A line is (sometimes) a good way to describe the relationship between two quantitative variables.

1. The data: \((x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\).
2. Scatterplots.
3. Examples.
4. Models for the relationship between \(x\) and \(y\): \(y = f(x)\).

5. Simplest model: \(y = \alpha + \beta x\).

<table>
<thead>
<tr>
<th>Observation = Model + Error</th>
<th>Observation = Fitted + Residual</th>
</tr>
</thead>
<tbody>
<tr>
<td>(y = \alpha + \beta x + \varepsilon)</td>
<td>(y = a + bx + e)</td>
</tr>
</tbody>
</table>

6. How to choose \(a\) and \(b\)? Choose \(a\) and \(b\) to minimize sums of squares of residuals.

(a) \(\hat{y_i} = a + bx_i\) (a “hat” is pronounced “fitted”)

(b) \(e_i = y_i - \hat{y}_i\) (\(e_i\) is the \(i^{th}\) residual)

(c) Choose \(a\) and \(b\) so that the sum of squares of residuals is minimized:

\[
SS_{Resid} = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2
\]

> plot(loss~Fe,data=corrosion,main='Mass Loss for various Iron Levels (mg/dm^2)')
> plot(W~R,data=al2003,main='R vs W for AL Teams in 2003')
> plot(Seconds~Meters,data=mentrack,main='World Track Records (Men)')
> plot(stackloss)
> l=lm(loss~Fe,data=corrosion)
> l

Call:
\texttt{lm(formula = loss ~ Fe, data = corrosion)}

Coefficients:

\begin{align*}
\text{(Intercept)} & \quad \text{Fe} \\
129.79 & \quad -24.02
\end{align*}

> abline(l)
Homework

1. Read Devore and Farnum Section 3.1 and (especially) pages 114–117.

2. Do problems 3.1.8 and 3.3.24abc