NATIONAL UNIVERSITY OF SINGAPORE
DEPARTMENT OF STATISTICS & APPLIED PROBABILITY

ST2334 PROBABILITY AND STATISTICS

SEMESTER I, AY 2022/2023

Tutorial 09: Solution

This set of questions will be discussed by your tutors during the tutorial in Week 12.

Please work on the questions before attending the tutorial.

1. Let $X \sim N(0, \sigma^2)$. Derive the formula for $E(X^{2k+1})$ for $k = 0, 1, 2, \ldots,$.

   SOLUTION
   For any real valued x, by symetricity,

   $$P(-X \le x) = P(X \ge -x) = P(X \le x),$$

   therefore $-X \sim N(0, \sigma^2)$.
   For any $k = 0, 1, 2, \ldots$, we have

   $$E\left((-X)^{2k+1}\right) = E\left(X^{2k+1}\right),$$

   or equivalently

   $$-E\left(X^{2k+1}\right) = E\left(X^{2k+1}\right),$$

   which immediately implies $E\left(X^{2k+1}\right) = 0$.

2. The random variable X representing the number of cherries in a cherry puff, has the following probability distribution:

x	4	5	6	7
$f(x)$	0.2	0.4	0.3	0.1

   (a) Find the mean μ and the variance σ^2 of X.

   (b) Find the mean $\mu_{\overline{X}}$, and the variance $\sigma^2_{\overline{X}}$ of the mean $\overline{X}$ for random samples of 36 cherry puffs from the above probability distribution.

   (c) Find the probability that the average number of cherries in 36 cherry puffs will be less than 5.5.

   SOLUTION

   (a) $\mu = \sum x f(x) = 4(0.2) + 5(0.4) + 6(0.3) + 7(0.1) = 5.3.$
   $\sigma^2 = \sum (x-\mu)^2 f(x) = (4-5.3)^2(0.2) + (5-5.3)^2(0.4) + (6-5.3)^2(0.3) + (7-5.3)^2(0.1) = 0.81.$

   (b) With $n = 36$, $\mu_{\overline{X}} = \mu = 5.3$; $\sigma^2_{\overline{X}} = \dfrac{\sigma^2}{n} = \dfrac{0.81}{36} = 0.0225.$

   (c) Applying the Central Limit Theorem, $\overline{X}$ approx $\sim N(5.3, 0.0255)$,

   $$P(\overline{X} < 5.5) = P\left(Z < \frac{5.5 - 5.3}{\sqrt{0.0225}}\right) = P(Z < 1.33) = 0.9082.$$

3. The chemical benzene is highly toxic to humans. However, it is used in the manufacture of many medicine dyes, leather, and many coverings. In any production process involving benzene, the water in the output of the process must not exceed 7950 parts per million (ppm) of benzene because of government regulations. For a particular process of concern the water sample was collected by a manufacturer 25 times randomly and the sample average was 7960 ppm. It is known from historical data that the standard deviation σ is 100ppm.

   (a) What is the probability that the sample average in this experiment would exceed the government limit if the population mean is equal to the limit? Use the central limit theorem.

   (b) Is an observed sample average of 7960 in this experiment firm evidence that the population mean for the process exceeds the government limit? Answer by computing $P\left(\overline{X} \geq 7960|\mu = 7950\right)$. Assume that the distribution of benzene concentration is normal.

   SOLUTION
   Let X = amount of benzene. $E(X) = \mu$ and $\text{var}(X) = 100^2$.

   (a) $n = 25$. By the Central Limit Theorem, $\overline{X} \sim N\left(\mu, \frac{100^2}{25}\right)$ approximately.

   $$P\left(\overline{X} > 7950\middle|\mu = 7950\right) = P(\overline{X} > \mu) = 0.5.$$

   (b) $X \sim N(\mu, 100^2)$. Hence $\overline{X} \sim N\left(\mu, \frac{100^2}{25}\right)$ approximately. We have

   $$P\left(\overline{X} \geq 7960\middle|\mu = 7950\right) = P\left(Z > \frac{7960 - 7950}{100/\sqrt{25}}\right) = P(Z > 0.5) = 0.3085.$$

   Therefore, there is no strong evidence that the population mean exceeds the government limit as it is likely to see a sample mean is equal to or greater than 7960 if the population mean equals to the government limit 7950.

