Review :

1. Error Rates for Multiple Comparisons
   (a) Comparison-wise error rate : $P($ Reject $H_o | H_o$ is true $)$ for a single contrast.
   (b) Experiment-wise error rate : Overall error rate for all contrasts.

2. General Contrasts
   (a) Protected test : Only if we reject the F-test, we proceed with the evaluation of contrast.
   (b) Bonferroni test : Each contrast is tested using the comparison-wise error rate of α/r, r = number of contrasts

3. Comparing All Means ($H_o : \mu_i = \mu_j$) - The Balanced Case
   (a) The LSD Approach : Reject H_o if $|\bar{x_{i.}} - \bar{x_{j.}}| > LSD$

$$LSD = T_{N-k,\alpha/2} \times \sqrt{s_p^2(\frac{2}{n})}$$

   (b) The Q Method (Tukey's) : Reject H_o if $|\bar{x_{i.}} - \bar{x_{j.}}| > QD$

$$QD = Q_{k,N-k,\alpha} \times \sqrt{s_p^2(\frac{1}{n})}$$

4. Regression
   (a) Model : Let Y_i be the Y-values corresponding to $X_i, i = 1, 2, \ldots, n$. Then

$$Y_i = b_0 + b_1 X_i + e_i$$

   The e_i are assumed to be independent and normally distributed, with a mean of 0 and a variance of σ_e^2.

   (b) Estimates for the Best Fitting Line by the Method of Least Squares

$$\begin{aligned}\hat{b_1} &= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n x_i y_i - \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n}}{\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n}} \\ \hat{b_0} &= \bar{y} - \hat{b_1}\bar{x}\end{aligned}$$

   (c) ANOVA Table for Simple Linear Regression

$$\begin{aligned}\text{SSTotal} &= \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2/n \\ \text{SSRegression} &= \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = \hat{b_1}[\sum_{i=1}^n x_i y_i - \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n}] \\ \text{SSError} &= \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \text{SSTotal} - \text{SSRegression}\end{aligned}$$

Source	df	SS	MS
Regression	1	SSRegression	SSRegression/1
Error	n-2	SSError	SSError/(n-2)
Total	n-1	SSTotal	

Practice Problem

1. An article reported the following data on survival times for rats exposed to nitrogen dioxide (70 ppm) via different injection regimens. There were 10 rats in each group.

regimen	A	B	C	D	E	F	G
mean	31.9	31.3	29.2	33.8	34.3	29.1	33.7
stdev	2.13	1.89	2.04	2.04	1.77	1.91	2.58

   (a) The total sum of squares is 550.30. Fill in the ANOVA table and test the hypothesis of no difference beween population means of 6 treatment.

   (b) State the assumptions you are making (again defining all symbols used.) Where possible, verify the assumptions. (You don't need to verify formally the homogeneity of variance assumption.)

   (c) Use the LSD and Tukey's Q method (at $\alpha = 0.05$) to compare all pairs of means.

   (d) Assuming that A is a control, use the Bonferroni procedure to test for equality of the mean of A with the mean of each of the other regimens. Use an experimentwise error rate of 0.05.

2. To investigate the influence of nitrogen (N) on early growth of red oak seedlings, the following data were collected : X=unit of N supplement, and Y=shoot elongation per seedling(cm)

X	0	0	0	0	1	1	1	1	2	2	2	2	3	3	3	3
Y	5.8	4.5	5.9	6.2	6.0	7.5	6.1	8.6	8.9	8.2	14.2	11.9	11.1	11.5	14.5	14.8

   The summary statistics are: $\sum_{i=1}^{16} y_i = 145.7$, $\sum_{i=1}^{16} y_i^2 = 1505.01$, $n = 16$.

   (a) Find the least squares estimates of slope and intercept.

   (b) Find 95% confidence intervals for each estimate above.

   (c) Estimate with 95% confidence limits the variance of Y about the theoretical straight line.

