Problem Set - 10

① The null & Alternate Hypothesis:

H_0 : The starting position affects horse's chance of winning

H_1 : the starting position doesn't affect horse's winning chance

Test - Stat

$$\chi^2 = \Sigma \frac{(\text{observed} - \text{Expected})^2}{\text{Expected}}$$

Expected Freq = $n * p$ $\left[n = 144,\ p = 1/8,\ E = 144 \times \frac{1}{8} = 18\right]$

O	P	E	(O-E)	$(O-E)^2$	$(O-E)^2/E$
29	0.125	18	11	121	6.7222
19	0.125	18	1	1	0.0555
18	0.125	18	0	0	0
25	0.125	18	7	49	2.7222
17	0.125	18	-1	1	0.0555
10	0.125	18	-8	64	3.5555
15	0.125	18	-3	9	0.5
11	0.125	18	-7	49	2.7222
					16.333

$\chi^2 = 16.333$

degree of freedom = 8 - 1 = 7

p - value = 1 - pchisq (16.333, df = 7) = 0.02229

∴ p-value < 0.05, we reject the null hypothesis & conclude that the starting position does not affect the outcome of the horse race.

3. Null Hypothesis (H_0): Sex ratio of Panamanian sand flies do not vary with height above ground.

Alternate Hypothesis (H_1): Sex ratio of Panamanian sand flies varies with height above ground.

observed:

	3ft	35ft	Total
Male	173	125	298
Female	150	73	223
Total	323	198	521

Expected: Male 3ft : $521 \times 323/521 \times 298/521$

Male 35ft : $521 \times 198/521 \times 298/521$

Female 3ft : $323/521 \times 298/521 \times 521$

Female 35ft : $198/521 \times 223/521 \times 521$

$\Rightarrow$

	3ft	35 ft
Male	184.75	113.25
Female	138.25	84.75

$$X^2 = \sum \frac{(O-E)^2}{E} = \frac{(173-184.75)^2}{184.75} + \frac{(150-138.25)^2}{138.25} + \frac{(125-113.25)^2}{113.25} + \frac{(73-84.75)^2}{84.75}$$

$$= 0.747 + 1.219 + 0.998 + 1.493 = 4.457$$

degrees of freedom = (no. of rows -1) $\times$ (no. of cols -1) = $(2-1) \times (2-1) = 1$

p-value = 1-pchisq(4.457, df = 1) = 0.0347

$\therefore$ p-value < 0.05, there's enough evidence to reject the null hypothesis

$\therefore$ Sex ratio of panamanian sand flies varies with height above ground.

7.

H0: Response to treatment does not vary by histological type for Hodgkin's disease

H1: Response to treatment varies by histological type for Hodgkin's disease

Expected:

	Positive	Partial	None
LP	(314 × 104)/538 = 60.698	(98 × 104)/538 = 18.94	(126 × 104)/538 = 24.356
NS	(314 × 96)/538 = 56.029	(98 × 96)/538 = 17.49	(126 × 96)/538 = 22.48
MC	(314 × 266)/538 = 155.25	(98 × 266)/538 = 48.45	(126 × 266)/538 = 62.29
LD	(314 × 72)/538 = 42.022	(98 × 72)/538 = 13.11	(126 × 72)/538 = 16.86

$$\chi^2 = \sum \frac{(O-E)^2}{E} = \frac{(74-60.7)^2}{60.7} + \frac{(18-18.94)^2}{18.94} + \frac{(12-24.36)^2}{24.36} +$$

$$\frac{(68-56.03)^2}{56.03} + \frac{(16-17.49)^2}{17.49} + \frac{(12-22.48)^2}{22.48} + \frac{(154-155.25)^2}{155.25} +$$

$$\frac{(54-48.45)^2}{48.45} + \frac{(58-62.3)^2}{62.3} + \frac{(18-42.02)^2}{42.02} + \frac{(10-13.12)^2}{13.12} +$$

$$\frac{(44-16.86)^2}{16.86} = 75.89$$

$df = (r-1)(c-1) = (4-1)(3-1) = 6$

p-value $= 1 -$ pchisq$(75.89, df = 6) = 2.52 \times 10^{-14}$

$\because$ p-value is very small, we reject H0.

