Monsoon Semester 2021-22: Tutorial 2

MA 2001D: Mathematics 3

Department of Mathematics
National Institute of Technology
Calicut

Question 1

Question 2

A manufacturer of flashlight batteries claims that its batteries will last an average of $\mu = 34$ hours of continuous use. Of course, there is some variability in life expectancy with $\sigma = 3$ hours. During consumer testing, a sample of 30 batteries lasted an average of only 32.5 hours. How likely is it to obtain a sample that performs this badly if the manufacturer's claim is true?

Answer

Given

$$\mu = 34, \quad \sigma = 3, \quad n = 30, \quad \bar{x} = 32.5$$

We need to find $P(\bar{X} \leq 32.5)$. Assuming the population is normally distributed, we have $\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

$$\begin{aligned} P(\bar{X} \leq 32.5) &= P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq \frac{32.5 - \mu}{\sigma/\sqrt{n}}\right) \\ &= P\left(Z \leq \frac{32.5 - 34}{3/\sqrt{30}}\right) \\ &= P(Z < -2.7386) \\ &= 0.0032 \end{aligned}$$

Assuming the claim, there is 0.32% chance to obtain a sample that performs this badly (Type-I error at 1% level of significance).

Question 3

Let $x_1, x_2, \ldots, x_7$ denote a random sample from a population having mean μ and variance σ_2. Consider the following estimators of μ
$\hat{\theta}_1 = \frac{x_1+x_2+\cdots+x_7}{7}, \hat{\theta}_2 = \frac{2x_1-x_6+x_4}{2}$. Are they unbiased estimators of μ? If so which one is more efficient? Give reason.

Answer: Let random variables $x_1, x_2, \ldots, x_7$ represent a random sample from a population with a mean μ and variance σ^2. Let us check for whether estimators are biased or not.

$$\hat{\theta}_1 = \frac{x_1 + x_2 + \cdots + x_7}{7}, \hat{\theta}_2 = \frac{2x_1 - x_6 + x_4}{2}$$

$$\begin{aligned} E\left(\hat{\theta}_1\right) &= E\left(\frac{x_1 + x_2 + \cdots + x_7}{7}\right) \\ &= \frac{1}{7}E\left(x_1\right) + \frac{1}{7}E\left(x_2\right) + \cdots + \frac{1}{7}E\left(x_7\right) \\ &= \frac{1}{7}\mu + \frac{1}{7}\mu + \cdots + \frac{1}{7}\mu = \mu \end{aligned}$$

Therefore, $E\left(\hat{\theta}_1\right)$ is an unbiased estimator of μ.

$$E\left(\hat{\theta}_2\right) = E\left(\frac{2x_1 - x_6 + x_4}{2}\right)$$
$$= E(x_1) - \frac{1}{2}E(x_6) + \frac{1}{2}E(x_4)$$
$$= \mu - \frac{1}{2}\mu + \frac{1}{2}\mu = \mu$$

Therefore, $E\left(\hat{\theta}_2\right)$ is an unbiased estimator of μ. Let us determine the variances of estimators. The variance of the estimator $\hat{\theta}_1$:

$$\operatorname{Var}\left(\hat{\theta}_1\right) = \operatorname{Var}\left(\frac{x_1 + x_2 + \cdots + x_7}{7}\right)$$
$$= \frac{1}{49}\operatorname{Var}(x_1) + \frac{1}{49}\operatorname{Var}(x_2) + \cdots + \frac{1}{49}\operatorname{Var}(x_7)$$
$$= \frac{1}{49}\sigma^2 + \frac{1}{49}\sigma^2 + \frac{1}{49}\sigma^2 + \frac{1}{49}\sigma^2 + \frac{1}{49}\sigma^2 + \frac{1}{49}\sigma^2 + \frac{1}{49}\sigma^2$$
$$= \frac{1}{7}\sigma^2$$

The variance of the estimator $\hat{\theta}_2$:

$$\begin{aligned}
\text{Var}\left(\hat{\theta}_2\right) &= \text{Var}\left(\frac{2x_1 - x_6 + x_4}{2}\right) \\
&= \text{Var}\left(x_1\right) + \frac{1}{4}\text{Var}\left(x_6\right) + \frac{1}{4}\text{Var}\left(x_4\right) \\
&= \sigma^2 + \frac{1}{4}\sigma^2 + \frac{1}{4}\sigma^2 \\
&= \frac{3}{2}\sigma^2
\end{aligned}$$

The best estimator is the one with a smaller variance because it is more likely to produce an estimate close to the true value μ. Therefore, $\hat{\theta}_1$ is the best estimator.

