STAC67H3F: Quiz 2, written part: solutions

TUT0001 (Tue 3 Nov) and TUT0002 (Wed 4 Nov), Fall 2026

Version A: TUT0001, greenhouse tomatoes

$n = 25$, $p = 3$, $n - p = 22$, so $MSE = SSE/(n-p) = 88/22 = 4$.

(a) Interpretation, fitted value and residual

For plots receiving the **same amount of irrigation**, each additional gram of fertilizer per m^2 goes with an increase of 1.2 kg in the mean yield.

For the plot with $x_1 = 4$ and $x_2 = 6$:

$$\hat{Y} = 10 + 1.2(4) + 0.6(6) = 10 + 4.8 + 3.6 = 18.4 \text{ kg}, \qquad e = Y - \hat{Y} = 20.1 - 18.4 = 1.7 \text{ kg}.$$

(b) The t test for irrigation

$H_0 : \beta_2 = 0$ against $H_a : \beta_2 \neq 0$. With $c_{22} = 0.01$, the diagonal entry of $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ for x_2:

$$s\{\hat{\beta}_2\} = \sqrt{MSE\, c_{22}} = \sqrt{4 \times 0.01} = 0.2, \qquad t^* = \frac{\hat{\beta}_2}{s\{\hat{\beta}_2\}} = \frac{0.6}{0.2} = 3.00.$$

Under H_0, $t^* \sim t(22)$, and $t(0.975;\, 22) = 2.074$. Since $|t^*| = 3.00 > 2.074$, **reject** H_0 at the 5% level: once fertilizer is taken into account, irrigation is related to the mean yield (more water, higher mean yield).

(c) A confidence interval for a mean response

With $\boldsymbol{X}_h = (1, 8, 9)'$, the estimated mean is $\hat{Y}_h = 10 + 1.2(8) + 0.6(9) = 10 + 9.6 + 5.4 = 25$ kg. Next,

$$(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_h = \begin{bmatrix} -0.11 \\ 0.025 \\ 0 \end{bmatrix}, \qquad \boldsymbol{X}_h'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_h = -0.11 + 8(0.025) + 9(0) = 0.09,$$

so $s^2\{\hat{Y}_h\} = MSE \times 0.09 = 0.36$ and $s\{\hat{Y}_h\} = 0.6$. With $t(0.975;\, 22) = 2.074$:

$$\hat{Y}_h \pm t(0.975;\, 22)\, s\{\hat{Y}_h\} = 25 \pm 2.074 \times 0.6 = 25 \pm 1.2443 = (23.76,\ 26.24).$$

Version B: TUT0002, iced drinks

$n = 20$, $p = 3$, $n - p = 17$.

(a) The estimates

$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$, so $\hat{\beta}_1$ and $\hat{\beta}_2$ are the second and third rows of $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ times $\boldsymbol{X}'\boldsymbol{Y}$:

$$\hat{\beta}_1 = (-0.26)(1600) + (0.012)(40\,380) + (-0.004)(16\,140) = -416 + 484.56 - 64.56 = 4,$$
$$\hat{\beta}_2 = (-0.08)(1600) + (-0.004)(40\,380) + (0.018)(16\,140) = -128 - 161.52 + 290.52 = 1.$$

Interpretation: on days with the **same number of hours of sunshine**, each additional degree of maximum temperature goes with 4 more iced drinks sold, on average.

(b) A confidence interval for β_2

$MSE = SSE/(n-p) = 340/17 = 20$ and $c_{22} = 0.018$, the diagonal entry of $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ for x_2, so

$$s\{\hat{\beta}_2\} = \sqrt{20 \times 0.018} = \sqrt{0.36} = 0.6.$$

With $t(0.975;\, 17) = 2.110$:

$$\hat{\beta}_2 \pm t(0.975;\, 17)\, s\{\hat{\beta}_2\} = 1 \pm 2.110 \times 0.6 = 1 \pm 1.266 = (-0.27,\ 2.27).$$

With 95% confidence, on days with the same maximum temperature, each additional hour of sunshine changes the mean number of iced drinks sold by between -0.27 and 2.27. The interval contains 0: once temperature is taken into account, the data do not show that sunshine is related to sales (at the 5% level).

(c) Adjusted R^2

The first column of $\boldsymbol{X}$ is a column of ones, so the first entry of $\boldsymbol{X}'\boldsymbol{Y}$ is $\sum_i Y_i = 1600$, and $\boldsymbol{Y}'\boldsymbol{Y} = \sum_i Y_i^2$. Hence

$$SST = \sum_i Y_i^2 - n\bar{Y}^2 = 130\,000 - \frac{1600^2}{20} = 130\,000 - 128\,000 = 2000,$$

$$R_a^2 = 1 - \frac{SSE/(n-p)}{SST/(n-1)} = 1 - \frac{340/17}{2000/19} = 1 - \frac{20 \times 19}{2000} = 1 - 0.19 = 0.81.$$