.

Midterm-II Exam (Part A)
CE 362: Statistics and Inference
Dr. Shafayat Abrar

Timing: 02:25 PM - 03:40 PM
Nov 28, 2022
Duration: 75 min

Q1 [25 points]: According to the National Association of Colleges and Employers, the average hourly wage of an undergraduate college student working as a co-op is \$17.3 and the average hourly wage of a college student working as an intern is \$16.6. Assume that such wages are normally distributed in the population and that the population variances are equal. Suppose these figures were actually obtained from the data below.

(a) Use these data and $\alpha = 0.10$ to test to determine if there is a significant difference in the mean hourly wage of a college co-op student and the mean hourly wage of a college intern.

(b) Using these same data, construct a 90% confidence interval to estimate the difference in the population mean hourly wages of college co-ops and interns.

Co-ops	Interns
16.97	16.23
16.38	15.58
17.51	17.34
18.55	16.04
18.47	14.93
19.20	17.25
15.68	17.38
17.04	17.02
18.37	15.12
16.08	17.21
16.88	16.98
16.27	17.55

Q2 [20 points]: The vice president of marketing brought to the attention of sales managers that most of the company's manufacturer representatives contacted clients and maintained client relationships in a disorganized, haphazard way. The sales managers brought the reps in for a three-day seminar and training session on how to use an organizer to schedule visits and recall pertinent information about each client more effectively. Sales reps were taught how to schedule visits most efficiently to maximize their efforts. Sales managers were given data on the number of site visits by sales reps on a randomly selected day both before and after the seminar. Use the following data to test whether significantly more site visits were made after the seminar ($\alpha = .05$). Assume the differences in the number of site visits are normally distributed.

Rep	Before	After
1	2	4
2	4	5
3	1	3
4	3	3
5	4	3
6	2	5
7	2	6
8	3	4
9	1	5

Q3 [25 points]: Using the given sample information, test the following hypotheses:

(a) $\mathcal{H}_0 : p_1 - p_2 = 0 \quad \mathcal{H}_a : p_1 - p_2 \neq 0$. Let $\alpha = 0.05$.

	Sample 1	Sample 2
c	$n_1 = 350$	$n_2 = 410$
	$x_1 = 160$	$x_2 = 190$

Note that x is the number in the sample having the characteristic of interest.

(b) $\mathcal{H}_0 : p_1 - p_2 = 0 \quad \mathcal{H}_a : p_1 - p_2 > 0$. Let $\alpha = 0.1$.

	Sample 1	Sample 2
c	$n_1 = 700$	$n_2 = 600$
	$\widehat{p_1} = 0.4$	$\widehat{p_2} = 0.25$

Q4 [20 points]: How long are resale houses on the market? One survey by the Houston Association of Realtors reported that in Houston, resale houses are on the market an average of 112 days. Of course, the length of time varies by market. Suppose random samples of 13 houses in Houston and 11 houses in Chicago that are for resale are traced. The data shown here represent the number of days each house was on the market before being sold. Use the given data and a 1% level of significance to determine whether the population variances for the number of days until resale are different in Houston than in Chicago. Assume the numbers of days resale houses are on the market are normally distributed.

$$\text{data}_{\text{Houstan}} = [132 \;\; 138 \;\; 131 \;\; 127 \;\; 99 \;\; 126 \;\; 134 \;\; 126 \;\; 94 \;\; 161 \;\; 133 \;\; 119 \;\; 88]$$
$$\text{data}_{\text{Chicago}} = [118 \;\; 85 \;\; 113 \;\; 81 \;\; 94 \;\; 93 \;\; 56 \;\; 69 \;\; 67 \;\; 54 \;\; 137];$$

Q5 [30 points]: Sketch a scatter plot from the following data, and determine the equation of the regression line.

x	12	21	28	8	20
y	17	15	22	19	24

Test the slope of the regression line. Use $\alpha = 0.05$.

Note: Data is also available at LMS.