Question 4

Distinguish between point estimation and interval estimation of a population parameter. Give examples.

Answer

- A **point estimate** is a single value estimate of a parameter. For instance, a **sample mean** is a point estimate of a population mean.
- An **interval estimate** gives you a range of values where the parameter is expected to lie. A **confidence interval** is the most common type of interval estimate.

Question 5

If $X_1, X_2, \cdots, X_N$ is a random sample from an infinite population with mean μ, obtain a condition under which $A_1X_1 + A_2X_2 + \cdots + A_NX_N$ will become an unbiased estimator of μ.

Answer

$E(X_i) = \mu, i = 1, 2, \cdots, N$
$A_1X_1 + A_2X_2 + \cdots + A_NX_N$ ia an unbiased estimator of μ if

$$E(A_1X_1 + A_2X_2 + \cdots + A_NX_N) = \mu.$$
$$A_1E(X_1) + A_2E(X_2) + \cdots + A_NE(X_N) = \mu$$
$$A_1\mu + A_2\mu + \cdots + A_N\mu = \mu$$
$$(A_1 + A_2 + \cdots + A_N) = 1$$

Question 6

Suppose we have a random sample $x_1, x_2, \cdots, x_{2n}$ of size 2n from a population having mean μ and variance σ^2. Given $\bar{X}_1 = \frac{1}{2n}\sum_{i=1}^{2n} x_i$ and $\bar{X}_2 = \frac{1}{n}\sum_{i=1}^{n} x_i$ be two estimates of μ

- (a) Are they unbiased estimators of μ?
- (b) If so which one is more efficient?

Answer

Given $\bar{X_1} = \frac{1}{2n}\sum_{i=1}^{2n} x_i$ and $\bar{X_2} = \frac{1}{n}\sum_{i=1}^{n} x_i$

$$
\begin{aligned}
E(\bar{X_1}) &= E\left(\frac{1}{2n}\sum_{i=1}^{2n} x_i\right) \\
&= \frac{1}{2n} \times E\left(\sum_{i=1}^{2n} x_i\right) \\
&= \frac{1}{2n} \times \sum_{i=1}^{2n} E(x_i) \\
&= \frac{2n\mu}{2n} \\
&= \mu
\end{aligned}
$$

$\therefore \bar{X_1}$ is an unbiased estimator of μ

तमसो मा ज्योतिर्गमय

Answer 6 cont.

$$
\begin{aligned}
E(\bar{X_2}) &= E\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) \\
&= \frac{1}{n} \times E\left(\sum_{i=1}^{n} x_i\right) \\
&= \frac{1}{n} \times \sum_{i=1}^{n} E(x_i) \\
&= \frac{n\mu}{n} \\
&= \mu
\end{aligned}
$$

$\therefore \bar{X_2}$ is an unbiased estimator of μ

Answer 6 cont.

$$
\begin{aligned}
Var(\bar{X_1}) &= Var\left(\frac{1}{2n}\sum_{i=1}^{2n} x_i\right)\\
&= \frac{1}{4n^2} \times Var\left(\sum_{i=1}^{2n} x_i\right)\\
&= \frac{1}{4n^2} \times \sum_{i=1}^{2n} Var(x_i)\\
&= \frac{2n\sigma^2}{4n^2}\\
&= \frac{\sigma^2}{2n}
\end{aligned}
$$

Answer 6 cont.

$$
\begin{aligned}
Var(\bar{X_2}) &= Var\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) \\
&= \frac{1}{n^2} \times Var\left(\sum_{i=1}^{n} x_i\right) \\
&= \frac{1}{n^2} \times \sum_{i=1}^{n} Var(x_i) \\
&= \frac{n\sigma^2}{n^2} \\
&= \frac{\sigma^2}{n}
\end{aligned}
$$

$Var(\bar{X_1}) < Var(\bar{X_2}) \implies \bar{X_1}$ is more efficient.

Question 7

Suppose that we observe a random variable X having Binomial distribution and get 'x' defectives in 'n' trials. Show that $\frac{x}{n}$ is an unbiased estimate and $\frac{x+1}{n+1}$ is not an unbiased estimate of the Binomial parameter p.

