Monsoon Semester 2021-22: Tutorial 2

MA 2001D: Mathematics 3

Department of Mathematics
National Institute of Technology
Calicut

Question 37

An examination was given to two classes consisting of 40 and 50 students respectively. For the first class, the mean mark was 74 with a standard deviation of 8, while in the second class, the mean mark was 78 with a standard deviation of 7. Is there a significant difference between the performance of the two classes at a level of significance of

- (a) 0.05
- (b) 0.01

Answer

Sample size: $n_1 = 40$, $n_2 = 50$

Sample mean: $\bar{x_1} = 74$, $\bar{x_1}2 = 78$

Sample standard deviation: $\sigma_1 = 8$, $\sigma_2 = 7$

$H_0 : \mu_1 = \mu 2$. i.e, there is no significant difference between the performance of the two classes.

$H_1 : \mu_1 \neq \mu 2$. i.e, there is significant difference between the performance of the two classes.

$$z = \frac{\bar{x_1} - \bar{x_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{74 - 78}{\sqrt{\frac{8^2}{40} + \frac{7^2}{50}}} = \frac{-4}{1.606} = -2.4907$$

Answer 37 cont.

- (a) Tabulated value of z at 5% level of significance is 1.96.

  i.e, $z_{\alpha/2} = 1.96 < |z|$

  Therefore, H_0 is accepted and hence there is no significant difference in the performance of two classes.

- (b) Tabulated value of z at 1% level of significance is 2.58.

  i.e, $z_{\alpha/2} = 2.58 > |z|$

  Therefore, H_0 is rejected and hence there is a significant difference in the performance of two classes.

Question 38

If $\{X_1, X_2, ..., X_n\}$ is a random sample of size n from Bernoulli distribution with parameter p, obtain the MLE for p.

Answer

The likelihood function is

$$L(x_1, x_2, ..., x_n/p) = p^{x_1}(1-p)^{1-x_1}p^{x_2}(1-p)^{1-x_2}.....p^{x_n}(1-p)^{1-x_n}$$
$$= p^{n\bar{x}}(1-p)^{n-n\bar{x}}$$

To determine p that maximizes the likelihood function:
Taking log on both sides
$logL = n\bar{x}logp + (n - n\bar{x})log(1-p)$
MLE is the solution of $\frac{\partial logL}{\partial p} = 0$ and $\frac{\partial^2 logL}{\partial p^2} < 0$

Question 39

Let $\{X_1, X_2, \ldots, X_n\}$ be a random sample of size n taken from a Normal population with mean μ and variance σ^2.
(a) If μ is known, obtain the MLE for σ^2.
(b) If σ^2 is known, obtain the MLE for μ.

Answer

(a) The likelihood function is

$$L\left(\sigma^2 \mid x_1, \ldots, x_n\right) = \prod_{i=n}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_i-\mu)^2/2\sigma^2} = \frac{1}{\left(2\pi\sigma^2\right)^{n/2}} e^{-\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^2}}$$

The function $z^b e^{-(cz)}$ has a maximum at $z = \mathrm{b}/\mathrm{c}$ for $\mathrm{z} \geq 0$ when b and c are positive. Taking $\mathrm{z} = \frac{1}{\sigma^2},\ \mathrm{b} = \frac{n}{2}$, and $\mathrm{c} = \sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2}$, we obtain the maximum likelihood estimator $\widehat{\sigma^2} = \sum_{i=1}^{n} \frac{(x_i-\mu)^2}{n}$, when μ is known.

(b) We have $(x_i-\mu)^2=(x_i-\bar{x}+\bar{x}-\mu)^2$, the likelihood function is

$$
\begin{aligned}
L(\mu \mid x_1,\ldots,x_n) &= \prod_{i=n}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_i-\mu)^2/2\sigma^2} \\
&= e^{-n(\bar{x}-\mu)^2/2\sigma^2}\left[\frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\sum_{i=1}^{n}\frac{(x_i-\bar{x})^2}{2\sigma^2}}\right]
\end{aligned}
$$

This likelihood is maximized overall values of μ when the exponent $n(\bar{x}-\mu)^2/2\sigma^2$ is minimized. Therefore, the MLE $\widehat{\mu}=\bar{x}$

Question 40

Let X be the random variable with p.d.f
$f(x) = \frac{1}{\theta^2} x e^{-\frac{x}{\theta}}, 0 < x < \infty, 0 < \theta < \infty$. Given a random sample of size $n, \{X_1, X_2, \ldots, X_n\}$, find the MLE for θ.