Q1. $\mathcal{H}_0: \mu_1 - \mu_2 = 0$

$\mathcal{H}_1: \mu_1 - \mu_2 \neq 0$

Co-op	Intern
$n_1 = 12$	$n_2 = 12$
$\bar{X}_1 = 17.2833$	$\bar{X}_2 = 16.5525$
$S_1 = 1.1322$	$S_2 = 0.9350$

(a) For two-tail test $\alpha/2 = 0.05$
$df = 12 + 12 - 2 = 22$

Critical $t_{0.05, 22} = \pm 1.717$

$$\text{Observed } t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{S_1^2(n_1-1) + S_2^2(n_2-1)}{n_1+n_2-2}}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$$

$$t = 1.5812$$

Since $t = 1.5812 < 1.717 = t_{0.05, 22}$,
the decision is to fail to reject the null hyp.

Q1 (b) $t_{0.05,22} = \pm 1.717$ ②

$$\bar{x}_1 - \bar{x}_2 \pm t\sqrt{\frac{s_1^2(n_1-1) + s_2^2(n_2-1)}{n_1+n_2-2}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

$$= 0.7308 \pm 1.717 * 0.4622$$

$$= 0.7308 \pm 0.7936$$

$$-0.0628 \leq \mu_1 - \mu_2 \leq 1.5244$$

Q2. $H_0 : D = 0$

$H_1 : D < 0$

Let $d = X_1 - X_2 = X_{before} - X_{after}$.

$n = 9$, $\bar{d} = -1.7778$

$S_d^2 = 2.9444 \Rightarrow S_d = 1.7159$

$\alpha = 0.05$, $df = n - 1 = 9 - 1 = 8$

For onetail test, $t_{0.05,8} = -1.86$

$$t_{observed} = \frac{\bar{d} - D}{S_d/\sqrt{n}} = \frac{-1.7778 - 0}{1.7159/\sqrt{9}}$$

$t_{obs} = -3.1082$

Since $t_{obs} < t_{crit}$

$-3.11 < -1.86$

The decision is to reject the null hypothesis

Q3. Sample 1

$n_1 = 350$

$x_1 = 160$

Sample 2

$n_2 = 410$

$x_2 = 190$

$\Rightarrow \hat{p}_1 = \frac{x_1}{n_1} = 0.4571 \qquad \hat{p}_2 = \frac{x_2}{n_2} = 0.4634$

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{160+190}{350+410} = 0.4605$$

$$\bar{q} = 1 - \bar{p} = 0.5395$$

$H_0: \; p_1 - p_2 = 0$

$H_a: \; p_1 - p_2 \neq 0$

For two-tail test, $\frac{\alpha}{2} = 0.025$

$$Z_{0.025} = \pm 1.96$$

$$Z = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\bar{p}\bar{q}\left(1/n_1 + 1/n_2\right)}} = \frac{0.4571 - 0.4634}{\sqrt{0.4605 \times 0.5395 \left(\frac{1}{350} + \frac{1}{410}\right)}}$$

$Z_{obs} = -0.1737, \quad |Z_{obs}| < 1.96$

Fail to reject H_0.

Q3 (b)

Sample 1	Sample 2
$\hat{p}_1 = 0.4$	$\hat{p}_2 = 0.25$
$n_1 = 700$	$n_2 = 600$

$H_0: p_1 - p_2 = 0$

$H_a: p_1 - p_2 > 0$

Let $\alpha = 0.1$. $\Rightarrow Z_{0.1} = 1.28$ (single-tail)

$$\bar{p} = \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2} = 0.3308, \quad \bar{q} = 0.6692$$

$$Z_{obs} = \frac{0.4 - 0.25}{\sqrt{\bar{p} \cdot \bar{q} \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = 5.7304 > 1.28$$

Reject the null hypothesis.

Q4. Let Houston be group 1,
& Chicago be group 2.

$$H_0: \sigma_1^2 = \sigma_2^2$$

$$H_a: \sigma_1^2 \neq \sigma_2^2$$

$$\alpha = 1\% = 0.01.$$

$$df_1 = 13-1 = 12$$
$$df_2 = 11-1 = 10$$

This is a two-tail problem. $\frac{\alpha}{2} = \frac{0.01}{2} = 0.005$

$$F_{0.005,12,10} = 5.66$$

$$F_{0.995,10,12} = 0.177 = \frac{1}{5.66}$$

If the observed value is greater than 5.66 or less than 0.177, we reject the null hypo.

$$\left.\begin{aligned} s_1^2 &= 393.3974 \\ s_2^2 &= 702.6909 \end{aligned}\right\} \Rightarrow F = \frac{s_1^2}{s_2^2} = 0.5598$$

Since, $0.177 < 0.5598 < 5.66$ is true
We accept H_0.