Answer

Since $E(X) = np$, we have

$$E\left(\frac{x}{n}\right) = \frac{1}{n} * E(X) = \frac{1}{n} * (np) = p$$

Also,

$$E\left(\frac{x+1}{n+1}\right) = \frac{E(X)+1}{n+1} = \frac{np+1}{n+1} \neq p$$

Hence, $\frac{x}{n}$ is an unbiased estimator of p and $\frac{x+1}{n+1}$ is a biased estimator of p.

Question 8

When we sample from an infinite population, what happens to the standard error of the mean of the sample, if the sample size (n) is
a) increased from 100 to 200
b) decreased from 360 to 90 .

Answer

When we sample from an infinite population, the standard error of the sample mean is $SE(\bar{X}) = \frac{\sigma}{\sqrt{n}}$.

a)$n = 100 \Rightarrow SE(\bar{X}) = \frac{\sigma}{10}$

$n = 200 \Rightarrow SE(\bar{X}) = \frac{\sigma}{10\sqrt{2}}$

Thus $SE(\bar{X})$ is decreased $\frac{1}{\sqrt{2}}$ times.

b)$n = 360 \Rightarrow SE(\bar{X}) = \frac{\sigma}{6\sqrt{10}}$

$n = 90 \Rightarrow SE(\bar{X}) = \frac{\sigma}{3\sqrt{10}}$

Thus $SE(\bar{X})$ is increased 2 times.

Question 9

If $\bar{X}$ is the mean of random sample of size n taken from a Normal population with mean μ and variance 100, find n such that $P(\mu - 5 < \bar{X} < \mu + 5) = 0.954$.

Answer

$\bar{X}$- sample mean
$\sigma^2 = 100 \Rightarrow \sigma = 10$
By CLT $\frac{X-\mu}{\sigma/\sqrt{n}} - \mathrm{N}(0,1)$
$P(\mu - 5 < \bar{X} < \mu + 5) = 0.954$
$P(-5 < \bar{X} - \mu < 5) = 0.954$
$P\left(\frac{-\sqrt{n}}{2} < \frac{\bar{X}-\mu}{10/\sqrt{n}} < \frac{\sqrt{n}}{2}\right) = 0.954$
$2P\left(0 < Z < \frac{\sqrt{n}}{2}\right) = 0.954$
$P\left(0 < Z < \frac{\sqrt{n}}{2}\right) = 0.477$
$\frac{\sqrt{n}}{2} = 1.995$
$\sqrt{n} = 3.99$
$n = 15.92$
$n \approx 16$

Question 10

Let X and S^2 be the mean and variance of a random sample of size 25 from a normal population with $\mu = 3$ and $\sigma^2 = 100$. Evaluate $P(0 < X < 6)$ and $P(55.2 < S^2 < 145.6)$

Answer

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) \implies \bar{X} \sim N(3, 2)$$

$$P(0 < \bar{X} < 6) = P\left(\frac{0-\mu}{\frac{\sigma}{\sqrt{n}}} < \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < \frac{6-\mu}{\frac{\sigma}{\sqrt{n}}}\right)$$

$$= P(-1.5 < z < 1.5)$$

$$= 2P(0 < z < 1.5)$$

$$= 2 \times 0.4332 = 0.8664$$

$$\frac{nS^2}{\sigma^2} \sim \chi^2_{n-1}$$

$$\begin{aligned} P\left(55.2 < S^2 < 145.6\right) &= P\left(\frac{25 \times 55.2}{100} < \frac{nS^2}{\sigma^2} < \frac{25 \times 145.6}{100}\right) \\ &= P\left(13.8 < \chi^2_{24} < 36.4\right) \\ &= 0.98 - 0.05 \\ &= 0.93 \end{aligned}$$

Question 11

A random sample of size 100 is taken from a population having $\sigma = 5.1$ and the sample mean $\bar{X} = 21.6$. Obtain the 95% confidence interval for μ.

Answer

If you don't know your population mean (μ) but you do know the standard deviation (σ), you can find a confidence interval for the population mean, with the formula: $\bar{X} \pm Z\frac{\sigma}{\sqrt{n}}$.

Substituting the given values of n, $\bar{X}$, σ and $Z_{0.025} = 1.96$ into the confidence interval formula, we get

$21.6 - 1.96(\frac{5.1}{\sqrt{100}}) < \mu < 21.6 + 1.96(\frac{5.1}{\sqrt{100}})$ Or $20.6 < \mu < 22.6$.

