STAT 217
Assignment #5

Review of Binomial instructions for the computer.

1. The probability that a person who undergoes a kidney operation will recover is 0.6. Find the probability that of 5 patients who undergo similar operations,
(a) none will recover $P(x=0)$ (b) not more than one will recover $P(X \leq 1)$ (c) at least 3 will recover $P(X \geq 3)$

From the MENU BAR select CALC>PROBABILITY DISTRIBUTIONS>BINOMIAL...... a dialog box appears. There are various ways in which you can find the required answers

ii. If a single probability is needed as in part a) of the example:
Select PROBABILITY from the options listed (clicking on the circle, enters a dot.)
In the box by NUMBER OF TRIALS, type the number **5**.
In PROBABILITY OF SUCCESS, insert **0.6**.
As INPUT CONSTANT, specify the “x” value from the question (**0** for part a)). Click on OK
The probability then comes up on the SEESION screen. [0.0102]

iii. When a sum of probabilities is involved, as in part b), begin in the same way {CALC>PROB DISTR>BINOM}
Select CUMULATIVE PROBABILITY from the options.
Enter NUMBER OF TRIALS and PROBABILITY OF SUCCESS (5 and 0.6) as before.
As INPUT CONSTANT, specify the ”x” from the question (**1** for part b)). Click on OK
The number appearing on the SESSION screen will be the $P(X \leq x)$. {$P(X \leq 1)$ for b)}
[0.08704]

This method could also be used for part c), but the required probabilities would have to be expressed in terms of the cumulative probability and the complement........$P(X \geq 3) = 1 - P(X \leq 2)$ [0.68256]

iv. When several different questions are asked about the same distribution, it would be helpful to have all of the probabilities for each individual “x” calculated at once.
Enter the values 0,1,2,3,4,5 into c1, the first column of the worksheet.
Then proceed as in the previous examples:
From the menu bar, select CALC>PROBABILITY DISTRIBUTIONS>BINOMIAL; select PROBABILITY from options; enter NUMBER OF TRIALS and PROBABILITY OF SUCCESS (5 and 0.6).
In the box associated with INPUT COMUMN indicate C1 (where you have listed the possible values of X) and request OPTIONAL STORAGE in C2. CLICK on OK>
On the worksheet, the probabilities for each value of X appear in column C2.

Repeating the procedure with CUMULATIVE PROBABILITY and choosing OPTIONAL STORAGE in C3 will complete a table of x values, individual probabilities, and cumulative probabilities in the WORKSHEET.

When we use the sign test, p = .5

<u>Non-Parametric Tests</u>

1. In 1995, the median age of Canadian residents was 34 years, as reported by the Census Bureau. A random sample taken this year of 10 Canadian residents yielded the following ages, in years,

40	60	12	55	34	43	47	37	9	34

At the 5% significance level, do the data provide sufficient evidence to conclude that the majority of today’s Canadian residents are older than 34 years of age? Use the sign test and the Wilcoxon {Ho: median = 34, Ha: median >34 S- =2, S+ = 6, n=8, p-value = $P(X \geq 6) = 0.1445$, T =13, $T\alpha = 6$, Fail to Rho}

2. Twenty years ago, the U.S. Bureau of Justice Statistics reported that the median educational attainment of jail inmates was 10 years. Ten current inmates are randomly selected and found to have the following educational attainments, in years,

14	10	5	6	8	10	10	8	9	9

Assume that the educational attainments of current jail inmates have a symmetric, non-normal distribution. At the 5% significance level, do the data provide sufficient evidence to conclude that the majority of this year's educational attainment has decreased from 10 years? Use the sign test and the Wilcoxon.{Ho:median =10yrs, Ha: median <10yrs S+=1, S- = 6 n=7 p-value =P(X≥6)=.0625, T =5.5, Tα = 4, Fail to Rho}

3. Several batches of fruit flies are exposed to each treatment, and the mortality percent is recorded as a measure of toxicity. The following data are obtained:

Treatment 1	40	28	31	38	43	46	29	18
Treatment 2	36	49	56	25	37	30	41	

Determine if the data strongly indicate different toxicity levels among the treatments at $\alpha = 0.05$ Assume non-normal distributions. {Ho: both treatment have the same toxicity level, Ha: the treatments do not have the same toxicity level T1=62, TL=39, TU=73, Fail to RHo}

4. Sample data were collected to compare the ages of CEOs of top growth companies in western Canada and Quebec, shown in table below.

