.5$
$R^2 = 0.49$

For statistic b_1 :

95% C.I. = [0.05, 1.05]
p-value = 0.036

Objective: The purpose of this observational study was to demonstrate if, and to what extent, age is associated with money.

Design and Methods: We collected a random sample of individuals and for each determined their age **(recorded in years)** and the amount of money (in dollars) in their accounts. Analysis of the data was done using **<u>linear regression</u>**.

Results: We obtained a random sample of $n = 9$ subjects. There is a statistically significant association between age and money (*p*-value =0.036). **For every additional year in age, an individual's amount of money increases on average by an estimated of $0.55** (95% C.I. = [$0.05, $1.05]).

Conclusions: We found that, as hypothesized, age is associated with money. In our sample age accounted for about half of the variability observed in money (R^2=0.49). We **<u>predict</u>** that a 50 year old will have $45.1 (95% P.I. = [$5.6, $84.5]), whereas a 40 year old will have $39.6 (95% P.I. = [$0.8, $78.4]).

Small Print: The analysis rests on the following assumptions:

- the observations are independently and identically distributed.
- the **response** variable, money, is normally distributed.
- Homoscedasticity of residuals or equal variance.
- the <u>relationship</u> between **response** and **predictor** variables is linear.

DEUTSCHE BUNDESBANK
EUROSYSTEM

"Our research (using linear regression) indicates that older people hold and use more cash."

5.3.1 Age, Income, and Education

The role of age is of interest because one could argue that the enduring importance of cash could be due to habit persistence. Indeed, previous literature indicates that older people hold and use more cash while young consumers are more likely to use new payment technologies (e.g., Daniels and Murphy, 1994; Boeschoten, 1998; Carow and Staten, 1999; Stavins, 2001; Hayashi and Klee, 2003).

Our results in Figure 3 reveal that "older" people use significantly more cash than younger people except for US, where younger individuals use more cash than older individuals. Note again that these descriptive statistics assume all other factors to be fixed. These figures regarding age do not control for differences in expenditure patterns or other personal characteristics, for example, younger consumers may buy different product and/or services and at different venues than older individuals. Therefore, a final answer on the role of age can only be given with estimations that control for these other variables, which will be focus of the next section.[18]

Income and education have been cited in the literature as important factors, with cash usage declining with higher income and education (e.g. Arango et al. (2011) for CA, von Kalckreuth et al. (2014b) for DE, and Schuh and Stavins (2010) as well as Cohen and Rysman (2013) for US). Figure 3 confirms differences along income terciles with less cash usage by higher income respondents. Even stronger differences are found along education. Notably, these differences

---

[18] von Kalckreuth et al. (2014a) find no evidence in favour of strong habit persistence. Instead, they attribute higher cash usage of older people to their differential characteristics, e.g. lower opportunity costs of time or lower income.

5.3.1 Age, Income, and Education

The role of age is of interest because one could argue that the enduring importance of cash could be due to habit persistence. Indeed, previous literature indicates that older people hold and use more cash while young consumers are more likely to use new payment technologies (e.g., Daniels and Murphy, 1994; Boeschoten, 1998; Carow and Staten, 1999; Stavins, 2001; Hayashi and Klee, 2003).

Our results in Figure 3 reveal that "older" people use significantly more cash than younger people except for US, where younger individuals use more cash than older individuals. Note again that these descriptive statistics assume all other factors to be fixed. These figures regarding age do not control for differences in expenditure patterns or other personal characteristics, for example, younger consumers may buy different product and/or services and at different venues than older individuals. Therefore, a final answer on the role of age can only be given with estimations that control for these other variables, which will be focus of the next section.[18]

Income and education have been cited in the literature as important factors, with cash usage declining with higher income and education (e.g. Arango et al. (2011) for CA, von Kalckreuth et al. (2014b) for DE, and Schuh and Stavins (2010) as well as Cohen and Rysman (2013) for US). Figure 3 confirms differences along income terciles with less cash usage by higher income respondents. Even stronger differences are found along education. Notably, these differences

[18] von Kalckreuth et al. (2014a) find no evidence in favour of strong habit persistence. Instead, they attribute higher cash usage of older people to their differential characteristics, e.g. lower opportunity costs of time or lower income.

Stat 306: Finding Relationships in Data.

Lecture 7

Section 3.1 Least squares with two or more explanatory variables

simple linear regression

Age vs. Money

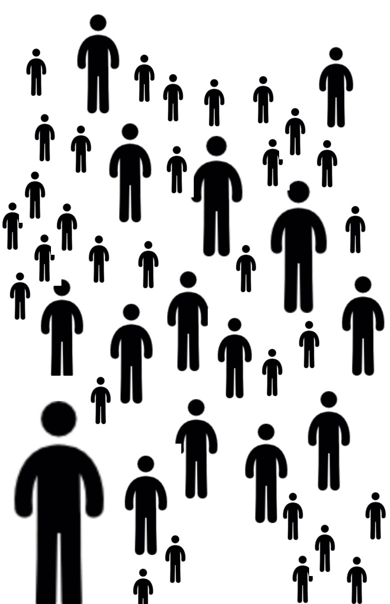
PREDICTOR *variable*

$x \longrightarrow$ Age in Years

RESPONSE *variable*

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

Population



Population parameters

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

Hypothesis Test

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

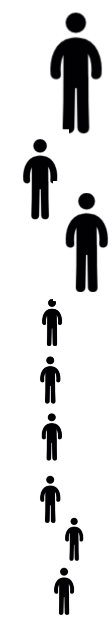
Sample statistics

$b_0 = 17.7$
$b_1 = 0.55$
$s = 15.5$
$R^2 = 0.49$

For parameter β_1 :

