Chapter 10: Inference for regression models

Introductory Statistics for Engineering Experimentation

Peter R. Nelson, Marie Coffin and Karen A.F. Copeland

Slides by Douglas Bates

Outline

10.1 Inference for a regression line

Section 10.1: Inference for a regression line

- Recall that a simple linear regression model is

$$\mathcal{Y}_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, \ldots, n$$

- The least squares estimates, $\widehat{\beta}_0$ and $\widehat{\beta}_1$, of the coefficients are functions of the data and hence are random variables. We associate *standard errors* with these estimates.
- The text derives formulas for the variance of the estimators. The formulas can be interesting but do not easily extend to more complex models. It is easier to simply read the standard error from the output.
- In the `R` output each coefficient estimate is accompanied by a `Std. Error` (standard error), a `t value` (the ratio of the estimate and its standard error) and a `Pr(>|t|)`, which is the p-value for the two-sided hypothesis test. The `confint` extractor can be used to determine confidence intervals.

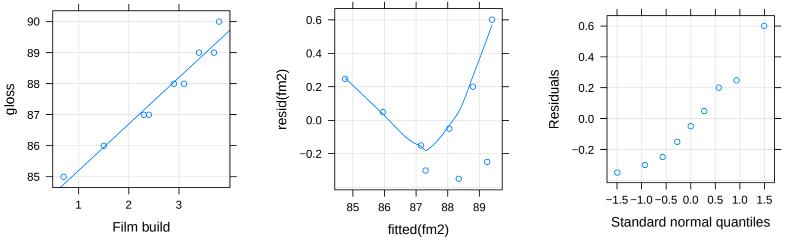
Examples 10.1.1 and 10.1.2

```
> summary(fm1 <- lm(time ~ temp, timetemp,
+                    subset = type == "Repaired"))
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -37.4902     2.7833  -13.47 3.52e-08
temp         -1.8643     0.1069  -17.45 2.30e-09
Residual standard error: 0.7699 on 11 degrees of freedom
Multiple R-squared: 0.9651, Adjusted R-squared: 0.9619
F-statistic: 304.3 on 1 and 11 DF,  p-value: 2.303e-09
> confint(fm1)

                 2.5 %     97.5 %
(Intercept) -43.616328 -31.364111
temp         -2.099460  -1.629053
```

- The confidence interval ($[-2.099, -1.629]$) on β_1, the slope, is of interest. The other confidence interval is not of interest because β_0 is meaningless for these data.

Example 10.1.3




```
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  83.7031     0.3184  262.86 3.04e-15
build         1.4988     0.1130   13.27 3.23e-06
Residual standard error: 0.3306 on 7 degrees of freedom
Multiple R-squared: 0.9618, Adjusted R-squared: 0.9563
F-statistic:   176 on 1 and 7 DF,  p-value: 3.234e-06
> confint(fm2)

                2.5 %    97.5 %
(Intercept) 82.950116 84.456061
build        1.231687  1.765977
```

Inference for coefficients

- As seen in the previous slides, we can evaluate confidence intervals on the coefficients, β_0 and β_1, with the `confint` extractor function.
- The formula for the $(1-\alpha)$ confidence interval on β_1 is

$$\widehat{\beta}_1 \pm t(\alpha/2, \nu)\, s_{\beta_1}$$

  where ν is the degrees of freedom for residuals ($n-2$ for a simple linear regression) and s_{β_1} is the standard error for the coefficient.
- The observed t statistic, $\widehat{\beta}_1/s_{\beta_1}$, is used to perform tests of the hypothesis $H_0 : \beta_1 = 0$. The p-value for the two-sided alternative is given in the coefficient table. The p-value for the one-sided alternative that is indicated by the data will be half this value. By "indicated by the data" I mean the alternative $H_a : \beta_1 > 0$, if $\widehat{\beta}_1 > 0$ and vice versa.