Student Information

Full Name : Mustafa Ozan ALPAY
Id Number : 2309615

Exercise 9.10

Question: 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:

(a)

First, we need to find the sample proportion;

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

The confidence interval can be found by the following formula;

$$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \tag{1}$$

Since we are interested in a confidence interval of 96%, $\alpha = 1 - 0.96$, we can get $\alpha = 0.04$. By looking at the **Student's T Distribution** table, we can get that $z_{\alpha/2} = z_{0.04/2} = 2.054$.
Finally by replacing everything in Formula 1, we get;

$$
\begin{aligned}
&= \hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \\
&= 0.12 \pm 2.054 \cdot \sqrt{\frac{0.12(1-0.12)}{200}} \\
&= 0.12 \pm 0.047 \\
&= [0.073, 0.167]
\end{aligned}
$$

(b)

We need to test two hypotheses to prove or disprove these claims. Namely,

- $H_0 : p = 0.1$
- $H_A : p > 0.1$

To disprove the manufacturer's claim, we need to reject H_0 in favour of H_A. By using the two sample Z test,

$$
\begin{aligned}
Z &= \frac{\hat{p} - p_0}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \\
&= \frac{0.12 - 0.1}{\sqrt{\frac{0.1(1-0.1)}{200}}} \\
&= 0.94
\end{aligned}
$$

By using the **Student's T Distribution** table, we get the critical values;

- $z_{0.04} = 1.751$
- $z_{0.15} = 1.036$

Since the z values are greater than the calculated Z value, we cannot say that we have significant evidence to disprove the manufacturer's claim at 4% and 15% levels of confidence.

Exercise 9.12

Question: An electronic parts factory produces resistors. Statistical analysis of the output suggests that resistances follow an approximately Normal distribution with a standard deviation of 0.2 ohms. A sample of 52 resistors has the average resistance of 0.62 ohms.

(a) Based on these data, construct a 95% confidence interval for the population mean resistance.

(b) If the actual population mean resistance is exactly 0.6 ohms, what is the probability that an average of 52 resistances is 0.62 ohms or higher?

Solution:

(a)

To calculate the confidence interval, we can use the following formula;

$$\begin{aligned} &= \overline{X} \pm t_{\alpha/2} \frac{s}{\sqrt{n}} \\ &= 0.62 \pm t_{0.05/2} \frac{0.2}{\sqrt{52}} \end{aligned}$$

where $t_{0.05/2}$ is equal to 1.96 by using the **Student's T Distribution** table.
The calculation gives us the confidence interval; $[0.57, 0.67]$.

(b)

To calculate the probability, we can use the following formula;

$$\begin{aligned} P(\overline{X} > 0.62) &= P\left(Z > \frac{0.62 - 0.6}{0.2\sqrt{52}}\right) \\ &= P(Z > 0.72) \end{aligned}$$

By using the **Standard Normal Distribution** table, we can get that $P(Z > 0.72) = 0.7642$.
Therefore $P(\overline{X} > 0.62) = 1 - 0.7642 = 0.2358$

Exercise 10.3

Question: The following sample is collected to verify the accuracy of a new random number generator (it is already ordered for your convenience).

-2.434	-2.336	-2.192	-2.010	-1.967	-1.707	-1.678	-1.563	-1.476	-1.388
-1.331	-1.269	-1.229	-1.227	-1.174	-1.136	-1.127	-1.124	-1.120	-1.073
-1.052	-1.051	-1.032	-0.938	-0.884	-0.847	-0.846	-0.716	-0.644	-0.625
-0.588	-0.584	-0.496	-0.489	-0.473	-0.453	-0.427	-0.395	-0.386	-0.386
-0.373	-0.344	-0.280	-0.246	-0.239	-0.211	-0.188	-0.155	-0.149	-0.112
-0.103	-0.101	-0.033	-0.011	0.033	0.110	0.139	0.143	0.218	0.218
0.251	0.261	0.308	0.343	0.357	0.463	0.477	0.482	0.489	0.545
0.590	0.638	0.652	0.656	0.673	0.772	0.775	0.776	0.787	0.969
0.978	1.005	1.013	1.039	1.072	1.168	1.185	1.263	1.269	1.297
1.360	1.370	1.681	1.721	1.735	1.779	1.792	1.881	1.903	2.009

(a) Apply the χ^2 goodness-of-fit test to check if this sample comes from the Standard Normal distribution.

(b) Test if this sample comes from the Uniform(-3,3) distribution.

(c) Is it theoretically possible to accept both null hypotheses in (a) and (b) although they are contradicting to each other? Why does it make sense?

Solution:

(a)

To apply the χ^2 goodness-of-fit test, first we need to calculate the mean of the generated values, and the standard deviation.

- $\overline{X} = -0.05773$
- $\sigma = 1.052660628$

Ideally, distributing between 5 to 8 bins is enough for a Chi-Square test. I decided to use 10 bins, which would give me a degree of freedom 9. However if a group has less than 5 counts in a bin, this group should be merged with another one. Because of that, in the end I used 8 bins, with a degree of freedom 7.

