$30 = 15$ (1)

Question 1

Statistic :- is the practice or science of collecting and analysing numerical data in large quantities,
⇒ It is the science of collecting, analyzing, presenting and interpreting data.

Stratified Sampling :- is a method of sampling from a population which can be partitioned in sub population
⇒ It is a type of sampling method in which the total population is divided into smaller group or strata to complete the sampling proces.

Usefullness of statistic in Engineering
- It helps in improvement of product design & testing product performance
- It helps in determining reliability and maintainability
- It helps in work out safe system of flight control for airports etc

---

- ✓ It is a critical tool for Robustness analysis
- ✓ It is use for measurement of system error analysis
- It is use for test data analysis.
- It is use for probabilistic risk assessment

statistic = a fact or piece of data obtained from a study of a large quantity of numerical data

(b) 3 red, 4 white 5 black Total = 3 + 4 + 5 = 12

P(3 balls of same colour) = P(3 red) + P(3 white) + P(3 black)

$P(3\text{ red}) = \frac{3}{12} \times \frac{3}{12} \times \frac{3}{12} = \frac{1}{64}$

$P(3\text{ white}) = \frac{4}{12} \times \frac{4}{12} \times \frac{4}{12} = \frac{1}{27}$

$P(3\text{ black}) = \frac{5}{12} \times \frac{5}{12} \times \frac{5}{12} = \frac{125}{1728}$

$P(3\text{ balls of same colour}) = \frac{1}{64} + \frac{1}{27} + \frac{125}{1728} = \frac{1}{8}$

Question 2

(2) Observed Data

	Yellow	Green	Total
Round	315	108	423
Wrinkled	101	32	133
Total	416	140	556

chi-square $\chi^2 = \sum \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}$

Calculating Expected

H_0 (Null hypothesis) :- There are no relationship btw the categorical variable.

H_A (Alternative hypothesis) There are relationship btw the categorical variable

Scanned by CamScanner

Expected Value

	Yellow	Green	total
Round	$\frac{410 \times 423}{556} = 312$	$\frac{140 \times 423}{556} = 107$	
Wrinkled	$\frac{410 \times 133}{556} = 98$	$\frac{140 \times 133}{556} = 33$	
Total			

$$\chi = \sum \frac{(O-E)^2}{E}$$

$$\chi^2 = \frac{(315-312)^2}{312} + \frac{(108-107)^2}{107} + \frac{(101-98)^2}{98} + \frac{(32-33)^2}{33} = \underline{0.16}$$

Degree of freedom $df = \sum n - 1$

$$df = (2-1)(2-1) = \underline{1}$$

we accept H_0 since $0.16 < 3.841$, We have statistically significant evidence at $\alpha = 0.05$ to show that H_0 is True

OR

Do not reject H_0
There is not enough evidence to suggest that there is no relationship btw the Categorical variable

Question 3 Binomial distribution

(i) $n = 10$ $p = \frac{1}{2} = 0.5$

$q = (1-p) = 0.5$

(i) $x = 3$

$P(3\text{ heads}) = {}^nC_x \cdot p^x \cdot q^{n-x}$

$10C3 \times 0.5^3 \times 0.5^{10-3}$

$\frac{10!}{(10-3)!\,3!} \times 0.5^3 \times 0.5^7 = \underline{0.117}$

or $\underline{11.7}\%$

(ii) Less than 3 heads

$P(<3H) = P(0) + P(1) + P(2)$

$x = 0$

$P(0\text{ Heads}) = 10C0 \times 0.5^0 \times 0.5^{10-0} = 0.00098$

$P(1\text{ Heads}) = 10C1 \times 0.5^1 \times 0.5^{10-1} = 0.0098$

$P(2\text{ Heads}) = 10C2 \times 0.5^2 \times 0.5^{10-2} = 0.044$

$P(<3\text{ head}) = 0.044 + 0.0098 + 0.00098$

$= \underline{0.0548}$

(iii) $P(>2\text{heads}) = \overset{0.117}{P(3H)} + \overset{0.205}{P(4H)} + \overset{0.246}{P(5H)} + \overset{0.205}{P(6H)} + \overset{0.117}{P(7H)} + \overset{0.044}{P(8H)}$

