

Statistics 213 Assignment 5

Note: answers may vary slightly due to rounding

1. A firm establishes a committee to investigate the amount each contract costs over and above the amount quoted in the original contract (overruns). The committee has determined that the standard deviation of overruns is \$17,500.
 - (a) The average overrun for a random sample of 50 contracts is \$12,000. Determine a 98% confidence interval estimate of the true mean overrun based on this sample. $\{\$6233.54, \$17,766.45\}$
 - (b) How large a sample should they use if they want to be 95% confident that the mean overrun is in error by no more than \$2000 $\{295\}$
 - (c) A random sample of 36 contracts is selected to estimate the average overrun. What is the probability that the sample mean will over-estimate the population mean overrun by at least \$5000 $\{.0436\}$

The directions for the t distribution for MINITAB is the same as the directions for the standard normal. The only difference is that you have to also plug in degrees of freedom.

2. (a) $P(T \leq -1.44)$ $df = 6$ (0.1)
- (a) $P(T \geq 0.52)$ $df = 20$ (.3044)
- (b) $P(T \geq -1.52)$ $df = 10$ (.9203)
- (c) $P(-0.33 \leq T \leq 0)$ $df = 15$ (.1270)
- (d) $P(-0.62 \leq T \leq 0.62)$ $df = 10$ (.4508)
- (e) $P(-0.50 \leq T \leq 1.98)$ $df = 12$ (.6513)
- (f) $P(0.34 \leq T \leq 2.33)$ $df = 16$ (.3525)

3. Determine the value of T_o if it is known that:

- (a) $P(-T_o \leq T \leq T_o) = 0.90$ $df = 5$ (2.015)
- (b) $P(-T_o \leq T \leq T_o) = 0.10$ $df = 10$ (.1289)
- (c) $P(T \geq T_o) = 0.20$ $df = 25$ (.8562)
- (d) $P(-1.66 \leq T \leq T_o) = 0.25$ $df = 12$ (-.5049)
- (e) $P(T \leq T_o) = 0.40$ $df = 15$ (-.2579)

Note: If you want to calculate the mean and standard deviation of a data set,

1. input all the data into one column.
2. Click on the header Calc>Column Statistics.
3. Click on the statistic that you are interested in (mean, st.dev etc)
4. Type the column in which the data is in, in the input variable box (or click on the input variable box and then double click on the column where the data is located.
5. Hit enter or click on OK

Note: You should familiarize yourself with some of the other functions of MINITAB. Check them out. You may find some time saving techniques.

4. The earnings per share for a random sample of technology stocks listed on the NYSE were (in \$'s):
1.90 2.15 2.01 0.89 1.53 1.89 2.12 2.05 1.75 2.22 3.44
 - (a) Assuming that earnings per share are normally distributed, determine a 95% confidence interval estimate of the average earnings per share of the NYSE technology stocks. $\{\bar{x} = 1.9955, s = 0.608, t = 2.228\}$ $\{\$1.58, \$2.41\}$
 - (b) A broker stated that the NYSE technical average earning was \$1.25 per share. Do the data confirm this or not. Use the results of (a) only.
 - (c) How large a sample should be used if the estimate of the mean NYSE technology stock earnings is to be in error by no more than \$.10. You want to have a 97% level of confidence in your sample. $\{175\}$
 - (d) How can we decrease/increase the error?

