Homework 4 – solutions 06 - 713

1. (a) Probability of winning each game = 1/2.

Let p : probability that one team will win all 17 games

q : probability that one team will lose at least 1 game

$$p = {}_{17}C_{17} \times \left(\frac{1}{2}\right)^{17} = \left(\frac{1}{2}\right)^{17}$$

$$q = 1 - p = 1 - \left(\frac{1}{2}\right)^{17}$$

The probability that all 32 teams will lose at least 1 game

is $\quad {}_{32}C_0 \times p^0 \times q^{32}$

Then, the probability that one or more teams will win all 17 games

is $\quad 1 - {}_{32}C_0 \times p^0 \times q^{32} = 1 - {}_{32}C_0 \times \left[\left(\frac{1}{2}\right)^{17}\right]^0 \times \left[1 - \left(\frac{1}{2}\right)^{17}\right]^{32}$

$$= \underline{0.000244}$$

Therefore, we can expect a "perfect season" to happen

once in ~ $\underline{4096}$ years.

(b) The probability that a team which is good enough to win 65% of its games is

$$(0.65)^{17}(1-0.65)^0 = (0.65)^{17} = \underline{0.000660}$$

(C) Oct 1

Win : 68

lose : 94

Let win = 1 and lose = 0.

$$\text{Mean} = \bar{x} = \frac{68}{162} = 0.420$$

$$\therefore \text{Variance} = s^2 = \frac{(1-0.42)^2 \times 68 + (0-0.42)^2 \times 94}{162-1}$$

$$= 0.2451$$

For a 95% confidence level, $z_{0.975} = 1.967$

$$|\bar{x} - \eta| < z_{0.975}\sqrt{\frac{s^2}{n}}$$

$$\rightarrow |0.420 - \eta| < 0.0765.$$

$$0.3435 < \eta < 0.4965$$

2. (a) Define $b = P(B)$ and $c = P(C)$. Also define $A_1 = B \cap C$, $A_2 = B \cap \bar{C}$, $A_3 = \bar{B} \cap C$, and $A_4 = \bar{B} \cap \bar{C}$

If events B and C are independent, then

$$P_1^\circ = bc, \quad P_2^\circ = b(1-c), \quad P_3^\circ = (1-b)c, \quad P_4^\circ = (1-b)(1-c)$$

The hypothesis that B and C are independent at a confidence level of $1-\alpha$ can be acceptable iff

$$= \sum_{i=1}^{4} \frac{(k_i - nP_i^\circ)^2}{nP_i^\circ} < \chi^2_{1-\alpha}(3)$$

(b) Using the notation in (a), Let B = probability that a new hire is a ChemE, C = probability that new hire leaves in first year.

Hence, $b = P(B) = 0.70$, $c = P(C) = 0.39$

$p_1^\circ = 0.273$ $\quad$ $p_2^\circ = 0.427$ $\quad$ $p_3^\circ = 0.117$ $\quad$ $p_4^\circ = 0.183$

and $\quad K_1 = 42$ $\quad$ $K_2 = 84$ $\quad$ $K_3 = 6$ $\quad$ $K_4 = 30$

Applying the χ^2 test with these values,

$$q = \sum_{i=1}^{4} \frac{(k_i - np_i^\circ)^2}{np_i^\circ} = 11.82$$

From table, $\chi^2_{0.95}(3) = 7.81$. Hence $q > \chi^2_{0.95}(3)$, so B and C are not independent at a 5% significance level

$\therefore$ My advice for CEO is to reduce the number of ChemE hired.

3. This problem depends upon the definition of the bins and then chi-squared test becomes more or less accurate depending upon that. For example, bins can be defined as $A_i = \{ i-1$ defective units observed in 1 hour$\}$ for $i = 1, 2, \cdots 7$ and then $A_8 = \bar{A}_1 \cap \bar{A}_2 \cap \bar{A}_3 \cdots \cap \bar{A}_7$