$+ \underset{0.0098}{P(9H)} + \underset{0.00098}{P(10H)}$

or

$P(>2\text{heads}) = 1 - P(<3\text{heads})$

$1 - 0.0548 = \underline{0.9452}$

(6) Poison Distribution

$$P(x;\mu) = \frac{[e^{-\mu}][\mu^{x}]}{x!}$$

$\mu = 4$ days

(i) $x = 6$ days

$$P(6) = \frac{[e^{-4}][4^{6}]}{6!}$$

$e = 2.718$

$= \underline{0.1042}$

(ii) $P(\leq 2) = P(0\text{ days}) + P(1\text{ day}) + P(2\text{ days})$

$P(0) = \frac{e^{-4} \times 4^{0}}{0!} = 0.0183$

$+$

$P(1) = \frac{e^{-4} \times 4^{1}}{1!} = 0.073$

$+$

$P(2) = \frac{e^{-4} \times 4^{2}}{2!} = 0.147$

$\rightarrow P(\leq 2) = \underline{0.2383}$

Question 4

x	y	xy	x^2	y^2
4	8	32	16	64
8	7	56	64	49
2	11	22	4	121
10	5	50	100	25
6	9	54	36	81
$\Sigma x =$ 30	$\Sigma y =$ 40	$\Sigma xy =$ 214	$\Sigma x^2 =$ 220	$\Sigma y^2 =$ 340

$n = 5$

$y = a + bx$

$\Sigma y = an + b\Sigma x$

$\Sigma xy = a\Sigma x + b\Sigma x^2$

$40 = 5a + 30b$

$214 = 30a + 220b$

$a = 11.9$

$b = -0.65$

$\boxed{y = 11.9 - 0.65x}$

$x = a + by$

$\Sigma x = an + b\Sigma y$

$\Sigma xy = a\Sigma y + b\Sigma y^2$

$30 = 5a + 40b$

$214 = 40a + 340b$

$a = 16.4$

$b = -1.3$

$\boxed{x = 16.4 - 1.3y}$

Scanned by CamScanner

(b) (i) $Y = ?$ $x = 12$

$y = 11.9 - 0.65x$ $11.9 - [0.65 \times 12] = \underline{\underline{4.1}}$

(ii) $x = ?$ $y = 6$

$x = 16.4 - 1.3y$ $\Rightarrow$ $16.4 - [1.3 \times 6] = 8.6$

(c) $K = 5$ car $N = 52$ cars $n = 13$ cars

P (3 of them sold will be white car)

Question 5

(a) Sample distribution

is a probability distribution of a statistic obtained from a larger number of samples drawn from a specific population.

(b) $N = 3000$ male $\quad \mu = 172.72 \quad \sigma = 7.62$ cm

$n = 80$

$s = ? \quad \bar{x}$

normal distr. $\quad Z = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}}$

(a) $s = \dfrac{\sigma}{\sqrt{n}} = \quad \sigma = 7.62$ cm

$n = 25$ students

$s = \dfrac{7.62}{\sqrt{25}} = 1.524$ cm

$\bar{x} = \mu = 172.72$ cm

(b) $\bar{x} = \mu = 172.72$ cm $\quad s = \dfrac{\sigma}{\sqrt{n}} \times \sqrt{\dfrac{N-n}{N-1}}$

$\dfrac{7.62}{\sqrt{25}} \times \sqrt{\dfrac{3000-25}{300-1}} = 1.518$ cm

975

Question 6

Interval estimate is the use of sample data to calculate an interval of possible values of an unknown population parameter; this is in contrast to point estimation which gives a single value.

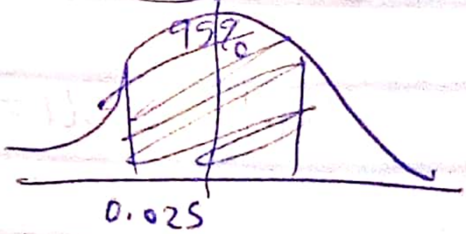
(b) $n = 100$ voters   55% supporters = 55 voters

$P = 0.55$

(a) 95% confidence limits $Z = 1.960$

Confidence limit

$$\hat{p} \pm Z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$



$$0.55 \pm 1.960 \times \sqrt{\frac{0.55(1-0.55)}{100}}$$

$0.025 + 0.95 = \boxed{0.975}$

$$0.55 \pm 0.098$$

95% confident that the proportion is between

(0.452, 0.648)

Question 7

$n = 6$

$\mu = 15700kg$ $\sigma = 320kg$

$\bar{x} = 16210kg$

(a) 0.05

(b) 0.01

We need check the hypothesis

H_0 $\mu = 15700$ vs $H_a = \mu > 15700$

we use t-test (n < 30)

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{16210 - 15700}{320/\sqrt{6}}$$

$= \underline{3.904}$

0.95 Z score for 0.05 = -1.645 — 95% CL

0.99 Z score for 0.01 = -2.325 99% CL

Scanned by CamScanner

Since $3.904 > 1.63$

we reject H_0 and conclude new technique can improve the breaking strength

(a) Hypothesis testing is an act in statistics whereby an analyst tests an assumption regarding a population parameter

Alternative hypothesis is a ~~new theory~~ position that states something is happening, a new theory is preferred instead of an old one.

Type I error is a kind of fault that occurs during the hypothesis testing process when a null hypothesis is rejected.