Answer

We first obtain the likelihood by multiplying the probability density function for each X_i. We then simplify this expression.

$$L(\theta) = \prod_{i=1}^{n} f(x_i; \theta) = \prod_{i=1}^{n} \theta^{-2} x_i e^{-x_i/\theta} = \theta^{-2n} \left(\prod_{i=1}^{n} x_i \right) \exp\left(\frac{-\sum_{i=i}^{n} x_i}{\theta} \right)$$

Instead of directly maximizing the likelihood, we instead maximize the log-likelihood.

$$\log L(\theta) = -2n \log \theta + \sum_{i=i}^{n} \log x_i - \frac{\sum_{i=i}^{n} x_i}{\theta}$$

Answer Continues

To maximize this function, we take a derivative with respect to θ.

$$\frac{d}{d\theta}\log L(\theta) = \frac{-2n}{\theta} + \frac{\sum_{i=i}^{n} x_i}{\theta^2}$$

We set this derivative equal to zero, then solve for θ.

$$\frac{-2n}{\theta} + \frac{\sum_{i=i}^{n} x_i}{\theta^2} = 0$$

Solving gives our estimator, which we denote with a hat.

$$\hat{\theta} = \frac{\sum_{i=i}^{n} x_i}{2n} = \frac{\bar{x}}{2}.$$

Question 41

An urn contains red and white marbles in an unknown proportion. A random sample of 60 marbles selected with replacement from the urn showed that 70% were red. Find: (i) 95% and 99% confidence limits for the actual proportion of red marbles in the urn.
(ii) size of the sample of marbles in order to be 95% and 99% confident that the true sample proportions do not differ more than 5%.

Answer

(i)Here 95% confidence limit is given by
$\bar{p} \pm z_{\alpha/2}\sqrt{\frac{\bar{p}\bar{q}}{n}}$
From the table $z_{\alpha/2} = 1.96$
$\bar{p} = \frac{70}{100}$; n=60 ; $\bar{q} = \frac{30}{100}$
$\therefore$ confidence limit is $0.7 \pm 1.96\sqrt{\frac{0.7x0.3}{60}}$
Similarly 99% confidence limit is given by $0.7 \pm 2.576\sqrt{\frac{0.7x0.3}{60}}$
$(z_{\alpha/2} = 2.576)$
(ii) Size of sample n is given by $n = (\frac{z_{\alpha/2}}{E})^2\bar{p}(1-\bar{p})$, where error E= $|p - \bar{p}| = 0.05$
$\therefore n = (\frac{1.96}{0.05})^2 0.7x0.3$ in order to be 95% confident
Similarly $n = (\frac{2.576}{0.05})^2 0.7x0.3$ in order to be 99% confident

Question 42

The standard deviation of the breaking strengths of 10 cables tested by a company was 815 kilograms. Find 95%, and 99% confidence limits for the standard deviation of all cables produced by the company.

Answer

S = 815 Kg   n = 10

95 % CI for $\sigma^2 = (\frac{nS^2}{u_2}, \frac{nS^2}{u_1})$

$P(u_1 < \chi^2_9 < u_2) = 0.95$

$P(u_1 < \chi^2 < u_2) = P[\chi^2 > u_1] - P[\chi^2 > u_2]$

$P[\chi^2 > u_1] = 0.975 \implies P[\chi^2 > 2.7] = 0.975$

(from table of χ^2 distribution)

$P[\chi^2 > u_2] = 0.025 \implies P[\chi^2 > 19.023] = 0.023$

(from table of χ^2 distribution)

$P(2.7 < \chi^2_9 < 19.023) = 0.95$

CI for $\sigma^2 = (\frac{10\times815^2}{19.023}, \frac{10\times815^2}{2.7}) = (349169.43, 2460092.59)$ therefore CI for $\sigma = (\sqrt{349169.43}, \sqrt{2460092.59}) = (590.91, 1568.47)$

Conti...

Similarly,
99 % CI for $\sigma^2 = (\frac{nS^2}{u_2}, \frac{nS^2}{u_1})$
$P(u_1 < \chi^2_9 < u_2) = 0.99$
$P(1.735 < \chi^2_9 < 23.589) = 0.99$
95% CI $\sigma^2 = (\frac{10\times 815^2}{23.589}, \frac{10\times 815^2}{1.735}) = (281582.517, 3828386.167)$
therefore CI for $\sigma = (\sqrt{281582.517}, \sqrt{3828386.167} = (530.64, 19956.63)$

Question 44

A manufacturer claimed that at least 95% of the equipment, which he supplied to a factory, confirmed to specifications. An examination of a sample of 200 pieces of equipment revealed that 18 were faulty. Test his claim at a significance level of (a) 1% and (b) 5%.

Answer

$H_0 : p \geq 0.95$ (claim)

$H_1 : p < 0.95$

$\hat{p} = \frac{182}{200} = 0.91$

$z = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{0.91-0.95}{\sqrt{\frac{0.95\times0.05}{200}}} = -2.5955$

p-value $=$ P$(z < -2.5955) = 0.0047$

(a) Since p-value is less than 0.01, reject H_0; reject claim.

(b) Since p-value is less than 0.05, reject H_0; reject claim.

Question 46

A usual complaint of interactive system users is the large variance of the response time. While contemplating the purchase of a new interactive computer system, we measure 30 random samples of response time, and the sample variance is computed to be 25 ms^2 . Assuming the response times are approximately normally distributed, find a 95% confidence interval for the population variance.

