ST305 HW10 Solution

(Total points: 33)

April 25, 2019

11.9 (3 points, one for each)

(a) The null hypothesis H_0 must refer to the parameter β_2 instead of the estimator b_2.

(b) The *square* of multiple correlation, R^2, gives the proportion of the variation in the response variable that is explained by the explanatory variables.

(c) A small P-value implies that *at least one* of the explanatory variables is statistically different from zero.

11.10 (4 points, one for each)

(a) Only the deviations ϵ_i are assumed to be Normal.

(b) The explanatory power of the model increases as F increases.

(c) If an explanatory variable has a statistically significant correlation with the response variable, the variable may not be a statistically significant predictor in multiple regression.

(d) The multiple correlation coefficient is the correlation between observations y_i and predicted values $\hat{y}_i$ and does not give the average correlation between the response variable and each explanatory variable.

11.14 (1 point) It occurs when there are highly correlated explanatory variables.
12.5 (4 points, one for each)

(a) It is between-group variations instead of within-group variation.

(b) SST=SSG+SSE.

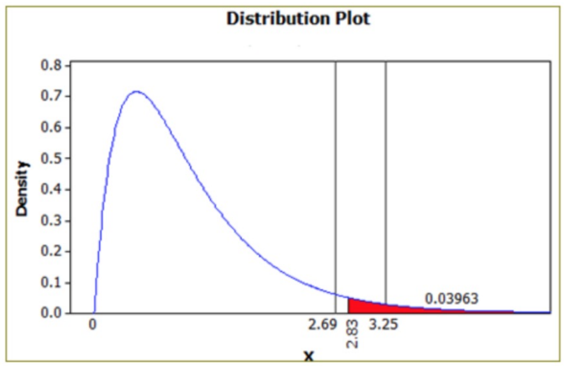
(c) σ is the parameter of the ANOVA model.

(d) The population means are not all the same does not mean that the distributions of values are far apart.

12.9 (4 points, one for each)

(a) For groups, $df = k - 1 = 5 - 1 = 4$. For observations, $df = N - k = 35 - 5 = 30$. $2.69 < F < 3.25$.

(b) The plot is



(c) From Table E, $0.025 < P < 0.05$.

(d) We can conclude that at least one mean is different.

12.13 (3 points, one for each)

(a) Response: egg cholesterol level. Populations: chickens with different diets or drugs. $I = 3$, $n_1 = n_2 = n_3 = 25$, $N = 75$.

(b) Response: rating on five-point scale. Populations: the three groups of students. $I = 3$, $n_1 = 31$, $n_2 = 18$, $n_3 = 45$, $N = 94$.

(c) Response: quiz score. Populations: students in each TA group. $I = 3$, $n_1 = n_2 = n_3 = 14$, $N = 42$.

12.14 (3 points, one for each)

(a) Response: time to complete a designed VR path. Populations: children randomly assigned to use joystick, wand, dancemat or gesture-based pinch gloves. $I = 4$, $n_1 = n_2 = n_3 = n_4 = 10$, $N = 40$.

(b) Response: length of bone from each chick. Populations: birds with different diets. $I = 5$, $n_1 = n_2 = n_3 = n_4 = n_5 = 13$, $N = 65$.

(c) Response: number of sandwich orders. Populations: customers which are offered by a free drink, free chips, a free cookies or nothing. $I = 4$, $n_1 = n_2 = n_3 = n_4 = 5$, $N = 20$.

12.23 (3 points, one for each)

(a) Based on the sample means, fiber is cheapest and cable is most expensive.

(b) Yes; the ratio $= 40.39/26.09 = 1.55 < 2$.

(c) For groups, $df = k - 1 = 3 - 1 = 2$. For observations, $df = N - k = 47 - 3 = 44$. $0.025 < P < 0.050$, or $P = 0.0427$.

12.24 (1 point) By 12.23(b), it is reasonable to use the pooled standard deviation as

$$s_p^2 = \frac{\sum_{i=1}^3 (n_i - 1)s_i^2}{\sum_{i=1}^3 (n_i - 1)} = 1143.59.$$

Let μ_1, μ_2 and μ_3 are the population means of three categories. The contrast is,

$$\psi = \mu_3 - \left(\frac{\mu_1 + \mu_2}{2}\right).$$

The estimate of ψ is

$$c = \bar{x}_3 - \left(\frac{\bar{x}_1 + \bar{x}_2}{2}\right) = -28.365.$$

The standard error is

$$SE_c = s_p \sqrt{\sum_{i=1}^{3} \frac{a_i^2}{n_i}} = 443.88.$$

The hypotheses are $H_0 : \psi = 0$ vs $H_a : \psi > 0$. The test statistic is

$$t = \frac{c}{SE_c} = -0.0639.$$

The corresponding P-value is $0.4767 > 0.05$. Thus, we fail to reject the null hypothesis when $\alpha = 0.05$.

12.33 (2 points, one for each)

(a) $\psi_1 = \mu_2 - (\mu_1 + \mu_4)/2$.

(b) $\psi_2 = (\mu_1 + \mu_2 + \mu_4)/3 - \mu_3$.

12.34 (5 points, one for each)

(a) $H_0 : \psi_i = 0$ vs $H_a : \psi_i \neq 0$ for $i = 1, 2$.

(b) $c_1 = \bar{x}_2 - (\bar{x}_1 + \bar{x}_4)/2 = 0.197$. $c_2 = (\bar{x}_1 + \bar{x}_2 + \bar{x}_4)/3 - \bar{x}_3 = 0.486$.

(c) $SE_{c_1} = s_p \sqrt{\sum_{i=1}^4 \frac{a_i^2}{n_i}} = 0.3098$. $SE_{c_2} = s_p \sqrt{\sum_{i=1}^4 \frac{a_i^2}{n_i}} = 0.2933$

(d) For ψ_1: $t_1 = \frac{c_1}{SE_{c_1}} = 0.64$, $df = N - k = 222 - 4 = 218$, two-sided P-value $= 0.5228$.
For ψ_2: $t_2 = \frac{c_2}{SE_{c_2}} = 1.66$, $df = N - k = 222 - 4 = 218$, two-sided P-value $= 0.0983$.

(e) $df = 222 - 1 = 221$. The 95% confidence intervals for ψ_1 is $(c_1 - tSE_{c_1}, c_1 - tSE_{c_1}) = (-0.41, 0.81)$. The 95% confidence intervals for ψ_2 is $(c_2 - tSE_{c_2}, c_2 - tSE_{c_2}) = (-0.09, 1.06)$.