95% C.I. $= [0.05, 1.05]$
p-value $= 0.036$

Sample, n=9



x	y
82	71
45	54
71	43
22	45
29	21
9	11
12	30
18	45
24	10

multiple linear regression

Age vs. Money

PREDICTOR *variables*

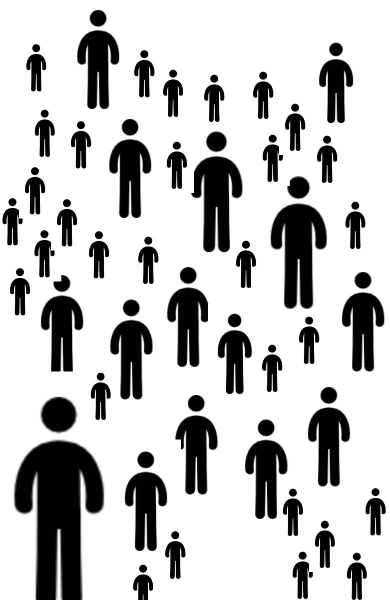
x_1 → Age in Years

x_2 → Income in thousands of $.

RESPONSE *variable*

Y → Cash in pocket dollars ($)

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

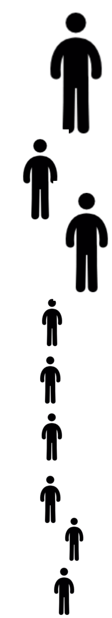
b_0 = 23.26
b_1 = 0.68
b_2 = -0.28
s = 13.9
R^2 = 0.65

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

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

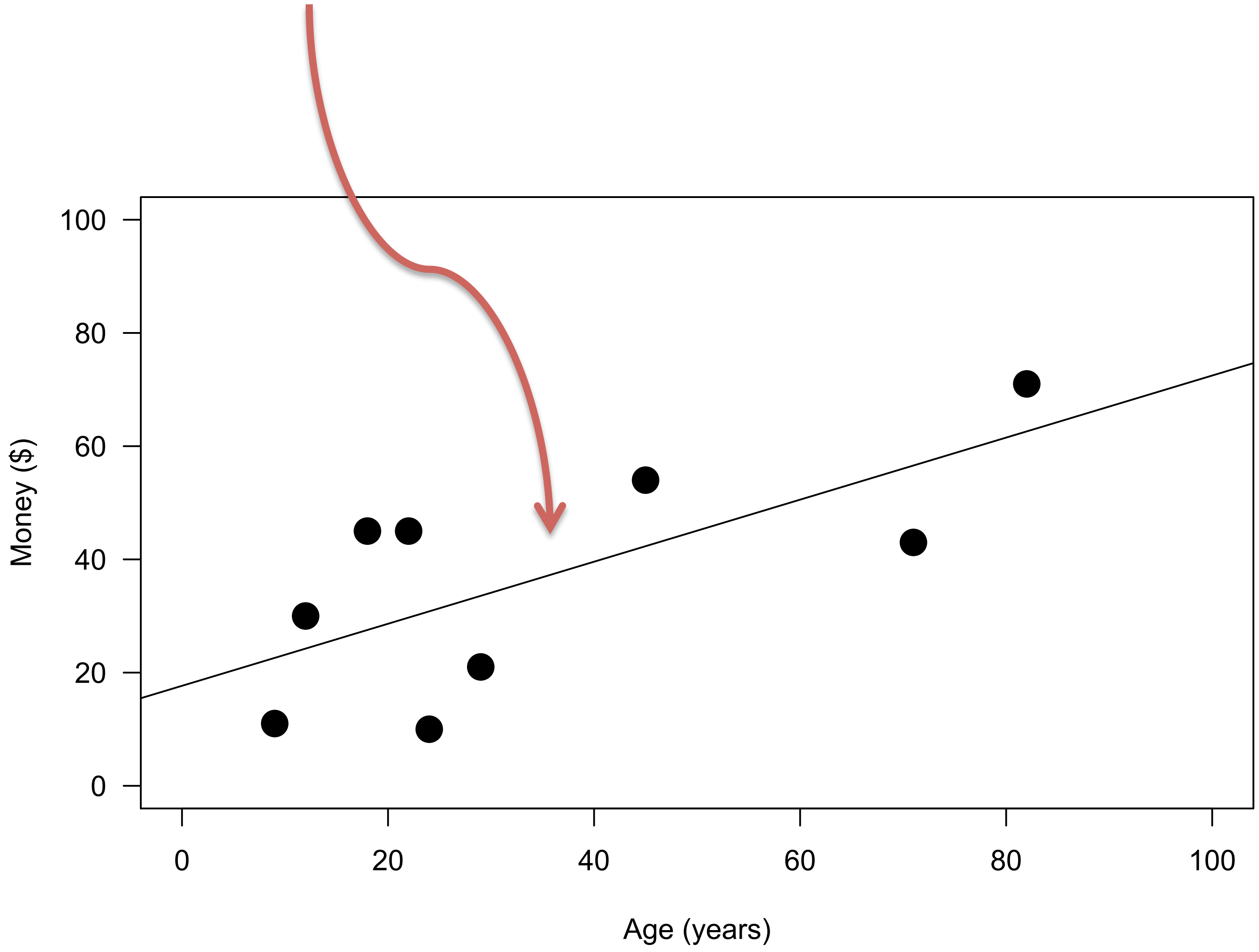
3.1 Least squares with two or more explanatory variables