Answer

We have $S^2 = 25$, $n = 30$, $1 - \alpha = 0.95$. Then

$$P\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2,(n-1)}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,(n-1)}}\right) = 0.95.$$

The $(1-\alpha)100\%$ confidence interval for σ^2 is

$$\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2,(n-1)}}, \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,(n-1)}}\right)$$

We have $\alpha = 0.05$. From the table, $\chi^2_{0.05/2,(30-1)} = 45.722$ and $\chi^2_{1-0.05/2,(30-1)} = 16.047$.
Thus the 95% confidence interval for σ^2 is $(15.857, 45.179)$.

Question 47

Execution times (in seconds) of 30 jobs processed by a computing centre were measured
and found to be:

and found to be:	10	19	90	40	15	11	32	17	4	152	23
	13	36	101	2	14	2	23	34	15	27	1
	57	17	3	30	50	4	62	48			

Find the 90% confidence intervals for the mean and the variance of execution time of a job. Assume that the execution time is approximately normally distributed.

Answer

From the data, we calculate $\sum x_i = 952$ and $\sum x_i^2 = 63090$. Hence

$$(n-1)S^2 = 63090 - \frac{952^2}{30} = 32879.8666667$$

(S^2 is the sample variance)
There are $n - 1 = 29$ degrees of freedom and the critical values of the χ^2_{29} -distribution are

$$\chi^2_{0.95,19} = 17.71 \quad \text{and} \quad \chi^2_{0.05,19} = 42.56$$

Note

1. If $x_1, x_2, x_3, \cdots, x_n$ is a random sample with variance S^2 taken from a normal population with variance σ^2 then a $100(1-\alpha)\%$ confidence interval for σ^2 is

$$\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}$$

   where $\chi^2_{\alpha/2,n-1}$ and $\chi^2_{1-\alpha/2,n-1}$ are the appropriate right-hand and left-hand values respectively of a chi-squared distribution with $n-1$ degrees of freedom.

2. The formula for a $100(1-2\alpha)\%$ confidence interval for one population mean in this case is

$$\bar{x} - \mathrm{t}^*_{n-1,\ \alpha}\frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + \mathrm{t}^*_{n-1,\ \alpha}\frac{s}{\sqrt{n}},$$

   where $t^*_{n-1,\ \alpha}$ is the critical t^* -value from the t -distribution with $n-1$ degrees of freedom (where n is the sample size).

- The confidence interval for σ^2 is

$$\frac{32879.87}{42.56} < \sigma^2 < \frac{32879.87}{17.71} \equiv 772.5532 \text{ Sec} < \sigma^2 \leq 1856.5706 \text{ Sec}$$

- The confidence interval for μ is

$$31.733 - 1.699 \cdot \frac{33.67177}{\sqrt{30}} \leq \mu \leq 31.733 + 1.699 \cdot \frac{33.67177}{\sqrt{30}}$$

Question 48

In the past the standard deviation of the response time to edition commands of a timesharing system was 25 milliseconds and the mean response time was 400 milliseconds. A new version of the operating system was installed, and with this random sample of 21 editing commands experienced a standard deviation of response times of 32 milliseconds. Is this increase in variability significant at a 5% level of significance? Is it significant at a 1% level of significance?

Answer

We wish to test

$$H_0 : \sigma^2 = (25)^2 \quad \text{versus} \quad \sigma^2 > (25)^2$$

The observed value of the chi-square statistic is:

$$\frac{(n-1)s^2}{\sigma_0^2} = \frac{20(32)^2}{(25)^2} = 32.8$$

Since $\chi^2_{20;0.05} = 31.41$, we conclude at the 5 percent level of significance that the new version of the system is unfair to time-sharing users. On the other hand, since $\chi^2_{20,001} = 37.566$, we cannot reject H_0 at the 1 percent level of significance.

Question 50

The manufacturer of a patent medicine claimed that it was 90% effective in relieving an allergy for a period of 8 hours. In a sample of 200 people who had allergy, the medicine provided relief for 160 people. Determine whether the manufacturer's claim is legitimate.

Answer

In test of significance of a sample proportion, population is divided into two mutually disjoint classes representing the qualitative characteristic (attribute) in such a way that one possesses a particular attribute (i.e. success) and other does not possess that attribute (i.e. failure)
To test whether there is any significant difference between sample proportion and the population proportion we set up the null hypothesis as

$$H_0 : P = P_0$$

i.e. the population proportion has a specified value P_0 or the sample has been drawn from the population with proportion P_0 or there is NO significant difference between the sample proportion (p) and population proportion (P) against

$$H_1 : P \neq P_0$$

i.e. the population proportion value is not P_0 or the sample has NOT been drawn from the population with proportion P_0 or there is a significant difference between the sample proportion (p) and population proportion (P)

Or

$$H_1 : P < P_0$$
$$H_1 : P > P_0$$

Test Statistic: Under H_0 the test statistic is

$$Z = \frac{Difference}{S.E(p)} = \frac{p - P}{\sqrt{\frac{P(1-P)}{n}}}$$

where
p = observed sample proportion of success = $\frac{x}{n}$
x = number of success possessing the given attribute
n = sample size or number of trials
P = population proportion of success
Q = population proportion of failure such that P+Q=1

- If the sample proportion is less than the population proportion i.e. if $p < P$ we assume left tailed test
- If the sample proportion is greater than the population proportion i.e. if $p > P$ we assume right tailed test

Let p denote the probability of obtaining relief from the allergy by using the medicine.

