Assignment on Chi-Square

1. A poker-dealing machine is supposed to deal cards at random, as if from an infinite deck.

In a test, you counted 1600 cards, and observed the following:

Spades	404
Hearts	420
Diamonds	400
Clubs	376

Could it be that the suits are equally likely? Or are these discrepancies too much to be random?

(1) $n=1600$

	observed	Expected	O-E	$(O-E)^2$
Spades	404	400	4	16
Hearts	420	400	20	400
Diamonds	400	400	0	0
Clubs	376	400	-24	+576
				$\Sigma(O-E)^2 =$

$$\chi^2 = \sum_{i=0}^{n} \frac{(O_i-E_i)^2}{E_i} = \frac{992}{400} = 2.48$$

$df = 4-1 = 3, \alpha = 0.05$ $\quad \chi^2_{critical} = 7.815$

$H_0: S_1 = S_2 = S_3 = S_4 = P(x) = 0.25$

H_1: H_0 is false

P-value $= 0.478$

Hence, we fail to reject null hypothesis and accept that Suits are equally likely.

2. Same as before, but this time jokers are included, and you counted 1662 cards, with these results:-

Spades	404
Hearts	420
Diamonds	400
Clubs	356
Jokers	82

a) How many jokers would you expect out of 1662 random cards? How many of each suit?

b) Is it possible that the cards are really random? Or are the discrepancies too large?

② $n = 1662$

type	observed	Expected	(O-E)	$(O-E)^2$
Spades	404	332.4	71.6	5126.56
Hearts	420	332.4	87.6	7673.76
Diamonds	400	332.4	67.6	4569.76
clubs	356	332.4	23.6	556.96
Jokers	82	332.4	-250.4	62700.16

$\Sigma(O-E)^2 = 80627.2$

$\alpha = 0.05$ $\quad df = 4$

(a) Jokers Count = 332.4 (Each Suit : 332.4)
Each Suit split will have 88 Jokers.

(b) $H_0 : P_S = P_H = P_D = P_C = P_J$

H_1 : H_0 is false.

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} = \frac{80627.2}{332.4} = 242.56$$

$\chi^2_{critical} = 9.488$ $\quad$ P-value ≤ 0.0001

Hence we reject null hypothesis
So the discrepancy is too large.

3. A genetics engineer was attempting to cross a tiger and a cheetah. She predicted a phenotypic outcome of the traits she was observing to be in the following ratio 4 stripes only: 3 spots only: 9 both stripes and spots. When the cross was performed and she counted the individuals she found 50 with stripes only, 41 with spots only and 85 with both. According to the Chi-Square test, did she get the predicted outcome?

(3) ratio = 4 : 3 : 9

	Actual (O)	Expected (E)	O-E	$(O-E)^2$
Stripes	50	44	6	36
Spots	41	33	8	64
both	85	99	-14	196
	$\Sigma O = 176$	$\Sigma E = 176$		$\Sigma (O-E)^2 = 296$

H_0 : ratio is 4:3:9

$df = 2$

$\alpha = 0.05$

H_1 : H_0 is false

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} = \frac{36}{44} + \frac{64}{33} + \frac{196}{99}$$

$$= 4.73$$

$\chi^2_{critical} = 5.991$

P value = 0.939

Hence, we accept Null hypothesis.

Yes, She got the predicted outcome.

4. In the garden pea, yellow cotyledon colour is dominant to green and inflated pod shape is dominant to the constricted form. Considering both of these traits jointly in self-fertilized dihybrids, the progeny appeared in the following numbers:-

193 green inflated, 184 yellow constricted, 556 yellow inflated, 61 green constricted

Do these genes assort independently? Support your answer using Chi-square analysis.

Note: - Genes assort independently if they follow the 9:3:3:1 rule (on the 16 square Punnett Square) resulting from a dihybrid cross.

4)

traits	observed	Expected	(O-E)	$(O-E)^2$
G I	193	559.125	-366.125	134047.52
Y C	184	186.375	-2.375	5.64
Y I	556	186.375	369.625	136622.64
G C	61	62.125	-1.125	1.266
	$\Sigma O = 994$			

ratio = 9:3:3:1 , $\alpha = 0.05$, $n = 4 - 1 = 3$.

H_0: ratio is 9:3:3:1 (no relation exists)

H_1: H_0 is false.

$$\chi^2 = \frac{134047.52}{559.125} + \frac{5.64}{186.375} + \frac{136622.64}{186.375} + \frac{1.266}{62.125}$$

$$= 972.85$$

$\chi^2_{critical} = 7.815$

p-value < 0.0001 (significant)

No, the genes don't assort independently

5. A department store, A, has 4 competitors: B, C, D and E. Store A hires a consultant to determine if the percentage of shoppers who prefer each of the five stores is the same. A survey of 1100 randomly selected shoppers is conducted and the results about which one of the stores shoppers prefer are as below. Is there enough evidence using a significance level α = 0.05 to conclude that the proportions are really the same?

Store	A	B	C	D	E
No of Shoppers	262	234	204	190	210

(5)

Store	Observed	Expected	O-E	$(O-E)^2$
A	262	220	42	1764
B	234	220	14	196
C	204	220	-16	256
D	190	220	-30	900
E	210	220	-10	100
	$\Sigma O = 1100$	$\Sigma E = 1100$		$\Sigma (O-E)^2 = 3216$

$H_0: P_A = P_B = P_C = P_D = P_E$

H_1: H_0 is false

$\alpha = 0.05$ $\quad df = 5-1 = 4$

$$\chi^2 = \frac{3216}{220} = 14.618$$

$\chi^2_{critical} = 9.488$

p-value $= 0.0055$

Hence, null hypothesis is rejected
→ Conclusion is that proportions are not same.

6. In the titanic Dataset, do a crosstab for embarked and survival rate. Using chi-square test, determine whether both of them are dependent or independent.

(6)

Embarked / Survived	C	Q	S	Total
0	75 (103.75)	47 (47.55)	427 (263.7)	549
1	93 (64.25)	30 (29.45)	217 (246.3)	340
Total	168	77	644	889

$df = 2$

$\alpha = 0.05$

$\chi^2_{critical} = 5.991$

H_0: no assoc

H_1: H_0 is false

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} = \frac{(75-103.75)^2}{103.75} + \frac{(47-47.55)^2}{47.55} + \frac{(427-263.7)^2}{263.7}$$

$$+ \frac{(93-64.25)^2}{64.25} + \frac{(30-29.45)^2}{29.45} + \frac{(217-246.3)^2}{246.3}$$

$$= 125.46 \qquad \therefore \chi^2 > \chi^2_{critical}.$$

Hence, the two variables are dependent (Rejected Null hypothesis)

7. Repeat the same experiment above with age bins and survival rate.

Note: - Age column and survival cannot be used for Chi-Square.

(7)

Age_bins / Survived	Toddler	Child	Adult	Elderly	Total
0	9 (14.25)	43 (52.85)	365 (352.15)	7 (4.75)	424
1	15 (9.75)	46 (36.15)	228 (240.85)	1 (3.25)	290
Total.	24	89	593	8	714

H_0 : no association (independent)
H_1 : association (Dependent)

$\alpha = 0.05$, $df = (4-1)(2-1) = 3$ $\chi^2_{critical} = \cancel{7}.815$

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} = 13.04$$

$$\chi^2 > \chi^2_{critical}$$

Hence, null hypothesis is rejected and conclude that age-bins and survival are dependent.