Question 9.7

In order to ensure efficient usage of a server, it is necessary to estimate the mean number of concurrent users. According to records, the average number of concurrent users at 100 randomly selected times is 37.7, with a standard deviation σ = 9.2.

- a. *Construct a 90% confidence interval for the expectation of the number of concurrent users.*
- b. *At the 1% significance level, do these data provide significant evidence that the mean number of concurrent users is greater than 35?*

Solution

Given:

$n = 100$
$\bar{X} = 37.7$
$\sigma = 9.2$

a. $\propto = 1 - 0.9 = 0.1$

We know that, confidence interval is given by $\bar{X} \pm z_{\propto/2} * \frac{\sigma}{\sqrt{n}}$

Therefore, $37.7 \pm Z_{0.05} * \frac{9.2}{\sqrt{100}} = 37.7 \pm (1.645 * 0.92) = 37.7 \pm 1.5134$

Thus, the confidence interval is [36.1866, 39.2134]

b. $\propto = 0.01$

We know that $H_0: \mu = 35$ and $H_A: \mu > 35$, thus a right-tail alternative is considered.

The test statistics is given by: $Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$

$$Z = \frac{37.7 - 35}{9.2/\sqrt{100}} = \frac{2.7}{0.92} = 2.9347$$

We know that $z_{0.1} = 2.326$

For a right tail alternative: $\begin{cases} reject\ H_0 & if\ Z \geq z_\alpha \\ accept\ H_0 & if\ Z < z_\alpha \end{cases}$

Since 2.9347 > 2.326, we reject H_0.

Question 9.8

Installation of a certain hardware takes random time with a standard deviation of 5 minutes.

- a. *A computer technician installs this hardware on 64 different computers, with the average installation time of 42 minutes. Compute a 95% confidence interval for the population mean installation time.*
- b. *Suppose that the population mean installation time is 40 minutes. A technician installs the hardware on your PC. What is the probability that the installation time will be within the interval computed in (a)?*

Solution

Given:

$\sigma = 5$

a. $\bar{X} = 42$
$n = 64$
$\alpha = 1 - 0.95 = 0.05$

We know that, confidence interval is given by $\bar{X} \pm z_{\propto/2} * \frac{\sigma}{\sqrt{n}}$
Therefore, $42 \pm Z_{0.025} * \frac{5}{\sqrt{64}} = 42 \pm (1.960 * 0.625) = 42 \pm 1.225$
Thus, the confidence interval is [40.775, 43.225]

b. $\mu = 40$

$$\begin{aligned} P(40.775 \leq \mu \leq 43.225) &= P\left(\frac{40.775 - \mu}{\sigma} \leq Z \leq \frac{43.225 - \mu}{\sigma}\right) \\ &= P\left(\frac{40.775 - 40}{5} \leq Z \leq \frac{43.225 - 40}{5}\right) \\ &= P\left(\frac{0.775}{5} \leq Z \leq \frac{3.225}{5}\right) \\ &= P(0.155 \leq Z \leq 0.645) \\ &= 0.7422 - 0.5636 \\ &= 0.1786 \end{aligned}$$

Question 9.10

We have to accept or reject a large shipment of items. For quality control purposes, we collect a sample of 200 items and find 24 defective items in it.

a. Construct a 96% confidence interval for the proportion of defective items in the whole shipment.

b. The manufacturer claims that at most one in 10 items in the shipment is defective. At the 4% level of significance, do we have sufficient evidence to disprove this claim? Do we have it at the 15% level?

Solution

Given:

$n = 200$

$\hat{p} = \frac{24}{200} = 0.12$

a. $\alpha = 1 - 0.96 = 0.04$

We know that, confidence interval is given by $\hat{p} \pm z_{\propto/2} * \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

Therefore, $0.12 \pm Z_{0.02} * \sqrt{\frac{0.12(1-0.12)}{200}}$
$= 0.12 \pm (2.045 * 0.0229) = 0.12 \pm 0.0470$
Thus, the confidence interval is [0.073, 0.167]

b. $p = 0.1$
$\alpha = 0.04$
We know that $H_0: p = 0.1$ and $H_A: p > 0.1$, thus a right-tail alternative is considered.