Western Canada	38	54	40	41	34	59	37	35	48	
Quebec	40	31	38	40	34	36	35	33	39	38

Test to see if the age for CEOs in Quebec is less than that for Western Canada at $\alpha = 0.05$.
(Ho: The age of CEOs is the same is Western Canada and Quebec, Ha: the age of CEO in Western Canada is higher than in Quebec, T1=111, TU = 111,RHo)

5. Human beta-endorphin (HBE) is a hormone secreted by the pituitary gland under conditions of stress. An exercise physiologist measured the resting (unstressed) blood concentration of HBE in two groups of mean: Group 1 consisted of 11 men who had been jogging regularly for some time, and group 2 consisted of 15 men who had just entered a physical fitness program.

Joggers	Fitness Program Entrants
39 40 32 60	70 47 54 27 31
19 52 41 32	42 37 41 9 18
13 37 28	33 23 49 41 59

(a) He believes that the HBE should be higher in group 2. Test this claim using $\alpha = 0.05$. Assume that data is not normal.
(Ho: HBE is the same for both groups, Ha: HBE is lower for the joggers , T1=137.5, $\mu = 148.5$, $\sigma = 19.2678$, zcrit = -1.645 zcalc= -.5709, fail to RHo)

(b) Find the p-value. (.2843)

6. Provincial governments have been re-examining their delivery of services, such as health care. The accompanying tables is based on one random sample of hospitals in New Brunswick and another in Nova Scotia, and show the total number of beds in each hospital.

New Brunswick		Nova Scotia	
23	47	15	13
153	47	27	136
397	15	8	26
12	500	85	311
15	56	12	132
398		64	

(a) At the 0.05 significance level, test the claim that the two provinces have the same distribution of hospital-bed numbers. (Ho: NB has the same number of beds as NS, Ha: NB does not have the same number of beds as NS, T1 = 139.5, μ = 126.5, σ= 15.2288, zcalc=.8536 , or T1=113.5, zcal=-.8536 zcrit = -1.96,+1.96, Fail to RHo)

(b) What is the p-value? (.3954)

(c) At what levels of significance would you come to a different conclusion in (a)? {α>.3954

7. Two critics rate the service at seven award winning restaurants on a continuous 0 to 10 scale. Is there a difference between the critics' ratings at a 0.05 significance level? Use the sign test and the Wilcoxon. {Ho: critics' ratings are the same, Ha: the critics' ratings are not the same, S+=1, S- = 6, n=7, p-value =2×P(X≥6) .125, T= 5, Tα = 2, Fail to RHo}

Restaurant	1	2	3	4	5	6	7
Critic 1	6.1	5.2	8.9	7.4	4.3	9.7	5.5
Critic 2	7.3	5.5	9.1	7.0	5.1	9.8	5.7

8. 13 people were given a pill. Their blood pressure was measured before and after they took a pill. At the 5% significance level, determine if blood pressure has decreased after taking the pill. Use the sign and the sign test and the Wilcoxon. {Ho: Blood pressure is the same Ha: Blood pressure is higher before the pill than after, S- = 1, S+= 10, n=11 p-value = P(X≥10)= .0059, T = 4.5, Tα14, Rho}

Before	70	80	72	76	76	76	72	78	82	64	74	92	74
After	68	72	62	70	58	66	68	52	64	72	74	60	74

9. In a data set of 108 subjects, 64 had temperatures below 37.0°C, 42 had temperatures above 37.0 °C and 2 had temperatures equal to 37.0°C. Test the claim that the median temperature is 37.0 °C at the 5% significance level. [Ho: median = 37.0°C Ha: median≠37.0°C, n=106, S+=42, S- = 64, P-value = 2×P(X≥64)=.0409 , Rho, or μ = 53 σ = 5.1478, zcalc= (63.5.5-53)/5.1478= 2.0398 zcrit = ±1.96, Rho, p-value = 2×P(z≥2.0398)=.04137]