Question 12

Let a random sample of size 17 from $N(\mu, \sigma^2)$ gave $\bar{X} = 4.7$ and $S^2 = 5.76$. Detarmine 90% confidence interval for μ

Answer

If $\bar{x}$ is the sample mean of a random sample of size n from a normal population with sample standard deviation s, a $100(1-\alpha)\%$ confidence interval on μ is given by

$\bar{x} - t_{\frac{\alpha}{2},n-1}\frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\frac{\alpha}{2},n-1}\frac{s}{\sqrt{n}}$, (1)

where $t_{\frac{\alpha}{2},n-1}$ is the upper $100\frac{\alpha}{2}$ percentage point of the t distribution with $n-1$ degrees of freedom.

A 90% CI with sample variance $s^2 = 5.76$, sample size $n = 17$ and sample mean $\bar{x} - 4.7$. Therefore,

$\alpha = 1 - 0.90 = 0.1 \Rightarrow t_{\frac{\alpha}{2},n-1} = t_{0.05,16} = 1.746$, (2)

From (1) and (2) we get

$4.7 - 1.746\frac{\sqrt{5.76}}{\sqrt{17}} \leq \mu \leq 4.7 - 1.746\frac{\sqrt{5.76}}{\sqrt{17}}$,

A 90% confidence interval for μ is,

$3.68 \leq \mu \leq 5.72$

Question 13

Past experience indicated that wire rods purchased from a company having a mean breaking strength of 400 pounds and S.D of 15 pounds. How many rods should you select so that you could be so certain with a probability of 0.950 that resulting mean would not be in error by more than 2 pounds?

Answer

Let X be the breaking strength of rods purchased with mean μ and variance σ^2.

Given $\mu = 400$, $\sigma = 15$

The $\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$

Confidence interrval: $\left(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$

Error bound for the mean, EBM $= 2$

(lower bound, upper bound)=(point estimate - EBM, point estimate + EBM)

95% C.I, $z_{\alpha/2} = 1.96$

$z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 2$

$\sqrt{n} = \frac{z_{\alpha/2}\sigma}{2} = \frac{1.96\times 15}{2} = 14.7 \implies n = 216$

Question 14

A random sample of size 25 from a Normal population has a mean $\bar{X} = 47.5$ and standard deviation $S = 8.4$. Does the information support or refute the claim that the mean of the population $\mu = 42.1$

Answer

Given

$n = 25, \bar{x} = 47.5, \mu = 42.1, s = 8.4$

We have from t-distribution

$t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} = \frac{47.5-42.1}{\frac{8.4}{\sqrt{25}}} = 3.21$

This value of t has 24 degrees of freedom. From the of t distribution for $v = 24$, we get the probability that t will exceed 2.797 is 0.005. Then the probability of getting a value greater than 3.21 is negligible.

Hence we conclude that the information given in the data of this example tend to refute the claim that the mean of the population is $\mu = 42.1$.

Question 15

The following are the timer between six calls for an ambulance in a certain city and patients arrival at the hospital: 27, 15, 20, 32, 18 and 26 minutes. Use this information to judge the reasonableness of the ambulance service's claim that it take on the average 20 minutes between the call for an ambulance and patient's arrival at the hospital.

Answer

Mean of the data, $\overline{x} = 23$ and sample standard deviation
$s = \frac{\sum_{i=1}^{6}(x_i - \overline{x})^2}{6-1} = 6.39.$
$H_0 : \mu = 20$
$H_1 : \mu \neq 20.$
The statistic $t = \frac{\overline{x}-\mu}{s/\sqrt{n}}$ follows t-distribution with $6 - 1 = 5$ degrees of freedom.
Then $t = \frac{\overline{x}-\mu}{s/\sqrt{n}} = \frac{23-20}{6.39/\sqrt{6}} = 1.15.$
Accept H_0 if $\mid t \mid \leq t_{\alpha/2,n-1}$ and reject H_0 otherwise.
From H_1, we look into a two sided t table with $\alpha = 0.05$.
Hence $t_{\alpha/2,n-1} = t_{0.025,5} = 2.5705.$
Thus, $t = 1.15 < t_{\alpha/2,n-1} = 2.5705$ and hence by our hypothesis we accept H_0.

Question 16

If we want to determine the average mechanical aptitude of large group of workers, how large a random sample will we need to be able to assert with a probability 0.95 that the sample mean will not differ from the true mean by more than 3 points? Assume that it is known from past experience that $\sigma = 20$.

