Stat 333
Lab #6

One-Way ANOVA

A firm developing a new citrus-flavoured soft drink conducted an experiment to study customer preferences for the colour of the drink. Four colours were considered: colourless, pink, orange, and lime green. Twenty test localities, which were similar in sales potential and representative of the target market for this product, were selected. Each colour was randomly assigned to five localities for test marketing. The number of cases sold per 1000 population during the test period are recorded below:

Colourless	Pink	Orange	Lime Green
26.5	31.2	27.9	30.8
28.7	28.3	25.1	29.6
25.1	30.8	28.5	32.4
29.1	27.9	24.2	31.7
27.2	29.6	26.5	32.8

Are the sales the same for all colours of the drink? Test at the 5% significance level.

Minitab

Method 1

1. Enter the data into the MINITAB worksheet one sample per column- *ie.* The 5 values from the 'colourless' sample are entered in c1, the 5 values of the 'pink' sample in c2, and so forth.
2. From the MENU BAR, select STAT>ANOVA>ONEWAY(UNSTACKED)
   When the dialog box appears, list all the columns in which you entered data.
3. Note the table and other details are printed in the SESSION window.

Method 2

1. Enter all the sample values into one column of the worksheet. In a corresponding position in a second column, enter the number of the sample (1,2,3, etc.) from which each value came. For instance, if you entered all the 'colourless' observations, then pink, then orange, then lime green in column c1, c1 would contain 26.5, 28.7, 25.1, 29.1, 27.2, 31.2, 28.3, ... 32.8. You would enter into c2: 1, 1, 1, 1, 1, 2, 2,..., 4. NOTE: These entries in c2 are most efficiently accomplished with CALC>MAKE PATTERENED DATA>SIMPLE SET OF NUMBERS. In the dialog box, store patterned data in c2. FROM 1st VALUE box type **1**, TO LAST VALUE box type **4,** LIST EACH VALUE box type **5**, LIST THE WHOLE SEQUENCE box type **1.**
2. From the MENU BAR, select STAT>ANOVA>ONEWAY
   As the RESPONSE, specify the column in which you entered the observations, as the FACTOR specify the column in which you placed the sample number.
3. The output appears in the SESSION window.

1. A large consulting firm hires a West Coast university to provide an M.B.A program for its employees. The basic statistics course is taught at four locations of the firm. After completion of the course, standardized tests are given to the participating employees at each location. The results are:

	Location			
Observation	A	B	C	D
1	96	65	66	60
2	88	74	90	72
3	92	77	88	66
4	75	82	73	75
5	81	70	85	78
6	62	78	87	68
7	86	84	94	
8	86		74	
9	90			
10	85			

Can it be concluded that there is no difference in test results between location? Use $\alpha = 0.05$
{Fcalc = 4.130, p-value = .016, Rho}

2. The sales research division of a large corporation is conducting research on sales methods for selling one of the products of the corporation. The division has designed a completely randomized one-factor analysis of variance model to investigate the efficiency of three sales methods. The responses are measured in units of $100 sales, and are listed below.

	Sales Method		
Response	A	B	C
1	20	13	31
2	23	18	28
3	22	16	48
4	22	29	28
5	36	14	30
6	45		25
7	21		
8	26		

Are there differences between the three different sales methods? Use $\alpha = 0.10$. What assumptions did you make? {Fcalc= 3.899, p-value = .042, Rho, Sales are normally distributed and share a common variance}

Two-Way ANOVA

The data in the following table represent the milliequivalents of sodium excreted by six subjects 2 hours after treatment with one of four diuretics assigned at random by a clinician over a 6-day period. Using significance level of 0.05, analyze the data to determine whether or not there are any differences between patients and any differences in the effectiveness of diuretics.

	Treatments (Diuretics)			
Subjects	A	B	C	D
1	3.9	30.6	25.2	4.4
2	5.6	30.1	33.5	7.9
3	5.8	16.9	25.5	4.0
4	4.3	23.2	18.9	4.4
5	5.9	26.7	20.5	4.2
6	4.3	10.9	26.7	4.4

{F_{Bl} = 1.5758,p-value =.2266, Fail to RHo}
(F_{Tr} = 37.1768, p-value ~ 0, Rho}

Method

1. Enter all of the data into one column. You should first enter all the values for treatment A, then treatment B, and so forth.

2. In the second column, enter from which treatment each value came from *(ie.* You would enter A, A, A, A, A, A, B, B,..., D) These entries are most efficiently accomplished with
   CALC>MAKE PATTERNED DATA > TEXT VALUES
   Store patterned data in **C2**
   In the dialog box, type A B C D (Note: leave a space between each letter)
   In the list each value box, type **6**
   In the list the whole sequence box, type **1.**
3. In the third column, we want to enter the subject number (*ie.* 1,2,3,4,5,6,1,2,3,4,5,6,...
   Go to CALC>MAKE PATTERNED DATA>SIMPLE SET OF NUMBERS
   (a) Store patterned data in **C3**
   (b) From 1st value, type **1**
   (a) To last value, type **6**
   (b) In steps of , type **1**
   (c) List each value, type **1**
   (d) List the whole sequence, type **4**

   This will assign the numbers 1 through 6 four times.

