Hypothesis Tests

Recall the `lm()` regression output with two independent variables:

Call:
lm(formula = Strength ~ Length + Height, data = wireBond)

Residuals:

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>1Q</th>
<th>Median</th>
<th>3Q</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-3.865</td>
<td>-1.542</td>
<td>-0.362</td>
<td>1.196</td>
<td>5.841</td>
</tr>
</tbody>
</table>

Coefficients:

|                | Estimate | Std. Error | t value | Pr(>|t|) |
|----------------|----------|------------|---------|---------|
| (Intercept)    | 2.263791 | 1.060066   | 2.136   | 0.044099 * |
| Length         | 2.744270 | 0.093524   | 29.343  | < 2e-16 *** |
| Height         | 0.012528 | 0.002798   | 4.477   | 0.000188 *** |

---

Signif. codes:  0 ***  0.001 **  0.01 *  0.05 .  0.1   1
Residual standard error: 2.288 on 22 degrees of freedom
Multiple R-squared: 0.9811, Adjusted R-squared: 0.9794
F-statistic: 572.2 on 2 and 22 DF, p-value: < 2.2e-16

The statistical model is

\[ Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon. \]

The first test (“test for significance of regression”) is of the null hypothesis that neither variable has any effect; \( H_0 : \beta_1 = \beta_2 = 0. \)

The test statistic is the \( F \)-statistic on the last line,

\[ F \text{-statistic: 572.2 on 2 and 22 DF, p-value: < 2.2e-16, which is large and highly significant: we reject this null hypothesis.} \]
We now test the individual coefficients, using the $t$-statistics on the respective rows of the output:

- $H_0 : \beta_1 = 0, \; t = 29.343$;
- $H_0 : \beta_2 = 0, \; t = 4.477$.

Both have small $P$-values, so both null hypotheses are rejected, and we conclude that both predictors are needed in the regression model.
The second-order model

The output for the second-order model in one independent variable is similar:

Call:
`lm(formula = Strength ~ Length + I(Length^2), data = wireBond)`

Residuals:

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>1Q</th>
<th>Median</th>
<th>3Q</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-5.104</td>
<td>-2.092</td>
<td>0.564</td>
<td>1.807</td>
<td>4.870</td>
</tr>
</tbody>
</table>

Coefficients:

|                  | Estimate  | Std. Error  | t value  | Pr(>|t|)   |
|------------------|-----------|-------------|----------|------------|
| (Intercept)      |  8.83254  |  1.49359    |  5.914   |  5.96e-06  *** |
| Length           |   1.7432  |   0.36951   |  4.718   |  0.000105  *** |
| I(Length^2)     |   0.0609  |   0.01871   |  3.255   |  0.003626  ** |

---

Signif. codes:  0 ***  0.001 **  0.01 *  0.05 . 0.1  1
Residual standard error: 2.598 on 22 degrees of freedom
Multiple R-squared: 0.9757, Adjusted R-squared: 0.9735
F-statistic: 441.2 on 2 and 22 DF, p-value: < 2.2e-16

The statistical model is

\[ Y = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \epsilon. \]

Again, first test the significance of the overall model,
\[ H_0 : \beta_1 = \beta_2 = 0; \]
F-statistic: 441.2 on 2 and 22 DF, p-value: < 2.2e-16 so we reject \( H_0 \).
Next, test $H_0 : \beta_2 = 0$, using the $t$-statistic; $t = 3.255$, so again we reject $H_0$.

Because we have decided that $\beta_2 \neq 0$, we do not test the significance of $\beta_1$.

- When the second-order term is needed in the model, we want to include the first-order term even if the coefficient is small.
- In the second-order model, the null hypothesis $\beta_1 = 0$ rarely if ever corresponds to any question of practical interest, so there is no point in testing it.
Confidence Intervals

Individual coefficient

Confidence intervals for individual coefficients are constructed in the usual way:

\[ \hat{\beta}_j \pm t_{\alpha/2, \nu} \times \text{estimated standard error of } \hat{\beta}_j \]

for a 100(1 − \(\alpha\))% confidence interval, where \(\nu = n − (k + 1)\) is the residual degrees of freedom.
Mean response