Practice Problem Solutions

1a.
```
ANALYSIS OF VARIANCE
SOURCE     DF        SS        MS
TREATMENT   6      281.5      49.92
ERROR      63      268.8      4.27
TOTAL      69      550.3

F = 49.92/4.27 = 11.69, while F(6,63,.05) = 2.26 from the table
and p-value < .001.  Null hypothesis for equality of treatment means is rejected.
```

1b. Assumptions: independent samples from seven normal populations with equal variances.
To verify the normality, we can make stem-and-leaf plots or normal score plots for **each** group(separately).
We can analyze assumption of equal variances with Levene's test.

1c. LSD method: $LSD = t_{N-k,\alpha/2} s_p \sqrt{\frac{2}{n}} = 1.9983(2.066)\sqrt{\frac{2}{10}} = 1.846$

```
Reg F  Reg C  Reg B  Reg A  Reg G  Reg D  Reg E
29.1   29.2   31.3   31.9   33.7   33.8   34.3
-----------
              -----------
                     ------------
                            -------------------
```

Q method: $QD = Q_{k,N-k,\alpha} s_p \sqrt{\frac{1}{n}} = 4.31(2.066)\sqrt{\frac{1}{10}} = 2.82$

```
Reg F  Reg C  Reg B  Reg A  Reg G  Reg D  Reg E
29.1   29.2   31.3   31.9   33.7   33.8   34.3
-------------------------
              -------------------------
                     -------------------------
```

1d. Test $H_{01}: \mu_A = \mu_B$, $H_{02}: \mu_A = \mu_C$, $H_{03}: \mu_A = \mu_D$, $H_{04}: \mu_A = \mu_E$, $H_{05}: \mu_A = \mu_F$, $H_{06}: \mu_A = \mu_G$.
Using Bonferroni method with $\alpha = .05$, $r = 6$, we can compute
$t_{N-k,\frac{\alpha}{2r}} s_p \sqrt{\frac{2}{n}} = 2.7241(2.066)\sqrt{\frac{2}{10}} = 2.517$.

```
Reg F  Reg C  Reg B  Reg A  Reg G  Reg D  Reg E
 29.1   29.2   31.3   31.9   33.7   33.8   34.3
               --------||----------------------
```

Only regimens F and C are significantly different from race A.

2a. The least square estimates are:

$$\hat{b_1} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n x_i y_i - \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n}}{\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n}}$$

$$
\begin{aligned}
&= 2.60 \\
\hat{b_0} &= \bar{y} - \hat{b_1}\bar{x} \\
&= 5.21
\end{aligned}
$$

2b. 95% Confidence Interval for b_1

$$\hat{b_1} - T_{n-2,\alpha/2}\frac{\sqrt{MSE}}{\sqrt{\sum(x_i-\bar{x})^2}} \leq b_1 \leq \hat{b_1} + T_{n-2,\alpha/2}\frac{\sqrt{MSE}}{\sqrt{\sum(x_i-\bar{x})^2}}$$

$$2.60 - 2.1451.784.47 \leq b_1 \leq 2.60 + 2.1451.784.47$$

$$1.75 \leq b_1 \leq 3.45$$

95 % Confidence Interval for b_0

$$\hat{b_0} - T_{n-2,\alpha/2}\sqrt{MSE}\sqrt{\frac{1}{n}+\frac{\bar{x}^2}{\sum(x_i-\bar{x})^2}} \leq b_0 \leq \hat{b_0} + T_{n-2,\alpha/2}\sqrt{MSE}\sqrt{\frac{1}{n}+\frac{\bar{x}^2}{\sum(x_i-\bar{x})^2}}$$

$$5.21 - 2.145\sqrt{3.166}\sqrt{116+1.5^24.47} \leq b_0 \leq 5.21 + 2.145\sqrt{3.166}\sqrt{116+1.5^24.47}$$

$$3.63 \leq b_0 \leq 6.82$$

2c. 95 % Confidence Interval for the variance of Y about the theoretical straight line

$$\frac{SSE}{\chi^2_{n-2,\alpha/2}} \leq \sigma_e^2 \leq \frac{SSE}{\chi^2_{n-2,1-\alpha/2}}$$

$$44.32626.12 \leq \sigma_e^2 \leq 44.3265.63$$

$$1.697 \leq \sigma_e^2 \leq 7.873.$$