

Probability and Statistics Assessed Coursework 5

Solution 1

a) Since the distribution is uniform and discrete, we can calculate mean and variance using the corresponding formula of uniform distribution

$$\mu = \frac{a+b}{2} = \frac{0+100}{2} = 50$$
$$\sigma^2 = \frac{(b-a)^2}{12} = \frac{(100-0)^2}{12} = \frac{2500}{3}$$

Solution 3

The 95% level confidence interval for μ , with degree of freedom 63, is

$$\sigma = \sqrt{\frac{4238}{64-1}} \approx 8.202$$
$$\bar{x} \pm 2 \cdot \frac{\sigma}{\sqrt{n}} = \frac{700}{64} \pm 1.96 \cdot \frac{8.202}{8} = 10.9375 \pm 2 \cdot 1.025 = 10.94 \pm 2.01$$

Solution 4

a) The 95% level confidence interval

$$\bar{x} = \frac{250}{100} = 2.5$$
$$\sigma = \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{\sum_i (x_i^2 - 2x_i\bar{x} + \bar{x}^2)}{n-1}} = \sqrt{\frac{725000 - 5 \cdot 250 + 625}{100-1}} = 85.54$$
$$\bar{x} \pm 1.96 \cdot \frac{\sigma}{\sqrt{n}} = \frac{250}{100} \pm 1.96 \cdot \frac{85.54}{10} = 2.5 \pm 16.77$$

Notice that the degree of freedom is beyond 40 so that we have chosen to use the normal distribution Φ table.

b) the 95% level confidence interval is

$$\bar{x} \pm 1.96 \cdot \frac{s_{n-1}}{\sqrt{n}} = 83.2 \pm 1.96 \cdot \frac{6.4}{10} = 83.2 \pm 1.25$$

We are using normal distribution Φ table for the same reason as in a).

Solution 8

a) For a two-sided hypothesis test, we have the following null hypothesis H_0 and alternative hypothesis H_1

$$H_0 : \mu = 0, H_1 : \mu \neq 0$$

The null distribution of sample mean is as follows

$$Z = \frac{\bar{D}}{s_{14}/\sqrt{15}} = \frac{\mu}{s_{14}/\sqrt{15}} \sim t(n-1)$$

where s_{14} is the bias-corrected standard deviation with degree of freedom 14.

b) With $\alpha = 0.1$, the rejection region is as follows

$$R = (-\infty, -t_{14, 1-\frac{\alpha}{2}}) \cup (t_{14, 1-\frac{\alpha}{2}}, \infty) = (-\infty, -1.7613) \cup (1.7613, \infty)$$

where the value of 1.76 is obtained from the t-distribution table with 5% and degree of freedom being 14.

c) The calculated value of test statistic is

$$z = \frac{D}{s_{14}/\sqrt{15}} = \frac{\bar{d}}{s_{14}/\sqrt{15}} = \frac{314/15}{1424.64/\sqrt{15}} = \frac{20.93}{367.84} \approx 0.057$$

where the calculation of bias-corrected standard deviation is as follows

$$s_{14} = \frac{\sqrt{\sum_{i=1}^{15} (d_i - \bar{d})^2}}{15 - 1} = \frac{19944.93}{14} \approx 1424.64$$

The z score of the test statistic is within the acceptance region. Therefore, we will accept the null hypothesis that there is no difference between the heights of progeny of cross-fertilised and self-fertilised plant and reject the alternative hypothesis.

Solution 9

a) The null hypothesis H_0 and alternative hypothesis H_1 are as follows

$$H_0 : \mu = 0.6, H_1 : \mu \neq 0.6$$

b) The Poisson Distribution with a small value of λ fits the description in the question since it monotonically and rapidly decreases and it is far from symmetric. It also fits the definition of this experiment as Poisson Distribution normally would associate with the number of events in a unit time or space.

c) We estimate the mean and variance of the distribution by using the Maximum Likelihood Estimator in the form of \ln :

$$\begin{aligned}
l(\lambda) &= \sum_{i=1}^{400} \ln(p_{X|\lambda}(x_i)) = \sum_{i=1}^{400} \ln\left(\frac{e^{-\lambda} \lambda^{x_i}}{x_i!}\right) \\
&= \ln\left(\prod_{i=1}^{400} \frac{e^{-\lambda} \lambda^{x_i}}{x_i!}\right) \\
&= \ln\left(\frac{e^{-400\lambda} \lambda^{\sum_{i=1}^{400} x_i}}{\prod_{i=1}^{400} x_i!}\right) \\
&= \ln(e^{-400\lambda}) + \ln(\lambda^{\sum_{i=1}^{400} x_i}) - \ln\left(\prod_{i=1}^{400} x_i!\right) \\
&= -400\lambda + \ln(\lambda) \sum_{i=1}^{400} x_i - \ln\left(\prod_{i=1}^{400} x_i!\right)
\end{aligned}$$