Then we must decide between the two hypothesis

$$H_0 : \text{p} = 0.9, \text{ and the claim is correct}$$
$$H_1 : \text{p} < 0.9, \text{ and the claim is false}$$

We choose a one tailed test since we are interested in determining whether the proportion of people relieved by the medicine is too low.

$$\text{Sample Proportion p} = \bar{x} = \frac{x}{n} = \frac{160}{200} = 0.8$$

$$\text{Test Statistic}$$
$$Z = \frac{0.8 - 0.9}{\sqrt{\frac{0.9(1-0.9)}{200}}} = -4.715$$

Considering α =0.05 the critical point is -1.96 and hence z = -4.715 lies in the critical region
Therefore we reject Ho.
The manufacturer's claim is not supported by the test results.

Question 52

A sample poll of 300 voters from district A and 200 voters from district B showed that 56% and 48% respectively were in favor of a given candidates. At 5% levels of significance test the hypothesis that (a) there is a difference between the districts, (b) the candidate is preferred in district A.

Answer

Given $n_1 = 300,\ n_2 = 200,\ \hat{p}_1 = 0.56,\ \hat{p}_2 = 0.48$
$\hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1+n_2} = 0.52$
Then $Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = 1.7541$
(a)$H_0 : P_1 = P_2,\ H_a : P_1 \neq P_2$
Since H_a contains $\neq$ we have to use two tailed test.
$\alpha = 0.05 \Rightarrow Z_{\frac{\alpha}{2}} = 1.96.$
Since calculated Z-value 1.7541<1.96, we fail to reject H_0.
(b)$H_0 : P_1 = P_2,\ H_a : P_1 > P_2$
Since H_a contains $>$ we have to use right tailed test.
$\alpha = 0.05 \Rightarrow Z_{\alpha} = 1.645.$
Since calculated Z-value 1.7541>1.645, we reject H_0.

Question 53

Random samples of 200 bolts manufactured by a machine A and 100 bolts manufactured by machine B showed 19 and 5 defective bolts respectively. Test the hypothesis that (a) the two machines are showing different qualities of performances and, (b) Machine B is performing better than Machine A. Use 0.05 level of significance.

Given $n_1 = 200,\ n_2 = 100,\ x_1 = 19,\ x_2 = 5.$
$\hat{p_1} = \frac{x_1}{n_1} = 0.095,\ \hat{p_2} = \frac{x_2}{n_2} = 0.05$
$\hat{p} = \frac{x_1+x_2}{n_1+n_2} = 0.08$
Then $Z = \frac{\hat{p_1}-\hat{p_2}}{\sqrt{\hat{p}(1-\hat{p})}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = 1.3543.$
(a)$H_0 : P_1 = P_2,\ H_a : P_1 \neq P_2$
Since H_a contains $\neq$ we have to use two tailed test.
$\alpha = 0.05 \Rightarrow Z_{\frac{\alpha}{2}} = 1.96.$
Since calculated Z-value 1.3543<1.96, we fail to reject H_0.

Question 55

Eight- meter sections of copper wire produced at two different plants have a maximum resistance specification of 1 ohm. Historically, the defective rates at each plant have been 10% at plant A and 11% at plant B, not meeting this specification. Recent samples of 150 and 100 wires at plant A and B respectively each had 12 wires that did not meet the specification.
(a) Are the defective rates equal?
(b) If the historical defective rates are neglected, is the decision in part (a) different?

Answer

If p_1 and p_2 denote the defective rates of the plants A and B, then we have to decide between the hypothesis ,
$H_0 : p_1 = p_2$, the defective rates are equal and $H_1 : p1 \neq p_2$, there is a significant difference between the defective rates.
Here we use an estimate of p, the average defective rate of two sample plants given by $p = \dfrac{(12+12)}{250} = .096$ and $q = .904$ Then under the hypothesis H_0,
$\mu_{p_1 - p_2} = 0$

Answer

$\sigma_{p_1-p_2} = \sqrt{pq\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} = \sqrt{.096 * .904 * \left(\frac{1}{50}+\frac{1}{100}\right)}$
$=.038301$
Then $z = \dfrac{p_1 - p_2}{\sigma_{p_1-p_2}} = \frac{.04}{.038031} = 1.0517$ On the basis of a two-tailed test at a 0.05 level of significance, we would reject the hypothesis H_0 only if the Z score were greater than 1.96. Since the Z score is only 1.05, we must conclude that the results are due to chance at this level of significance. On the basis of a two-tailed test at a 0.01 level of significance, we would reject the hypothesis H_0 only if the Z score were greater than 2.576. Since the Z score is only 1.05, we must conclude that the results are due to chance at this level of significance.

Question 56

The following table gives the observed values for weekly defects by three shifts and two production lines. Use this data to determine if there is any relationship between the shift and the production line.

