Test 1

Total marks: 100 Time: 2 hours

Part-I: **20 X 2 = 40**

1. Following are the heights of 7 Indian males in cm: 170, 168, 178, 175, 182, 180, 181. What is the sample standard deviation of heights?

a) 5.5 b) 6.0

c) 5.0 d) 4.5

Correct Answer: a

Detailed Explanation:

The mean of the given data $= \frac{170+\cdots+181}{7} = 176.29$

Sample standard deviation $= \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2} = \sqrt{\frac{1}{7-1}[(170-176.29)^2 + \cdots + (181-176.29)^2\ ]} =$ $5.499 \approx 5.5$

2. What is the probability of getting a total of 5 when a pair of fair dice is tossed?

a) $\frac{2}{9}$ b) $\frac{1}{9}$

c) $\frac{5}{18}$ d) $\frac{1}{36}$

Correct Answer: b

Detailed Explanation:

Outcomes in favour of sum equal to 5 is {(1,4), (2,3), (3,2), (4,1)}

Total number of outcomes= $6 \times 6 = 36$

Probability $= \frac{4}{36} = \frac{1}{9}$

3. If $P(M) = 0.4$, $P(N) = 0.8$, and $P(N|M) = 0.6$, then $P(M \cup N)$ =?

a) 0.72 b) 1.2

c) 0.84 d) 0.96

Correct Answer: d

Detailed Explanation:

We know, for two events M and N, $P(M \cup N) = P(M) + P(N) - P(M \cap N)$

Again, the conditional probability $P(N|M) = \frac{P(M\cap N)}{P(M)}, where, P(M) > 0$

So, $P(M \cup N) = P(M) + P(N) - P(M \cap N) = 0.4 + 0.8 - 0.6 \times 0.4 = 0.96$

4. A probability distribution can be represented in the form of-

a) table, b) graph,

c) mathematical formula d) All of the above

Correct Answer: d

Detailed Explanation:

A probability distribution can be in the form of a table, graph, or mathematical formula.

5. Which of the following are correct for a normal distribution?

a) Median and mode are equal to mean b) Mode is equal to the variance

c) Median is greater than mean d) Mode is equal to standard deviation

Correct Answer: a
Detailed Explanation:
The mean, median and mode are all measures of central tendency in the distribution. Since the normal distribution has one central value that is symmetric about the mean, they will tend to be the same.

6. What do ANOVA calculate
   a) Z-scores
   b) T-scores
   c) F ratios
   d) Chi square

Correct Answer: c
Detailed Explanation:
The F-ratio is the ratio of the between group variance to the within group variance. The F-ratio is used in an ANOVA (Analysis of Variance) that provides more insight into data compared to using only the mean or median.

7. In a study, subjects are randomly assigned to one of three groups: experimental A, experimental B, or experimental C. After treatment, the mean scores for the three groups are compared. The appropriate statistical test for comparing these means is:
   a) The correlation coefficient
   b) Chi square
   c) The t - test
   d) The analysis of variance

Correct Answer: d
Detailed Explanation:
The analysis of variance (ANOVA) is used to compare differences of means among more than two groups. Specifically, ANOVA compares the amount of variation between groups with the amount of variation within groups.

8. Choose the figure which shows cumulative uniform density function for

$$f(x) = \begin{cases} \dfrac{1}{(b-a)} & for\ a \le x \le b \\ 0, & elsewhere \end{cases}$$

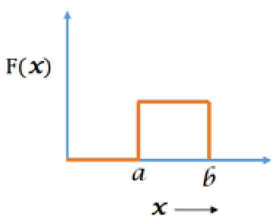


Figure (a)

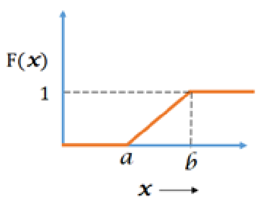


Figure (b)

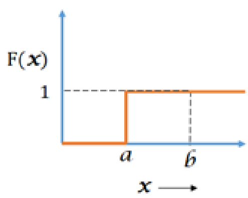


Figure (c)

   a) Figure (a) is the cumulative uniform distribution function
   b) Figure (b) is the cumulative uniform distribution function
   c) Figure (c) is the cumulative uniform distribution function
   d) None of the above

Correct Answer: b
Detailed Explanation:
The cumulative distribution function of a random variable to be calculated at a point x is represented as F(X). It is the probability that the random variable X will take a value less than or equal to x.

9. In the following context, which denotes a random variable?
There is a box containing 100 balls: 30 red, 20 blue and 50 black balls.
   a) Probability that we draw two blue balls or two red balls from the box.
   b) Drawing any 5 balls at random.
   c) The number of red, blue and black balls drawn from the box.
   d) Drawing five red and six black balls from the box.

