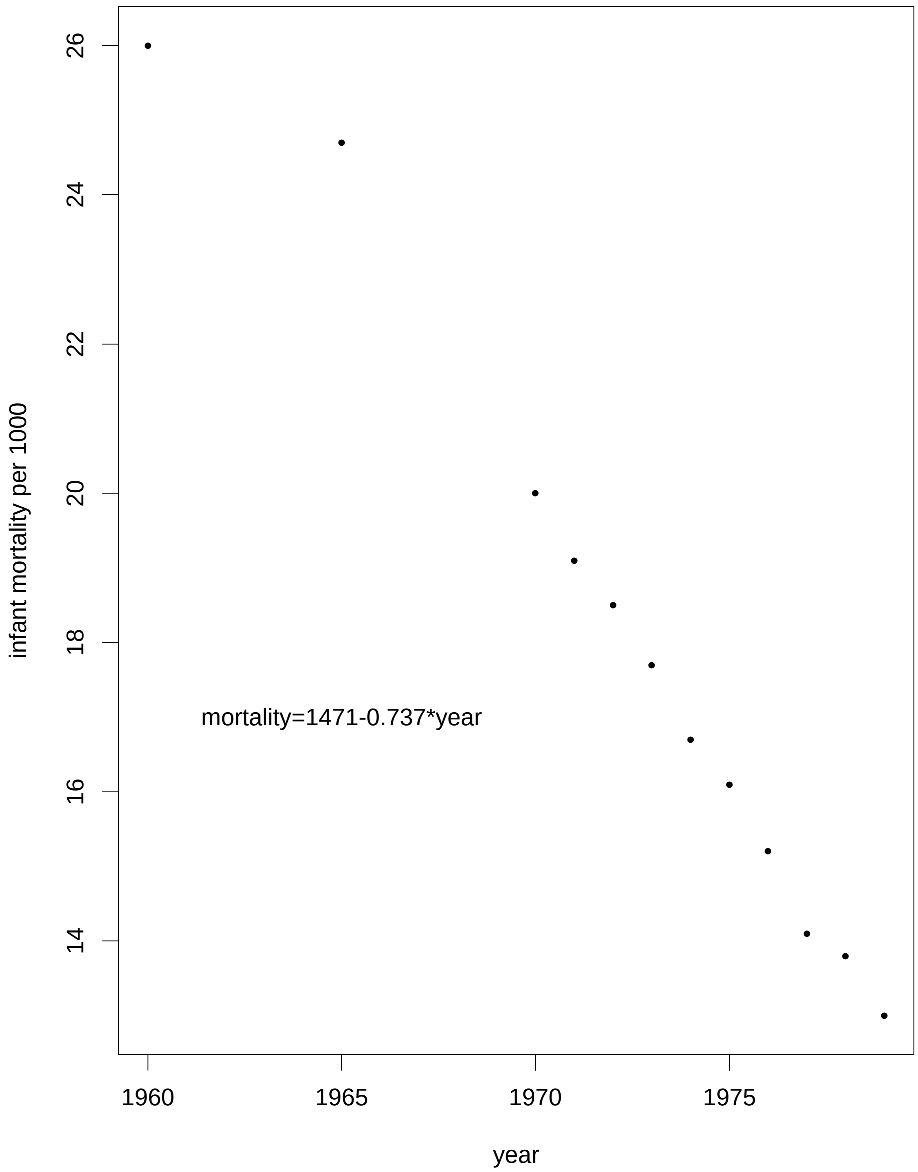
1. Figure 1 displays a scatter plot of US infant mortality rate per 1000 versus year, for $n = 12$ years from 1960-1979. A least-squares regression line is included. The sample Pearson correlation coefficient between mortality rate and year is -0.99.

a. Find the predicted infant mortality rate in 1989, assuming that the same trend continued for 10 more years after 1979.
i. 1471-0.737
ii. 1471-0.737(10)
iii. 1471-0.737(1989)

b. Is it possible that this same trend expressed by the estimated regression line could still hold true in the year 2002.
i. yes, because of continuing medical advances
ii. no, because it would imply negative mortality rates for 2002

Figure 1: Scatter plot of US infant mortality rate per 1000 versus year from 1960-1979.



2. Two models are proposed for the regression of a a dependent variable Y on a predictor X.
model 1: $y_i = \beta_0 + \beta_1 x_i + e_i$,
model 2: $y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + e_i$,
For each model, we assume that the e_is are independent and distributed $N(0, \sigma^2)$.
Both models are fitted on a collection of n=10 observations, and the results are summarized below.

Model 1 results :

Coefficients:

	Value	Std. Error	t value	Pr(>|t|)
(Intercept)	2.4672	1.6146	1.5280	0.1650
x	0.7552	0.2602	2.9022	0.0198

$\hat{\sigma}^2 = 5.588$
$SYY = 91.73$

Model 2 results

Coefficients:

	Value	Std. Error	t value	Pr(>|t|)
(Intercept)	-1.6576	4.2024	-0.3945	0.7069
x	5.0334	3.1499	1.5980	0.1612
x^2	-1.0284	0.6497	-1.5829	0.1645
x^3	0.0667	0.0390	1.7131	0.1375

$\hat{\sigma}^2$=4.687
$SYY = 91.73$

a. Test the null hypothesis that model 1 is the correct model versus the alternative hypothesis that model 2 is correct.

b. Compute the coefficient of determination R^2 for both model 1 and model 2. Show that in all datasets R^2 for model 2 will always be at least as great as R^2 for model 1.

3. Suppose that the effects of two drugs given in varying dosages (log(dose)=x) are studied by administering them to separate samples of n subjects each. The model for the effect of the first drug on the response variable y is

$$y_{i1} = \beta_0 + \beta_1 x_{i1} + e_{i1}$$

while for the second drug it is

$$y_{i2} = \beta_0 + \beta_2 x_{i2} + e_{i2}.$$

In each case $i = 1, 2, ..., n$, and $\bar{x}_1 = \bar{x}_2 = 0$. Assume that all observations are independent and e_1 and e_2 are normally distributed with mean 0 and common unknown variance σ^2.

a. Obtain the least squares estimate of $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$.

b. Construct an unbiased estimator of σ^2.

4. Eight observations of a dependent variable Y and two predictors X_1 and X_2 were collected. The data are given below.

Y	X_1	X_2
2	1	1
-2	-1	1
5	2	2
-2	-2	2
7	3	3
-3	-3	3
15	4	4
-4	-4	4

The following multiple regression model will be used.

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + e_i$$

for $i = 1, 2, ..., 8$. Assume $E[e_i] = 0$ and $var(e_i) = \sigma^2, cov(e_i, e_j) = 0$.

a. Find the least squares estimates $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$.

b. How would removing x_2 from the model affect the estimate of β_1?