Line	Shift		
	1	2	3
1	10	12	13
2	14	9	12

Answer

Observed Frequency (f_o)

Line	Shift			
	1	2	3	Total
1	10	12	13	35
2	14	9	12	35
Total	24	21	25	70

Answer Contd

Expected Frequency (f_e)

Line	Shift			
	1	2	3	Total
1	$\frac{35\times24}{70}=12$	$\frac{35\times21}{70}=10.5$	$\frac{35\times25}{70}=12.5$	35
2	$\frac{35\times24}{70}=12$	$\frac{35\times21}{70}=10.5$	$\frac{35\times25}{70}=12.5$	35
Total	24	21	25	70

Define Null and Alternating Hypothesis

H_0: The values for weekly defects by three shifts and two production lines are not related.

H_a: There is a significant relation between the values for weekly defects by three shifts and two production lines are related.

Calculate Test statistic

$$\chi^2_{cal} = \sum \frac{(f_e-f_o)^2}{f_e}$$
$$= \frac{(12-10)^2}{12} + \frac{(10.5-12)^2}{10.5} + \frac{(12.5-13)^2}{12.5} + \frac{(12-14)^2}{12} + \frac{(10.5-9)^2}{10.5} + \frac{12.5-12)^2}{12.5}$$
$$= 0.3333 + 0.2143 + 0.02 + 0.3333 + 0.2143 + 0.02$$
$$= 1.1352.$$

Answer Contd

If $\chi^2_{cal} > \chi^2_{\alpha}$, We reject null Hypotheses.
Here df=(3-1)*(2-1)=2. That is, $1.1352 > \chi^2_{\alpha,df}$ gives a $\alpha = 0.5669$ significant relation between the values for weekly defects by three shifts and two production lines are related.

Question 57

The personnel director at your company has asked you to verify the statement that absenteeism is twice as bad on Mondays as the rest of the week. You are given personnel records for 3 months covering 890 days of lost work. Test the statement at the 5% level.

Day	Monday	Tuesday	Wednesday	Thursday	Friday
Days lost	304	176	139	141	130

Solution

Day	Monday	Tuesday	Wednesday	Thursday	Friday
Obs. Frequency	304	176	139	141	130
Exp.Frequency	296.67	148.33	148.33	148.33	148.33

Total lost work days $= 890$

$\frac{890}{6} = 148.33$

H_0 : Absenteeism is twice as bad on Monday as the rest of the week

H_1 : Absenteeism is not twice as bad on Monday as the rest of the week

ie, H_0 : Expected frequency on Monday $= 2 \times 148.33 = 296.67$

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$
$$= \frac{53.729}{296.67} + \frac{765.629 + 87.049 + 53.729 + 335.929}{148.33} = 8.556$$

Degrees of freedom $= n - 1 = 4$ and $\chi^2_{(4,0.05)} = 9.49$

$\chi^2 = 8.556 < 9.49 = \chi^2_\alpha$

Hence, accept H_0.

Question 59

The response time T of a circuit activator varies between 0.01 and 0.06in discrete 0.01 units. The observed frequencies in 100 tests are given. Using level of significance of 10% determine if the observed data follows a discrete uniform distribution

Response time t	0.01	0.02	0.03	0.04	0.05	0.06
Frequency	18	18	15	14	19	16

Solution

Response time t	0.01	0.02	0.03	0.04	0.05	0.06
Observed Frequency	18	18	15	14	19	16
Expected Frequency	16.67	16.67	16.67	16.67	16.67	16.67

$N = 100, n = 6, \alpha = 0.1$
$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$
$= \frac{1}{16.67}((18 - 16.67)^2 + \cdots + (16 - 16.67)^2) = \frac{19.3334}{16.67} = 1.1597$
$(\chi^2)_(\alpha = 0.1, n - 1 = 5) = 9.24$
$\therefore \chi^2 < (\chi^2)_(\alpha)$
Hence the observed data follows discrete uniform distribution.

Question 61

The number of flaws per $100m^2$ of cloth is to be checked by a quality control chart which assumes an underlying Poisson population. For the last 75 samples the number of flaws have been recorded as under:

Number of flaws	0	1	2	3	4	5	6
Frequency	20	30	15	7	2	1	0

If the level of significance is 5%,is the Poisson population assumption is a reasonable one?