3.1 Least squares with two or more explanatory variables

This section extends the ideas in Section 2.1. The data are (y_i, x_{i1}, x_{i2}), $i = 1, \ldots, n$. One can fit the hyperplane equation $y = b_0 + b_1 x_1 + b_2 x_2$ via least squares. This includes the special case where $x_2 = x_1^2$ in which case the equation is the least squares quadratic in one explanatory variable.

The steps for fitting a line to (y_i, x_i), $i = 1, \ldots, n$, can be generalized, and are outlined below. The least squares criterion is to minimize the sum of squares of vertical deviations. With b_0, b_1, b_2 as function arguments, the objective function is:

prediction equation : $y = b_0 + b_1x$



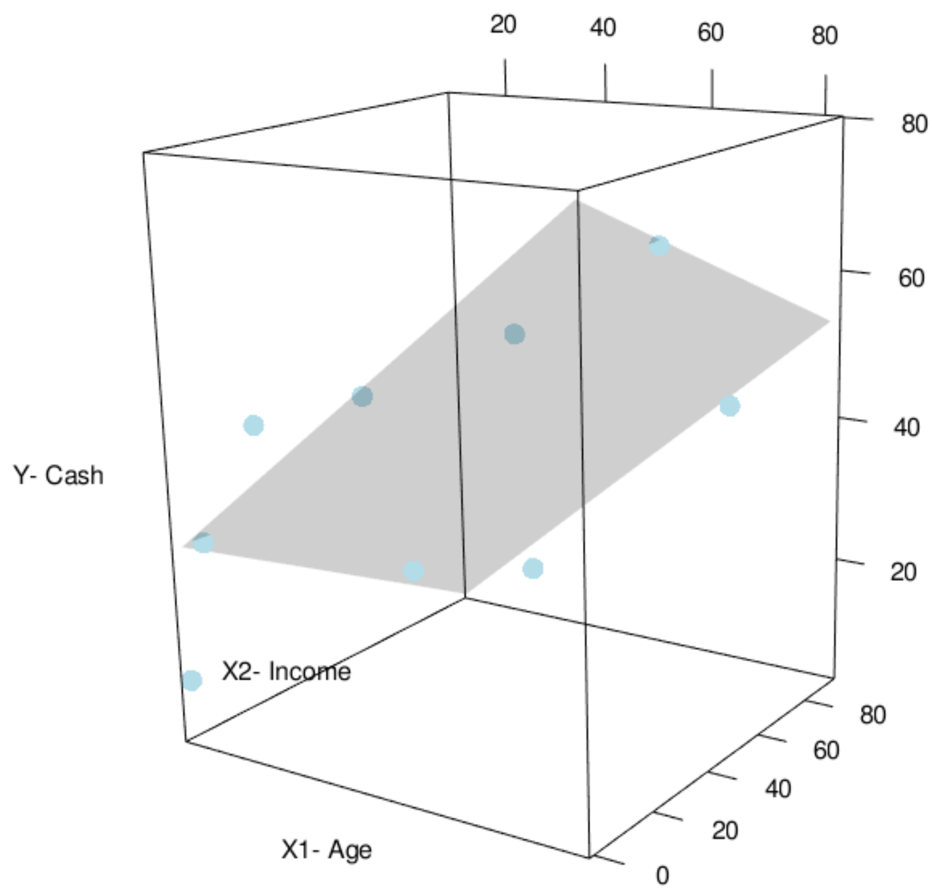
3.1 Least squares with two or more explanatory variables

"hyperplane equation"

```
> library(rgl)
> f <- function(x1, x2){  23.27 + 0.68*x1 - 0.28*x2}
> n <- 9
>
> 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)
>
> plot3d(x1,x2,y, type="p", col="lightblue", xlab="X1- Age", ylab="X2-
Income", zlab="Y- Cash", site=5, lwd=15, size=12)
> my_surface(f, alpha=.2 )
>
```

3.1 Least squares with two or more explanatory variables

"hyperplane equation"



3.1 Least squares with two or more explanatory variables

Once again we can minimize the least squares with simple calculus:

$$S(b_0, b_1, b_2) = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_{i1} - b_2 x_{i2})^2. \tag{3.1}$$