Naman Jain Vimal Kumar
1001670153

The test statistics is given by: $Z = \frac{\hat{p}-p}{\sqrt{p(1-p)/_n}}$

$$Z = \frac{0.12 - 0.1}{\sqrt{0.1(1-0.1)/_{200}}} = \frac{0.02}{0.0212} = 0.9428$$

We know that $z_{0.04} = 2.045$

For a right tail alternative: $\begin{cases} reject\ H_0 & if\ Z \geq z_\alpha \\ accept\ H_0 & if\ Z < z_\alpha \end{cases}$

Since 0.9428 < 2.045, we accept H_0.

p-value = $P(Z \geq Z_0) = 1 - P(Z < Z_0)$
$= 1 - P(Z < 0.9428) = \phi(0.9428) = 0.1736$

Yes, we have more than 15% level.

Question 9.11

Refer to Exercise 9.10. Having looked at the collected sample, we consider an alternative supplier. A sample of 150 items produced by the new supplier contains 13 defective items. Is there significant evidence that the quality of items produced by the new supplier is higher than the quality of items in Exercise 9.10? What is the P-value?

Solution

Given:

$Old\ Supplier: \widehat{p_1} = 0.12$

$New\ Supplier: \widehat{p_2} = \frac{13}{150} = 0.0867$

Therefore,

Pooling proportion: $\hat{p} = \frac{24+13}{200+150} = \frac{37}{350} = 0.1057$

Let,

$H_0: p_1 \leq p_2$
$H_A: p_1 > p_2$
$\alpha = 0.05$

Test Statistic: $Z = \frac{\widehat{p_1}-\widehat{p_2}}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1}+\frac{1}{n_2})}}$

$$Z = \frac{0.12 - 0.0867}{\sqrt{0.1057(1-0.1057)\left(\frac{1}{200}+\frac{1}{150}\right)}}$$

$$Z = \frac{0.033}{\sqrt{0.1057(0.8943)(0.0116)}} = \frac{0.033}{\sqrt{0.0011}} = \frac{0.033}{0.033} = 1$$

p-value = $P(Z > Z_0) = 1 - P(Z \leq Z_0) = 1 - 0.8413 = 0.1587$

The p-value is greater than the given significance level, so we fail to reject the null hypothesis and accept that there is no significant proof to say that the quantity of items produced by new suppliers is higher than those of the old suppliers.

Question 9.17

A news agency publishes results of a recent poll. It reports that candidate 'A' leads candidate 'B' by 10% because 45% of the poll participants supported Ms.A whereas only 35% supported Mr.B. What margin of error should be reported for each of the listed estimates, 10%, 35%, and 45%? Notice that 900 people participated in the poll, and the reported margins of error typically correspond to 95% confidence intervals.

Solution

Given:

$\propto = 0.05$

$n = 900$

We know that,

$$margin = z_{\propto/2} * \sqrt{\frac{p(1-p)}{n}} \quad ; z_{0.025} = 1.960$$

$$margin = z_{\propto/2} * \sqrt{\frac{\widehat{p_1}(1-\widehat{p_1})}{n_1} + \frac{\widehat{p_2}(1-\widehat{p_2})}{n_2}}$$

Therefore,

$$margin_A = z_{0.025} * \sqrt{\frac{0.45(1-0.45)}{900}} = 1.96 * \sqrt{\frac{0.45(0.55)}{900}} = 0.0325$$

$$margin_B = z_{0.025} * \sqrt{\frac{0.35(1-0.35)}{900}} = 1.96 * \sqrt{\frac{0.35(0.65)}{900}} = 0.0311$$

$$margin_{10} = z_{0.025} * \sqrt{\frac{0.45(1-0.45)}{900} + \frac{0.35(1-0.35)}{900}}$$

$$= 1.96 * \sqrt{\frac{0.45(0.55)}{900} + \frac{0.35(0.65)}{900}} = 0.0450$$

Question 9.18

Consider the data about the number of blocked intrusions in Exercise 8.1, p. 233.