By calculating the areas under Normal Distribution, we can get the expected values for the intervals. (An online calculator is available at http://onlinestatbook.com/2/calculators/normal_dist.html)

The distribution of bins, expected values and Chi-Square results can be summarized in the tables below;

Interval	Observed	Expected	Expected for 100	χ^2
$x < -2.0$	4	0.0325	3.25	0.173076923
$-2.0 < x < -1.5$	4	0.0528	5.28	0.31030303
$-1.5 < x < -1.0$	15	0.1	10	2.5
$-1.0 < x < -0.5$	9	0.1518	15.18	2.515968379
$-0.5 < x < 0.0$	22	0.1847	18.47	0.674656199
$0.0 < x < 0.5$	15	0.18	18	0.5
$0.5 < x < 1.0$	12	0.1406	14.06	0.301820768
$1.0 < x < 1.5$	11	0.088	8.8	0.55
$1.5 < x < 2.0$	7	0.0442	4.42	1.505972851
$2.0 < x$	1	0.0253	2.53	0.925256917
Total	100	1	100	9.957055068

Table 1: Initial Chi-Square Test

Interval	Observed	Expected	Expected for 100	χ^2
$x < -1.5$	8	0.0853	8.53	0.032931
$-1.5 < x < -1.0$	15	0.1	10	2.5
$-1.0 < x < -0.5$	9	0.1518	15.18	2.515968379
$-0.5 < x < 0.0$	22	0.1847	18.47	0.674656199
$0.0 < x < 0.5$	15	0.18	18	0.5
$0.5 < x < 1.0$	12	0.1406	14.06	0.301820768
$1.0 < x < 1.5$	11	0.088	8.8	0.55
$1.5 < x$	8	0.0695	6.95	0.158633
Total	100	1	100	7.234009

Table 2: Chi-Square Test with two bins merged with others

From the table, we can see that the summed Chi-Square value is 7.23. If we look at the Chi-Square Distribution table, on the row of $v = 7$, we can see that our value falls between $\alpha = 0.2$ and $\alpha = 0.8$, which is greater than the significance value of 0.05. Therefore we can conclude that we couldn't find any significant evidence that against our sample coming from the Standard Normal Distribution.

(b)

Similarly, we can generate the Chi-Square table by using the Expected values of the Uniform Distribution.

Interval	Observed	Expected	Expected for 100	χ^2
$-3.0 < x < -1.5$	8	0.25	25	11.56
$-1.5 < x < -1.0$	15	0.083333333	8.333333333	5.333333333
$-1.0 < x < -0.5$	9	0.083333333	8.333333333	0.053333333
$-0.5 < x < 0.0$	22	0.083333333	8.333333333	22.41333333
$0.0 < x < 0.5$	15	0.083333333	8.333333333	5.333333333
$0.5 < x < 1.0$	12	0.083333333	8.333333333	1.613333333
$1.0 < x < 1.5$	11	0.083333333	8.333333333	0.853333333
$1.5 < x < 3.0$	8	0.25	25	11.56
Total	100	1	100	58.72

Table 3: Chi-Square Test for Uniform Distribution with merged bins

From the table, we can see that the summed Chi-Square value is 58.72. Since our degree of freedom is 8, if we look at the Chi-Square Distribution table, on the row of $v = 7$, we can see that our value falls below $\alpha = 0.001$, which is way less than the significance value of 0.05. Therefore we can conclude that there is strong evidence that our sample does not follow Uniform Distribution.

(c)

According to the central limit theorem, with very large sample sizes, it is possible to have samples following both Uniform and Standard Normal Distributions; however it is not the case with our sample size. By selecting a significance value that satisfies both hypotheses, we can accept them both. This clearly shows the importance of selecting the p values properly.

Exercise 10.9

Question: The Probability and Statistics course has three sections - S01, S02, and S03. Among 120 students in section S01, 40 got an A in the course, 50 got a B, 20 got a C, 2 got a D, and 8 got an F. Among 100 students in section S02, 20 got an A, 40 got a B, 25 got a C, 5 got a D, and 10 got an F. Finally, among 60 students in section S03, 20 got an A, 20 got a B, 15 got a C, 2 got a D, and 3 got an F. Do the three sections differ in their students' performance?

Solution:
To calculate the expectations for each grade in the sections, we can use the following formula;

$$\widehat{Exp}(i,j) = \frac{n_i \cdot n_j}{n}$$

By calculating each expected value and Chi-Square values, we get the following table;

Sections / Grades		A	B	C	D	F	Total
Section 1	Observed	40	50	20	2	8	120
	Expected	34.2857	47.1429	25.7143	3.8571	9	120
	χ^2	0.9524	0.1732	1.2698	0.8942	0.1111	3.4007
Section 2	Observed	20	40	25	5	10	100
	Expected	28.5714	39.2857	21.4286	3.2143	7.5	100
	χ^2	2.5714	0.0130	0.5952	0.9921	0.8333	5.0051
Section 3	Observed	20	20	15	2	3	60
	Expected	17.1429	23.5714	12.8571	1.9286	4.5	60
	χ^2	0.4762	0.5411	0.3571	0.0026	0.5	1.8771
Total		80	110	60	9	21	280

Table 4: Chi-Square Test for independence

From the table, we can understand that;

- Total $\chi^2 = 10.2828$
- Degrees of freedom $(5-1)\cdot(3-1) = 8$

By looking at the Chi-Square Distribution table, on the row of $v = 8$, we can see that our value falls between $\alpha = 0.20$ and $\alpha = 0.80$, which is way greater than the significance value of 0.05. Therefore we can conclude that there is no significant evidence supports that the three sections perform differently.