1. H_0 : start position affects horse's chances
H_1 : start position doesn't affect horse's chances

Let's calculate our test statistic

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

Let's tabulate the test statistic

O	P	E	(O-E)	$(O-E)^2$	$(O-E)^2/E$
29	0.125	18	11	121	6.72
19	0.125	18	1	1	0.05
18	0.125	18	0	0	0
25	0.125	18	7	49	2.72
17	0.125	18	-1	1	0.05
10	0.125	18	-8	64	3.55
15	0.125	18	-3	9	0.5
11	0.125	18	-7	49	2.72
					16.33

Our test statistic $\chi^2 = 16.33$

D.O.F $= 8-1 = 7$

P-value $=$ 1- pchisq(16.33, df = 7) $= 0.022$

$\therefore$ we reject the null hypothesis which means that the starting position doesn't affect the horse's chances

2. Tall -T Dwarf -D Cut - C Potato - P

We know that the probability for a recessive trait is $\frac{1}{4}$. Then, probability of a dominant trait will be $\frac{3}{4}$.

There are 4 possible combinations

$E_1 = TC$

$P(E_1) = P(T) \times P(C) = \frac{3}{4} \times \frac{3}{4} = \frac{9}{16}$

$E_2 = TP$

$P(E_2) = P(T) \times P(P) = \frac{3}{4} \times \frac{1}{4} = \frac{3}{16}$

$E_3 = DC$

$P(E_3) = P(D) \times P(C) = 1/4 \times 3/4 = \frac{3}{4}$

$E_4 = DP$

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

b. H_0 : Observed and expected frequencies are similar

H_1 : observed and expected frequencies are not similar

$$O_1 = 928, \quad O_2 = 288, \quad O_3 = 293, \quad O_4 = 104$$

We are also given $n = 1611$

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

We have the observed but we can calculate the expected value by np.

$$e_1 = \frac{9}{16} \times 1611 = 906.19$$

$$e_2 = \frac{3}{16} \times 1611 = 302.06$$

$$e_3 = \frac{3}{16} \times 1611 = 302.06$$

$$e_4 = \frac{1}{16} \times 1611 = 100.69$$

$$X^2 = \frac{(926-906.19)^2}{906.19} + \frac{(288-302.06)^2}{302.06} + \frac{(293-302.06)^2}{302.06} + \frac{(104-100.69)^2}{100.69}$$

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

$$D.O.F = \text{No. of categories} - 1 = 4 - 1 = 3$$

p-value = 1-pchisq(1.466, df = 3) = 0.6901

We fail to reject the null hypothesis. $\therefore$,

3. H_0 : Sex ratio of sandflies does not vary with ground height.

H_1 : Sex ratio of sandflies varies with height of ground

Given :

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

Expected $= n \times p$

$$P(\text{male } 3\text{ft}) = P(\text{Male}) \times P(3\text{ft})$$
$$= \frac{298}{521} \times \frac{323}{521}$$

$$E(\text{male } 3\text{ft}) = 521 \times \frac{298}{521} \times \frac{323}{521} = 184.75$$

$$P(\text{male } 35\text{ ft}) = P(\text{male}) \times P(35\text{ft})$$
$$= \frac{298}{521} \times \frac{198}{521}$$

$$E(\text{male } 35\text{ ft}) = 521 \times \frac{298}{521} \times \frac{198}{521} = 113.25$$

$$E(\text{Female } 3ft) = \frac{323}{521} \times \frac{298}{521} \times 521 = 138.25$$

$$E(\text{Female } 35ft) = \frac{198}{521} \times \frac{223}{521} \times 521 = 84.75$$

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

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

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

$$D.O.F = (\text{no. of rows} - 1)(\text{no. of columns} - 1)$$
$$= (2-1)\times(2-1) = 1$$

P-value = 1-pchisq(4.457, df=1) = 0.03

$\therefore$ We reject the null hypothesis and say that the sex ratio of sandfies varies with height of the ground.

d. H_0 : Response to treatment does not vary by histological type of Hodgkin's disease

H_1 : Response to treatment varies by histological type of Hodgkin's disease

Expected:	positive	Partial	None
LP	$\frac{314 \times 104}{538} = 60.698$	$\frac{98 \times 104}{538} = 18.94$	$\frac{126 \times 104}{538} = 24.356$
NS	$\frac{314 \times 96}{538} = 56.029$	$\frac{98 \times 96}{538} = 17.49$	$\frac{126 \times 96}{538} = 22.48$
MC	$\frac{314 \times 266}{538} = 155.25$	$\frac{98 \times 266}{538} = 48.45$	$\frac{126 \times 266}{538} = 62.29$
LD	$\frac{314 \times 72}{538} = 42.022$	$\frac{98 \times 72}{538} = 13.11$	$\frac{126 \times 72}{538} = 16.86$

$$X^2 = \frac{\Sigma(O-E)^2}{E} = \frac{(74-60.7)^2}{60.7} + \frac{(18-18.94)^2}{18.94} +$$

$$\frac{(12-24.36)^2}{24.36} + \frac{(68-56.03)^2}{56.03} + \frac{(16-17.49)^2}{17.49} + \frac{(12-22.48)^2}{22.48}$$

$$+\frac{(154-155.25)^2}{155.25}+\frac{(54-48.45)^2}{48.45}+\frac{(58-62.3)^2}{62.3}+$$

$$\frac{(18-42.02)^2}{42.02}+\frac{(10-13.12)^2}{13.12}+\frac{(44-16.86)^2}{16.86}=75.89$$

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

p-value $=$ 1-pchisq(75.87, df = 6) $=2.52\times10^{-14}$

$\therefore$ We reject H_0 and say there is a difference in the response to the treatment varies by histological type of Hodgkin's disease.

5. H_0 : There's no association b/w anger and heart disease

H_1 : There's association b/w anger and heard disease

Observed

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

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	$\frac{633 \times 190}{8474} = 14$	$\frac{633 \times 8284}{8474} = 619$

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

$$+ \frac{(110-106)^2}{106} + \frac{(4621-4625)^2}{4625} + \frac{(27-14)^2}{14}$$

$$+ \frac{(606-619)^2}{619}$$

$$= 4.128 + 0.095 + 0.151 + 0.003 + 12.07 + 0.273$$

$$= 16.72$$

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

p-value = 1- pchisq(16.72, df=2) = 0.0002

$\therefore$ We reject the null hypothesis and we say there is association b/w anger & heart disease.

(b) We cannot prove causality just by a statistical model it only helps quantify the relationship but doesn't necessarily control for every possible confounding variables in our problem. So its highly likely that there are other factors that are affecting this relations-hip. Just because there is a correlation b/w 2 aspects like anger and heart disease it isn't

to say definitely that there is a causality.
More quality data will be required to make a better conclusion on the causality.