We also need to estimate the mean response for new predictor values $x_{1,\text{new}}, x_{2,\text{new}}, \ldots, x_{k,\text{new}}$:

$$\hat{Y}_{\text{new}} = \hat{\beta}_0 + \hat{\beta}_1 x_{1,\text{new}} + \cdots + \hat{\beta}_k x_{k,\text{new}}.$$ 

The confidence interval is of course

$$\hat{Y}_{\text{new}} \pm t_{\alpha/2,\nu} \times \text{estimated standard error}$$

but the calculation is complicated, and best handled by software:

```r
code
wireBondLm1 <- lm(Strength ~ Length + Height, wireBond)
predict(wireBondLm1, data.frame(Length = 8, Height = 275), interval = "confidence")
```

"Multiple Linear Regression"
Prediction Interval

As before, we may want a *prediction* interval, for a *single* new observed response $Y_{\text{new}}$:

```
predict(wireBondLm1, data.frame(Length = 8, Height = 275),
         interval = "prediction")
```

The prediction interval at the new predictors is always wider than the confidence interval at the same new predictors.
Interpretation

Suppose that a new item will be manufactured with semiconductors of die height 275 attached with wires of length 8.

- We are 95% *confident* that the mean pull-off strength for items from this process will be between 26.66 and 28.66.
- If a single prototype is produced and tested, there is a 95% *probability* that its pull-off strength will be between 22.81 and 32.51.
Coefficient of Determination

How well does the model fit the observed data?

If we had no predictors, we would predict \( Y \) by \( \beta_0 \), estimated by \( \bar{y} \); the sum of squared residuals would be just the total sum of squares,

\[
SS_T = \sum_{i=1}^{n} (y_i - \bar{y})^2.
\]

Using the regression model, we predict \( Y \) by \( \hat{Y} \); the sum of squared residuals is the residual sum of squares,

\[
SS_E = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2.
\]
The coefficient of determination, or the “fraction of variability explained by the model”, is

\[ R^2 = 1 - \frac{SS_E}{SS_T} \]

\[ = \frac{SS_T - SS_E}{SS_T} \]

\[ = \frac{SS_R}{SS_T} \]

where

\[ SS_R = SS_T - SS_E \]

is the sum of squares for the regression.
The $R^2$ values for the two-variable model and the second-order model are 98.11% and 97.57%, respectively.

Both explain around 98% of the variability in pull-off strength; the two-variable model explains a little more than the second-order model.

We can combine them:

```r
summary(lm(Strength ~ Length + I(Length^2) + Height, wireBond))
```

and $R^2$ increases to 98.64%.
A problem

$R^2$ always increases when we add a new predictor to a model.

For instance, if we add Height$^2$ to the model, it is not significant, but $R^2$ increases to 98.66%.

One solution is to use the *adjusted* $R^2$,

$$R^2_{\text{adj}} = 1 - \frac{SS_E}{SS_T/n - (k + 1)} = 1 - \frac{MS_E}{MS_T}$$

which increases only if the new predictor reduces $MS_E$, the residual mean square; however, $R^2_{\text{adj}}$ can be negative.

Both $R^2$ and $R^2_{\text{adj}}$ are reported by most software.
Another problem

\( R^2_{\text{adj}} \) increases when we add a new predictor to a model, if and only if the \( t \)-ratio for the new predictor has \(|t| > 1\).

So the new \( \hat{\beta} \) may not be significantly different from 0, but still \( R^2_{\text{adj}} \) increases.

Suppose we change the question: how well will the model predict new observations?

Ideally, we collect new data and test the model on them: a validation exercise.
Cross validation

If we have no new data, we cannot carry out a true validation, but we can use *cross* validation:

- Leave out the $i^{th}$ observation, and refit the model to the remaining observations;
- Use the refitted model to predict the left-out response $y_i$, writing $\hat{y}_{(i)}$ for the prediction;
- The Prediction Error Sum of Squares is

$$PRESS = \sum_{i=1}^{n} (y_i - \hat{y}_{(i)})^2.$$
A statistic that corresponds to $R^2$ is

$$P^2 = 1 - \frac{\text{PRESS}}{SS_T}.$$ 

In R

```r
library("qpcR")
wireBondLm2 <- lm(Strength ~ Length + I(Length^2) + Height, wireBond)
PRESS(wireBondLm2)$P.square
wireBondLm3 <- lm(Strength ~ Length + I(Length^2) + Height + I(Height^2), wireBond)
PRESS(wireBondLm3)$P.square
```