4. From MENU BAR select STAT>ANOVA>TWO-WAY
   (a) Response, type **C1**
   (b) Row factor , type **C2** (MINITAB takes row factor to represent treatment effect))
   (c) Column factor, type **C3** (MINITAB takes column factor to represent block effect)
   Note: do not take "row factor" and "column factor" literally
5. Output appears in the SESSION window.

3. A county employs 3 assessors who are responsible for determining the values of residential property in the county. To see whether or not these assessors differ in their appraisal, 5 houses were selected and each assessor determined the market value of each house. The data (assessors are the treatment since your main concern is to see if there is difference in appraisals) was then analyzed using a two-way ANOVA routine giving the following (partially completed) ANOVA table.

   (a) Complete the ANOVA table

Source	SS	DF	MS	F
Treatment	45.9			
Block	141.7			
Error				
Total	250.8			

   (b) Is there any indication of a difference between appraisors? Use a 5% significance level for testing.
   {Fcalc = 2.905,p-value = .1126, fail to RHo}

4. For the following data, present the ANOVA table. What conclusions can you draw from the two F tests? Use an $\alpha = 0.05$

	Block			
Treatment	1	2	3	4
1	14	8	10	4
2	7	7	4	2
3	12	6	16	6

   {F_{Bl} = 3.462, p-value= .0915, fail to RHo}
   { F_{Tr} = 3.2299, p-value = .1117, fail to RHo}

5. Three loaves of bread, each made according to a different recipe, are baked in one oven at the same time. Because of possible uncontrolled variations in oven performance, each baking is treated as a block. This procedure is repeated five times and the following measurements of density are obtained.

	Block				
Recipe	1	2	3	4	5
1	.95	.86	.71	.72	.74
2	.71	.85	.62	.72	.64
3	.69	.88	.51	.73	.44

(a) How should the three oven positions of the three loaves be selected for each trial?
{Randomize the position of the loaves in the oven}

(b) Perform an analysis of variance for these data using a 5% significance level
{F_{Bl} = 5.309, p-value = .0219, RHo}
{ F_{Tr} = 3.92 , p-value = .065, fail to RHo}

Practice more questions from the text.

The F-distribution and Simple Linear Regression
Note: there may be slight differences in the answers due to rounding.

1. Here is a set of data showing the historic yearly rates of return in seven randomly selected years, for Stock Y and the New York Stock Exchange Index (the predictor variable).

Year	1	2	3	4	5	6	7
Stock Y	2.0%	7.9%	-6.0%	-9.5%	13.5%	7.5%	1.2%
NYSE Index	4.9%	13.0%	-2.5%	-10.6%	11.0%	14.5%	4.3%

(a) Write down the linear regression model expressing the yearly rate of return on Stock Y as a linear function of the yearly rate of return of the NYSE Index.
(b) Estimate the intercept and slope term in the model. (Note: the slope term is referred to as Stock Y's "beta", or β. This is a measure that stock analysts uses to evaluate the past performance of a stock. Stocks possessing β's greater than 1 tend to have larger expected rates of return compared to stocks with smaller β's. [βo=-1.7470, $\beta 1$= 0.8332]

Minitab instructions

1. Enter Stock Y data in column C1
2. Enter NYSE Index data in column C2
3. Click on **STAT>Regression>Regression**
4. Enter C1 in the **Response** box
5. Enter C2 in the **Predictor** box
6. Click on **Graphs,** click on **residuals versus fits**, click **OK**
7. Click **OK**

You will now get a graph of the residuals vs fits for the data. There will also be a printout of the regression equation and the ANOVA table on the screen. If you want to see a scatter plot of the data with the fitted line,

Click on **STAT>Regression>Fitted Line Plot**
Enter C1 in **Response (Y)**
Enter C2 in **Predictor (X)**
Highlight **Linear** for **Type of Regression**.
Click **OK**.