Take partial derivatives with respect to b_0, b_1, b_2. This leads to:

$$\frac{\partial S}{\partial b_0} = -2 \sum_{i=1}^{n} (y_i - b_0 - b_1 x_{i1} - b_2 x_{i2}), \tag{3.2}$$

$$\frac{\partial S}{\partial b_1} = -2 \sum_{i=1}^{n} x_{i1}(y_i - b_0 - b_1 x_{i1} - b_2 x_{i2}), \tag{3.3}$$

$$\frac{\partial S}{\partial b_2} = -2 \sum_{i=1}^{n} x_{i2}(y_i - b_0 - b_1 x_{i1} - b_2 x_{i2}). \tag{3.4}$$

3.1 Least squares with two or more explanatory variables

Set the equations to 0, divide by -2 and use the algebraic steps as in (2.16)–(2.17). $(\hat{b}_0, \hat{b}_1, \hat{b}_2)$, rewrite as

$$n\hat{b}_0 + \hat{b}_1 \sum_{i=1}^{n} x_{i1} + \hat{b}_2 \sum_{i=1}^{n} x_{i2} = \sum_{i=1}^{n} y_i, \tag{3.5}$$

$$\hat{b}_0 \sum_{i=1}^{n} x_{i1} + \hat{b}_1 \sum_{i=1}^{n} x_{i1}^2 + \hat{b}_2 \sum_{i=1}^{n} x_{i1}x_{i2} = \sum_{i=1}^{n} x_{i1}y_i, \tag{3.6}$$

$$\hat{b}_0 \sum_{i=1}^{n} x_{i2} + \hat{b}_1 \sum_{i=1}^{n} x_{i1}x_{i2} + \hat{b}_2 \sum_{i=1}^{n} x_{i2}^2 = \sum_{i=1}^{n} x_{i2}y_i. \tag{3.7}$$

3.1 Least squares with two or more explanatory variables

Set the equations to 0, divide by −2 and use the algebraic steps as in (2.16)–(2.17). $(\hat{b}_0, \hat{b}_1, \hat{b}_2)$, rewrite as

$$n\hat{b}_0 + \hat{b}_1 \sum_{i=1}^{n} x_{i1} + \hat{b}_2 \sum_{i=1}^{n} x_{i2} = \sum_{i=1}^{n} y_i, \tag{3.5}$$

$$\hat{b}_0 \sum_{i=1}^{n} x_{i1} + \hat{b}_1 \sum_{i=1}^{n} x_{i1}^2 + \hat{b}_2 \sum_{i=1}^{n} x_{i1}x_{i2} = \sum_{i=1}^{n} x_{i1}y_i, \tag{3.6}$$

$$\hat{b}_0 \sum_{i=1}^{n} x_{i2} + \hat{b}_1 \sum_{i=1}^{n} x_{i1}x_{i2} + \hat{b}_2 \sum_{i=1}^{n} x_{i2}^2 = \sum_{i=1}^{n} x_{i2}y_i. \tag{3.7}$$

3.1 Least squares with two or more explanatory variables

Set the equations to 0, divide by −2 and use the algebraic steps as in (2.16)–(2.17). $(\hat{b}_0, \hat{b}_1, \hat{b}_2)$, rewrite as

$$n\hat{b}_0 + \hat{b}_1 \sum_{i=1}^{n} x_{i1} + \hat{b}_2 \sum_{i=1}^{n} x_{i2} = \sum_{i=1}^{n} y_i, \tag{3.5}$$

$$\hat{b}_0 \sum_{i=1}^{n} x_{i1} + \hat{b}_1 \sum_{i=1}^{n} x_{i1}^2 + \hat{b}_2 \sum_{i=1}^{n} x_{i1}x_{i2} = \sum_{i=1}^{n} x_{i1}y_i, \tag{3.6}$$

$$\hat{b}_0 \sum_{i=1}^{n} x_{i2} + \hat{b}_1 \sum_{i=1}^{n} x_{i1}x_{i2} + \hat{b}_2 \sum_{i=1}^{n} x_{i2}^2 = \sum_{i=1}^{n} x_{i2}y_i. \tag{3.7}$$

Put in matrix form:

$$\begin{pmatrix} n & \sum x_{i1} & \sum x_{i2} \\ \sum x_{i1} & \sum x_{i1}^2 & \sum x_{i1}x_{i2} \\ \sum x_{i2} & \sum x_{i1}x_{i2} & \sum x_{i2}^2 \end{pmatrix} \begin{pmatrix} \hat{b}_0 \\ \hat{b}_1 \\ \hat{b}_2 \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum x_{i1}y_i \\ \sum x_{i2}y_i \end{pmatrix}. \tag{3.8}$$

