Introduction to Statistics for Engineers
Homework 5

Name: Rhea Mae Edwards

OSU-ID: 932-389-303

Instructions

- The homework is due on Friday, June. 3rd and must be submitted on Canvas before midnight. (Please read the policies on late homeworks on the syllabus)
- Homeworks must be submitted to Canvas as a Word or PDF document. Any other format (including JPEG) will not be accepted and your homework will be considered late. (Please read the policies on late homeworks on the syllabus)
- You must show your work and provide complete answers in order to receive full credit. Solutions restricted only the final numerical values that do not reflect your statistical reasoning will not receive full credit. The homework is worth 25 points.
- If not using the space assigned for each question, you must clearly indicate the problem that you are working.
- You must include your name and OSU-ID number in your homework document.
- Failing to follow any of these instructions may result in a delay in the grading or a penalization in your final score.

$P_0 = 0.20$
$n = 400$
$\alpha = 0.01$
$\hat{p} = 61/400$
$= 0.1525$
$H_0: p = 0.20$
$H_1: p < 0.20$

1. Suppose that a pharmaceutical company claims that side effects will be experienced by fewer than 20% of the patients who use a particular medication. In a clinical trial with 400 patients, they find that 61 patients experienced side effects. Conduct an hypothesis test to determine whether the company's claim is reasonable. Use a significance level $\alpha = 0.01$.

   (a) State the null and alternative hypothesis for this test.

   Null Hypothesis $H_0: p = 0.20$

   Alternative Hypothesis $H_1: p < 0.20$

   (b) Compute the appropriate test statistic for this test.

   $np_0 \geq 15$ and $n(1-p_0) \geq 15$?

   $(400)(0.20) \geq 15$ yes.

   80 ✓

   $(400)(1-0.20) \geq 15$

   320 ✓

   $$z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} \quad = SE_{H_0}$$

   $$z = \frac{(0.1525 - 0.20)}{\sqrt{(0.20)(1-0.20)/400}} = \frac{-0.0475}{0.02}$$

   $$\boxed{z = -2.375}$$

   (c) Obtain the P-value for this test. What are your conclusions based on the significance level of the test?

   Looking at the z-table, we find that the P-value is $\boxed{0.0088}$

   Since the P-value of 0.0088 is less than the significance level of 0.01, there is strong enough evidence against/reject the null hypothesis of $p = 0.20$, and therefore we can conclude (based on the data) that the company's claim is reasonable, and that their company's pharmaceutical side effects will be experienced by fewer than 20% of the patients who use a particular medication.

2. A packaging device is set to fill detergent packets with a mean weight of 150 g. From historical data, the standard deviation is known to be 5.0 g. It is important to check the machine periodically, because if it is over-filling it increases the cost of the materials, whereas if it is under-filling the firm is liable to prosecution. Suppose that samples of 10 packets are taken every hour to monitor the packaging process. On Monday the following sample means were observed:

$\sigma = 5.0g$
$M_c = 150g$
$n = 8$

$$\underset{\bar{x}_1}{152.5}, \underset{\bar{x}_2}{153.0}, \underset{\bar{x}_3}{153.5}, \underset{\bar{x}_4}{154.5}, \underset{\bar{x}_5}{148.4}, \underset{\bar{x}_6}{150.1}, \underset{\bar{x}_7}{152.5}, \underset{\bar{x}_8}{151.8}$$

(a) Compute the appropriate contra chart limits and identify the target value for the $\bar{X}$-chart. (Use $z_{\alpha/2} = 1.96$)

Upper Control Limit (UCL) $= M_0 + z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$

