Chapter 6

Gaussian Response Models

6.1 Gaussian Response Models Part I

6.1.1 Introduction

Example: STAT 230 and 231 Final Grades

No.	S230	S231
1	76	76
2	77	79
3	57	54
4	75	64
5	74	64
6	60	60
7	81	85
8	86	82
9	96	88
10	79	72

No.	S230	S231
11	87	76
12	71	50
13	63	75
14	77	72
15	96	84
16	65	69
17	71	43
18	66	60
19	90	96
20	50	50

No.	S230	S231
21	98	83
22	80	88
23	67	52
24	78	75
25	100	99
26	94	94
27	83	83
28	51	37
29	77	90
30	77	67

- Why might we be interested in collecting data such as these?
- What might be a reasonable choice for the target and study population?
- What are the variates? What type are they?
- What is the explanatory variate? What is the response variate?
- How do we summarize these data numerically and graphically?
- What model could we use to analyse these data?

6.1.2 Sample Correlation

Recall that the sample correlation is a numerical measure of the linear relationship between two variates. It is defined as

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}},$$

where

$$
\begin{aligned}
S_{xx} &= \sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n(\bar{x})^2. \\
S_{xy} &= \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}. \\
S_{yy} &= \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - n(\bar{y})^2.
\end{aligned}
$$

Recall that $-1 \le r \le 1$.

Sample Correlation for STAT 230/231 Final Grades

Let x be the STAT 230 final grade, and y be the STAT 231 final grade.
For these data, we have

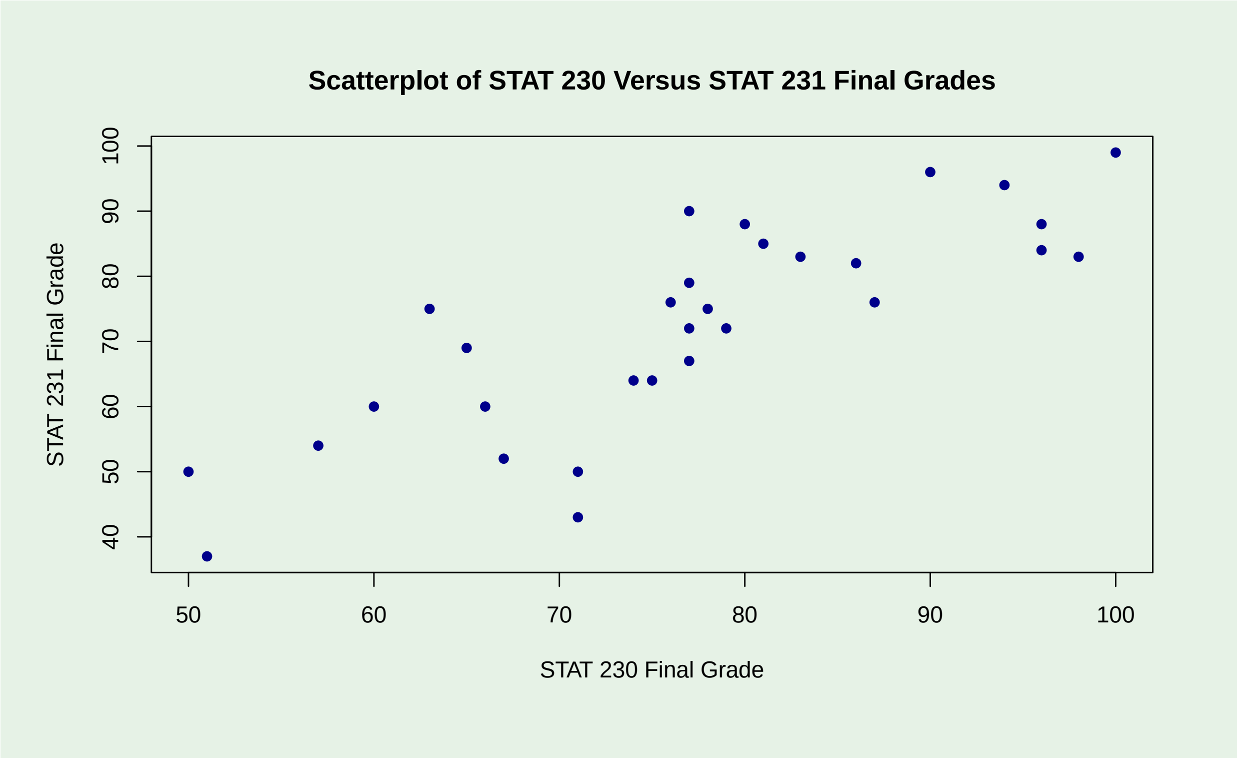
$$S_{xx} = 5135.8667, \qquad S_{xy} = 5106.8667, \qquad S_{yy} = 7585.3667.$$