Solution

To determine the expected poisson frequencies we have to estimate λ using the sample data, as

$$\bar{\lambda} = \bar{x} = \frac{\sum f_i x_1}{N}$$

$$= \frac{0\times20+1\times30+2\times15+3\times7+4\times2+5\times1+6\times0}{75} = \frac{94}{75} = 1.253$$

The Poisson frequency function is given by

$$f(x) = \frac{Ne^{-\lambda}\lambda^x}{x!}, x : 0, 1, 2 \ldots 6$$

$f(0) = 75 \times \frac{e^{-1.253}(1.253)^0}{0!} = 21.42 \approx 21$

$f(1) = 75 \times \frac{e^{-1.253}(1.253)^1}{1!} = 26.84 \approx 27$

$f(2) = 75 \times \frac{e^{-1.253}(1.253)^2}{2!} = 16.81 \approx 17$

$f(3) = 75 \times \frac{e^{-1.253}(1.253)^3}{3!} = 7.02 \approx 7$

$f(4) = 75 \times \frac{e^{-1.253}(1.253)^4}{4!} = 2.2 \approx 2$

$f(5) = 75 \times \frac{e^{-1.253}(1.253)^5}{5!} = 0.55 \approx 1$

$f(6) = 75 \times \frac{e^{-1.253}(1.253)^6}{6!} = 0.11 \approx 0$

Here we set up the hypothesis as

$\mathrm{H}_0 : \mathrm{PD}$ is a good fit to the giver data against

$\mathrm{H}_1 : \mathrm{PD}$ is not a good fit to the given data.

The test statistic is $\chi^2 = \sum_{i=1}^{n} \frac{(\mathrm{O_i - E_i})^2}{\mathrm{E}_i}$

O_i	E_i	$\mathrm{O}_i - \mathrm{E}_i$	$(\mathrm{O}_i - \mathrm{E}_i)^2$	$(\mathrm{O_i} - \mathrm{E_i})^2 / \mathrm{E_i}$
20	21	-1	1	0.037
30	27	3	9	0.333
15	17	-2	4	0.235
10	10	0	0	0
75	75			0.605

$\therefore \chi^2 = 0.605$ (Not that the last 3 cells are pooled)
The table value χ^2 for $7 - 1 - 3 - 1 = 2$ df (1 df for $\Sigma O_i = \Sigma E_i$, 1 df for estimating λ, 3 df for pooling 3 cell frequencies) and probability $\alpha = 0.05$ is 5.991. Since the calculated value is less than the table value, H_0 is accepted. That is, there is a great agreement with the observed and theoretical frequencies, which says that PD is a good fit to the given data.

Question 62

The number of taxis waiting at a cab stand is thought to follow a uniform distribution. The number present was recorded at 50 random times:

Number of taxis	0	1	2	3	4	5	6	7	8
Frequency	4	6	8	5	7	6	6	5	3

Use the appropriate goodness of fit test to determine if the uniform distribution is a good choice at the 5% significance level.

Solution

Number of taxis	0	1	2	3	4	5	6	7	8
Observed Frequency	4	6	8	5	7	6	6	5	3
Expected Frequency	5.55	5.55	5.55	5.55	5.55	5.55	5.55	5.55	5.55

$N = 50, n = 8, \alpha = 0.05$
$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} = 3.151$
$(\chi^2)_{(\alpha=0.05, n-1=8)} = 7.22$
$\therefore \chi^2 < (\chi^2)_{(\alpha)}$
Hence the uniform distribution is a good choice at the 5% significance level.

Question 68

The table shows the relation between the performance of students in Mathematics and Physics. Test the hypothesis that the performance in Physics is independent of performance in Mathematics, using a

1. 0.05 and
2. 0.01 significance level.

		MATHEMATICS		
		High Grades	Medium Grades	Low Grades
PHYSICS	High Grades	23	42	15
	Medium Grades	78	37	37
	Low Grades	61	25	25

Solution 68

Observed frequency:

		Grades in Mathematics			
		High	Medium	Low	Total
Grades in Physics	High	23	42	15	80
	Medium	78	37	37	152
	Low	61	25	25	111
	Total	162	104	77	343

Expected frequency:

		Grades in Mathematics			
		High	Medium	Low	Total
Grades in Physics	High	$\frac{80\times162}{343}=38$	$\frac{80\times104}{343}=24$	80-38-24 =18	80
	Medium	72	46	34	152
	Low	52	34	25	111
	Total	162	104	77	343

Solution 68 cont.

Step 1:

Null Hypothesis (H_0): The performance in Mathematics is independent of the performance in Physics.

Alternative Hypothesis : The performance in Mathematics is not independent of the performance in Physics.

Step 2 :

L.O.S : 5%

Degree of freedom $= (r-1)(c-1) = (3-1) \times (3-1) = 4$

Hence, Critical value $(X_\alpha^2) = 9.4877$

Solution 68 cont.

Observed frequency (O)	Expected frequency(E)	$X^2 = \frac{(O-E)^2}{E}$
23	38	5.921
42	24	13.5
15	18	0.5
78	72	0.5
37	46	0.195
37	34	0.264
61	52	1.557
25	34	2.382
25	25	0
	Total,X^2_{cal}	24.819

Step 4: Decision :
Since $X^2_{cal} > X^2_{\alpha}$, H_0 is rejected.
Hence , the performance in Mathematics is not independent of the performance in Physics.

Question 69

The results of a survey made to determine whether the age of a driver 21 years of age or older has any effect on the number of automobile accidents in which he is involved are indicated in the table. At a level of significance of (a)0.05 and (b)0.01, test the hypothesis that the number of accidents is independent of the age of the driver. What possible sources of difficulty in sampling techniques as well as other considerations could affect your conclusions?

