

Stat 213
Assignment #6
Hypothesis testing and Regression

1. 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.
 - (a) Do records indicate that the mean length of a job is not 27 days as believed by Noah? Set up the hypothesis.
 - (b) 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}
 - (c) Determine the p-value of this test. { ~ 0 }

2. 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.
 - (a) 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}
 - (b) Determine the P-value of the test carried out in (a) {0.0097}

3. 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
 - (a) 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, we Rho}
 - (b) 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}
 - (c) Same question as (b) except now assume that $\alpha = 0.01$. {p-value $> \alpha$, Fail to RHo}
 - (d) At what levels of significance would you come to a different conclusion in (a) (Fail to RHo when $\alpha < 0.0469$)

4. 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?
 - (a) Carry out the appropriate hypothesis test to answer this question, using a 5% significance level. { $t_{\text{calc}} = 2.312 > 2.001$, Rho}
 - (b) Determine the p-value of the test carried out in (a) {0.0242}
 - (c) At what levels of significance would you come to a different conclusion in (a) (Fail to RHo when $\alpha < 0.0242$)

5. 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.
 - (a) 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}

6. 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)

7. 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.
 - (a) 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}
 - (b) 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}
 - (c) Determine the p-value of the test carried out in (b). Interpret {0.0005}
 - (d) 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}
 - (e) What change should you consider to the specifications made in (d) that will reduce the sample size for either Case (i) or (ii)

8. 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]

9. 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.
 - (a) At the 0.05 level, is the editor's statement likely to be true? ($z_{\text{calc}}=-.8165 > -1.96$ Fail to RHo)
 - (b) Find the p-value (.4142)
 - (c) If $\alpha = .1$, would the conclusion in (a) be different?
 - (d) At what levels of significance would you conclude that the editor's claim is false? ($\alpha > .4142$)

10. 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.
 - (a) 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)
 - (b) Find the p-value and interpret. (.1794)
 - (c) At what levels of significance would you conclude that the machine is not operating properly? ($\alpha > .1794$)

Minitab instructions for regression.

The following are two test scores for 10 individuals.

Test X	75	78	88	92	95	67	55	73	74	80
Test Y	81	73	85	85	89	73	66	81	81	81

Find the least squares regression line for predicting the likely score on test **Y** based on their score for test **X**.

1. Input all the data for Test X in column C1 and all the data for Test Y in column C2.
2. Go to the menu bar and select Stat<Regression.

Click the mouse on C2 Test Y in the rectangle on the left of the box: then press SELECT (Or type C1 in the Response cell). This enters the name of your variable as the response variable (dependent variable).

Click the mouse on C1 Test X in the rectangle on the left of the box, then press SELECT. This enters the name of your variable as the predictor variable (independent variable). Then click OK.

You will get a printout of the regression equation as well as the r^2 value.

$$\hat{y} = 39.87 + 0.51x \quad r^2 = 0.79$$

3. To see the scatter plot of the data with the fitted line, select Stats<Regression<Fitted Line Plot... Enter the response and predictor variables and click OK.
4. In order to show your work, you should make another column for $X \times Y$. To do this, click on Calc. Type C3 in the **store result in** box. To multiply columns, type C1*C2 in the Expression box and click OK. You will now have a column C3 in your worksheet.
5. To find the mean of a column, click Calc<Column Statistic. Select the column that you are interested in and then click mean, then OK. You do the same for the Sum and Sum of Squares.

$$\begin{array}{ll} n = 10 & \Sigma x^2 = 61,661 \\ \Sigma x = 777 & \Sigma^2 = 63,629 \\ \Sigma y = 795 & \Sigma xy = 62,432 \end{array}$$

Additional questions.

1. A morning newspaper lists the following used-car prices for a foreign compact, with age measured in years and selling price measured in thousands of dollars.

Age	1	2	2	3	3	4	6	7	8	10
Price	9.45	8.4	8.6	6.8	6.5	5.6	4.75	3.89	2.7	1.47

- (a) Find the least squared regression line for predicting price based on age. [$\hat{y} = 9.81 - 0.868(x)$]
 - (b) Find the Se^2 . [0.255]
 - (c) From the fitted regression line, determine the predicted value for the average selling price of a 5-year-old compact. [5.47]
 - (d) Find r^2 and comment on it. [.967]
 - (e) Find the predicted price for a car that is 11 years old. (.262-> \$2600) Why should we not try and predict this?
2. In an experiment designed to determine the relationship between the doses of a compost fertilizer x and the yield of a crop y , the following summary statistics are recorded:

$$n = 15, \quad \bar{x} = 10.8, \quad \bar{y} = 122.7, \quad S_{xx} = 70.6, \quad S_{yy} = 98.5, \quad S_{xy} = 68.3$$

Assume a linear relationship.

- (a) Find the equation of the least squares regression line. [$\hat{y} = 112.2521 + .9674(x)$]

- (b) Compute the error sum of squares and estimate δ^2 (Se^2). [32.4251, 2.4942]
- (c) Find r^2 and comment on it. (.6708)
- (d) Estimate the yield of crop for a dose of 10. [121.9261]

3. Many college students obtain college degree credits by demonstrating their proficiency on exams developed as part of the College Level Examination Program (CLEP). Based on their scores on the College Qualification Test (CQT), it would be helpful if students could predict their scores on a corresponding portion of the CLEP exam. The following data are for Total CQT score and mathematical CLEP score.

Total CQT	170	147	166	125	182	133	146	125	136
Total CLEP	698	518	725	485	745	538	485	625	471
Total CQT	179	174	128	152	157	174	185	171	102
Total CLEP	798	645	578	625	558	698	745	611	458

- (a) Find the least squares fit of a straight line. [$\hat{y} = 59 + 3.61(x)$]
 - (b) Find the Se^2 [4094]
 - (c) Find the r and comment on it. [0.8130]
4. Students' scores on the mathematics portion of the ACT examination, x , and on the final examination in first-semester calculus (200 points possible), y , are given and the following summary statistics are given.
- $n = 15$, $\bar{x} = 26.6$, $\bar{y} = 125$, $\Sigma xx = 10781$, $\Sigma yy = 249992$, $\Sigma xy = 50919$

Assume a linear relationship.

- (a) Find the equation of the least squares regression line. [$\hat{y} = -40.6941 + 6.2291(x)$]
- (b) Compute the error sum of squares and estimate δ^2 (Se^2). [9113.8019, 701.0617]
- (c) Find r^2 and comment on it. (.4164)
- (d) Estimate the final examination score for a student who scored 30 on the ACT examination. [146.1789]