a. *Construct a 95% confidence interval for the difference between the average number of intrusions attempts per day before and after the change of firewall settings (assume equal variances).*
b. *Can we claim a significant reduction in the rate of intrusion attempts? The number of intrusions attempts each day has approximately Normal distribution. Compute P-values and state your conclusions under the assumption of equal variances and without it. Does this assumption make a difference?*

Solution

We know that

$$\bar{X} = \sum x_i / n \qquad s^2 = \sum (x_i - \bar{X})^2 / n-1$$

From the working.xlsx,

	N	Mean	Std Deviation
Before change of Firewall Settings	14	50	7.62
After change of Firewall Settings	20	40.2	7.96

Degrees of freedom: $df = n_1 + n_2 - 2 = 14 + 20 - 2 = 32$

a. $\propto = 1 - 0.95 = 0.05$

Confidence Interval of T-tabulation of $(\propto, df)$: $\overline{X_1} - \overline{X_2} \pm t_{critical} * \left(\frac{s_1{}^2}{n_1} + \frac{s_2{}^2}{n_2}\right)$

$$= 50 - 40.2 \pm t_{0.025} * \left(\frac{7.62^2}{14} + \frac{7.96^2}{20}\right) = 9.8 \pm (2.037 * 2.7047)$$

$$= 9.8 \pm 5.5095 = [4.2905, 15.3095]$$

b. Let,

H_0: $no\ significance\ difference\ between\ the\ avg.intrusions\ per\ day$

H_A: $Significant\ difference\ between\ the\ avg.intrusions\ per\ day$

both based on changes made before firewall settings.

$\propto = 1 - 0.95 = 0.05$

$z_{\alpha/2} = 2.307$

Then, test statistic 'T' is given by:

Unequal Variances

$$t = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{S_1{}^2}{n_1} + \frac{S_2{}^2}{n_2}}} = \frac{50 - 40.2}{\sqrt{\frac{7.62^2}{14} + \frac{7.96^2}{20}}} = \frac{9.8}{2.7047} = 3.6233$$

$$\nu = \frac{\left(\frac{s_1{}^2}{n_1} + \frac{s_2{}^2}{n_2}\right)^2}{\left(\frac{s_1{}^2}{n_1{}^2(n_1-1)} + \frac{s_2{}^2}{n_2{}^2(n_2-1)}\right)} = 29$$

$$p - index = TDIST(3.6233, 29, 2) = 0.000978$$

Equal Variances

$$s_p{}^2 = \frac{\sum(x_i - \bar{x})^2 + \sum(y_i - \bar{y})^2}{n + m - 2} = 7.8206$$

$$t = (\overline{X_1} - \overline{X_2}) / \left(s_p{}^2 * \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right) = 3.5962$$

$$\nu = n_1 + n_2 - 2 = 32$$

$$p - index = TDIST(3.6233, 29, 2) = 0.000978$$

In both cases we reject H0.

Question 9.20

A manager questions the assumptions of Exercise 9.8. Her pilot sample of 40 installation times has a sample standard deviation of s = 6.2 min, and she says that it is significantly different from the assumed value of σ = 5 min. Do you agree with the manager? Conduct the suitable test of a standard deviation. [Use 2% Significance]

Solution

Given:

$n = 40$
$s = 6.2$
$\propto = 0.02$

Let,

$H_0 : \sigma = 5$
$H_A : \sigma \neq 5$

Therefore, test Statistic:

$$\chi^2 = \frac{(n-1) * s^2}{\sigma^2} = \frac{(40-1) * 6.2^2}{5^2} = 59.9664$$

Degrees of freedom: $df = n - 1 = 40 - 1 = 39$

$p - value = CHIDIST(59.9664{,}39) = 0.017043$

$t_{critical} = t_{0.02}(39) = 2.426$

Conclusion, the p-index is less than the given significance value of 0.02, thus we reject the null hypothesis and accept the alternative hypothesis, which is in favor of the manager.