Correct Answer: c
Detailed Solution:
The different number of red, blue, and black balls drawn from the box can be represented as a random variable.

10. Assume that the set of the ordered pairs $(x, f(x))$ is probability distribution of a discrete random variable X. Which of the following statement is NOT true?
   a) $f(x) \geq 0$
   b) $-\infty < f(x) < +\infty$
   c) $\sum_x f(x) = 1$
   d) $P(X = x) = f(x)$

Correct Answer: b
Detailed Solution:
From the definition of the probability distribution of a discrete random variable, we know $f(x) \geq 0, \sum_x f(x) = 1, and\ P(X = x) = f(x)$. But $f(x)$ can't be negative.

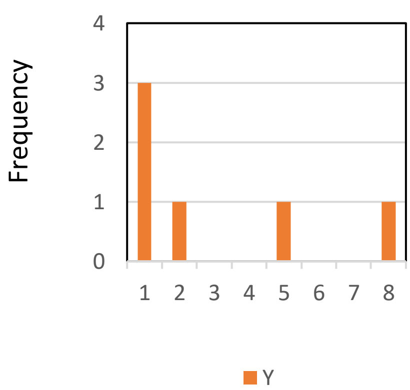
11. Let X be a discrete random variable with probability distribution $f(x)$. The mean, or expected value of X is
   a) $\mu = E(X) = \sum_x x\, f(x)$
   b) $\mu = E(X) = \int_{-\infty}^{\infty} x f(x)$
   c) $\mu = E(X) = \sum_x \frac{x}{f(x)}$
   d) $\mu = E(X) = \sum_x \frac{f(x)}{x}$

Correct Answer: a
Detailed Explanation:
The mean (also called the expected value) of a discrete random variable X is the number

$$\mu = E(X) = \sum_x x\, f(x)$$

The mean of a random variable may be interpreted as the average of the values assumed by the random variable in repeated trials of the experiment.

12. A distribution of data is shown in the figure given below.



Select the mean and median of the sample from the option given below?
   a) Mean(Y) = 2.0, Median(Y) = 2.5
   b) Mean(Y) = 2.5, Median(Y) = 2.0
   c) Mean(Y) = 3.0, Median(Y) = 1.5
   d) Mean(Y) = 3.5, Median(Y) = 1.0

Correct answer: c

Detailed Explanation:

Mean $= \frac{(1\times3+1\times(2+5+8))}{6} = 3.0$

Median of the list {1, 1, 1, 2, 5, 8} is $= \frac{(1+2)}{2} = 1.5$

13. In the following Table, Column A lists some sampling distributions, whereas Column B lists the name of sampling distributions. All symbols bear their usual meanings.

Column A		Column B	
(A)	$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$	(W)	Normal distribution
(B)	$\frac{\bar{X}-\mu}{S/\sqrt{n}}$	(X)	Chi-squared distribution
(C)	$\frac{{S_1}^2/{\sigma_1}^2}{{S_2}^2/{\sigma_2}^2}$	(Y)	t-distribution
(D)	$\frac{(n-1)S^2}{\sigma^2}$	(Z)	F distribution

Some matchings from Column A and Column B are given below. Select the correct matching?

a) (A)-(Z), (B)-(X), (C)-(Y), (D)-(Z)

b) (A)-(X), (B)-(Z), (C)-(W), (D)-(W)

c) (A)-(Y), (B)-(W), (C)-(X), (D)-(Y)

d) (A)-(W), (B)-(Y), (C)-(Z), (D)-(X)

Correct answer: d

Detailed Explanation:

χ^2 : Describes the distribution of variance.

t : Describes the distribution of normally distributed random variable standardized by an estimate of the standard deviation.

F : Describes the distribution of the ratio of two variables.

14. If the value of α, the significant level of a hypothesis test is increased, then which of the following will occur?

a) Type I error increases while Type-II error decreases.

b) Type I error decreases while Type-II error increases.

c) Both Type I and Type II errors increase.

d) Both Type I and Type II errors decrease.

Correct Answer: a

Detailed Explanation:

α denotes the probability of making Type I error and β denotes the probability of making Type II error. When α increases, the probability of Type-I errors increases. Further, when one increases other decreases. Hence, (a) is correct.

15. The variance of a binomial distribution is given by? (consider n = number of trials, p = the probability of success)

a) np^2

b) $\frac{n}{p}$

c) $np(1+p)$

d) $np(1-p)$

Correct Answer: d

Detailed Explanation:

The variance of a binomial distribution is $\sigma^2 = npq$, where $q =$ the probability of failure = $1 - p$

16. Which of the following statement is/are false as per Central Limit Theorem?
    a) The mean of the sampling distribution of the mean is the population mean.
    b) The variance of the distribution of the sample mean is $\frac{\sigma^2}{n}$.
    c) The sampling distribution becomes less variable with increased sample size.
    d) For populations with larger variances, the sample mean is a reliable estimate of the population mean.