		Age of the driver				
		21-30	31-40	41-50	51-60	61-70
No of Accidents	0	748	821	786	720	672
	1	74	60	51	66	50
	2	31	25	22	16	15
	more than 2	9	10	6	5	7

Answer

Observed frequency:

		Age of the driver					
		21-30	31-40	41-50	51-60	61-70	total
No of Accidents	0	748	821	786	720	672	3747
	1	74	60	51	66	50	301
	2	31	25	22	16	15	109
	more than 2	9	10	6	5	7	37
	total	862	916	865	807	744	4194

Expected frequency:

		Age of the driver					
		21-30	31-40	41-50	51-60	61-70	total
No of Accidents	0	770	818	773	721	665	3747
	1	62	66	62	58	53	301
	2	22	24	22	21	20	109
	more than 2	7.6	8.1	7.6	7.1	6.6	37
	total	862	916	865	807	744	4194

Hint:
$\frac{3747\times862}{4194} = 770$ $\frac{3747\times916}{4194} = 818$ $\frac{3747\times865}{4194} = 773$ $\frac{3747\times807}{4194} = 721$
$3747 - 770 - 818 - 773 - 721 = 665$

Step 1:
Null Hypothesis (H_0): The number of accidents is independent of the age of the driver.
Alternative Hypothesis : The number of accidents is not independent of the age of the driver.
Step 2 :
case 1
L.O.S : 5%
Degree of freedom $= (r-1)(c-1) = (4-1) \times (5-1) = 12$
Hence, Critical value $(\chi^2_\alpha) = 21.026$. (In χ^2 table value compared to 0.05 and for 12 degrees of freedom)
case 2
L.O.S : 1%
Critical value $(\chi^2_\alpha) = 26.22$. (In χ^2 table value compared to 0.01 and for 12 degrees of freedom)

Observed frequency (O)	Expected frequency(E)	$\chi^2 = \frac{(O-E)^2}{E}$
748	770	0.629
821	818	0.011
786	773	0.219
720	721	0.0013
672	665	0.073
74	62	2.322
60	66	0.545
51	62	1.952
66	58	1.103
50	53	0.169
31	22	3.681
25	24	0.041
22	22	0
16	21	1.190
15	20	1.25

Observed frequency (O)	Expected frequency(E)	$\chi^2 = \frac{(O-E)^2}{E}$
9	7.6	0.258
10	8.1	0.446
6	7.6	0.337
5	7.1	0.621
7	6.6	0.024
	Total, χ^2_{cal}	14.872

Step 4: Decision :

Case 1:L.O.S 5%

Since $\chi^2_{cal} < \chi^2_{\alpha}$, We fail to reject H_0.

Hence, the number of accidents is independent of the age of the driver.

Case 2:L.O.S 1%

Since $\chi^2_{cal} < \chi^2_{\alpha}$, We fail to reject H_0.

Hence, the number of accidents is independent of the age of the driver.

Question 70

The owner of a machine shop must decide which of two snack-vending machines to install in his shop. If each machine is tested 250 times, the machine fails to work (neither delivers the snack nor returns the money) 13 times, and the second machine fails to work 7 times, test at the 0.05 level of significance whether the difference between the corresponding sample proportions is significant, using the χ^2 statistic.

Answer

$H_0 : p_1 = p_2$
$H_1 : p_1 \neq p_2$
Consider the table:

	Machine 1	Machine 2	Total
No. of success	237	243	480
No. of failures	13	7	20
Total	250	250	500

$$e_{11} = \frac{480 * 250}{500} = 240$$

$$e_{12} = \frac{480 * 250}{500} = 240$$

$$e_{21} = \frac{20 * 250}{500} = 10$$

$$e_{22} = \frac{20 * 250}{500} = 10$$

$\chi^2 = \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$
$\chi^2 = .0375 + 0.0375 + .9 + .9 = 1.875$
From tables, $\chi^2_{0.05,1} = 3.84$
$\chi^2 < \chi^2_{\alpha,d.f}$
Accept H_0.
Hence difference between the proportions is insignificant.

Question 71

In a random sample of 160 workers exposed to a certain amount of radiation, 24 experienced some ill effects. Construct a 99% confidence interval for the corresponding true percentage using the large sample confidence interval formula.

Answer

$100(1-\alpha)\%$ confidence interval for proportion p is
$(p' - z_{\alpha/2}\sqrt{\frac{p'q'}{n}}, p' + z_{\alpha/2}\sqrt{\frac{p'q'}{n}})$
$100(1-\alpha) = 99 \Rightarrow \alpha = 0.01$
From the normal table $z_{\alpha/2} = 2.575$
$n = 160, p' = \frac{24}{160}, q' = 1 - \frac{24}{160}$
Substituting we get the 99% CI for proportion as
$(\frac{24}{160} - 2.575\sqrt{\frac{\frac{24}{160}\frac{136}{160}}{160}}, \frac{24}{160} + 2.575\sqrt{\frac{\frac{24}{160}\frac{136}{160}}{160}}) = (0.0773, 0.2227).$

Question 72

If 26 of 200 brand A tyres fail to last 20,000 miles, whereas the corresponding figures for 200 tyres each of brands B, C and D are 23, 15 and 32, use the 0.05 level of significance to test the null hypothesis that there is no difference in the quality of the four kinds of tyres with regard to their durability.