Question 9.23

Anthony says to Eric that he is a stronger student because his average grade for the first six quizzes is higher. However, Eric replies that he is more stable because the variance of his grades is lower. The actual scores of the two friends (presumably, independent and normally distributed) are in the table.

	Quiz 1	Quiz 2	Quiz 3	Quiz 4	Quiz 5	Quiz 6
Anthony	85	92	97	65	75	96
Eric	81	79	76	84	83	77

a. *Is there significant evidence to support Anthony's claim? State H0 and HA. Test equality of variances and choose a suitable two-sample t-test. Then conduct the test and state conclusions.*
b. *Is there significant evidence to support Eric's claim? State H0 and HA and conduct the test.*

For each test, use the 5% level of significance.

Solution

From working.xlsx:

	Mean	Std Deviation
Anthony	85	12.7593
Eric	80	3.2249

Also given:

$\alpha = 0.05$
$n = 6$

Therefore

$z_{\propto} = 1.943$

a. Let

$H_0: \mu_A \leq \mu_E$
$H_A: \mu_A > \mu_E$

To find, pooled variance:

$S_P{}^2 = \frac{(n_A-1)s_A{}^2+(n_B-1)s_B{}^2}{n_A+n_B-2} = \frac{(6-1)(12.76)^2+(6-1)(3.23)^2}{6+6-2} = 86.625$

Finding, the test statistic

$$t = \frac{X_A - X_B}{\sqrt{S_P{}^2 * \left(\frac{1}{n_A} + \frac{1}{n_B}\right)}} = \frac{85 - 80}{\sqrt{(86.593)^2 * \left(\frac{1}{6} + \frac{1}{6}\right)}} = 0.9297$$

Degrees of freedom: $df = n_A + n_B - 2 = 6 + 6 - 2 = 10$

$p - index = TDIST(0.9297, 10, 1) = 0.187138$

Here, the p-index is less than the given significance value 0.05, hence we reject H0 and accept that there is no significant evidence to support Anthony's statement.

b. Let

$H_0: \sigma_A{}^2 \leq \sigma_E{}^2$
$H_A: \sigma_A{}^2 > \sigma_E{}^2$

Finding, the test statistic

$$F = \frac{s_A{}^2}{s_B{}^2} = \frac{12.76^2}{3.23^2} = 15.70325$$

Degrees of Freedom: $\upsilon_A = \upsilon_B = n - 1 = 6 - 1 = 5$

$p - index = FDIST(15.70325, 5, 5) = 0.004464$

Conclusion, the p-index is lesser than the given significance value, so we reject the null hypothesis and accept that there is sufficient evidence to support Eric's statement.

Question (9)

You are given three dice. The first dice is unbiased. The second dice follows the following pmf:

x	*1*	*2*	*3*	*4*	*5*	*6*
P(x)	*1/9*	*1/9*	*1/9*	*2/9*	*2/9*	*2/9*

The third dice follows the following pmf:

x	*1*	*2*	*3*	*4*	*5*	*6*
P(x)	*1/9*	*1/9*	*2/9*	*2/9*	*2/9*	*1/9*

i. *Calculate the average amount of information (in bits) encoded by each dice*
ii. *Calculate the relative entropy (in bits) of:*
    a. *The second dice w.r.t the first dice*
    b. *The third dice w.r.t to first dice*
    c. *The second dice w.r.t to third dice.*