3.1 Least squares with two or more explanatory variables

Put in matrix form:

$$\begin{pmatrix} n & \sum x_{i1} & \sum x_{i2} \\ \sum x_{i1} & \sum x_{i1}^2 & \sum x_{i1}x_{i2} \\ \sum x_{i2} & \sum x_{i1}x_{i2} & \sum x_{i2}^2 \end{pmatrix} \begin{pmatrix} \hat{b}_0 \\ \hat{b}_1 \\ \hat{b}_2 \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum x_{i1}y_i \\ \sum x_{i2}y_i \end{pmatrix}. \tag{3.8}$$

```
> matrix(c(n,sum(x1),sum(x2),sum(x1),sum(x1^2),sum(x1*x2),sum(x2),sum(x1*x2),sum(x2^2)),
nrow=3, ncol=3)
     [,1]  [,2]  [,3]
[1,]    9   312   332
[2,]  312 16240 14119
[3,]  332 14119 20390
>
> matrix(c(b0hat,b1hat,b2hat),3,1)
Error in matrix(c(b0hat, b1hat, b2hat), 3, 1) : object 'b0hat' not found
>
> matrix(c(sum(y), sum(x1*y), sum(x2*y)),3,1)
      [,1]
[1,]   330
[2,] 14413
[3,] 11695
```

3.1 Least squares with two or more explanatory variables

For the matrix form, write

$$(3.9) \qquad \mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} \\ 1 & x_{21} & x_{22} \\ \vdots & \vdots & \vdots \\ 1 & x_{i1} & x_{i2} \\ \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} \end{pmatrix}, \quad \mathbf{X}^T = \begin{pmatrix} 1 & 1 & \cdots 1 \\ x_{11} & x_{21} & \cdots x_{n1} \\ x_{12} & x_{22} & \cdots x_{n2} \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{pmatrix}, \quad \hat{\mathbf{b}} = \begin{pmatrix} \hat{b}_0 \\ \hat{b}_1 \\ \hat{b}_2 \end{pmatrix}.$$

The dimensions are $n \times 3$ for $\mathbf{X}$, $3 \times n$ for $\mathbf{X}^T$, $n \times 1$ for $\mathbf{y}$, and 3×1 for $\hat{\mathbf{b}}$.

```
> 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
```

3.1 Least squares with two or more explanatory variables

Note that the entry in row s and column t of $\mathbf{X}^T\mathbf{X}$ is the inner product of row s of $\mathbf{X}^T$ and column t of $\mathbf{X}$, or equivalently the inner product of columns s and t of $\mathbf{X}$. The sth diagonal entry of $\mathbf{X}^T\mathbf{X}$ is the inner product of column s of $\mathbf{X}$ with itself, or the sum of squares of column s of $\mathbf{X}$. Then, equation (3.8) becomes

$$(\mathbf{X}^T\mathbf{X})\hat{\mathbf{b}} = \mathbf{X}^T\mathbf{y} \tag{3.10}$$

$$\text{or} \quad \hat{\mathbf{b}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}. \tag{3.11}$$

Equation (3.11) assumes that $\mathbf{X}^T\mathbf{X}$ is a non-singular matrix so that its inverse is defined. The discussion of a condition for non-singularity is given in Section 3.11.

```
> bhat <- solve(Xt %*% X) %*% Xt %*% y
> bhat
           [,1]
[1,] 23.2660767
[2,]  0.6814606
[3,] -0.2771398
```

3.1 Least squares with two or more explanatory variables

Alternatively, to solve for the $\hat{b}_j$'s, from the first partial derivative:

$$n\hat{b}_0 + \hat{b}_1 \sum_{i=1}^{n} x_{i1} + \hat{b}_2 \sum_{i=1}^{n} x_{i2} = \sum_{i=1}^{n} y_i,$$

write

(3.12) $$\hat{b}_0 = \overline{y} - \hat{b}_1\overline{x}_1 - \hat{b}_2\overline{x}_2,$$

substitute into the other two equations (3.6) and (3.7), and show that $\hat{b}_1, \hat{b}_2$ can be expressed in terms of the sample covariances and variances

(3.13) $$s_{x_1y},\ s_{x_2y},\ s_{x_1x_2},\ s^2_{x_1},\ s^2_{x_2}.$$

The resulting expressions for $\hat{b}_1, \hat{b}_2$ are generalizations of (2.22). The details are left as an exercise.

3.1 Least squares with two or more explanatory variables

Alternatively, to solve for the $\hat{b}_j$'s, from the first partial derivative:

$$n\hat{b}_0 + \hat{b}_1 \sum_{i=1}^{n} x_{i1} + \hat{b}_2 \sum_{i=1}^{n} x_{i2} = \sum_{i=1}^{n} y_i,$$

write

(3.12) $$\hat{b}_0 = \overline{y} - \hat{b}_1 \overline{x}_1 - \hat{b}_2 \overline{x}_2,$$

substitute into the other two equations (3.6) and (3.7), and show that $\hat{b}_1, \hat{b}_2$ can be expressed in terms of the sample covariances and variances

(3.13) $$s_{x_1 y},\ s_{x_2 y},\ s_{x_1 x_2},\ s^2_{x_1},\ s^2_{x_2}.$$

The resulting expressions for $\hat{b}_1, \hat{b}_2$ are generalizations of (2.22). The details are left as an exercise.