In this case, $k_1 = 2$, $k_2 = 1$, $k_3 = 2$, $k_4 = 3$, $k_5 = 5$, $k_6 = 4$, $k_7 = 2$, and $k_8 = 3$

$$q = \sum_{i=1}^{8} \frac{(k_i - np_i)^2}{np_i}$$

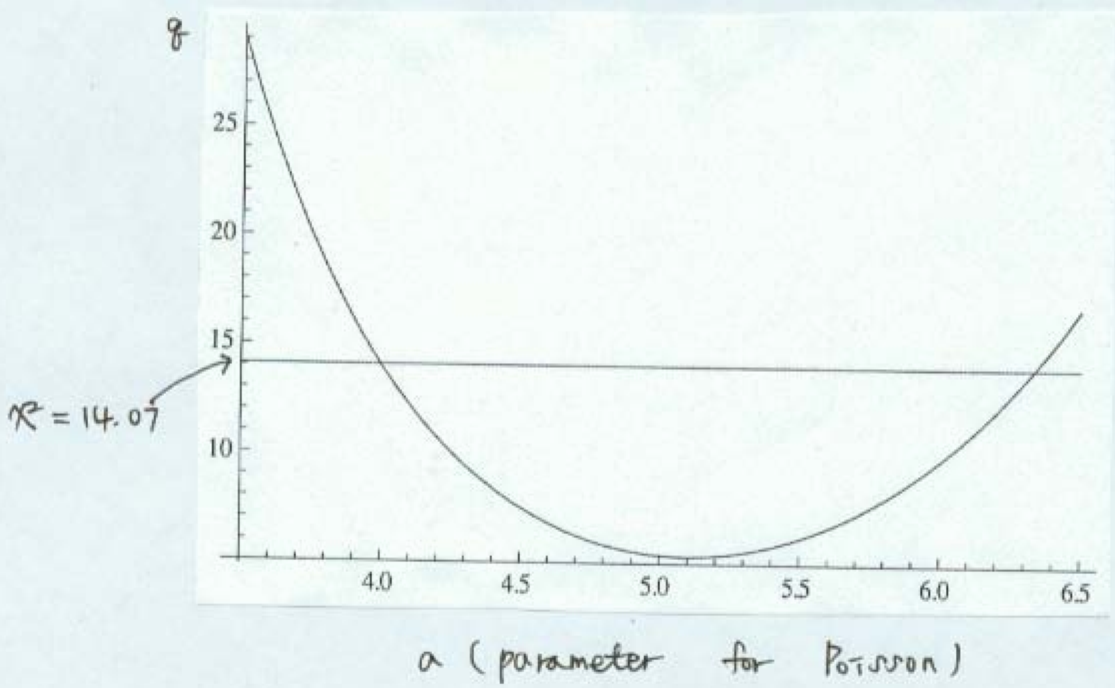
For 5% significance level, $\chi^2(7) = 14.07$. For hypothesis to be correct, q should be less than this value.

Hypothesis : Poisson distribution

$$p_i = \frac{e^{-a} a^i}{i!}$$

Thus, q can be calculated for different values of a and those values of a can be chosen for which q is less than 14.07. From the plot below, we can say that for the hypothesis to be acceptable