Correct Answer: d

Detailed Explanation:

The variance of the distribution of the sample mean is $\frac{\sigma^2}{n}$. Its square root, $\frac{\sigma}{\sqrt{n}}$, is the standard deviation (Standard error of the mean) of the sampling distribution of the mean, also called standard error of the mean. That is, the more variable the population, the more variable is the sampling distribution. Therefore, we can say for populations with larger variances, the sample mean will be less reliable estimate of the population mean.

17. If $X_1, X_2, X_3 \dots \dots .. X_n$ are mutually independent random variables and each have a chi-squared distribution with v degrees of freedom, then the random variable $Y = X_1 + X_2 + X_3 + \dots \dots .. + X_n$ has a chi-squared distribution with v degrees of freedom.
    a) True
    b) False

Correct Answer: b

Detailed Explanation:

The random variable Y will have a degree of freedom $v + v + \dots + v \ (total\ n\ terms) = n \times v$

18. In a hypothesis testing, a certain population parameter μ is compared with a specific value of 100. A set of null (H_0) and alternate hypothesis (H_1) is given below. Which of the following indicates a two-tailed test?
    a) $H_0: \mu = 100 \ \ H_1: \mu > 100$
    b) $H_0: \mu = 100 \ \ H_1: \mu < 100$
    c) $H_0: \mu = 100 \ \ H_1: \mu \neq 100$
    d) None of the above

Correct Answer: c

Detailed Explanation:

An alternative hypothesis that specifies that the parameter can lie on either side of the value specified by H_0 is called a two-sided (or two-tailed) test.

Example.

$H_0: \mu = 100$ and $H_1: \mu \neq 100$

Again, a statistical test in which the alternative hypothesis specifies that the population parameter lies entirely above or below the value specified in H_0 is called a one-sided (or one-tailed) test.

19. Which of the following statements is true about the maximum error of estimation (known as margin of error)?
    a) It is indicator of the precision of an estimate and defined as one-half the width of a confidence interval
    b) It is indicator of the precision of an estimate and is equal to the width of a confidence interval
    c) It is indicator of the precision of an estimate and defined as double the width of a confidence interval
    d) None of the above

Correct Answer: a
Detailed Explanation: The maximum error of estimation is indicator of the precision of an estimate and defined as one-half the width of a confidence interval.

20. Which of the following statements is true about supervised and unsupervised classification?
    a) In supervised classification, the class label is known, but in unsupervised classification the class label of the data is unknown.
    b) In unsupervised classification, the class label is known, but in supervised classification the class label of the data is unknown.
    c) In both cases, the class label of the data is unknown.
    d) In both cases, the class label of the data is known.

Correct Answer: a
Detailed Explanation:
In supervised classification, the class label is known, but in unsupervised classification the class label of the data is unknown.

Part-II: Subjective type questions **6 X 10 = 60**

1. For a student welfare scheme, three companies A, B and C supply 25%, 35% and 40% of the cycles for the students of a school. In a particular year, it is seen that 5%, 4% and 2% of the cycles produced by these companies are defective. Now, if a cycle is found to be defective, what is the probability that the cycle was supplied by company A?

Answer:

Let A, B and C be the events that the cycles are supplied by A, B and C respectively.

Let D be the event that the cycles are defective

Given,

$P(A) = 0.25, P(B) = 0.35, P(C) = 0.40, P(D|A) = 0.05, P(D|B) = 0.04 \; and \; P(D|C) = 0.02$

$$P(A|D) = \frac{\mathrm{P(D|A)\times P(A)}}{\mathrm{P(D|A)\times P(A)+P(D|B)\times P(B)+P(D|C)\times P(C)}}$$

$$= \frac{0.05\times0.25}{0.05\times0.25+0.04\times0.35+0.02\times0.40} = \frac{0.0125}{0.0125+0.0140+0.0080} = \mathbf{0.36}$$

2. A shipment of 20 similar laptop computers to a retail outlet contains 3 that are defective. If a school makes a random purchase of 2 of these computers, find the probability distribution for the number of defectives.

Answer:

Let X be a random variable whose values x are the possible numbers of defective computers purchased by the school. Then x can only take the numbers 0, 1, and 2.

$$f(0) = P(X = 0) = \frac{\binom{3}{0}\binom{17}{2}}{\binom{20}{2}} = \frac{136}{190}$$

$f(1) = P(X=1) = \frac{\binom{3}{1}\binom{17}{1}}{\binom{20}{2}} = \frac{51}{190}$,

and $f(2) = P(X=2) = \frac{\binom{3}{2}\binom{17}{0}}{\binom{20}{2}} = \frac{3}{190}$.

