MATH 244 (L1)　　　Applied Statistics

Quiz 4

Name ____________________　　Student ID _______________　　Tutorial section ____

Time allowed : 45 minutes

1. (*12 marks*) In an effort to improve the quality of recording tapes, the effects of four kinds of coatings A, B, C, D on the reproducing quality of sound are compared. The measurements of sound distortion obtained from tapes treated with the four coatings were summarized in the following table.

Coating	Sample size	Mean	Standard Deviation
A	6	11	1.9
B	6	19	1.6
C	6	16	1.2
D	6	14	1.3

(a) Construct the ANOVA table and test whether there is treatment effect. Use $\alpha = 0.05$.

$$\bar{\bar{Y}} = \frac{6\times 11 + 6\times 19 + 6\times 16 + 6\times 14}{24} = 15$$

$$SS_A = \sum_{i=1}^{4} n_i \left(\bar{Y}_i - \bar{\bar{Y}}\right)^2 = 6\left[(11-15)^2 + \cdots + (14-15)^2\right] = 204$$

$$SS_E = \sum_{i=1}^{4} (n_i - 1) S_i^2 = 5\left(1.9^2 + 1.6^2 + 1.2^2 + 1.3^2\right) = 46.5$$

ANOVA table :

Source	SS	d.f.	MS	F-ratio
Treatment	204	3	68	29.25
Error	46.5	20	2.325	
Total	250.5	23		

Test $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$ vs H_1 : not H_0.

$F(3,20,0.05) = 3.10 < 29.25$

Reject H_0 at $\alpha = 0.05$.

(b) From the result in (a), can we conclude that all the effects of the four kinds of coatings are different to each other? Why?

Rejecting $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$ only means not all the treatment effects are the same. However, we cannot conclude that all the treatment effects are different from each other because it may be possible that two of them are different but two of them are the same.

(c) Find the 90% confidence interval for each of the pairwise differences between the four treatment effects.

90% C.I. for $\alpha_1 - \alpha_2$:

$$(\overline{Y}_1 - \overline{Y}_2) \pm t_{20,0.05}\sqrt{MS_E\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

$$= (11-19) \pm (1.725)\sqrt{(2.325)\left(\frac{1}{6}+\frac{1}{6}\right)}$$

$$= -8 \pm 1.5186 = [-9.5186\,, -6.4814] \qquad (\alpha_1 < \alpha_2)$$

90% C.I. for $\alpha_1 - \alpha_3$: $(11-16) \pm 1.5186$

$$= -5 \pm 1.5186 = [-6.5186\,, -3.4814] \qquad (\alpha_1 < \alpha_3)$$

90% C.I. for $\alpha_1 - \alpha_4$: $(11-14) \pm 1.5186$

$$= -3 \pm 1.5186 = [-4.5186\,, -1.4814] \qquad (\alpha_1 < \alpha_4)$$

90% C.I. for $\alpha_2 - \alpha_3$: $(19-16) \pm 1.5186$

$$= 3 \pm 1.5186 = [1.4814\,, 4.5186] \qquad (\alpha_2 > \alpha_3)$$

90% C.I. for $\alpha_2 - \alpha_4$: $(19-14) \pm 1.5186$

$$= 5 \pm 1.5186 = [3.4814\,, 6.5186] \qquad (\alpha_2 > \alpha_4)$$

90% C.I. for $\alpha_3 - \alpha_4$: $(16-14) \pm 1.5186$

$$= 2 \pm 1.5186 = [0.4814\,, 3.5186] \qquad (\alpha_3 > \alpha_4)$$

(d) What conclusion can be drawn from the results in part (c)? Do you have 90% confidence that this conclusion is correct? Explain briefly.

We can conclude that $\alpha_2 > \alpha_3 > \alpha_4 > \alpha_1$, i.e. the ranking of the qualities of tapes with the four coatings is A, D, C, B, with coating A the best and coating B the worst.

This conclusion was drawn based on the six comparisons. For each of these comparisons we have only 90% confidence that it is correct. Hence the overall confidence that all these comparisons are correct is less than 90%. Therefore the overall confidence of this conclusion is less than 90%.

2. (*13 marks*) The Ego Cosmetics Company wants to determine if a relationship exists between advertising expenditures and sales. The following data were compiled for six major sales regions, where X represents advertising expenditures in thousands of dollars and Y represents sales in millions of dollars.

Region	**1**	**2**	**3**	**4**	**5**	**6**
X	450	550	800	350	250	600
Y	35	40	55	33	32	45

(a) Fit a regression line of *Y* on *X*. Please show all your steps.

$$\overline{X} = 500, \quad \overline{Y} = 40, \quad \sum_{i=1}^{6} X_i^2 = 1690000, \quad \sum_{i=1}^{6} Y_i^2 = 9988, \quad \sum_{i=1}^{n} X_i Y_i = 128300$$

$$S_{xx} = \sum_{i=1}^{6} X_i^2 - n\overline{X}^2 = 1690000 - (6)(500)^2 = 190000$$

$$S_{yy} = \sum_{i=1}^{6} Y_i^2 - n\overline{Y}^2 = 9988 - (6)(40)^2 = 388$$

$$S_{xy} = \sum_{i=1}^{6} X_i Y_i - n\overline{X}\overline{Y} = 128300 - (6)(500)(40) = 8300$$

$$b = \frac{S_{xy}}{S_{xx}} = \frac{8300}{190000} = 0.04368 \quad , \quad a = \overline{Y} - b\overline{X} = 40 - (0.04368)(500) = 18.1579$$

Fitted regression line : $\widehat{Y} = 18.1579 + 0.04368X$

(b) Construct the ANOVA table.

$$SS_R = \frac{S_{xy}^2}{S_{xx}} = \frac{8300^2}{190000} = 362.5789$$

$$SS_E = S_{yy} - SS_R = 400 - 362.5789 = 37.4211$$

ANOVA table :

Source	**SS**	**d.f.**	**MS**	**F-ratio**
Regression	362.5789	1	362.5789	38.76
Error	37.4211	4	9.3553	
Total	400	5		

(c) Find the coefficient of determination.

$$R^2 = \frac{SS_R}{SS_T} = \frac{362.5789}{400} = 90.64\%$$

(d) Find the 90% prediction interval for the sales in a region at which the advertising expenditures is $420000.

$X_0 = 420$, $\hat{Y}_0 = 18.1579 + (0.04368)(420) = 36.5035$

90% P.I. for Y_0 : $\hat{Y}_0 \pm t_{4,0.05}\sqrt{MS_E\left(1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{S_{xx}}\right)}$

$$= 36.5035 \pm (2.132)\sqrt{(9.3553)\left(1 + \frac{1}{6} + \frac{(420 - 500)^2}{190000}\right)}$$

$$= 36.5035 \pm 7.1445 = [29.3590 , 43.6480]$$

(e) Using the regression line obtained in (a), find the prediction of the sales in a region at which the advertising expenditures is five millions. Is this prediction reasonable? Explain briefly.

$X_0 = 5000$, $\hat{Y}_0 = 18.1579 + (0.04368)(5000) = 236.5579$

The predicted sales will be 236.5 millions. This prediction is not reasonable because $X_0 = 5000$ is far beyond the range of the observed values of *X*. The fitted regression line would not be adequate for modelling the relationship between *X* and *Y* outside this range.

< E N D >