3.1 Least squares with two or more explanatory variables

Next consider the extension to p explanatory variables, with data $(y_i, x_{i1}, \ldots, x_{ip})$ for $i = 1, \ldots, n$. One can fit the hyperplane equation $y = b_0 + b_1 x_1 + \cdots + b_p x_p$ via least squares by minimizing the sum of squares of vertical deviations:

$$S(b_0, b_1, \ldots, b_p) = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_{i1} - \cdots - b_p x_{ip})^2. \tag{3.16}$$

Exercise: With $\hat{b} = (b_0, b_1, \ldots, b_p)^T$, check that you can follow the same steps with $p = 3$ explanatory variables and general $p \geq 2$, and get

$$(\mathbf{X}^T\mathbf{X})\hat{\mathbf{b}} = \mathbf{X}^T\mathbf{y}, \tag{3.17}$$

3.1 Least squares with two or more explanatory variables

Put in matrix form:

$$\begin{pmatrix} n & \sum x_{i1} & \sum x_{i2} \\ \sum x_{i1} & \sum x_{i1}^2 & \sum x_{i1}x_{i2} \\ \sum x_{i2} & \sum x_{i1}x_{i2} & \sum x_{i2}^2 \end{pmatrix} \begin{pmatrix} \hat{b}_0 \\ \hat{b}_1 \\ \hat{b}_2 \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum x_{i1}y_i \\ \sum x_{i2}y_i \end{pmatrix}. \tag{3.8}$$

```
x3 <- c(2,1,4,6,8,3,12,3,4)
X<-matrix(c(rep(1,n),x1,x2,x3), n, 4)
X
Xt <- t(X)
Xt

matrix(c(n,sum(x1),sum(x2),sum(x3), sum(x1),sum(x1^2), sum(x1*x2), sum(x1*x3),
 sum(x2),sum(x1*x2),sum(x2^2), sum(x2*x3), sum(x3),sum(x1*x3),sum(x2*x3), sum(x3^3) ), nrow=4, ncol=4)

Xt%*%X
```

3.1 Least squares with two or more explanatory variables

(3.18)
$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{i1} & \cdots & x_{ip} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix}, \quad \hat{\mathbf{b}} = \begin{pmatrix} \hat{b}_0 \\ \hat{b}_1 \\ \vdots \\ \hat{b}_p \end{pmatrix}.$$

The dimensions are:

(3.19) $\dim(\mathbf{X}) =$ ________________

(3.20) $\dim(\mathbf{X}^T) =$ ________________

(3.21) $\dim(\mathbf{X}^T\mathbf{X}) =$ ________________

(3.22) $\dim(\mathbf{X}^T\mathbf{y}) =$ ________________

(3.23) $\dim(\hat{\mathbf{b}}) =$ ________________

3.1 Least squares with two or more explanatory variables

design matrix or *data matrix*

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{i1} & \cdots & x_{ip} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix}, \quad \hat{\mathbf{b}} = \begin{pmatrix} \hat{b}_0 \\ \hat{b}_1 \\ \vdots \\ \hat{b}_p \end{pmatrix}. \tag{3.18}$$

$$(\mathbf{X}^T\mathbf{X})\hat{\mathbf{b}} = \mathbf{X}^T\mathbf{y}$$

$$\text{or} \quad \hat{\mathbf{b}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

The *system of normal equations*

3.1 Least squares with two or more explanatory variables

As before, Y is a **random** vector and X is fixed.

(3.36) $$Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \epsilon_i, \quad i = 1, \ldots, n, \quad \epsilon_i \sim N(0, \sigma^2) \text{ independently.}$$

With $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^T$, the least square estimator is

(3.37) $$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_p \end{pmatrix} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y},$$

using (3.17) and (3.18).

3.1 Least squares with two or more explanatory variables

As before, Y is a **random** vector and X is fixed.

(3.36) $$Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \epsilon_i, \quad i = 1, \ldots, n, \quad \epsilon_i \sim N(0, \sigma^2) \text{ independently.}$$

With $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^T$, the least square estimator is

(3.37) $$\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_p \end{pmatrix} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y},$$

using (3.17) and (3.18).

```
> X <- matrix(c(rep(1,n),x1,x2),nrow=n,ncol=3)
> Xt <- t(X)
> y <- matrix(y, nrow=n, ncol=1)
>
> betahat <- solve(Xt %*% X) %*% Xt %*% y
> betahat
           [,1]
[1,] 23.2660767
[2,]  0.6814606
[3,] -0.2771398
```

multiple linear regression

Age vs. Money

PREDICTOR *variables*

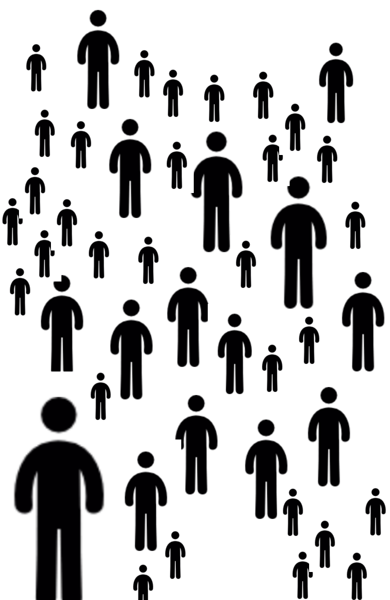
$x_1 \longrightarrow$ Age in Years

$x_2 \longrightarrow$ 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

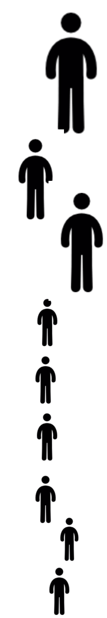
$b_0 = 23.26$
$b_1 = 0.68$
$b_2 = -0.28$
$s = 13.9$
$R^2 = 0.65$