Answer

We need to choose n such that

$$z_{,025}\frac{\sigma}{\sqrt{n}} = 3$$

Since $\sigma = 20$, and $z_{,025} = 1.96$,

$$\sqrt{n} = \frac{1.96 \cdot 20}{3} = 13.067$$

Thus, $n = 170.7 \simeq 171$. Thus, the required sample size is 171 .

Question 17

The dean of a college wants to use the mean of a random sample to estimate the average amount of time students to get from class to the next, and she wants to be able to assert with 99% confidence that the error is at most 0.25 minutes. If it can be presumed that $\sigma = 1.40$ minutes from past experience, how large a sample needs be taken?

Solution: $\sigma = 1.40$ and $Z_{0.005} = 2.575, E = 0.25$ into the formula for n we get $\mathrm{n} = \left(\frac{2.575\mathrm{X}1.40}{0.25}\right)^2 \approx 207.9 \approx 208$. Hence the solution.

Question 18

It is desired to estimate the mean time (in days) of continuous use until an answering machine will first require service. If it can be assumed that $\sigma = 60$ days, how large a sample is needed so that one will be able to assert until 90% confidence that the sample mean is off by at most 10 days?

Answer

Solution: $\sigma = 60$ and $Z_{0.05} = 1.6449, E = 10$ into the formula for n we get $\mathrm{n} = \left(\frac{1.6449 \times 60}{10}\right)^2 \approx 97.4051 \approx 97$. Hence the solution.

Question 19

10 bearings made by a certain process have a mean diameter of 0.5060 cm with S.D of 0.0040 cm. Assume that the data may be looked upon as a random sample from a normal population. Construct a 95% confidence interval for the actual average diameter of bearings made by this process.

Answer

Here, $n = 10, \bar{X} = 0.5060, s = 0.0040.$
$(1-\alpha) = \frac{95}{100} = 0.95$. Hence $\frac{\alpha}{2} = 0.025.$
Therefore, $t_{n-1,\frac{\alpha}{2}} = t_{9,0.025} = 2.262.$
Hence a 95% confidence interval for the average diameter

$$
\begin{aligned}
&= (\bar{X} - t_{n-1,\frac{\alpha}{2}}\frac{s}{\sqrt{n}}, \bar{X} + t_{n-1,\frac{\alpha}{2}}\frac{s}{\sqrt{n}}) \\
&= (0.5060 - 2.262(\frac{0.0040}{\sqrt{10}}), 0.5060 + 2.262(\frac{0.0040}{\sqrt{10}})) \\
&= (0.5031, 0.5089)
\end{aligned}
$$

Question 20

Inspecting ceramic tiles prior to their shipment, a quality control engineer detect $2, 3, 6, 0, 4$ and 9 defectives in 6 cartons each containing 144 tiles. Assuming the data may be looked upon as a random sample from a population which can be approximated closely by a Normal distribution, what can he assert with 99% confidence about the maximum error if he uses the mean of sample as a point estimate, of true average number of defective per carton?

Answer

For the given data, $\overline{\mathrm{X}} = \frac{2+3+6+0+4+9}{6} = 4$
$s^2 = \frac{1}{5}\left[2^2 + 1^2 + 2^2 + 4^2 + 0^2 + 5^2\right] = 10$
$t_{0.005}$ with 5 degrees of freedom is 4.032.
Therefore, maximum error with 99% confidence is
$\frac{s}{\sqrt{n}} \times t_{0.005(5)} = \frac{\sqrt{10}}{\sqrt{6}4.032} = 5.205$

Question 21

Show that the sample mean $\bar{x}$ is an unbiased estimator of $\frac{1}{\theta}$ for the distribution function, $f(x) = \theta(1-\theta)^{x-1}$, $x = 1, 2, \ldots$, where $0 < \theta < 1$.

Answer

We have to show that $E(\bar{x}) = \frac{1}{\theta}$.

$$E(\bar{x}) = E\left(\frac{\sum_{i=1}^{n} x_i}{n}\right) = \frac{1}{n}\sum_{i=1}^{n} E(x_i).$$

$$\begin{aligned} E(x) &= \sum_{x=1}^{\infty} x\theta(1-\theta)^{x-1} \\ &= \theta(1 + 2(1-\theta) + 3(1-\theta)^2 + \cdots) \\ &= \theta \times \frac{1}{(1-(1-\theta))^2} \\ &= \frac{1}{\theta} \end{aligned}$$

(Because, for $|x| < 1$, $1 + 2x + 3x^2 + \cdots = \frac{1}{(1-x)^2}$. Here, $0 < \theta < 1 \Rightarrow |1-\theta| < 1$.) Hence, $E(\bar{X}) = \frac{1}{n} \times n \times \frac{1}{\theta} = \frac{1}{\theta}$.