5. A pilot study has revealed that the standard deviation of workers' monthly earnings in the chemical industry is \$180. How large a sample must be chosen to obtain an estimator of the mean salary that, with 90% confidence, will be correct to within \pm \$20? {220}
6. A random sample of 41 quarters has a mean weight of 5.622g and a standard deviation of 0.068g.
 - (a) Construct a 98% confidence interval estimate of the population mean of all quarters in circulation. {t = 2.423} [5.5963g, 5.6477]
 - (b) The U.S. Department of the Treasury claims that it mints quarters to yield a mean weight of 5.640g. Is this claim consistent with the confidence interval? Explain why.
7. Assume that we want to estimate the mean IQ scores for the population of professors. How many professors must be randomly selected for IQ test if we want 95% confidence that the sample means is within 2 IQ points of the population mean? Assume the standard deviation is 15. [217]
8. It is found that a sample size of 843 is necessary to estimate the mean weight (in grams) of sugar in packets supplied by Domino. That sample size is based on a 95% degree of confidence and a population standard deviation that is estimated by the sample standard deviation of 0.074g. Find the margin of error. [.005]
9. Maximum heart rates during automated snow removal: $n=10$, $\bar{x}=124$, $s=18$. Find the 95% confidence interval estimate of the population mean for those who use the electric snow thrower. {t=2.262} [111.1245, 136.8755]
10. The 95% confidence interval for the true mean distance by male students in one year is 11,290 to 12,466. This was based on a sample of 121 randomly selected male students. Find the sample standard deviation that was used. [s=3266.6667]
11. You have been hired by the Ford Motor Company to do market research, and you must estimate the percentage of households in which a vehicle is owned. How many households must you survey if you want to be 94% confident that your sample percentage has a margin of error of three percentage points?
 - (a) Assume that a previous study suggested that vehicles are owned in 86% of households. [473]
 - (b) Assume that there is no available information that can be used to estimate the percentage of households in which a vehicle is owned. [982]
12. Because a proposed survey is time-consuming, an enterprising pollster posts it on the Internet and promises free software to everyone who responds by completing the survey. Results include 2250 responses, and 80% of them indicate that a fax machine is owned. Construct a 95% confidence interval for the percentage of all people who have a fax machine. Are the results valid? Why or why not? [78.3%, 81.7%] Results are not valid because the sample is self-selected (not a random sample)
13. You have just been hired by a marketing company to conduct a survey to estimate the mean amount of money spent by movie patrons (per movie) in Alberta. You need to know how many should be in your sample size that would correspond to 98% confidence and a 25¢ margin of error. It is reasonable to assume that typical amounts range from \$3 to about \$15. (782)
14. A random survey of 85 CEOs in British Columbia showed 70 respondents have a computer on their desk. Based on those results, construct a 98% confidence interval for the percentage of all CEOs in British Columbia who **do not** have a computer on their desk. (.0802, .2728)
15. The drug Ziac is used to treat hypertension. In a clinical test, 3.2% of 221 Ziac users experienced dizziness.
 - (a) Construct a 99% confidence interval estimate of the percentage of all Ziac users who experience dizziness. (.0015, .0625)

- (b) In the same clinical test, people in the placebo group didn't take Ziac but 1.8% of them reported dizziness. Based on the results in parts (a) and (b), what can we conclude about dizziness as an adverse reaction to Ziac?

Hypothesis testing.

16. A patio building company believes that it takes 27 days on average to complete a job. If the job is done in less time, the owner is afraid that the job may be rushed and will sacrifice quality. Further, a job that takes longer will escalate the costs. Records of fifty completed jobs are randomly selected. The mean length of job was found to be 25.3 days with a standard deviation of 2.1 days.
- Do records indicate that the mean length of a job is not 27 days as believed by Noah? Set up the hypothesis.
 - Suppose testing was carried out at the 1% level of significance (or $\alpha = .01$). What would your conclusion in (a) be { $t_{\text{calc}} = -5.724 < -2.68$, Rho}
 - Determine the p-value of this test. { ~ 0 }
17. Industrial espionage is a growing problem. It has been estimated that corporate extortion cost companies more than \$3.35 million on average. Sixty-five cases of this nature were examined and found to average \$3.71 million with a standard deviation of \$1.21 million.
- Do the data support the statement concerning the estimated cost of corporate extortion? Assume $\alpha = 0.01$ { $t_{\text{calc}} = 2.3986 > 2.38604$, Rho}
 - Determine the P-value of the test carried out in (a) {0.0097}
18. The owner of a small publishing firm thinks that business has improved lately. Last year the daily revenue for the firm was \$5,000 on average. A random sample of 20 recent days reveals a mean daily revenue of \$5,200 with a standard deviation of \$507
- Do the data support the owner's belief? Use the P-value for the appropriate statistical test. { $P(t \geq 1.7642) = 0.0469$, since p-value is small ($< .05$), we Rho}
 - Suppose that testing was carried out at a 5% significance level. What would be your conclusion? Use the results of (a)(i.e. the p-value) to make your decision. Explain why you reached your decision, and, hence, your conclusion. {p-value $< \alpha$, Rho}
 - Same question as (b) except now assume that $\alpha = 0.01$. {p-value $> \alpha$, Fail to RHo}
 - At what levels of significance would you come to a different conclusion in (a) (Fail to RHo when $\alpha < .0469$)
19. A cereal manufacturer sells boxes of cereal that list the weight as 19 oz. A random sample of 60 boxes had a mean fill of 19.2 oz with a standard deviation of 0.67 oz. Would this data indicate that the actual mean fill of all boxes of this cereal differs from the weight listed on the box?
- Carry out the appropriate hypothesis test to answer this question, using a 5% significance level. { $t_{\text{calc}} = 2.312 > 2.001$, Rho}
 - Determine the p-value of the test carried out in (a) {0.0242}
 - At what levels of significance would you come to a different conclusion in (a) (Fail to RHo when $\alpha < .0242$)
20. Big Burger claims that their deluxe special has at least 0.25 pounds of beef. A sample of 100 burgers made by this company had a mean of 0.237 pounds of beef with a variance of 0.0036.
- Is Big Burger guilty of false advertising? Suppose that you are prepared to accept a 5% chance of making a Type I error when carrying out the statistical test on the data. { $t_{\text{calc}} = -2.1667 < -1.6604$, Rho}