Answer

$H_0 : p_1 = p_2 = p_3 = p_4$
$H_1 : p_1,\ p_2,\ p_3$ and p_4 are not equal.

Consider the table: (No need of this table if we find e_{ij} by using $\widehat{p}$ and comes for use ,for the equation in brackets)

	A	B	C	D	Total
No. of failures (X_k)	26	23	15	32	X = 96
No. of success ($n_k - X_k$)	174	177	185	168	n - X = 704
Total (n_k)	200	200	200	200	n = 800

Level of significance, $\alpha = 0.05$.

Reject the null hypothesis if $\chi^2 > \chi^2_{\alpha,k-1} = \chi^2_{0.05,3} = 7.81$

$$\widehat{p} = \frac{X_1 + X_2 + X_3 + X_4}{n_1 + n_2 + n_3 + n_4} = \frac{96}{800} = 0.12$$

$e_{1j} = n_j.\widehat{p} \left(= \frac{n_j.X}{n} \right) = 200(0.12) = 24$

$e_{2j} = n_j.(1 - \widehat{p}) \left(= \frac{n_j.(n-X)}{n} \right) = 200(0.88) = 176$ for j = 1, 2, 3 and 4.

$$\chi^2 = \sum_{i=1}^{2} \sum_{j=1}^{4} \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$
$$= 7.102$$

$\therefore$ accept the null hypothesis since $\chi^2 = 7.102 < 7.81$.

Question 73

One method of seeding clouds was successful in 57 of 150 attempts while another method was successful in 33 of 100 attempts. At the 0.05 level of significance, can we conclude that the first method is better than the second? Find a large sample 95 % confidence interval for the true difference of probabilities.

Answer

Solution

$p_1^{'} = \frac{57}{150}, n_1 = 150$

$p_2^{'} = \frac{53}{100}, n_2 = 100$

Let $H_0 : p_1 = p_2$ and $H_1 : p_1 > p_2$

$Q = 0.05, w = z > 1.96$

$p^* = \frac{x_1 p_1^{'} + x_2 p_2^{'}}{x_1 + x_2} = \frac{57+33}{250} = \frac{90}{250} = \frac{9}{25}$

$q^* = 1 - p^* = \frac{16}{25}$

$z = \frac{p_1^{'} - p_2^{'}}{\sqrt{p^* q^* (\frac{1}{x_1} + \frac{1}{x_2})}}$

$\frac{\frac{57}{150} - \frac{33}{100}}{\sqrt{\frac{9}{25} * \frac{16}{25} * \frac{250}{150*100}}} = 0.8062$

cont..

since $z < 1.96$ we accept H_0
we cannot conclude that the first method is better than the second.
The 95% CI for $p_1 - p_2$ is

$$\left[p_1^{'} - p_2^{'} \pm 1.96 * \sqrt{\frac{p_1^{'} q_1^{'}}{x_1} + \frac{p_2^{'} q_2^{'}}{x_2}}\right]$$

$$= \left[(0.38 - 0.33) \pm 1.96 * \sqrt{\frac{0.38*0.62}{150} + \frac{0.33*0.64}{100}}\right]$$

$$= [-0.07053, 0.17053]$$

Question 75

The following is the distribution of the daily number of power failures reported in western city on 300 days. Test at the 0.05 level of significance whether the daily number of power failures in this city is a random variable having the Poisson distribution with $\lambda = 3.2$.

Number of power failures	0	1	2	3	4	5	6	7	8	9
Number of days	9	43	64	62	42	36	22	14	6	2

Solution

H_0 : The population is Poisson distribution.
H_1 : The population is not Poisson distribution.
$\alpha = 0.05$

x	f	f(x)=$\frac{e^{3.2}(3.2)^x}{x!}$	Expected frequency(300 $\times f(x)$)
0	9	0.04076	$12.228 \approx 12$
1	43	0.13044	$39.132 \approx 39$
2	64	0.20870	$62.61 \approx 63$
3	62	0.22262	$66.786 \approx 67$
4	42	0.17809	$53.427 \approx 54$
5	36	0.11398	$34.194 \approx 34$
6	22	0.06079	$18.237 \approx 18$
7	14	0.02779	8.337
8	6	0.01112	3.336 $\}$ $12.858 \approx 13$
9	2	0.00395	1.185

Solution 75 cont.

To test goodness of fit

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{k-r-1}$$
$$= \sum \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_{8-1}$$

O_i	9	43	64	62	42	36	22	22
E_i	12	39	63	67	54	34	18	13

$$\chi^2 = \frac{(9-12)^2}{12} + \frac{(43-39)^2}{39} + ... + \frac{(22-13)^2}{13}$$
$$= 11.4536$$

$\chi^2_{(0.05,7)} = 14.067$

$\chi^2 < \chi^2_{(0.05,7)}$

so accept null hypothesis, that is the population is Poisson.