Question 22

$X_1, X_2, X_3, \ldots X_n$ is a random sample from Bernoulli random variable X taking value 1 with probability θ and 0 with probability $1-\theta$. Show that $\frac{T(T-1)}{n(n-1)}$ is an unbiased estimator of θ^2 where $T = \sum_{i=1}^{n} X_i$

Answer: Here, $T = \sum_{i=1}^{n} X_i$, which means $T \sim B(n, \theta)$, binomial with parameters n and θ. So,

$$E(T) = n\theta$$
$$V(T) = n\theta(1-\theta)$$

and $E\left(T^2\right) = n\theta + n(n-1)\theta^2$

We have to prove that $\frac{T(T-1)}{n(n-1)}$ is an unbiased estimator of θ^2. That is we have to show that $E\left(\frac{T(T-1)}{n(n-1)}\right) = \theta^2$. Now, consider $E\left(\frac{T(T-1)}{n(n-1)}\right) = \frac{1}{n(n-1)} E(T(T-1))$

$$= \frac{E\left(T^2 - T\right)}{n(n-1)}$$
$$= \frac{E\left(T^2\right) - E(T)}{n(n-1)}$$
$$= \frac{n\theta + n(n-1)\theta^2 - n\theta}{n(n-1)} = \theta^2$$

Hence proved.

Question 23

Prove that the sample variance $S^2 = \sum \frac{(x_i - \bar{x})^2}{n-1}$ is an unbiased estimator of the population variance σ^2.

Answer

Let $x_1, x_2, ...x_n$ be n independent observations from a population with mean μ and variance σ^2 $E(x_i) = \mu$ $Var(x_i) = \sigma^2$.

We will use,

- (1) $E(\sum x_i) = \sum E(x_i)$.
- (2) $E(cx_i) = cE(x_i)$
- (3) $V(X) = E(X^2) - [E(X)]^2$
- (4) $V(\bar{x}) = E(\bar{x}^2) - [E(\bar{x})]^2$
- (5) $E(X) = \sigma^2 + \mu^2$
- (6) $E(\bar{x}^2) = \frac{\sigma^2}{n} + \mu^2$

We have to show that $E(S^2) = \sigma^2$

Now,

$E(\sum(x_i - \bar{x})^2) = E(\sum(x_i^2 - 2x_i\bar{x} + \bar{x}^2))$

$= E(\sum x_i^2 - \sum 2x_i\bar{x} + \sum \bar{x}^2)$

$= E(\sum x_i^2 - 2\bar{x}\sum x_i + n\bar{x}^2)$

since $\bar{x} = \frac{\sum x_i}{n}, \sum x_i = n\bar{x}$

$E(\sum(x_i - \bar{x})^2) = E(\sum x_i^2 - 2\bar{x}.n\bar{x} + n\bar{x}^2)$

$= E(\sum x_i^2 - n\bar{x}^2)$

Answer Contd

$= \sum E(x_i^2) - E(n\bar{x}^2)$
$= \sum E(x_i^2) - nE(\bar{x}^2)$
$= \sum(\sigma^2 + \mu^2) - n(\frac{\sigma^2}{n} + \mu^2)$
$= n\sigma^2 + n\mu^2 - \sigma^2 - n\mu^2$
$= (n-1)\sigma^2$
Now, $E(S^2) = E(\frac{\sum(x_i - \bar{x})^2}{n-1})$
$= \frac{1}{n-1}E(\sum(x_i - \bar{x})^2)$
$= \frac{1}{n-1}(n-1)\sigma^2 = \sigma^2$
Thus $E(S^2) = \sigma^2$
Hence proved.

Question 24

Consider the problem of testing hypothesis $H_0 : \mu = 880$ against $H_1 : \mu \neq 880$ and $\bar{X} = 871, S = 21$ and $n = 50$ using $\alpha = 0.01$. Find β when the actual mean μ is equal to 870 and 875.