4. A normal population with unknown variance has a mean of 20. Is one likely to obtain a random sample of size 9 from this population with a standard deviation being 4.1 and a mean being larger than or equal to 24? If not, what conclusion would you draw?

   SOLUTION
   If $\mu = 20$, then

   $$P\left(\overline{X} > 24\right) = P\left(\frac{\overline{X} - 20}{4.1/\sqrt{9}} > \frac{24 - 20}{4.1/\sqrt{9}}\right) = P(T_8 > 2.9268) = 0.00955.$$

   So we conclude that $\mu > 20$, since this probability is very small showing that it is very unlikely to get mean of 24 if the population mean is really 20.

5. Two different box-filing machines are used to fill cereal boxes on the assembly line. The critical measurement influenced by these machines is the weight of the product in the machines. Engineers are quite certain that the variance of the weight of product is $\sigma^2 = 1$ gram. Experiments are conducted using both machines with sample sizes of 36 each. The sample averages for machine A and B are $\bar{x}_A = 4.5$ grams $\bar{x}_B = 4.7$ grams. Engineers seemed surprised that the two sample averages for the filling machines were so different.

   (a) Use the central limit theorem to determine $P(\overline{X}_B - \overline{X}_A \geq 0.2)$ under the condition that $\mu_A = \mu_B$.

(b) Do the aforementioned experiments seem to, in any way, strongly support a conjecture that the two population means for the two machines are different? Explain using your answer in (a).

SOLUTION

(a) When $\mu_A = \mu_B$, $E(\overline{X}_B - \overline{X}_A) = 0$ and

$$V(\overline{X}_B - \overline{X}_A) = V(\overline{X}_B) + V(\overline{X}_A) = \frac{\sigma^2}{n} + \frac{\sigma^2}{n} = \frac{2}{36}.$$

Therefore $\overline{X}_B - \overline{X}_A$ approx $N(0, 2/36)$. We have

$$P\left(\overline{X}_B - \overline{X}_A \geq 0.2\right) = P\left(Z > \frac{0.2}{\sqrt{2/36}}\right) = 0.1981.$$

(b) Since the probability in Part (a) is not small, therefore it is not unlikely to observe $\overline{X}_B - \overline{X}_A \geq 0.2$ when $\mu_A = \mu_B$. Hence, the conjecture that $\mu_A \neq \mu_B$ is likely not true.

6. Find the probability that a random sample of 25 observations, from a normal population with variance $\sigma^2 = 6$, will have a variance S^2

(a) greater than 9.1.

(b) between 3.462 and 10.745.

SOLUTION

(a) $P(S^2 > 9.1) = P\left(\frac{(n-1)S^2}{\sigma^2} > \frac{(24)(9.1)}{6}\right) = P(\chi^2(24) > 36.4) = 0.05017.$

(b)

$$\begin{aligned} P(3.462 < S^2 < 10.754) &= P\left(\frac{(24)(3.462)}{6} < \frac{(n-1)S^2}{\sigma^2} < \frac{(24)(10.754)}{6}\right) \\ &= P\left(13.848 < \chi^2(24) < 42.98\right) = 0.94. \end{aligned}$$

7. If S_1^2 and S_2^2 represent the variances of independent random samples of size $n_1 = 8$ and $n_2 = 12$, taken from normal populations with equal variances, find $P\left(\frac{S_1^2}{S_2^2} < 4.89\right)$.

SOLUTION

Since σ_1^2 and σ_2^2 are equal, therefore S_1^2/S_2^2 follows an F distribution with $(7, 11)$ degrees of freedom. Hence $P(S_1^2/S_2^2 < 4.89) = 0.99$.

8. Let X be a binomial random variable with parameters n and p.

(a) Let $U = X/n$. Show that U is an unbiased estimator of p.

(b) Let $V = \frac{X + n/2}{3n/2}$. Show that V is a biased estimator of p.

SOLUTION

(a) $E(U) = E(X)/n = np/n = p$. Since $E(U) = p$, therefore U is an unbiased estimator of p.

(b) $E(V) = \frac{E(X+n/2)}{3n/2} = \frac{np+n/2}{3n/2} = \frac{p+1/2}{3} \neq p$ unless $p=1$. Since $E(V) \neq p$, therefore V is a biased estimator of p.