We then differentiate $l(\lambda)$ and then set it to equal 0

$$\begin{aligned}
l'(\lambda) &= \frac{d}{d\lambda}(-400\lambda + \ln(\lambda) \sum_{i=1}^{400} x_i - \ln(\prod_{i=1}^{400} x_i!)) \\
&= -400 + \frac{\sum_{i=1}^{400} x_i}{\lambda} \\
l'(\lambda) &= 0 \\
-400 + \frac{\sum_{i=1}^{400} x_i}{\lambda} &= 0 \\
400\lambda &= \sum_{i=1}^{400} x_i \\
\lambda &= \frac{\sum_{i=1}^{400} x_i}{400}
\end{aligned}$$

To check it's indeed the maximum value, we also take the second derivative

$$l''(\lambda) = \frac{d^2}{d^2\lambda} l(\lambda) = -\frac{\sum_{i=1}^{400} x_i}{\lambda^2}$$

Since the count should all be positive, $\sum_{i=1}^{400} x_i$ should be positive, and $\lambda^2 \geq 0$. Thus, $l''(\lambda) < 0$ and it's indeed the maximum point.

From the property of Poisson distribution, both the mean and variance equal to λ

$$\mu = \sigma^2 = \lambda = \frac{0 \cdot 213 + 1 \cdot 128 + 2 \cdot 37 + 3 \cdot 18 + 4 \cdot 3 + 5 \cdot 1}{400} = 0.6825$$

d) Since the population variance σ^2 is unknown, we need to use the Student's t-distribution to conduct hypothesis testing. However, since the sample size is big, the t value has a high degree of freedom and we will use normal distribution to approximate critical region. Thus, we will use the Φ table, or the cdf of normal distribution to find the value of z below. Since this hypothesis testing is a two-sided test, a 5% significance level indicates that we need to get the value of z where $\Phi(z) = \frac{5\%}{2} = 2.5\%$. Using the Φ table we can find that the value of z is 1.96.

By the fact that the normal distribution comes with mean close to population mean μ and variance $\frac{\sigma}{\sqrt{n}}$, we can easily find that the critical region would be $(-\infty, -1.96 \frac{\sigma}{\sqrt{n}}) \cup (1.96 \frac{\sigma}{\sqrt{n}}, \infty)$.
By the fact that for Poisson Distribution, $\sigma = \sqrt{\mu}$, we have $(-\infty, -1.96 \sqrt{\frac{\mu}{n}}) \cup (1.96 \sqrt{\frac{\mu}{n}}, \infty)$ as the critical region to reject the null hypothesis.

Notice that we have used union to include both sides in the distribution. This is due to the fact that our hypothesis test is a two-sided one. (i.e. it's testing whether the population mean is 0.6 or not, rather than taking one side and discard the other)

The test statistic is as follows

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{\mu - 0.6}{\sqrt{\frac{\mu}{n}}}$$

e) The limits of critical region are

$$\#limit = \pm 1.96 \cdot \sqrt{\frac{\mu}{n}} = \pm 1.96 \cdot \sqrt{\frac{0.6}{400}} = \pm 0.076$$

f) The test statistic value is calculated as below

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{\mu - 0.6}{\sqrt{\frac{\mu}{n}}} = \frac{0.6825 - 0.6}{0.6825/\sqrt{400}} = 2.42$$

Since the test statistic is outside our critical region (i.e. the calculated z value is greater than 1.96), we will reject the null hypothesis and accept that the mean number of cells per square is not 0.6.

Solution 10

We will perform a chi-square analysis using the chi-square statistic

Since male and female have the same chance of being born, the expectation should follow a binomial distribution with $p = 0.5$. So that the expected value in the table are

Boys/Girls	# of Families
5/0	$0.5^5 \cdot 320 = 10$
4/1	$5 \cdot 0.5^5 \cdot 320 = 50$
3/2	$10 \cdot 0.5^5 \cdot 320 = 100$
2/3	$10 \cdot 0.5^5 \cdot 320 = 100$
1/4	$5 \cdot 0.5^5 \cdot 320 = 50$
0/5	$0.5^5 \cdot 320 = 10$

Notice that if we postulate that male and female are equally probable, then the observed

$$\begin{aligned}
 X^2 &= \frac{(18 - 10)^2}{10} + \frac{(56 - 50)^2}{50} + \frac{(110 - 100)^2}{100} + \frac{(88 - 100)^2}{100} + \frac{(40 - 50)^2}{50} + \frac{(8 - 10)^2}{10} \\
 &= 6.4 + 0.72 + 1 + 1.44 + 2 + 0.4 = 11.96
 \end{aligned}$$

The degree of freedom is $6 - 1 - 1 = 4$ since there are 6 categories in total and we are estimating p in the binomial distribution.

Therefore, the results for 1% level is 13.277 and for 5% level is 9.488.

We reject null hypothesis under the 5% level and accept it under then 1% level.