Concepts

1. **Least squares estimation.** Given paired data \((x_1, y_1), \ldots, (x_n, y_n)\), we find a line that minimizes the sum of the squared errors (SSE):

\[
SSE = \sum_{j=1}^{n} r_j^2 = \sum_{j=1}^{n} (y_j - \hat{y}_j)^2 = \sum_{j=1}^{n} (y_j - \beta_0 - \beta_1 x_j)^2.
\]

Then the regression line is \(y = \hat{a}_0 + \hat{a}_1 x\).

2. By differentiating SSE with respect to \(\beta_0\) and \(\beta_1\), we get normal equations:

\[
\beta_0 + \beta_1 = \bar{y} \\
\beta_0 + \beta_1 \cdot \bar{x} = \bar{y} \\
\]

Solving these equations, we get \(\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}\) and \(\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}\), where the sample covariance \(S_{xy} = n(\bar{xy} - \bar{x}\bar{y})\).

3. Outliers are data points with unusually large residuals. Outliers might cause the lack of fit of a regression line.

4. \(\sigma^2\) determines the amount of variability inherent in the linear model. An unbiased estimator of \(\sigma^2\) is \(\hat{\sigma}^2 = \frac{SSE}{n-2}\).

**In-class problems**

**Example 1.** 10 students took two midterm exams.

<table>
<thead>
<tr>
<th>Student</th>
<th>01 02 03 04 05 06 07 08 09 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Midterm 1</td>
<td>80 75 60 90 99 60 55 85 65 70</td>
</tr>
<tr>
<td>Midterm 2</td>
<td>70 60 70 72 95 66 60 80 70 60</td>
</tr>
</tbody>
</table>

Let’s find the regression line.

```r
> x<-c(80, 75, 60, 90, 99, 60, 55, 85, 65, 70)
> y<-c(70, 60, 70, 72, 95, 66, 60, 80, 70, 60)
> lm(y˜x)
```

```
Call: lm(formula = y ˜ x)
Coefficients: (Intercept) x
29.4827 0.5523
```

**Example 2.** One student who got 99 felt he did not need to take the second exam. So he stayed at home playing PS2 and got 0 in the second midterm exam (PS2 = Sony’s play station). Let’s see how his score influence the regression line.

```r
> lm(y˜x)
```

```
Call: lm(formula = y ˜ x)
Coefficients: (Intercept) x
113.27 -0.71
```

**Example 3.** Estimate \(\text{Var}(Y_j)\) in the linear model in Example 1.

```r
> (cov(y,y)-beta1*cov(x,y))/8
```

```
[1] 6.498408
```

**Self-study problems**

Example 12.5., 12.6., 12.7., 12.8.

**Homework 6**

Due April 10 (Thursday) 11:00am.