Problem Set 11

(Vinay Kumar Ranganath Babu)

 # Problem1

observed = c(29,19,18,25,17,10,15,11)

expected = rep(sum(observed)/length(observed),8)

X2 = sum((observed-expected)^2/expected)

p.value = 1 - pchisq(X2,df=7)

Values	
expected	num [1:8] 18 18 18 18 18 18 18 18
observed	num [1:8] 29 19 18 25 17 10 15 11
p.value	0.0222394774623906
X2	16.3333333333333

Since our p-value is on the lower side of 0.02, we can reject the null hypothesis. Hence horse's starting position does not affect its chance of winning.

 # Problem2

observed = c(121,84,118,226,226,123)

prapos = c(0.13,0.14,0.13,0.24,0.20,0.16)

expected = prapos * sum(observed)

X2 = sum((observed-expected)^2/expected)

p.value = 1 - pchisq(X2,df=length(observed)-1)

Values	
expected	num [1:6] 117 126 117 216 180 ...
observed	num [1:6] 121 84 118 226 226 123
p.value	1.860202696502e-05
prapos	num [1:6] 0.13 0.14 0.13 0.24 0.2 0.16
X2	29.4874007766547

Since our p-value is too small, we can reject the null hypothesis. Hence the claimed proportions are not credible in light of the provided data.

 # Problem3

n=1000

observed=c(30, 93, 159, 184, 195, 171, 92, 45, 31)

expected=c(sum(dbinom(0:1, 16, .29)*n),dbinom(2, 16, .29)*n,dbinom(3, 16, .29)*n,dbinom(4, 16, .29)*n,dbinom(5, 16, .29)*n,dbinom(6, 16, .29)*n, dbinom(7, 16, .29)*n,
dbinom(8, 16, .29)*n, sum(dbinom(9:16, 16, .29)*n))

X2=sum((observed-expected)^2/expected)

G2=2 * sum(observed * log(observed/expected))

p.valueX = 1-pchisq(X2, 1)

p.valueG = 1-pchisq(G2, 1)

Values	
expected	num [1:9] 31.4 83.5 159.1 211.2 207.1 ...
G2	11.9693183070806
n	1000
observed	num [1:9] 30 93 159 184 195 171 92 45 31
p.valueG	0.000540837238129144
p.valueX	0.00038265779331359
X2	12.6150280615715

The Test static for LR is 11.96932 and P-value 0.000540

The Test static for Pearson's is 12.6150 and P-value 0.00038

Hence we can reject null hypothesis that the researchers' coloring algorithm was not performed as intended.

 # Problem4

a) log(x+1) – log(x)

Log(2) – log(1) + log(3) – log(2)+ .....+log(10) – log(9)-log(1) + log(10)

= 1

Hence verified.

b) observed = c(107,55,39,22,13,18,13,23,15)

n = sum(observed)

p = log10(1+1/(1:9))

expected = n*p

X2 = sum((observed-expected)^2/expected)

p.valueX = 1-pchisq(X2 , 8)

using likelihood ratio

G2 = 2 * sum(observed * log(observed/expected))

p.valueY = 1-pchisq(G2 , 8)

Values	
expected	num [1:9] 91.8 53.7 38.1 29.6 24.2 ...
G2	15.5558864933756
n	305
observed	num [1:9] 107 55 39 22 13 18 13 23 15
p	num [1:9] 0.301 0.1761 0.1249 0.0969 0.0792 ...
p.valueX	0.0639909378017312
p.valueY	0.0491962228990812
X2	14.7596477007757

We get values of 0.049 approximately 0.05.

This is not very clear, however we can say leading digits hold Bedford's law and doesn't say how the law can't be too sure with this one.

 # Problem5

observed = c(173,125,150,73)

dumm = sum(observed)

expected = c((323*298/dumm),(198*298/dumm),(323*223/dumm),(198*223/dumm))

X2 = sum((observed-expected)^2/expected)

p.valueX = 1 - pchisq(X2,df=1)

G2 = 2* sum(observed * log(observed/expected))

p.valueG = 1 - pchisq(G2,df=1)

Values	
dumm	521
expected	num [1:4] 184.7 113.3 138.3 84.7
G2	4.62306318005088
observed	num [1:4] 173 125 150 73
p.valueG	0.0315448566517489
p.valueX	0.0321033323560087
X2	4.59297035792607

P-value is less than 0.05 hence we reject the null hypothesis, the panama sandflies do not vary with height above ground.

 # Problem6

positive = c(74,68,154,18)

partial = c(18,16,54,10)

none = c(12,12,58,44)

n.positive = sum(positive)

n.partial = sum(partial)

n.none = sum(none)

n = n.positive + n.partial + n.none

disease = (positive + partial + none) / n

positive.expected = disease * n.positive

partial.expected = disease * n.partial

none.expected = disease * n.none

X2 = sum((positive-positive.expected)^2/positive.expected,(partial-partial.expected)^2/partial.expected,(none-none.expected)^2/none.expected)

p.value = 1 - pchisq(X2,df=3*2)

Values	
disease	num [1:4] 0.193 0.178 0.494 0.134
n	538
n.none	126
n.partial	98
n.positive	314
none	num [1:4] 12 12 58 44
none.expected	num [1:4] 24.4 22.5 62.3 16.9
p.value	2.52020626589911e-14
partial	num [1:4] 18 16 54 10
partial.expected	num [1:4] 18.9 17.5 48.5 13.1
positive	num [1:4] 74 68 154 18
positive.expected	num [1:4] 60.7 56 155.2 42
X2	75.8901487931342

Test static is 75.9 and P-value 2.5e-14. Hence we reject null hypothesis- patient's response to treatment for Hodgkin's disease does not vary by histological.