Solution

i. Average Amount of information: $H = -\sum p(x) * \log_2(p(x))$

$$H(Dice\ 1) = \left(-\frac{1}{6}\log_2\frac{1}{6}\right) * 6 = 0.1667 * 2.5849 * 6$$
$$H(Dice\ 1) = 2.5849 \approx 2.59$$

$$H(Dice\ 2) = \left(\left(-\frac{1}{9}\log_2\frac{1}{9}\right) * 3\right) + \left(\left(-\frac{2}{9}\log_2\frac{2}{9}\right) * 3\right)$$
$$H(Dice\ 2) = (0.1111 * 3.1669 * 3) + (0.2222 * 2.1699 * 3)$$
$$H(Dice\ 2) = 1.0566 + 1.4466$$
$$H(Dice\ 2) = 2.5032 \approx 2.50$$

$$H(Dice\ 3) = \left(\left(-\frac{1}{9}\log_2\frac{1}{9}\right) * 3\right) + \left(\left(-\frac{2}{9}\log_2\frac{2}{9}\right) * 3\right)$$
$$H(Dice\ 3) = (0.1111 * 3.1669 * 3) + (0.2222 * 2.1699 * 3)$$
$$H(Dice\ 3) = 1.0566 + 1.4466$$
$$H(Dice\ 3) = 2.5032 \approx 2.50$$

ii. To calculate the relative entropy: $d = \sum p(x) * \log_2\left(\frac{p(x)}{q(x)}\right)$

a. Second Dice w.r.t first dice:

$$d = \frac{1}{9}\log_2\frac{1/9}{1/6} + \frac{1}{9}\log_2\frac{1/9}{1/6} + \frac{1}{9}\log_2\frac{1/9}{1/6} + \frac{2}{9}\log_2\frac{2/9}{1/6} + \frac{2}{9}\log_2\frac{2/9}{1/6} + \frac{2}{9}\log_2\frac{2/9}{1/6}$$

$$d = \left(3 * \left(\frac{1}{9}\log_2\frac{6}{9}\right)\right) + \left(3 * \left(\frac{2}{9}\log_2\frac{12}{9}\right)\right)$$
$$d = (3 * 0.1111 * -05849) + (3 * 0.2222 * 0.4150)$$
$$d = -0.1950 + 0.2767$$
$$d = 0.0817$$

b. Third Dice w.r.t first dice:

$$d = \frac{1}{9}\log_2\frac{1/9}{1/6} + \frac{1}{9}\log_2\frac{1/9}{1/6} + \frac{2}{9}\log_2\frac{2/9}{1/6} + \frac{2}{9}\log_2\frac{2/9}{2/6} + \frac{2}{9}\log_2\frac{2/9}{2/6} + \frac{1}{9}\log_2\frac{1/9}{2/6}$$
$$d = \left(3 * \left(\frac{1}{9}\log_2\frac{6}{9}\right)\right) + \left(3 * \left(\frac{2}{9}\log_2\frac{12}{9}\right)\right)$$
$$d = (3 * 0.1111 * -05849) + (3 * 0.2222 * 0.4150)$$
$$d = -0.1950 + 0.2767$$
$$d = 0.0817$$

c. Second Dice w.r.t third dice:

$$d = \frac{1}{9}\log_2\frac{1/9}{1/9} + \frac{1}{9}\log_2\frac{1/9}{1/9} + \frac{1}{9}\log_2\frac{1/9}{2/9} + \frac{2}{9}\log_2\frac{2/9}{2/9} + \frac{2}{9}\log_2\frac{2/9}{2/9} + \frac{2}{9}\log_2\frac{2/9}{1/9}$$
$$d = \left(\frac{6}{9}\log_2 1\right) + \left(\frac{2}{9}\log_2 2\right) + \left(\frac{1}{9}\log_2\frac{1}{2}\right)$$
$$d = (0.6667 * 0) + (0.2222 * 1) + (0.1111 * -1)$$
$$d = 0.1111$$