Hence, we can conclude that there's a great difference & the response to treatment varies by histological type for Hodgkin's disease.

5. Ho: there's no association between anger & heart disease

H1: there's significant association b/w anger & heart disea

Observed Frequencies

Anger \ Heart Disease	Present	Absent	Total
Low	53	3110 - 53 = 3057	3110
moderate	110	4731 - 110 = 4621	4731
High	27	633 - 27 = 606	633
Total	190	8284	8474

Expected Frequencies

Anger \ Heart disease	Present	Absent
Low	$\frac{3110 \times 190}{8474} = 70$	$\frac{3110 \times 8284}{8474} = 3040$
Moderate	$\frac{4731 \times 190}{8474} = 106$	$\frac{4731 \times 8284}{8474} = 4625$
High	$\frac{633 \times 190}{8474} = 14$	$\frac{633 \times 8284}{8474} = 619$

$$\chi^2 = \frac{\Sigma (O-E)^2}{E} = \frac{(53-70)^2}{70} + \frac{(3057-3040)^2}{3040} + \frac{(110-106)^2}{106}$$

$$+ \frac{(4621-4625)^2}{4625} + \frac{(27-14)^2}{14} + \frac{(606-619)^2}{619}$$

$$= 4.128 + 0.095 + 0.151 + 0.003 + 12.07 +$$

$$0.273 = 16.72$$

$df = (r-1)(c-1) = (3-1)(2-1) = 2 \times 1 = 2$

p-value $= 1 - \text{pchisq}(16.72, df = 2) = 0.0002$

$\because$ p-value < 0.05, we reject H_0 and hence there's association b/w anger & heart disease

(b) The result from the above chi-squared test ~~value~~ indicates that there's a statistically significant association b/w anger & the development of heart disease but it cannot prove causation as there are other factors that could be influencing this relationship and correlation does not imply causation. we would require further research data & experiments to establish the same.

2. a) Height Trait - tall is denoted as T, dwarf as D
leaf Trait - cut leaves as C, potato leaves as P
where T & D are the dominant & recessive traits for height respectively. C & D are the dominant & recessive traits for leaf shape respectively

E_1 (tall cut leaf): parents could be TTCC, TTCP, TDCC, TDCP.

$\rightarrow P(E_1) = P(T) \times P(C) = 3/4 \times 3/4 = 9/16$

E_2 (tall potato leaf): parents = TTPP, TDPP

$\rightarrow P(E_2) = P(T) \times P(P) = 3/4 \times 1/4 = 3/16$

E_3 (dwarf cut leaf): parents = DDCC, DDCP

$\rightarrow P(E_3) = P(D) \times P(C) = 1/4 \times 3/4 = 3/16$

E_4 (dwarf potato leaf): parents = DDPP

$\rightarrow P(E_4) = P(D) \times P(P) = 1/4 \times 1/4 = 1/16$

c) Observed frequencies:

$O_1 = 926, \quad O_2 = 288, \quad O_3 = 293, \quad O_4 = 104$

Expected freq $= n \times p$

$e_1 = 9/16 \times 1611, \quad e_2 = 3/16 \times 1611, \quad e_3 = 3/16 \times 1611,$

$e_4 = \frac{1}{16} \times 1611 \Rightarrow \begin{bmatrix} e_1 = 906.19 & e_2 = 302.06 \\ e_3 = 302.06 & e_4 = 100.67 \end{bmatrix}$

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

$$= \frac{(926 - 906.19)^2}{906.19} + \frac{(288 - 302.06)^2}{302.06} + \frac{(293 - 302.06)^2}{302.06}$$

$$+ \frac{(104 - 100.69)^2}{100.69}$$

$$= 0.433 + 0.654 + 0.271 + 0.108 = 1.466$$

p-value = 1 - pchisq(1.466, df = 3) = 0.6907

∵ p-value > 0.05, we fail to reject the null hypothesis that the observed frequencies are consistent with the expected frequencies. Hence, the cell probabilities ~~computed~~ found in part (a) are correct.

Null H_0: observed freq is consistent with expected freq.

Alternate H_1: Observed Freq is inconsistent with expected freq

Degrees of freedom: total no. of categories $- 1 = 4 - 1 = 3$