UCL $= 150 + (1.96)\left(\frac{5.0}{\sqrt{8}}\right) = 153.46g$

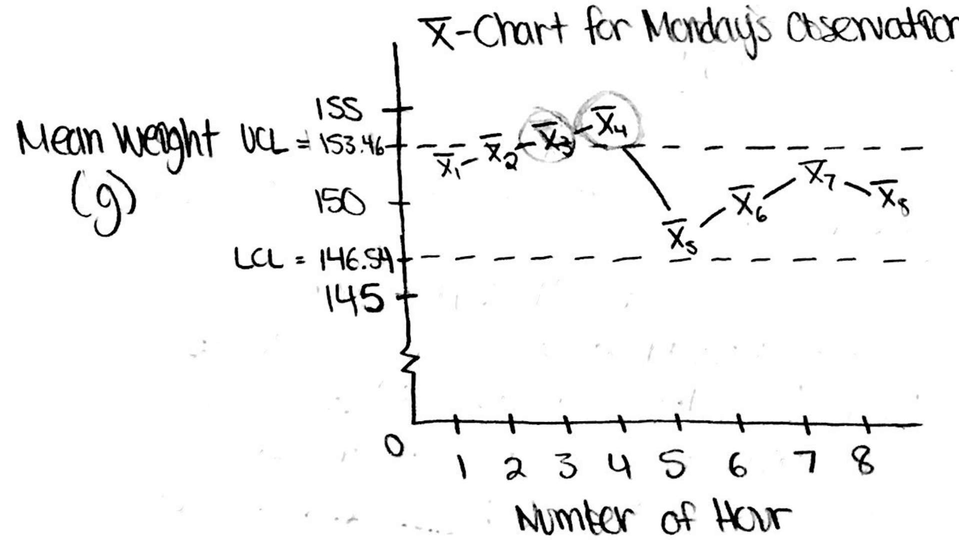
Lower Control Limit (LCL) $= M_0 - z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$

LCL $= 150 - (1.96)\left(\frac{5.0}{\sqrt{8}}\right) = 146.54$

Target Value = 150 g

$\boxed{\text{UCL} = 153.46g \quad \text{LCL} = 146.54g \quad \text{Target Value} = 150g}$

(b) Plot the $\bar{X}$-chart for Monday's observations and identify the in-control and out-of-control observations.



Out-of-Control Observations:

153.5 ($\bar{x}_3$), 154.5 ($\bar{x}_4$)

In-control Observations: 152.5g ($\bar{x}_1$), 153.0 ($\bar{x}_2$), 148.4 ($\bar{x}_5$), 150.1 ($\bar{x}_6$), 152.5 ($\bar{x}_7$), 151.8 ($\bar{x}_8$)

3. Airplanes approaching the runway for landing are required to stay within the localizer (a certain distance left and right the runway). When an airplane deviates form the localizer, it is sometimes referred to as an "exceedence". Consider two airlines at a large airport. During a three-week period, airline1 had 14 exceedences out of 156 and airline 2 had 11 exceedences out of 198 flights. Can we conclude that the probability of an ~~excellence~~ exceedence is the same for both airlines?

$\hat{p}_1 = \frac{14}{156} = 0.09$

$\hat{p}_2 = \frac{11}{198} = 0.06$

$x = 14$

$y = 11$

$n_1 = 156$

$n_2 = 198$

$\delta_0 = 0$

(a) Conduct an hypothesis test using a 0.05 significance level. $\alpha = 0.05$

$H_0: p_1 - p_2 = \delta_0$

$H_0: p_1 - p_2 = 0$

$H_1: p_1 - p_2 \neq 0$

$z = \frac{(\hat{p}_1 - \hat{p}_2) - \delta_0}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

$\hat{p} = \frac{x+y}{n_1+n_2}$

$\hat{p} = \frac{14+11}{156+198} = \frac{25}{354} = 0.07$

$z = \frac{(0.09 - 0.06) - 0}{\sqrt{(0.07)(1-0.07)\left(\frac{1}{156} + \frac{1}{198}\right)}} = 1.247$

P-Value using z-Table $\rightarrow (1 - 0.8944) = 0.1056$ $\times 2 \rightarrow 0.2112$

Since the P-Value of 0.2112 is greater than the significance level of 0.05, we fail to reject H_0, concluding (based on the data) that the probability of an exceedence is the same for both airlines.

(b) Construct a 95% confidence interval for the true difference between the proportion ~~if~~ of exceedences for the two airlines and compare with your results ~~form~~ from part (a).

$(\hat{p}_1 - \hat{p}_2) \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

$(0.09 - 0.06) \pm (1.96)\left(\sqrt{\frac{(0.09)(1-0.09)}{156} + \frac{(0.06)(1-0.06)}{198}}\right)$

$0.034 \pm 0.055 \rightarrow (-0.021, 0.089)$

So we are 95% confident that the true difference $p_1 - p_2$ between the proportion if exceedences lies between $(-0.021, 0.089)$.

Because the resulting confidence interval contains the value 0, we can conclude (based off the data) there are no differences for the probabilities of an exceedence between the two airlines, also saying that the probability of an exceedence is the same for both airlines. This 95% confidence interval conclusion and the hypothesis test using a 0.05 significance level conclusion from part (a), agree with one another that we can conclude that the probability of an exceedence is the same for both airlines. ✓

MSres = 6.686

4. Data on corn yield y (in bu/acre) and rainfall x (in in/year) in six U.S. corn-producing states (Iowa, Nebraska, Illinois, Indiana, Missouri and Ohio), was recorded for 12 consecutive years. The data is avialable in the file "HW5.csv":

   (a) Construct a scatter plot of the data. Does the relation between x and y seem to be linear?