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

3.1 Least squares with two or more explanatory variables

```
> X <- matrix(c(rep(1,n),x1,x2),nrow=n,ncol=3)
> Xt <- t(X)
> y <- matrix(y, nrow=n, ncol=1)
>
> betahat <- solve(Xt %*% X) %*% Xt %*% y
> betahat
           [,1]
[1,] 23.2660767
[2,]  0.6814606
[3,] -0.2771398
```

```
> yhat<-X%*%(betahat)
> yhat
          [,1]
 [1,] 71.94021
 [2,] 40.35196
 [3,] 50.58716
 [4,] 28.00404
 [5,] 31.94284
 [6,] 29.39922
 [7,] 30.88932
 [8,] 32.76097
 [9,] 14.12427
```

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

For later uses, k is the number of β's (here $p+1$) and column dimension of $\mathbf{X}$ but later it can be much more than the number of explanatory variables when binary dummy variables are created from categorical explanatory variables.

- Least squares estimates $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p$.
- Fitted or predicted values

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_p x_{ip}, \quad i = 1, \ldots, n. \tag{3.39}$$

- Residuals

$$e_i = y_i - \hat{y}_i, \quad i = 1, \ldots, n. \tag{3.40}$$

```
> 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

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

- Mean square of residuals or estimated σ^2:

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

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

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

3.4 Statistical software output for multiple regression

- Sum of squares of residuals

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

- Mean square of residuals or estimated σ^2:

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

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

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

```
> 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 numera

(3.44) $$SS(Total) = \sum_{i=1}^{n}(y_i - \bar{y})$$

- Multiple correlation coefficient or coefficient of detern

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

(3.46) $$\text{adj}R^2 \stackrel{\text{def}}{=} 1 - \frac{SS(Res}{SS(Tota}$$

```
> 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
```

R^2 measures the proportion of total variation in the y-variable about $\bar{y}$ 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.

multiple linear regression

Age vs. Money

PREDICTOR *variables*

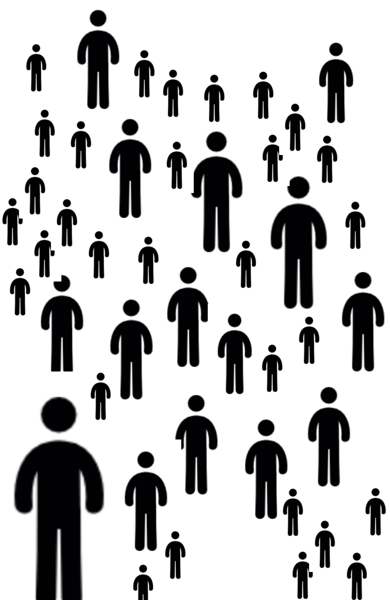
x_1 → Age in Years

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

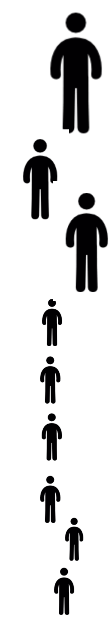
b_0 = 23.26
b_1 = 0.68
b_2 = -0.28
s = 13.9
R^2 = 0.65

For parameter β_1 :