Answer

$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{21}{\sqrt{50}} = 2.97$
$\frac{\alpha}{2} = 0.005$
We fail to reject the null hypothesis (occur a type II error) if we get a z statistic greater than -2.58.
$P(-2.58 \leq z \leq 2.58) = P(\bar{X} \geq \bar{X}_{critical} | \mu = 880, \sigma_{\bar{X}} = 2.97) = 0.995$
$$-2.58 = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}}$$
$\Rightarrow \bar{x}_{critical} = 872.3374$
If μ_1 is 870 or 875, $H_1 : \mu < 880$
Computing the probability of getting a sample mean greater than $\bar{x}_{critical} = 872.3374$
When $\mu_1 = 870$
$P(\bar{x} > 872.3374 | \mu = 870, \sigma_{\bar{x}} = 2.97) = P(z > \frac{872.3374 - 870}{2.97}$
$= P(z > 0.787) = 0.78230$
When $\mu_1 = 875$
$P(\bar{x} > 872.3374 | \mu = 875, \sigma_{\bar{x}} = 2.97) = P(z > \frac{872.3374 - 875}{2.97}$
$= P(z > 0.89) = 1 - 0.181673 = 0.81327$

Question 25

Consider the problem of testing the hypothesis $H_0 : \mu = 2.3$ against $H_1 : \mu > 2.3$ using $\alpha = 0.05$. Find the values of $\bar{x}$ for rejecting the true H_0 given $n = 36$ and $s = 1.2$. Calculate β when the actual mean is equal to $2.4, 2.5$ and 2.6.

Answer

Test the hypothesis, $H_0 : \mu = 2.3$ against $H_1 : \mu > 2.3$ using $\alpha = 0.05$.
Given $n = 36$ (Large sample) and $s = 1.2$
We have to find the value of $\bar{x}$ for rejecting the true H_0.
Since σ is unknown use the test statistic, $Z = \frac{\bar{x}-\mu}{s/\sqrt{n}} \sim N(0, 1)$
We reject H_0 if $Z > Z_\alpha$.
Here $Z = \frac{\bar{x}-2.3}{1.2/\sqrt{36}}$ and $Z_{0.05} = 1.645$

$$\frac{\bar{x} - 2.3}{1.2/\sqrt{36}} > 1.645$$

$$\bar{x} > 2.629$$

Answer Contd

$\beta = p(\text{type II error}) = p(\text{Accept } H_0 | H_1 \text{ is true})$
We have to calculate β for actual values of $\mu = 2.4, 2.5, 2.6.$
For $\mu = 2.4$
$\beta = p(\bar{x} \leq 2.629 | \mu = 2.4) = p(Z \leq \frac{2.629-2.4}{1.2/\sqrt{36}}) = p(Z \leq 1.145)$
For $\mu = 2.5$
$\beta = p(\bar{x} \leq 2.629 | \mu = 2.5) = p(Z \leq \frac{2.629-2.5}{1.2/\sqrt{36}}) = p(Z \leq 0.645)$
For $\mu = 2.6$
$\beta = p(\bar{x} \leq 2.629 | \mu = 2.6) = p(Z \leq \frac{2.629-2.6}{1.2/\sqrt{36}}) = p(Z \leq 0.145)$

Question 26

Suppose that 100 tires of a certain brand lasted on the average 21431 miles cost an SD of 1295 miles using 0.05 level of significance, test the null hypothesis $\mu = 22000$ miles against $\mu < 22000$.

Answer

Level of significance, $\alpha = 0.05$

n = 100, σ= 1295, $\bar{X}$ = 21431, $Z_{0.05}$=1.645

$$H_0 : \mu = 22000$$
$$H_1 : \mu < 22000$$

Here the test static, $Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$
Reject H_0 if $Z < -Z_\alpha$
Accept H_0 if $Z \geq -Z_\alpha$

$$Z = \frac{21431 - 22000}{129.5}$$
$$= -4.3938$$

$\therefore$ reject H_0, since $Z = -4.3938 < -1.645$.

Question 27

The burning rate of a rocket propellant is being studied. Specification requires that the mean burning rate must be 40 units and the variance of burning rate be 4 units. Test, $H0 : \mu = 40$ against $H1 : \mu \neq 40$ based on a random sample of size 25, $\bar{x} = 41.25$ and $\alpha = 0.05$.

Answer

The hypothesis is denoted by,

$$H0 : \mu = 40$$

$$H1 : \mu < 40 \quad \text{or} \quad \mu > 40$$

The alternative hypothesis here is two sided or two tailed as it goes to both the tails of the distribution of burning rates whatever be it.
The sample information,

$$\bar{x} = 41.25$$

$$\sigma = 2$$

Let us test the above hypothesis by a 95% CI. Since the sample size is large, an approximate z-interval would be the appropriate confidence interval.

Conti...