Thus,

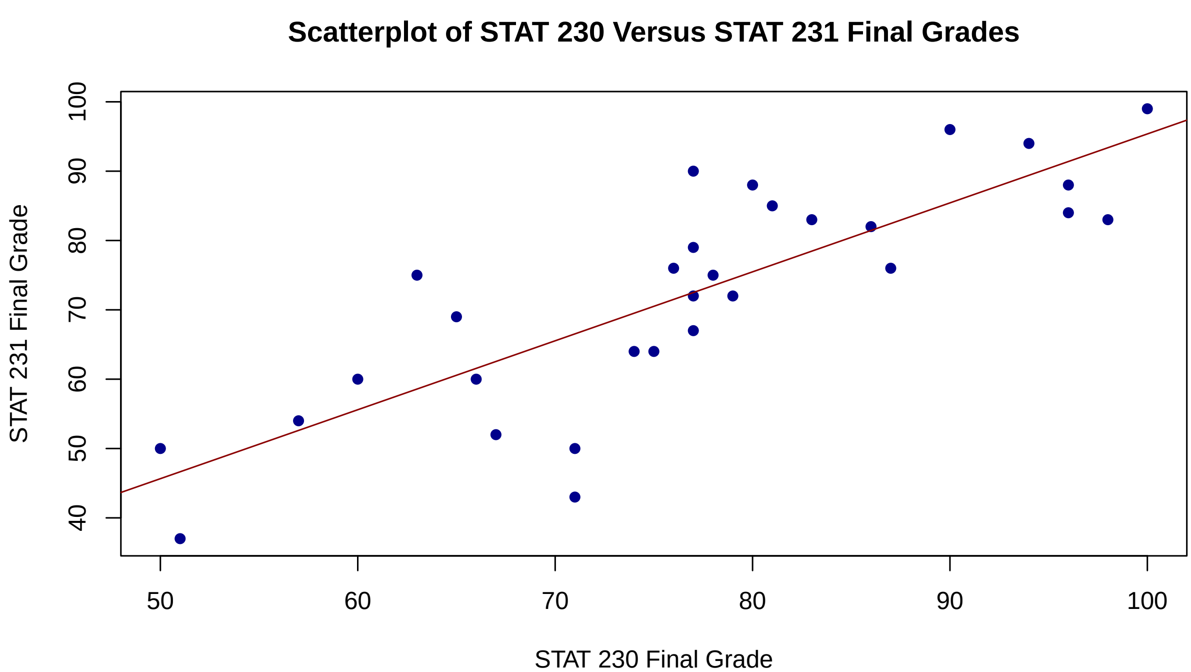
$$r = \frac{5106.8667}{\sqrt{(5135.8667)(7585.3667)}} = 0.82.$$

Since r is close to 1, we would say that there is a strong positive linear relationship between STAT 230 and STAT 231 final grades.

```r
dat <- read.table("data-6.1",header=T)
x <- dat$s230
y <- dat$s231
xbar <- mean(x)
ybar <- mean(y)
n <- length(x)
Sxx <- sum(x^2)-n*(mean(x))^2
Syy <- sum(y^2)-n*(mean(y))^2
Sxy <- sum(x*y)-n*mean(x)*mean(y)
Sxx

## [1] 5135.867

Syy

## [1] 7585.367

Sxy

## [1] 5106.867

r <- Sxy/sqrt(Sxx*Syy)
r

## [1] 0.8181998
```



6.1.3 Least Squares Estimates



To determine the fitted line $y = \alpha + \beta x$, which minimizes the sum of the squares of the distances between the observed points and the fitted line.

We need to find the values of α and β which minimize

$$g(\alpha, \beta) = \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2.$$

These values are determined by simultaneously solving the equations

$$\frac{\partial g}{\partial \alpha} = \frac{\partial}{\partial \alpha} \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2 = \sum_{i=1}^{n} 2(y_i - \alpha - \beta x_i)(-1) = 0,$$
$$\frac{\partial g}{\partial \beta} = \frac{\partial}{\partial \beta} \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2 = \sum_{i=1}^{n} 2(y_i - \alpha - \beta x_i)(-x_i) = 0.$$

These equations can be written as

$$\bar{y} - \alpha - \beta \bar{x} = 0, \tag{1}$$
$$\sum_{i=1}^{n} (y_i - \alpha - \beta x_i)(x_i) = 0. \tag{2}$$

From equation (1), we obtain $\alpha = \bar{y} - \beta \bar{x}$ which we can substitute into equation (2) to obtain

$$\sum_{i=1}^{n} x_i \big[y_i - \bar{y} - \beta(x_i - \bar{x}) \big] = 0,$$

or

$$\beta = \frac{\sum_{i=1}^{n} x_i (y_i - \bar{y})}{\sum_{i=1}^{n} x_i (x_i - \bar{x})} = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) y_i}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \frac{S_{xy}}{S_{xx}}.$$

Therefore, the least squares estimates are

$$\alpha = \hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}, \qquad \beta = \hat{\beta} = \frac{S_{xy}}{S_{xx}}.$$

And, the equation of the fitted line is

$$y = \hat{\alpha} + \hat{\beta} x.$$

6.1.4 STAT 231 Versus STAT 230 Final Grades

For the STAT 230/231 data, we have the following

$$\bar{x} = \frac{2302}{30} = 76.7333, \qquad \bar{y} = \frac{2167}{30} = 72.2333.$$
$$S_{xx} = 5135.8667, \qquad S_{xy} = 5106.8667, \qquad S_{yy} = 7585.3667.$$
$$\hat{\beta} = \frac{S_{xy}}{S_{xx}} = \frac{5106.8667}{5135.8667} = 0.9944.$$
$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} = 72.2333 - (0.9944)(76.7333) = -4.0667.$$

The fitted line is

$$y = -4.0667 + 0.9944x.$$