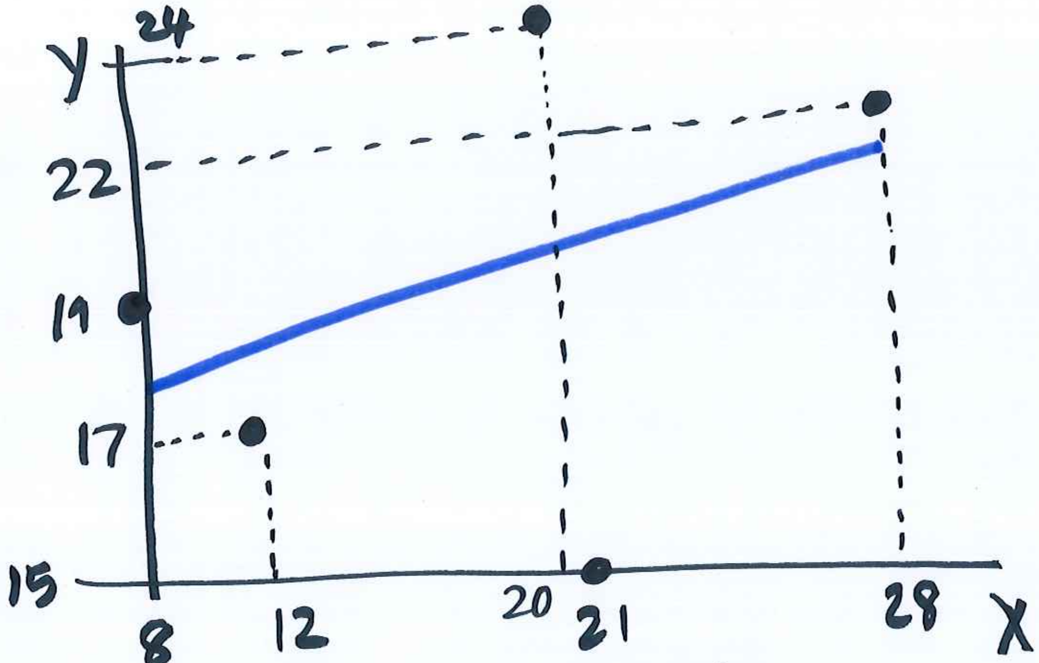
Q5. $X = [12, 21, 28, 8, 20]'$;

$Y = [17, 15, 22, 19, 24]'$;

$XX = [\text{ones}(5,1) \quad X]$;

$Bs = (XX' * XX)^{-1} * XX' * Y$

$$= \begin{bmatrix} 16.5096 \\ 0.1624 \end{bmatrix} = \begin{bmatrix} \hat{b}_0 \\ \hat{b}_1 \end{bmatrix}$$



$b_0 = 16.5096$; $b_1 = 0.1624$;

line([8, 28], b_1 * [8, 28] + b_0)

$$s_b = \frac{s_e}{\sqrt{\Sigma x^2 - (\Sigma x)^2/n}}$$

$\hat{Y} = b_0 * b_1 X = [18.46, 19.92, 21.06, 17.81, 19.76];$

$$S_e^2 = \frac{SSE}{n-2}$$

$$SSE = \Sigma_{i=1}^{5} (Y_i - \hat{Y}_i)^2 = \text{sum}((Y - \text{Yhat}).\hat{}2)$$
$$= 46.6399$$

$$S_e = \sqrt{\frac{SSE}{n-2}} = 3.9429.$$

$$S_b = \frac{3.9429}{\sqrt{1833 - 89^2/5}} = 0.25$$

$b_1 = 0.1624$

$\alpha = 0.05$

$H_0: \quad \beta_1 = 0$

$H_a: \quad \beta_1 \neq 0$

This is a two-tail problem

$$\alpha/2 = 0.025$$

$$df = n-2 = 5-2 = 3$$

$$t_{0.025,3} = \pm 3.182$$

$$t = \frac{b_1 - \beta_1}{S_b} = \frac{0.1624 - 0}{0.25} = 0.6496$$

Since $t = 0.6496 < t_{0.025,3} = 3.182$, the decision is to fail to reject H_0. (Slope is significant).