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

p-value = 0.016

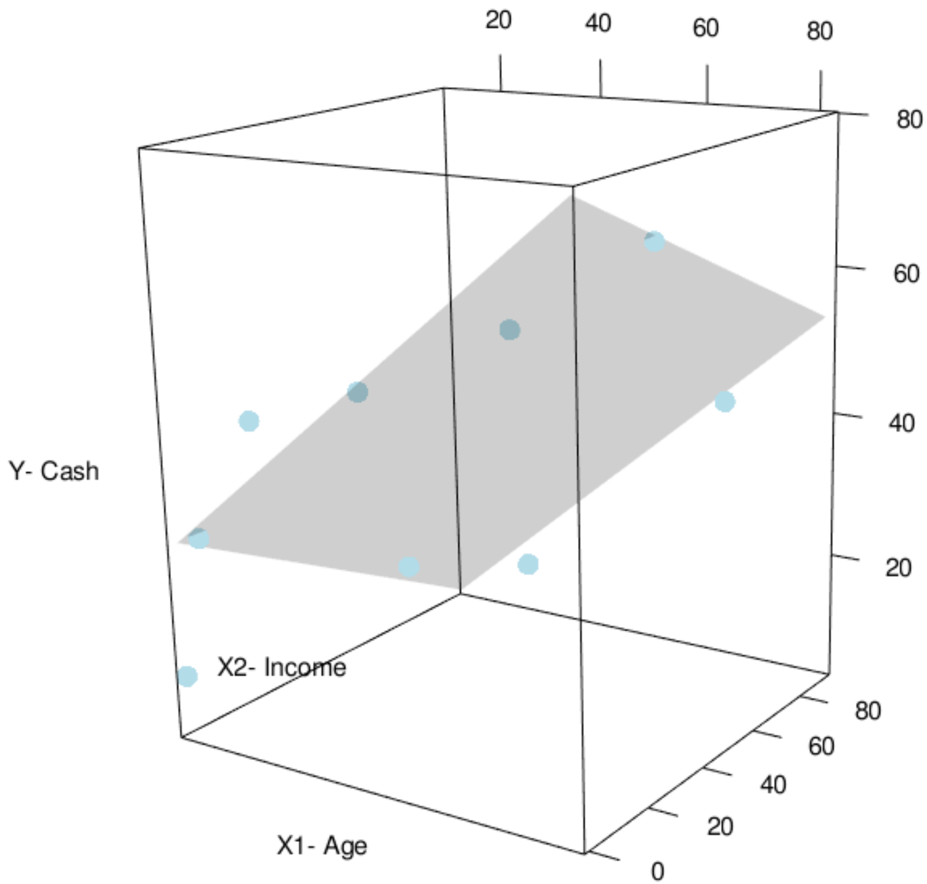
Sample, n=9



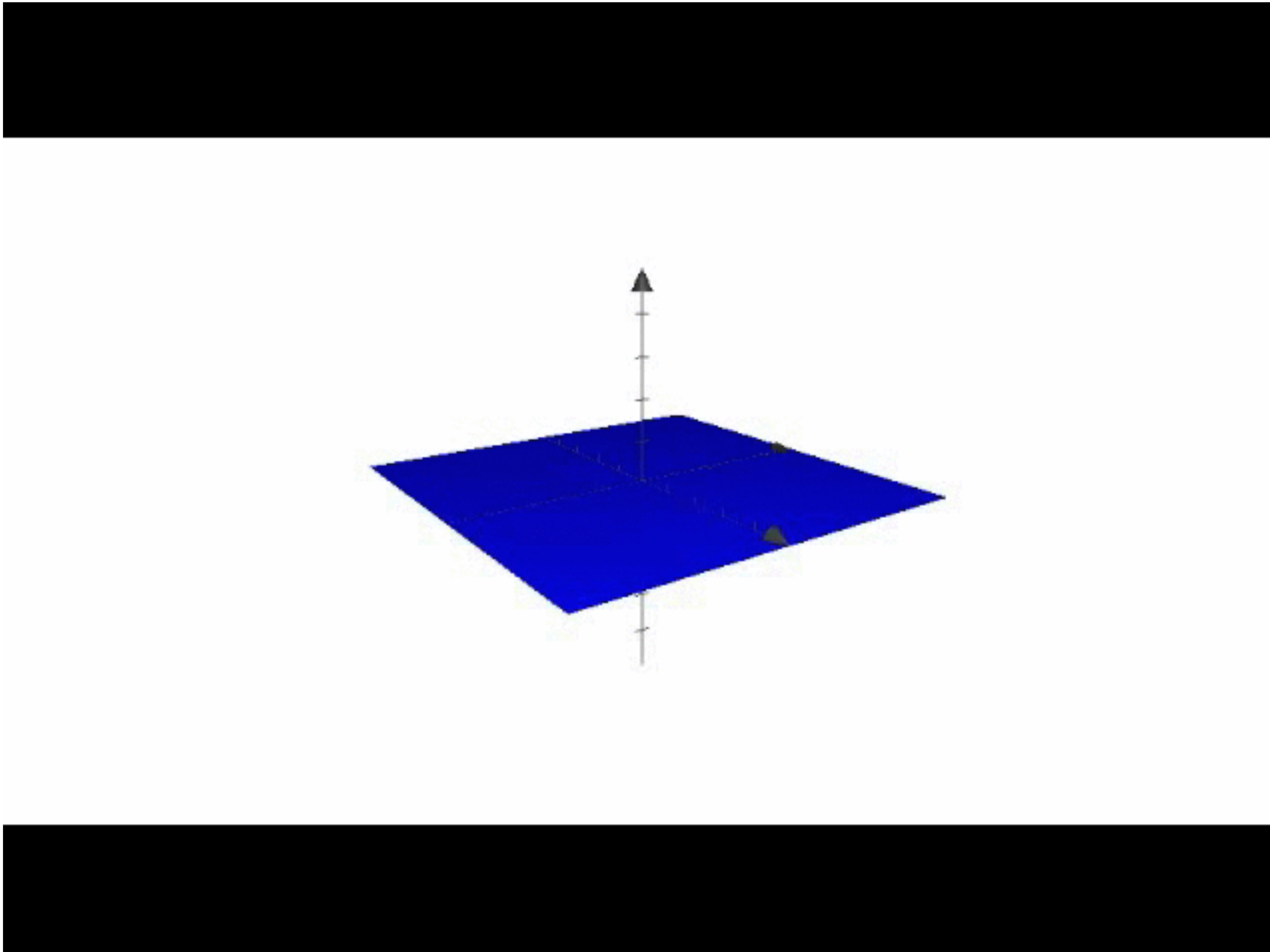
x_1	x_2	y
82	26	71
45	49	54
71	76	43
22	37	45
29	40	21
9	0	11
12	2	30
18	10	45
24	92	10

3.1 Least squares with two or more explanatory variables

"hyperplane equation"



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



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.