Math 323
Lab #6

Do questions 9.1,9.3, 9.9, 9.11, 9.13, 9.14, 9.72-9.82, from chapter 9, 10.27-10.34, 10.76-10.88 and 11.1-11.5.

1. A sporting goods manufacturer who produces both white and yellow tennis balls claims that more than 75% of all tennis balls sold are yellow. A marketing study of the purchases of white and yellow tennis balls at a number of stores showed that of 470 cans sold, 410 were yellow and 60 were white.
   (a) Is there sufficient evidence to support the manufacturer's claim at a 1% significance level?
   [Ho: p=.75, Ha: p>.75zcalc =6.1213, zcrit = 2.33, yes]
   (b) Calculate the probability of a Type II error if, in fact, 80% of the tennis balls sold are yellow.
   [$\hat{p}_c$ =.7965, P(z< -.1897) = .4247]
2. Consider the following hypothesis test:
   Ho:μ = 2400
   Ha:μ > 2400
   The rejection region has been defined as R: Sample mean > 2446.5
   Complete the table below indicating whether the given (population mean, sample mean) pair would result in acceptance or rejection of the null hypothesis and which type of error, if any, would result.

Population Mean	Sample Mean	Ho True or False	Accept or Reject Ho	Type of Error
2401	2400			
2401	2450			
2400	2450			
2400	2401			

3. Management of a shopping centre believes that on weekends, people spend more than an hour and a half on average at the centre. To see if this is the case, a parking survey will be taken for a random sample of 100 cars on weekends. It is assumed that the standard deviation of parking times is 30 minutes. Assume that testing is carried out at the 2.5% significance level. Determine the power of the test if the true mean parking time (for the population) is 97 minutes. [Ho: μ=90min, Ha: μ>90min
   $\overline{X}_c$ =95.88, P(z>-.3733) =0.6443]
4. The manufacturer of Grin toothpaste claims that children under the age of 10 years that use their toothpaste regularly have, on average, less than 2 cavities. A random sample of 25 children had an average of 1.94 cavities with a sample standard deviation of 0.13. Assume cavity rate is normally distributed. Do the data support the manufacturer's claim?
   (a) Carry out the test with a 5% significance level [[Ho: μ =2, Ha: μ <2, tcalc=-2.3077, tcrit = -1.711, Rho]
   (b) Determine the p-value for the test in (a) [0.0150]
   (c) Using a 5% significance level, find the probability of making a type II error if the true mean for cavities is 1.96. [$\overline{X}_c$ =1.9555, P(t>-.1731) =0.5678]
   (d) Suppose they want to test Ho: μ= 2 against μ= 1.96 with α=β= .025. Assume that σ= .15. Find the sample size that will ensure this accuracy. (217)

5. A random sample of 324 adults shows that 120 smoke. At the 1% significance level
   (a) test the claim that more than 1/3 of all adults smoke. [Ho: p =1/3, Ha: p>1/3, zcalc = 1.4167, zcrit = 2.33, Fail to RHo]
   (b) What's the probability of making a type I error? (0.01)
   (c) Using a 1% significance level, find the power of the test if the true population proportion was 40%.
   [$\hat{p}_c$ = .3943, P(z>-.2094) = 0.5832]

6. Assume that you are using a significance level of α= 0.05 to test the claim that μ < 2 and that your sample consists of 50 randomly selected values.
   (a) Find β given that the population actually has a normal distribution with μ = 1.5 and σ = 1. [P(z>1.8695) =0.0294]
   (b) Find the power of the test and describe what it means. [0.9706]

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. Find the estimated slope and intercept.[$\hat{\beta}_0$ =1.7470, $\hat{\beta}_1$ = 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**.

(b) Construct a 95% confidence interval estimate for the slope. Interpret the meaning of this interval. [$0.4560 \le \beta 1 \le 1.2105$]

(c) 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 slope. [Tcalc=5.676 >2.015, RHo]

(d) Perform the same test in (c) using the correlation coefficient. Are the results in (c) and (d) consistent? [tcalc = 5.676 >2.015, yes]

(e) Find the standard deviation of the error for the regression and interpret its significance. [S= 3.250]

(f) Find the coefficient of determination and interpret its meaning. [$r^2 = 0.8656$]

(g) 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]

(h) 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]

(i) 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 *s* and interpret this value. [s = 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? [t= 4.339>2.306, RHo]
(f) Set up a 95% confidence interval estimate of the true population slope. [$0.2109 \le \beta 1 \le 0.6891$]
(g) Set up a 95% confidence interval estimate of the true population intercept. [$-3.6374 \le \beta o \le 3.4604$]
(h) Set up a 95% confidence interval estimate of the average number of managers at companies where there are 10 programmers. [2.9711, 5.8559]
(i) Set up a 95% prediction interval estimate of the number of managers for a particular company in which there are 10 programmers. [0.8354, 8.1716]
(j) Construct a residual (error) 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) 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. [t=0.1677, p-value = .871, Fail to RHo]
(e) 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]
(f) Construct a 95% prediction interval for the salary of the president of the United Way in Calgary. [$74,227.12, $247,588.52]
(g) 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.

(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) 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. [t=18.16, RHo, p-value~0]