The Effect of Vitamin C on Tooth Growth in Guinea Pigs

Available in *R* via *ToothGrowth*. From the description:
The response is the length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2 mg) with each of two delivery methods (orange juice or ascorbic acid). Each combination is tested 10 times.
Source: C. I. Bliss (1952) The Statistics of Bioassay. Academic Press.
References: McNeil, D. R. (1977) Interactive Data Analysis. New York: Wiley.
$n = 60$
$Y_i =$ Tooth length
$S_i = \begin{cases} 0 & \text{orange juice} \\ 1 & \text{ascorbic acid} \end{cases}$
$d_{1i} = I(\text{dose} = 1)$
$d_{2i} = I(\text{dose} = 2)$

Model

$$Y_i = \beta_1 + S_i\beta_2 + d_{1i}\beta_3 + d_{2i}\beta_4 + \epsilon_i$$

Computing $\hat{\beta}$:

```
> Y = ToothGrowth$len
> X = cbind(1,ToothGrowth$supp=="VC",ToothGrowth$dose==1,ToothGrowth$dose==2)
> hbeta <- solve(t(X)%*%X)%*%t(X)%*%Y
> hbeta
```

```
       [,1]
[1,] 12.455
[2,] -3.700
[3,]  9.130
[4,] 15.495
```

We compute the fitted values $\hat{\mathbf{Y}}$ and residuals $\mathbf{e}$:

```
> P <- X%*%solve(t(X)%*%X)%*%t(X)
> Yhat <- P%*%Y
> e <- Y - Yhat
```

Is the delivery method important? Will test

$$H_0 : \beta_2 = 0 \text{ against } H_1 : \beta_2 \neq 0$$

```
> RSS <- t(e)%*%e
> c <- c(0,1,0,0)
> est <- c%*%hbeta
> est
```

```
     [,1]
[1,] -3.7
```

```
> sdhat <- sqrt(t(c) %*% solve(t(X)%*%X) %*% c*RSS/(60-4))
> sdhat
```

```
          [,1]
[1,] 0.9882795
```

A 95% confidence interval for $\mathbf{c}^T\boldsymbol{\beta}$:

```
> L <- est+sdhat*qt(0.025,df=60-4)
> U <- est+sdhat*qt(0.975,df=60-4)
> cat("[",L,U,"]\n")
```

```
[ -5.679762 -1.720238 ]
```

0 is not in the CI -> reject $H_0 : \beta_2 = 0$.

Next, we want to compute the p-value of the test that rejects for large values of $|T|$, where

$$T = \frac{\mathbf{c}^T\hat{\boldsymbol{\beta}}}{\sqrt{\mathbf{c}^T(X^TX)^{-1}\mathbf{c}\frac{\text{RSS}}{n-p}}}$$

Let t denote the observed value of T. Under H_0, $T \sim t_{60-4}$. Thus the p-value is

$$p = \text{P}(|T| \geq |t|) = \text{P}(T < -|t| \text{ or } T \geq |t|) = \underbrace{\text{P}(T < -|t|)}_{=\text{P}(T\geq|t|)} + \text{P}(T \geq |t|) = 2(1 - \text{P}(T \leq |t|))$$

```
> abst <- abs(est/sdhat)
> 2*(1-pt(abst,df=60-4))
```

```
             [,1]
[1,] 0.0004292793
```

One can get most of the above directly using the function *lm*

```
> summary(lm(len~supp+factor(dose), data=ToothGrowth))
```

```
Call:
lm(formula = len ~ supp + factor(dose), data = ToothGrowth)

Residuals:
   Min     1Q Median     3Q    Max 
-7.085 -2.751 -0.800  2.446  9.650 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)    12.4550     0.9883  12.603  < 2e-16 ***
suppVC         -3.7000     0.9883  -3.744 0.000429 ***
factor(dose)1   9.1300     1.2104   7.543 4.38e-10 ***
factor(dose)2  15.4950     1.2104  12.802  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.828 on 56 degrees of freedom
Multiple R-squared: 0.7623,	Adjusted R-squared: 0.7496 
F-statistic: 59.88 on 3 and 56 DF,  p-value: < 2.2e-16
```