10. The median time to perform a task is around 20 minutes. In a data set of 60 subjects, the sum of positive ranks is 1000 and the sum of negative ranks is 830. At the 5% significance level, is the median time 20 minutes? [Ho: median = 20min Ha: median≠20min, n=60, T+=1000, T- = 830, μ = 915 σ = 135.84, zcalc= -.6257 zcrit = ±1.96, Fail to Rho, p-value = 2×P(z<-.6257)=.5315]

11. The Life Trust Insurance Company funded a university study of drinking and driving. After 30 randomly selected drivers were tested for reaction times, they were given two drinks and tested again, with the result that 22 had slower reaction times, 6 had faster reaction times, and 2 received the same scores as before the drinks. At the 0.01 significance level, test the claim that the drinks slowed down the reaction times. Based on these very limited results, does it appear that the insurance company is justified in charging higher rates for those who drink and drive? . [Ho: The drinks had no effect on driving Ha: Drinks slow down reaction time n=28, S+=22, S- = 6, P-value = P(X≥22)=.0019 , Rho, or μ = 14 σ = 2.6458, zcalc= (21.5-14)/2.6458= 2.8346 zcrit = 2.33, Rho, p-value = P(z≥2.8346)=.0022]

12. Use the appropriate nonparametric method to perform the test in question 1 from assignment #4. (Ho:Test results are the same for all 4 locations Ha: test results are not all the same for the 4 locations H=9.68 Hα = 7.815, RHo)

13. Use the appropriate nonparametric method to perform the test in question 2 from assignment #4. (Ho: The 3 sales methods yield the same result Ha: Not all of the sales methods yield the same result H=7.56 Hα = 4.605, RHo)

14. Do as many questions as possible from the text for more practice.

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. [$0.4558 \le \beta 1 \le 1.2106$]

(d) Is the rate of return on Stock Y related to the rate of return on the NYSE Index? Test at a level of significance of 0.05. [Ho: $\beta 1=0$ Ha: $\beta 1 \neq 0$ Tcalc=5.676 >2.571, RHo]

(e) Construct an analysis of variance table for the above regression. In addition, perform the same test in (d) using the F-test. (again, α = 0.05). Are the results in (c) and (d) consistent? [Ho: $\beta 1=0$ Ha: $\beta 1 \neq 0$ Fcalc = 32.23 >6.61, RHo, yes]

(f) Find the standard error of the regression and interpret its significance. [Se= 3.250]

(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. [$\hat{y}$ = -0.0885 + 0.45x]

(b) Interpret the meaning of the slope and intercept.

(c) Compute Se 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 [Ho: $\beta1=0$ Ha: $\beta1\neq0$ T= 4.339>2.306, RHo]

(f) At the 0.05 level of significance, test for the appropriateness of the simple linear regression model. Use F-test [Ho: $\beta1=0$ Ha: $\beta1\neq0$ F=18.834 >5.32, Rho]

(g) Set up a 95% confidence interval estimate of the true population slope. [$0.2109 \leq \beta1 \leq 0.6891$]

(h) Set up a 95% confidence interval estimate of the true population intercept. [$-3.6374 \leq \beta o \leq 3.4604$]

(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? Use the p-value approach and interpret the p-value in the context of the question. [Ho: $\beta1=0$ Ha: $\beta1\neq0$ t=0.1677, p-value = .871, 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 both the t-test and the F-test. What is your conclusion? [Ho: $\beta1=0$ Ha: $\beta1\neq0$ F=0.0281,p-value = .871, Fail to RHo]

(f) This past year the United Way in Calgary raised $34.94 per capita. Construct a 99% confidence interval for the mean salary of the president of the United Way in Calgary. [$107,862.93, $213,922.71]

(g) Construct a 95% prediction interval for the salary of the president of the United Way in Calgary. [$74,227.12, $247,588.52]

(h) Estimate, with 90% level of reliability, the average (mean) salary of a United Way president who raised $29.00 per capita. [$130,189.84, $188,795.56]

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. [.982]

(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, 98.2%]

(d) Is there a significant linear relationship between output and aptitude test results? Use the p-value approach and interpret the p-value in the context of the question. [Ho: $\beta 1=0$ Ha: $\beta 1 \neq 0$ t=18.16, RHo, p-value~0, F=329.61,p-value~0 RHo]