(c) Construct a 95% confidence interval estimate for Stock Y's β ($\beta 1$). Interpret the meaning of this interval. [$0.4560 \le \beta 1 \le 1.2105$
(d) Is the rate of return on Stock Y positively related to the rate of return on the NYSE Index? Test at a level of significance of 0.05 using the t-test. [t=5.676, p-value = .0024, RHo]

(e) Construct an analysis of variance table for the above regression. In addition, perform the same test in (d) using a different test. (again, $\alpha = 0.05$). Are the results in (c) and (d) consistent? [Fcalc= 32.23, p-value = .002, RHo]
(f) Find the standard error of the regression and interpret its significance. [Se^2= 10.563]
(g) Find the coefficient of determination and interpret its meaning. [$r^2 = 0.8656$]
(h) Find a 94% confidence interval estimate for the mean rate of return on Stock Y if the rate of return on the NYSE Index is 4.6%. [-0.8916, 5.063]
(i) Find a 99% confidence interval estimate for this year's rate of return on Stock Y if the New York Stock Exchange Index has a rate of return of 8.1%. (or a 99% prediction interval). Interpret this interval. Would you invest in this stock, based on your interval? [-9.1316, 19.1354]
(j) Find the coefficient of correlation between the rate of return on Stock Y and the rate of return on the NYSE [r = +0.930]

2. The Director of Management Information Systems at a conglomerate must prepare his long-range forecasts for the company's 3-year budget. In particular, he must develop staffing ratios to predict the number of managers and project leaders based on the number of programmers. The results of a sample of the electronic data processing staffs of 10 companies within the industry are displayed below.

# of applications Programmers	15	7	20	12	16	20	10	9	18	15
# of Managers and Project leaders	6	2	10	4	7	8	4	6	7	9

(a) Find the regression coefficients. State the least squares linear regression equation. [y^ = -0.0885 + 0.45x]
(b) Interpret the meaning of the slope and intercept.
(c) Compute Se^2 and interpret this value. [Se = 1.42]
(d) Compute the coefficient of determination and interpret its meaning in this problem. [$r^2 = 0.7018$]
(e) At the 0.05 level of significance, is there a linear relationship between the number of managers and the number of application programmers? Use T-test [t= 4.339, p-value = .0024 Rho]
(f) At the 0.05 level of significance, test for the appropriateness of the simple linear regression model. Use F-test [F=18.834, p-value = .0025 Rho]
(g) Set up a 95% confidence interval estimate of the true population slope. [$0.2109 \le \beta 1 \le 0.6891$]
(h) Set up a 95% confidence interval estimate of the true population intercept. [$-3.6374 \le \beta o \le 3.4604$]
(i) Set up a 95% confidence interval estimate of the average number of managers at companies where there are 10 programmers. [2.9711, 5.8559]
(j) Set up a 95% prediction interval estimate of the number of managers for a particular cmpany in which there are 10 programmers. [0.8354, 8.1716]
(k) Construct a residual plot of the above data. What can you conclude from this residual plot? Does the linear model seem appropriate? Explain.

3. High salaries for presidents and high executives of charitable organizations have been in the news from time to time. Consider the information in the table below for the United Way in 10 major cities in Canada.

City	Salary of President	Money Raised (per capita)
Ottawa	$161,396	$17.35
Montreal	$189,808	$15.81
Toronto	$201,490	$16.74
Winnipeg	$171,798	$31.49
Halifax	$108,364	$15.51
St.John's	$126,002	$23.87
Regina	$146,641	$15.89
Saskatoon	$155,192	$9.32

Edmonton	$169,999	$29.84
Vancouver	$143,025	$24.19

(a) Find the least-squares regression equation that expresses the presidents' annual salary as a linear function of the amount of money raised (per capita). Interpret the meaning of the slope term in the context of the question. [$\hat{y} = 152657 + 235.71x$]

(b) This past year, the City of Lethbridge (with a population of approximately 70,000) raised a total of 1.9 million dollars. Estimate the salary of the president of the United Way Lethbridge Chapter. [$x=27.14$, $\hat{y} = \$159{,}024.95$]

(c) Find the ANOVA table. What percentage of the variation in presidents' salary is explained by the fact that some raised more money per capita than others? [$r^2 = 0.0035$, very small]

(d) Is there a significant linear relationship between the president's salary and per capita money raised? [t=0.1677, p-value = .8718, fail to RHo]

(e) Does there appear to be a significant linear relationship between the amount of money raised per capita and the presidents' salary? Conduct this test using the F-test. What is your conclusion? [F=0.0281,p-value = .871, fail to RHo]

4. The following is a MINITAB output for a random sample of 8 employees at Tackey Toy Manufacturing Company. The company wanted to see if there was a relationship between aptitude test results and output (dozens of units produced).

The regression equation is
Output= 1.03 +5.14 test results

Predictor	Coef	St.Dev	T	P
Constant		2.070		
Test result		0.2831		

S=1.695 R-sq= R-sq(adj)=97.9%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression					
Residual Error					
Total		968.00			

(a) Fill in the above tables and find R-sq.

(b) State the least squares regression equation. [$\hat{y} = 1.03 + 5.14x$]

(c) What percentage of the variation in output is explained by the fact that some had higher test results on the aptitude tests than others? [$r^2 = 0.982$, very large]

(d) Does there appear to be a significant linear relationship between the aptitude test results and output? Conduct this test using both the t-test and the F-test at the 5% significance level.. What is your conclusion? [t=18.16, F=329.61, p-value ~ 0,Rho]