1. **Coefficient of determination** measures the proportion of the total variations in $y$ that can be explained by the linear model and it is defined as

$$r^2 = 1 - \frac{SSE}{SST}$$

Since $SST = S_{yy}$ and $SSE = S_{yy} - S_{xy}^2/S_{xx}$, we can see that $r^2 = S_{xy}^2/(S_{xx}S_{yy})$. $r$ is called the **sample correlation coefficient**.

2. **Example.** Hooke’s law states that the length change $y$ in spring is proportional to applied force $x$. The following measurements are available.

<table>
<thead>
<tr>
<th>x (kg)</th>
<th>29.4</th>
<th>39.2</th>
<th>49.0</th>
<th>58.8</th>
<th>68.6</th>
<th>78.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>y (mm)</td>
<td>4.25</td>
<td>5.25</td>
<td>6.50</td>
<td>7.85</td>
<td>8.75</td>
<td>10.10</td>
</tr>
</tbody>
</table>

We fit the model $Y = \beta_0 + \beta_1 x + \epsilon$.

```r
> x<-c(29.4,39.2,49.0,58.8,68.6,78.4)
> y<-c(4.25,5.25,6.50,7.85,8.75,10.10)
> summary(lm(y~x))
```

Call: lm(formula = y ~ x)

Residuals:

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.06905</td>
<td>-0.10524</td>
<td>-0.02952</td>
<td>0.14619</td>
<td>-0.12810</td>
<td>0.04762</td>
</tr>
</tbody>
</table>

Coefficients:

| Estimate | Std. Error | t value | Pr(>|t|) |
|----------|------------|---------|----------|
| (Intercept) | 0.658095 | 0.163995 | 4.013 | 0.0160 * |
| x | 0.119825 | 0.002906 | 41.238 | 2.07e-06 *** |

---

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1191 on 4 degrees of freedom
Multiple R-Squared: 0.9977, Adjusted R-squared: 0.9971
F-statistic: 1701 on 1 and 4 DF, p-value: 2.067e-06
3. The least squares estimation for $\beta_1$ are given by $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$ where $S_{xy} = n(\bar{xy} - \bar{x}\bar{y})$. Since $\sum_{j=1}^{n}(x_j - \bar{x})Y = 0$, $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \sum_{j=1}^{n}c_jY_j$, where $c_j = (x_j - \bar{x})/S_{xx}$. From $\sum_{j=1}^{n}c_j = 0$, $\sum_{j=1}^{n}c_jx_j = 1$, we can show that $\mathbb{E}\hat{\beta}_1 = \beta_1$ showing unbiasedness. From $\sum_{j=1}^{n}c_j^2 = S_{xx}^{-1}$, $\forall \hat{\beta}_1 = \sigma^2/S_{xx}$. Since $\hat{\beta}_1$ is a linear combination of normal distributions, $\hat{\beta}_1 \sim N(\beta_1, \sigma^2/S_{xx})$. The test statistic

$$T = \frac{\hat{\beta}_1 - \beta_1}{S_{\hat{\beta}_1}} \sim t_{n-2},$$

where $S_{\hat{\beta}_1} = \hat{\sigma}/\sqrt{S_{xx}}$ and $\hat{\sigma} = SSE/(n-2)$. It can be shown that $SSE = S_{yy} - S_{xy}^2/S_{xx}$ and $S_{\beta_1} = \frac{1}{\sqrt{n-2}}\sqrt{\frac{S_{xx}}{S_{xx}} - (\frac{S_{xy}}{S_{xx}})^2}$. In a slightly different way, $SSE = \sum_{j=1}^{n}y_j^2 - \hat{\beta}_0\sum_{j=1}^{n}y_j - \hat{\beta}_1\sum_{j=1}^{n}x_jy_j$.

4. Based on the test statistic, a $100(1-\alpha)\%$ CI for $\beta_1$ is $\hat{\beta}_1 \pm t_{\alpha/2, n-2}S_{\hat{\beta}_1}$. For hypothesis testing $H_0: \beta_1 = 0$ vs. $H_1: \beta_1 \neq 0$

we reject $H_0$ if $|t| > t_{\alpha/2, n-2}$ at $100(1-\alpha)\%$ significance. A hypothesis testing on $\beta_0$ and a CI construction are similar. $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{x}$ is distributed as normal with mean $\beta_0$ and variance $\sigma^2x^2/S_{xx}$. Testing $H_0: \beta_0 = 0$ is based on $T = \frac{\hat{\beta}_0 - \beta_0}{S_{\beta_0}} \sim t_{n-2},$ where $S_{\beta_0} = \sigma^2x^2/S_{xx}$.

**Assigned problems.** Exercise 12.30., 12. 36. HW 6 due April 28th. Solve the assigned problems in Lecture 19-22.