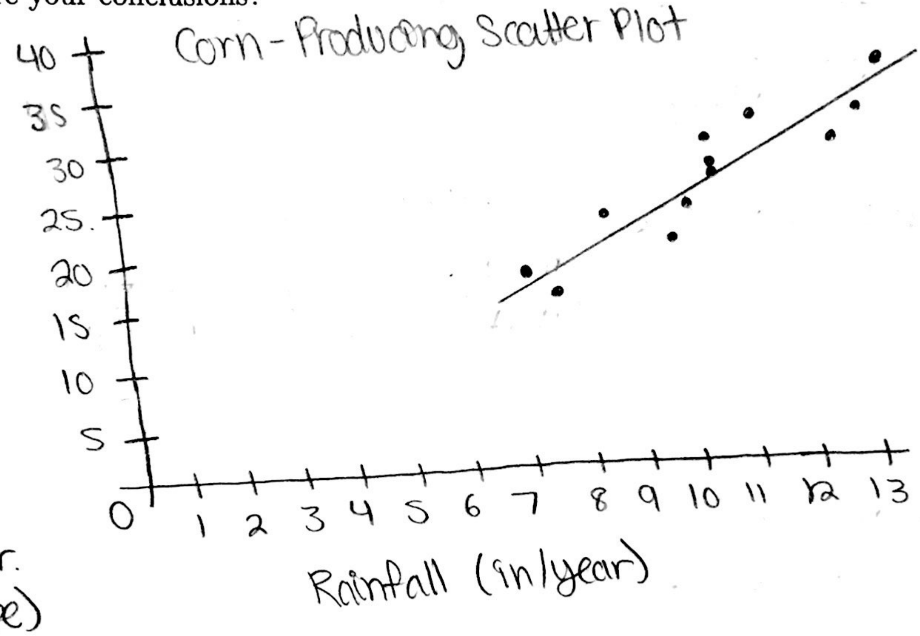
   (b) Obtain the least squares regression line.

   (c) Obtain and interpret the value of ~~R2~~ R^2, the coefficient of determination.

   (d) Conduct an hypothesis test to determine whether the slope is equal to 0. What are your conclusions?

a.)

Corn Yield (bu/acre)



Yes, the relation between x and y seem to be linear. (with a positive slope) (increasing)

b.) $y = 2.372x + 4.951$ (roughly) → Obtained by using R software

(Least squares Regression line drawn on scatterplot in part (a))

c.) Coefficient of Determination $R^2 = 0.7745$ → Obtained by using R Software

77.45% of the variation in the amount of corn yields can be explained by the amount of rainfall of that year, when using the least squares regression line calculated by the data given.

d.) $H_0: \beta_1 = 0$
$H_1: \beta_1 \neq 0$

$$t = \frac{b_1 - c}{\sqrt{\frac{MSres}{(n-1)S_x^2}}}$$

$df = n - 2$

$c = 0$
$MSres = 6.686$
$n = 12$
$b_1 = 2.372$
↳ Caclulated by R Program
$S_x = 1.927$
$\alpha = 0.05$

$$t = \frac{2.372}{\sqrt{\frac{6.686}{(12-1)(1.927)^2}}} = 5.8616$$

w/ $df = 10$ → P-value $= 2.151 \times 10^{-4}$

Calculated using TI-84 Plus Calculator

Since the calculated P-value of 2.151×10^{-4} is less than the significance level of 0.05, we can reject H_0, giving us strong enough evidence to conclude (based on the data) that the slope is not equal to 0.

From June 1st lecture:

Checking if the slope is equal to 2, would be a more reasonable approach.

$H_0: \beta_1 = 2$ vs $H_1: \beta_1 \neq 2$

$$t = \frac{2.372 - 2}{\sqrt{\frac{6.686}{(12-1)(1.927)^2}}} = 0.919$$

$df = 10$ → P-Value $= 0.498$

Calculated using TI-84 Plus Calculator

By using this approach instead with testing if the slope is equal to 2, we are able to conclude (based off the data) that the slope is equal to 2, because there is enough evidence resulting that we <u>cannot</u> reject the H_0, because the P-value of 0.498 is greater than the significance level of 0.05.