ST3131, Sem 2, 2021.

Mid term Solution

Q1 a. Range: 7 → 17.
Comment: quite symmetric, possibly no outlier.

b. Cor = 0.7900471.
Comment: quite strong, positive relationship.
possibly linear.

c. Fitted Model:

$$\hat{y} = 7.0943 + 1.1155 * flavor + 1.5335 * I(region = 2) + 1.2234 * I(region = 3)$$

d. $R^2 = 0.8242$.

$$SS_T = 96.615 + 30.960 + 27.213 = 154.788.$$

Q2. a. For intercept: t_value = 1383.4714 / 1255.2404 = 1.10216

For x: $se(\hat{\beta}_1) = \frac{10.6222}{65.378} = 0.1625$

p-value = p-value of t-test for $\beta_1 \Rightarrow 1.71e{-}14$.

F-statistic = $(t\text{-value for } x)^2$ bcz this is a simple model.

$= 65.378^2 = 4274.3$

$MS_{Res} = 2313^2 \Rightarrow SS_{Res} = 2313^2 * (n-2)$ where $n - 2 = 10$

$\Rightarrow SS_{Res} = 2313^2 * 10$.

$$\text{F-statistic} = \frac{MS_R}{MS_{Res}} \Rightarrow MS_R = SS_R = 4274.3 * 2313^2 =$$

$$\Rightarrow R^2 = \frac{SS_R}{SS_T} = \frac{SS_R}{SS_{Res} + SS_R} = 0.99765.$$

$$R_a^2 = 1 - \frac{(1-R^2)(n-1)}{n-p} = 1 - \frac{(1-R^2) * 11}{10}$$

$$= 0.9974$$

b. σ is estimated by $\sqrt{MS_{Res}} = 2313$.

c. Fitted model: $\hat{y} = 1383.4714 + 10.6222 * x$.

d. H_0: Model 1 is not significant or $H_0: \beta_1 = 0$

test statistic: $t = 65.378 \sim t_{10}$

p-value < 0.0001. Model 1 is highly significant.

e. when $x = 10000$ then $\hat{y} = 1383.4714 + 10.6222 * 10000$

$= 107605.4714$.

f. 95% CI for β_1 is: $\hat{\beta_1} \pm t_{10}(0.025) * se(\hat{\beta_1})$

$10.6222 \pm 2.228 * 0.1625 = (10.26, 10.984)$

when intercept = 1383.4714 then 95% CI of mean y is:

$1383.4714 + \text{CI of } \beta_1 * 10000 = (103983.4714 ; 11223.4714)$

Q_3. a. Fitted Model:

$$\hat{y} = 7.12079 - 0.01262 * \text{Aroma} + 1.12206 * \text{Flavor} - 1.53452 * I(\text{reg} = B) + 1.23557 * I(\text{reg} = C).$$

b. $H_0: \beta_{\text{Aroma}} = 0$ vs $H_1: \beta_{\text{Aroma}} \neq 0$

test statistic $t = \frac{-0.01262}{0.24029} = -0.053 \sim t_{n-p}$,

where $n - p = 33$ & $p = 5 \Rightarrow n = 38$.

Hence $t = -0.053 \sim t_{33}$ has p-value = 0.958.

$\Rightarrow$ Aroma is not significant.

c. $H_0: \beta_{\text{reg}B} = \beta_{\text{reg}C} = 0$ vs H_1: at least one of them $\neq 0$

test statistic: $F = \frac{25.641/2}{27.211/33} = 15.548 \sim F_{2,33}$

p-value = 1.7489 e-05 < 0.001.

Data provide very strong evidence that region is sig

d. Model 3: $y \sim$ Aroma.

H_0: Model 3 is not sig. or $H_0: \beta_{Aroma} = 0$.

test statistic $= \dfrac{SS_R(\beta_{Aroma} \mid \beta_0)}{SS_{Res}/(38-2)} = \dfrac{77.442}{(24.494 + 25.641 + 27.211)/36}$

$= \dfrac{77.442}{2.1485} = 36.044 \sim F_{1,36}$

p-value $= 6.871e{-7} \rightarrow$ very small.

$\Rightarrow$ Model is significant.

e. $R^2 = \dfrac{SS_R}{SS_T} = \dfrac{77.442 + 24.494 + 25.641}{27.442 + 24.494 + 25.641 + 27.211}$

$= 0.8242$.