9. Assume that the helium porosity (in percentage), Y, of coal samples taken from any seam is normally distributed.

   (a) If the true standard deviation of Y is 0.75, compute a 95% confidence interval for the average porosity of a certain seam if the average porosity for 20 specimens from the seam was 4.85.

   (b) How large a sample size is necessary if the length of the 95% interval is to be 0.40?

   (c) If the variance of Y is unknown, and the sample standard deviation for the sample in (a) is 0.75, compute a 95% confidence interval for the average porosity of a certain seam.

   SOLUTION
   Let $Y =$ helium porosity of a coal sample; then $Y \sim N(\mu, \sigma^2)$.

   (a) It is given that $\sigma = 0.75$, $n = 20$, and $\bar{y} = 4.85$. Hence a 95% confidence interval for μ is given by $\bar{y} \pm z_{0.025}\frac{\sigma}{\sqrt{n}} = 4.85 \pm 1.96\frac{0.75}{\sqrt{20}} = 4.85 \pm 0.3287 = (4.5213, 5.7187)$.

   (b) The length of a 95% confidence interval is $2z_{0.025}\frac{\sigma}{\sqrt{n}}$. Hence the length of 95% confidence interval being 0.4 implies that $2(1.96)\frac{0.75}{\sqrt{n}} = 0.4$. Therefore $n = 54$.

   (c) It is given that $s = 0.75$, $n = 20$ and $\bar{y} = 4.85$. Hence a 95% confidence interval for μ is given by $\bar{y} \pm t(19, 0.025)\frac{s}{\sqrt{n}} = 4.85 \pm 2.093\frac{0.75}{\sqrt{20}} = 4.85 \pm 0.351 = (4.499, 5.201)$.

10. A random sample of 12 shearing pins is taken in a study of the Rockwell hardness of the head on the pin. Measurements on the Rockwell hardness were made for each of the 12, yielding an average value of 48.50 with a sample standard deviation of 1.5. Assuming the measurements to be normally distributed, construct a 90% confidence interval for the mean Rockwell hardness.

    SOLUTION
    $n = 12$ is small, σ is unknown, and the data are normally distributed. Therefore a 90% confidence interval for μ is given by

    $$\bar{x} \pm t(11, 0.05)\frac{s}{\sqrt{n}} = 48.5 \pm (1.796)\frac{1.5}{\sqrt{12}} = 48.50 \pm 0.7777 = (47.722, 49.278).$$

11. A random sample of size $n_1 = 25$ taken from a normal population with a standard deviation $\sigma_1 = 5$ has a mean $\bar{x}_1 = 80$. A second random sample of size $n_2 = 36$ taken from a different normal population with a standard deviation $\sigma_2 = 3$ has a mean $\bar{x}_2 = 75$. Find a 94% confidence interval for $\mu_1 - \mu_2$.

    SOLUTION
    We have a two-sample problem. The data are normal; the population standard deviations σ_1 and σ_2 are known. Therefore the 94% confidence interval for $\mu_1 - \mu_2$ is given by

    $$(\bar{x}_1 - \bar{x}_2) \pm z_{0.03}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = (80-75) \pm (1.88)\sqrt{\frac{5^2}{25} + \frac{3^2}{36}} = 5 \pm 2.102 = (2.898, 7.102).$$

12. A study was conducted to determine if a certain metal treatment has any effect on the amount of metal removed in a pickling operation. A random sample of 100 pieces was immersed in a bath for 24 hours without the treatment, yielding an average of 12.2 millimeters of metal removed and a sample standard deviation of 1.1 millimeters. A second sample of 200 pieces was exposed to the treatment, followed by the 24-hour immersion in the bath, resulting in an average removal of 9.1 millimeters of metal with a sample standard deviation of 0.9 millimeters. Compute a 98% confidence interval estimate for the difference between the population means. Does the treatment appear to reduce the mean amount of metal removed?

SOLUTION

$n_1 = 100$, $n_2 = 200$; therefore the sample sizes are large. σ_1 and σ_2 are unknown. We can construct the 98% confidence interval is given by

$$(\bar{x}_1 - \bar{x}_2) \pm z_{0.01}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = (12.2 - 9.1) \pm (2.33)\sqrt{\frac{1.1^2}{100} + \frac{0.9^2}{200}} = 3.1 \pm 0.296 = (2.804, 3.396).$$

Since the 98% confidence interval does not cover 0 and is in the positive range, the treatment appears to reduce the mean amount of metal removed.