A 95% CI for μ is given by,

$$\begin{aligned}\bar{x} \mp \frac{Z_{\frac{\alpha}{2}}\sigma}{\sqrt{n}} &= 41.25 \mp \frac{2Z_{0.025}}{\sqrt{25}} \\ &= 41.25 \mp \frac{2 \times 1.96}{5} \\ &= 41.25 \mp 0.784 \\ &= [40.466, 42.034]\end{aligned}$$

The above interval provides possible values of μ. Since $\mu = 40$, the hypothesized value of the parameter is not in the 95% CI, we reject the null hypothesis. Equivalently, we say that the alternative hypothesis is accepted at 0.05 level of significance.

Question 28

A random sample of size 16 taken from a Normal population whose variance is known to be 900 is used to test $H_0 : \mu = 350$ against $H_1 : \mu > 350$. If the critical region is $\bar{x} > 365$,
a) Calculate the significance level of the test
b) Find the probability of Type II error if $\mu = 372.5$

Answer

Given $X\ N(\mu, 900)$. Then $\bar{x} \sim N(\mu, \frac{900}{16})$.
$a)$

$$
\begin{aligned}
\alpha &= P(Reject H_0/H_0) \\
&= P(\bar{x} > 365/\mu = 350) \\
&= P(z > \frac{365 - 350}{\sqrt{900/16}}) \\
&= P(z > 60/30) \\
&= 0.0228
\end{aligned}
$$

Answer

$b)$

$$
\begin{aligned}
P(TypeII) &= P(acceptH_0/H_1) \\
&= 1 - P(RejectH_0/H_1) \\
&= 1 - P(\bar{x} > 365/\mu = 372.5) \\
&= 1 - P(z > \frac{365 - 372.5}{30/4}) \\
&= 1 - P(z > -1) \\
&= 0.1587
\end{aligned}
$$

Question 29

Suppose that for a given population with $\sigma = 8.4$, we want to test the null hypothesis $\mu = 80$ against the alternative hypothesis $\mu < 80$ on the basis of a random sample of size n $= 100$. If the null hypothesis is rejected for $\bar{x} < 78$ and otherwise it is accepted, what is the probability of Type I error?

Answer

When n=100, the sample size is large.
Therefore by CLT,
$\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$
P[rej $H_0|H_0$ is true] $= \alpha$
p[$\bar{x} < 78|\mu = 80] = \alpha$
p[$z < \frac{78-80}{8.4/10}] = \alpha$
p[$z < -2.38] = \alpha$
$\therefore \alpha = 0.008634$

Question 30

A random sample of six steel beams has a mean compressive strength of $58,392$ psi (pounds per square inch) with a standard deviation of 648 psi. Use this information and the level of significance $\alpha = 0.05$ to test whether the true average compressive strength of the steel from which the sample came is 58000. Assume distribution of strengths to be normal.

Answer

We have $H_0 : \mu = 58000$, $H_1 : \mu \neq 58000$ and level of significance $\alpha = 0.05$. By assuming the population as normal we can use the t statistic

$$t = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}.$$

Since the alternative hypothesis is two sided, the critical region is defined by $t < -t_{0.025}$ or $t > t_{0.025}$, where $t_{0.025}$ with 5 degrees of freedom is 2.571. In this case $\mu_0 = 58000$, $\bar{X} = 58392$, $S = 648$ and $n = 6$ so

$$t = \frac{58392 - 58000}{648/\sqrt{6}} = 1.48$$

Since $1.48 < 2.571$, we can not reject the null hypothesis at the 0.05 level of significance.

Question 31

Given a random sample of 5 cups of ice cream from different production lots. We want to test whether the fat content of the certain kind of ice cream exceeds 14%. What can we conclude at 0.01 level of significance about the null hypothesis $\mu = 14\%$ if the sample has the mean $\overline{x} = 14.9\%$ and S.D $= 0.42\%$?

Answer

We have $H_0 : \mu = 0.14$
$H_1 : \mu > 0.14$.
Given level of significance is 0.01. Since the alternative hypothesis is one sided, the critical region is $t > t_{0.01}$ where $t_{0.01}$ with 4 degrees of freedom is 3.747.
Here $\mu_0 = 0.14$, $n = 5$, $s = 0.0042$, $\overline{x} = 0.14$.
$t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} = \frac{0.149 - 0.14}{0.0042/\sqrt{5}} = 4.79$.
Since the calculated t value $= 4.79 > 3.747 =$ table value, we reject the null hypothesis at 0.01 level of significance.