- (b) What is the minimum significance level for which you would conclude that Big Burger is guilty of false advertising? {0.0163}
21. A perfume distributor believes that the mean cost to process a sales order is \$13.25. The cost controller, fears that the average cost of processing is more than that. She is interested in taking action if costs are high, but she can accept the situation if the actual mean cost is below the hypothesized value. A random sample of 100 orders has a sample mean of \$13.35. Assuming the value of σ is its historical value of \$.50, conduct a test at the 0.01 level of significance. ($z_{\text{calc}}=2.00 < 2.33$, Fail to RHo)
22. An economist claims that the unemployment rate for non-English speaking people is at least 30% in a specific region of the country. In a random sample of 400 non-English-speaking people in this region 90 were found to be unemployed.
- Determine a 95% confidence interval estimate of the proportion of non-English-speaking people in the region that are unemployed. Do these data support the economist's claim? Explain why or why not {0.1841, 0.2659}
 - Do the data indicate that the economist's claim may not be true? Carry out the appropriate statistical test at a 2.5% significance level. { $z_{\text{calc}} = -3.2733 < -1.96$, Rho}
 - Determine the p-value of the test carried out in (b). Interpret {0.0005}
 - How large of a sample should be used to estimate the proportion of non-English-speaking unemployed people in the region if we want to be 98% sure that our estimate is in error by no more than 0.02(use z value from handout)? Determine the sample required if (i) you make use of the sample information given above, and if (ii) you want to know the maximum sample size required to meet the specifications. {2367, 3394}
 - What change should you consider to the specifications made in (d) that will reduce the sample size for either Case (i) or (ii)
23. In a survey of 1002 people, 701 said that they voted in the recent presidential election. Test to see if this is significantly different from the actual proportion who voted (.61 of the population). Use a 5% significance level and explain the conclusion. [$z_{\text{calc}}=5.81 > 1.96$, Rho]
24. The editor of a newspaper has written that 25 percent of the university students in the paper's circulation area read newspapers daily. A random sample of 200 of these university students shows that 45 of them are daily readers of newspapers.
- At the 0.05 level, is the editor's statement likely to be true? ($z_{\text{calc}}=-.8165 > -1.96$ Fail to RHo)
 - Find the p-value (.4142)
 - If $\alpha = .1$, would the conclusion in (a) be different?
 - At what levels of significance would you conclude that the editor's claim is false? ($\alpha > .4142$)
25. A supervisor assumes that the bottling machine is operating properly if at most 5 percent of the processed bottles are not full. A random sample of 100 bottles had 7 bottles that weren't full.
- Using a significance level of 0.01, conduct a test to see if the machine is operating properly. ($z_{\text{calc}}=.9177 < 2.33$, Fail to RHo)
 - Find the p-value and interpret. (.1794)
 - At what levels of significance would you conclude that the machine is not operating properly? ($\alpha > .1794$)

This part will be covered on the final but not on Quiz 5.

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. [$\beta_0 = -1.7470$, $\beta_1 = 0.8332$])

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**
- (c) Find the standard error of the regression and interpret its significance. [$s = 3.250$]
- (d) Find the coefficient of determination and interpret its meaning. [$r^2 = 0.8656$]
- (e) 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_e and interpret this value. [$s = 1.42$]
- (d) Compute the coefficient of determination and interpret its meaning in this problem. [$r^2 = 0.7018$]

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? $[r^2 = 0.0035, \text{very small}]$