Thus, the probability distribution of X is:

x	0	1	2
f(x)	$\frac{136}{190}$	$\frac{51}{190}$	$\frac{3}{190}$

3. An electrical firm manufactures light bulbs that have a lifetime that is approximately normally distributed with a mean of 800 hours and a standard deviation of 40 hours. Test the hypothesis that μ = 800 hours against the alternative, $\mu \neq 800$ hours, if a random sample of 30 bulbs has an average life of 788 hours. Use a P-value in your answer. Write the null and alternative hypothesis. What is the result of this hypothesis testing?

Answer:

The hypotheses are,

$$H_0: \mu = 800,$$

$$H_1: \mu \neq 800.$$

Now, $z = \frac{788-800}{40/\sqrt{30}} = -1.64$, and $P - value = 2P(Z < -1.64) = (2)(0.0505) = 0.1010$.

Hence, the mean is not significantly different from 800 for α < 0.101, null hypothesis is not rejected.

4. In the manufacture of car tires, a particular production process is known to yield 10 tires with defective walls in every batch of 100 tires produced. From a production batch of 100 tires, a sample of 4 is selected for testing to destruction.
Find:
   a) The probability that the sample contains 1 defective tire.
   b) The expectation of the number of defectives in samples of size 4.
   c) The variance of the number of defectives in samples of size 4.

Answer:

Given, $N = 100,\ n = 4,\ k = 10,\ x = 1$

a) $P(X=1) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} = \frac{\binom{10}{1}\binom{90}{3}}{\binom{100}{4}} = \mathbf{0.299}$

b) $E(X) = np = 4 \times 0.1 = \mathbf{0.4}$

c) $V(X) = np(1-p)\frac{N-M}{N-1} = 0.4 \times 0.9 \times \frac{90}{99} = \mathbf{0.33}$

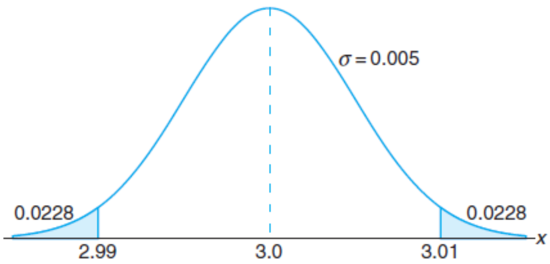
5. In an industrial process, the diameter of a ball bearing is an important measurement. The buyer sets specifications for the diameter to be 3.0 ± 0.01 cm. The implication is that no part falling outside these specifications will be accepted. It is known that in the process the diameter of a ball bearing has a normal distribution with mean $\mu = 3.0$ and standard deviation $\sigma = 0.005$. On average, how many manufactured ball bearings will be scrapped? (Note: $P(Z < -2.0) = 0.0228$)

Answer:

The values corresponding to the specification limits are $X1 = 2.99$ and $X2 = 3.01$

The corresponding z values are

$$z_1 = \frac{2.99-3.0}{0.005} = -2.0 \quad and \quad z_2 = \frac{3.01-3.0}{0.005} = +2.0$$



Hence,

$$P(2.99 < X < 3.01) = P(-2.0 < Z < 2.0)$$

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5	0.496	0.492	0.488	0.484	0.4801	0.4761	0.4721	0.4681	0.4641
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.2	0.4207	0.4168	0.4129	0.409	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.352	0.3483
-0.4	0.3446	0.3409	0.3372	0.3336	0.33	0.3264	0.3228	0.3192	0.3156	0.3121
|	|	|	|	|	|	|	|	|	|	|
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.025	0.0244	0.0239	0.0233
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-2.1	0.0179	0.0174	0.017	0.0166	0.0162	0.0158	0.0154	0.015	0.0146	0.0143
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.011
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.4	0.0082	0.008	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.5	0.0062	0.006	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048

From Table 4, $P(Z < -2.0) = 0.0228$

Due to symmetry of the normal distribution, we find that

$$P(Z < -2.0) + P(Z > 2.0) = 2(0.0228)$$
$$= 0.0456.$$

On average, 4.56% of manufactured ball bearings

will be scrapped.

6. The average zinc concentration recovered from a sample of measurements taken in 36 different locations in a river is found to be 2.6 grams per millilitre. Find the 95% confidence intervals for the mean zinc concentration in the river. Assume that the population standard deviation is 0.3 gram per millilitre. (Note: Z value for 95% confidence interval = 1.96)

Answer:

Given, estimate of μ , that is, $\bar{X} = 2.6$

Confidence co-efficient = 95%

∴ The significance level, $\alpha = 5\%$

∴ The confidence interval,

$$\left((2.6 - (1.96)\left(\frac{0.3}{\sqrt{36}}\right)\right) < \mu < \left(2.6 + (1.96)\left(\frac{0.3}{\sqrt{36}}\right)\right)$$

$$\Rightarrow 2.50 < \mu < 2.70$$

⇒ 95% Confidence Interval