MATH1001 Worksheet III-5

11.2.2

The accompanying table shows fictitious data for three samples.

	Sample 1	Sample 2	Sample 3
	23	18	20
	29	12	16
	25	15	17
	23		23
			19
Mean	25.00	15.00	19.00
SD	2.83	3.00	2.74

(a) Compute SS(between) and SS(within).

(b) Compute SS(total), and verify the relationship

between SS(between), SS(within), and SS(total).

(c) Compute MS(between), MS(within), and s_{pooled}.

We have n.. = 12, $\sum_{i=1}^{I}\sum_{j=1}^{n_i} y_{ij}$ = 240, and $\bar{\bar{y}}$ = 240/12 = 20.

(a) SS(between) = $(4)(25 - 20)^2 + (3)(15 - 20)^2 + (5)(19 - 20)^2 = 180$;

SS(within) = $(23 - 25)^2 + (29 - 25)^2 + \ldots + (19 - 19)^2 = 72$.

(b) SS(total) = $(23 - 20)^2 + (29 - 20)^2 + \ldots + (19 - 20)^2 = 252$;

SS(between) + SS(within) = 180 + 72 = 252 = SS(total).

(c) df(between) = 2; MS(between) = 180/2 = 90;

df(within) = 9; MS(within) = 72/9 = 8;

$s_{\text{pooled}} = \sqrt{8} = 2.83$.

11.4.2

It is thought that stress may increase susceptibility to illness through suppression of the immune system. In an experiment to investigate this theory, 48 rats were randomly allocated to four treatment groups: no stress, mild stress, moderate stress, and high stress. The stress conditions involved various amounts of restraint and electric shock. The concentration of lymphocytes (cells/ml * 10 -6) in the peripheral blood was measured for each rat, with the results given in the accompanying table. Calculations based on the raw data yielded SS(between) = 89.036 and SS(within) = 340.24.

	No stress	Mild stress	Moderate stress	High stress
$\bar{y}$	6.64	4.84	3.98	2.92
s	2.77	2.42	3.91	1.45
n	12	12	12	12

(a) Construct the ANOVA table and test the global null hypothesis at $\alpha = 0.05$.

(b) Calculate the pooled standard deviation, s_{pooled}.

(a) The hypotheses are

H_0: The stress conditions all produce the same mean lymphocyte concentration
($\mu_1 = \mu_2 = \mu_3 = \mu_4$)

H_A: Some of the stress conditions produce different mean lymphocyte concentrations
(the μ's are not all equal)

The number of groups is I = 4 and the total number of observations is $n_\bullet = 48$. Thus, df(between) = I - 1 = 3 and df(within) = $n_\bullet$ - I = 44.

Source	df	SS	MS
Between groups	3	89.036	29.68
Within groups	44	340.24	7.733
Total	47	429.28	

The test statistic is $F_s = \frac{\text{MS(between)}}{\text{MS(within)}} = 29.68/7.733 = 3.84$. With df = 3 and 40 (the closest value to 44), Table 10 gives $F_{0.02} = 3.67$ and $F_{0.01} = 4.31$. Thus, we have $0.01 < P < 0.02$. Since $P < \alpha$, we reject H_0. There is sufficient evidence ($0.01 < P < 0.02$) to conclude that some of the stress conditions produce different mean lymphocyte concentrations.

(b) $s_{pooled} = \sqrt{\text{MS(within)}} = \sqrt{7.733} = 2.78$ cells/ml x 10^{-6}.

12.2.3

In a study of natural variation in blood chemistry, blood specimens were obtained from 284 healthy people. The concentrations of urea and of uric acid were measured for each specimen, and the correlation between these two concentrations was found to be r = 0.2291. Test the hypothesis that the population correlation coefficient is zero against the alternative that it is positive. Let $\alpha = 0.05$.

The hypotheses are

H_0: There is no correlation between blood urea and uric acid concentration in the population of all healthy persons ($\rho = 0$)

H_A: Blood urea and uric acid concentration are positively correlated ($\rho > 0$)

The test statistic is

$$t_s = r\sqrt{\frac{n-2}{1-r^2}} = 0.2291\sqrt{\frac{282}{1-0.2291^2}} = 3.952.$$

The degrees of freedom are 284 - 2 = 282, so we consult Table 4 using df = 140. We find that $t(_{40,0.0005} = 3.361$. Thus, $P < 0.0005$, so we reject H_0. There is sufficient evidence ($P < 0.0005$) to conclude that blood urea and uric acid concentration are positively correlated.

13.2.5-6

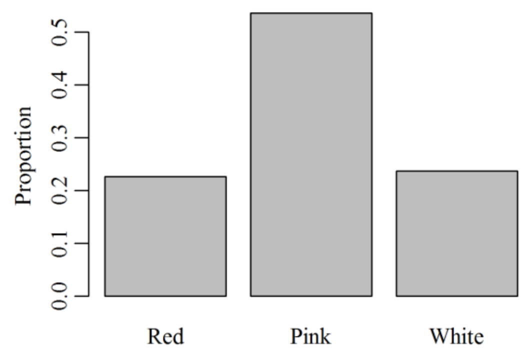
A geneticist self-pollinated pink-flowered snapdragon plants and produced 97 progeny with the following colors: 22 red plants, 52 pink plants, and 23 white plants. The purpose of this experiment was to investigate a genetic model that states that the probabilities of red, pink, and white are 0.25, 0.50, and 0.25.

(a) Identify the type of statistical method that is appropriate for these data.

(b) Conduct an appropriate complete analysis of the data that also includes a graphical display and discussion of how the data do or do not meet the necessary conditions for validity.

(a) The only variable in this study is color, which is a categorical variable. A chi-square goodness-of-fit test could be used to assess whether the genetic model is consistent with the data.

(b) The bargraph below displays the sample proportions for each color type.



The hypotheses are

H_0: The model is correct (Pr{red} = 0.25, Pr{pink} = 0.5, Pr{white} = 0.25)
H_A: The model is incorrect

The observed and expected frequencies (in parentheses) are:

Red	Pink	White	Total
22 (24.25)	52 (48.5)	23 (24.25)	97

$\chi^2_s = 0.526$; df = 2. From Table 9, we find $\chi^2_{2,0.20} = 3.22$. Thus, P > 0.20 and we do not reject H_0. There is little or no evidence (P > 0.20) that the model is not correct; the data are consistent with the model.

Checking the conditions for validity we note that this study considers a single categorical variable for a random samples of plants and that the sample size is sufficiently large (all expected values are greater than or equal to five).