Preview on Testing and Inference: Assessing Accuracy of Coefficient Estimates

Hypothesis Test on the Coefficients

- When you run a linear regression in R and look at the summary, you see more than just the coefficient estimates.

```
Call:
lm(formula = mpg ~ horsepower, data = Auto)

Residuals:
     Min       1Q   Median       3Q      Max 
-13.5710  -3.2592  -0.3435   2.7630  16.9240 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 39.935861   0.717499   55.66   <2e-16 ***
horsepower  -0.157845   0.006446  -24.49   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.906 on 390 degrees of freedom
Multiple R-squared:  0.6059,	Adjusted R-squared:  0.6049 
F-statistic: 599.7 on 1 and 390 DF,  p-value: < 2.2e-16
```

Hypothesis Test on the Coefficients

- Suppose we have a linear model $Y = \beta_0 + \beta_1 X + \epsilon$.
- H_0 (null hypothesis): There is no relationship between X and Y
- H_a (alternative hypothesis): There is some relationship between X and Y
- Mathematically, this corresponds to testing

$$H_0 : \beta_1 = 0$$

versus

$$H_a : \beta_1 \neq 0$$

Hypothesis Test on the Coefficients

- To test the null hypothesis, we need to determine whether $\hat{\beta}_1$, our estimate for β_1 is sufficiently far from zero so that we can be confident that β_1 is non-zero.
- We are using the concept of the *Standard Error*: When we use linear regression, we get an **estimate** of β, which we denote by $\hat{\beta}$. This **estimate** may be too high or too low. The standard error, roughly speaking, gives an indication about how much our estimate $\hat{\beta}$ differs on average from the true parameter β.
- If the standard error is small, then even small values of β_1 could provide enough evidence that $\beta_1 \neq 0$.
- If the standard error is large, then $\hat{\beta}_1$ must be large in absolute value in order for us to reject the null hypothesis.

Hypothesis Test on the Coefficients

- In practice, we compute the t-statistic

$$t = \frac{\hat{\beta}_1 - 0}{\mathsf{SE}(\hat{\beta}_1)}$$

- For now: We can compute the probability of observing a value equal to or larger than $|t|$ under the null hypothesis (i.e. assuming that $\beta_1 = 0$). This probability is called the **p value**.

Hypothesis Test on the Coefficients

- Broadly speaking: A small p-value indicates that it is unlikely to observe such a substantial association between the predictor and the response due to chance, in the absence of any real association between the predictor and the response.
- If we see a small p-value, we can *reject the null hypothesis*.
- Typical p-value cutoffs for rejecting the null hypothesis are 0.05 or 0.01.