$$3.996 < a < 6.335$$



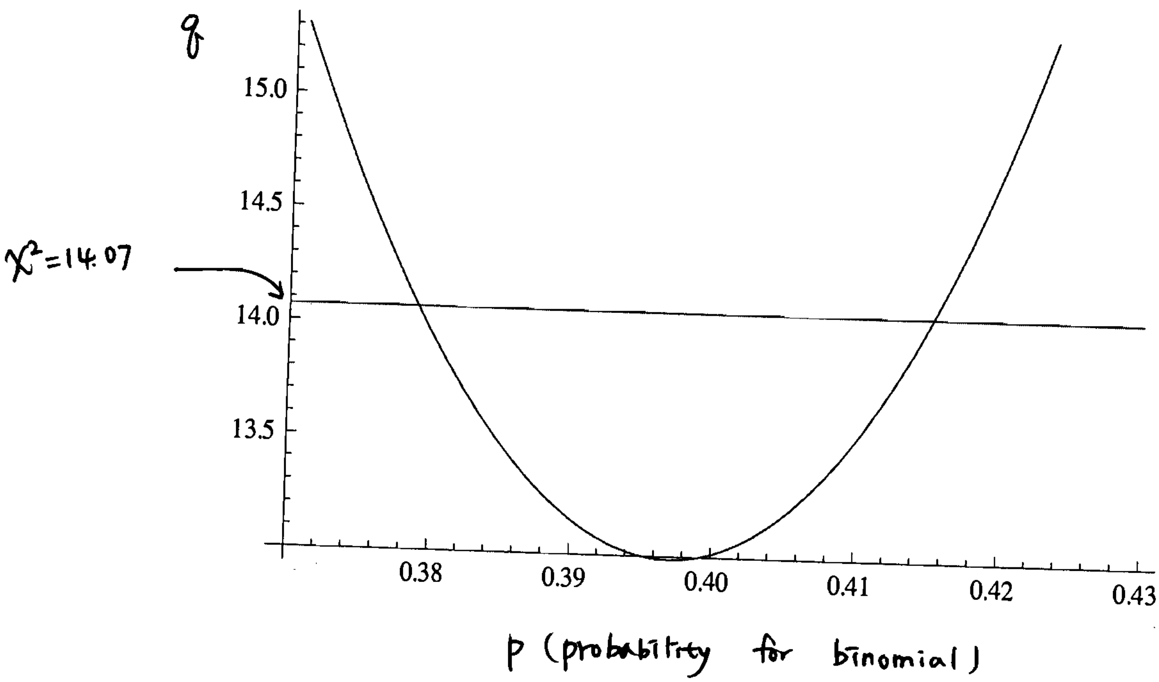
Hypothesis : Binomial distribution

In case of binomial distribution,

$$P(x=k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad k = 0, 1, \ldots, n$$

Thus, q can be calculated for different values of p and then those values of p can be chosen for which q is less than 14.07. From the plot below, we can say that for the hypothesis to be acceptable

$$0.377 < p < 0.419$$



4. (a) I analyzed these data by using "Data Analysis → Regression" of Excel

From the summary output (on the next page),

$A = 1.5031626\,B + 0.844081647\,C - 24.27895733$

$\therefore\ C_1 = 1.5031626,\ C_2 = 0.844081647,\ C_3 = -24.27895733$

◎ overall F test

⇒ $H_0 : C_1 = C_2 = 0$ (null hypothesis)

H_A : at least one $C_i \neq 0$ (alternative hypothesis)

if p value < 0.05, then H_0 is rejected at 5% level of significance

Therefore, at least one C_i is not equal to zero

H_0 means that there is no significant overall regression using all i independent variables in the model

⇒ That H_0 is rejected means that entire set of independent variables contributes significantly to the prediction of A

p value for F test is 0.0017, so H_0 is rejected

So, it seems that entire set of independent variables contributes significantly to the prediction of A.

However, the p value of C_1 is 0.167372089 > 0.05

Moreover, R^2 is only 0.25158304

The R^2 provides a quantitative measure of how well the fitted model and if the value is 1, we can say that the fit of the model is perfect $(0 \le R^2 \le 1)$

$\therefore$ the expression, $A = C_1 B + C_2 C + C_3$, is not justified by the data

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.501580805
R Square	0.251583304
Adjusted R Square	0.217564363
Standard Error	5.887945196
Observations	47

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	512.7652	256.3826	7.395389111	0.001702771
Residual	44	1525.38754	34.66789863		
Total	46	2038.15274			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-24.27895733	6.814831121	-2.7543304	0.008519249	-42.0440819	-6.513832754	-42.0440819	-6.513832754
X Variable 1	1.5031626	1.070716077	1.403885336	0.167372089	-0.654723835	3.661049034	-0.654723835	3.661049034
X Variable 2	0.844081647	0.245220972	3.442126674	0.001276949	0.349871259	1.338292035	0.349871259	1.338292035

RESIDUAL OUTPUT

Observation	*Predicted Y*	*Residuals*	*Standard Residuals*
1	10.07519127	25.81860773	4.483546082
2	9.486649824	13.00337218	2.258108531
3	12.24898778	6.978100219	1.211786252
4	7.305906202	10.0912548	1.75240301
5	5.710243256	6.696048744	1.162806431
6	4.670753027	6.788257973	1.178819081
7	5.683651004	4.428968996	0.769115314
8	10.67645631	-1.847073312	-0.320754643
9	11.1146815	-2.415802501	-0.419517658
10	6.501140399	2.040080601	0.354271442
11	8.58243656	-1.12260956	-0.194947448
12	7.83085526	-1.41435026	-0.245609858
13	6.56473279	-0.36094479	-0.062680088
14	6.171599922	-0.270637922	-0.046997794
15	5.58074277	0.17387523	0.030194409
16	8.463343961	-2.719509961	-0.472258162
17	12.12757948	-6.569521477	-1.140834261
18	2.447232789	3.061793211	0.531697569
19	3.97698764	1.12397036	0.195183759
20	7.355629933	-2.754226933	-0.478286959
21	6.424824417	-1.894642417	-0.329015285
22	8.236711719	-3.720941719	-0.646162405
23	6.303416113	-2.105348113	-0.365605511
24	5.177202016	-1.031280016	-0.17908756
25	3.929579615	-0.334993615	-0.058173521
26	4.409436351	-0.885683351	-0.153838614
27	5.485926718	-1.982322718	-0.344241461
28	7.308221907	-4.405255907	-0.764997403
29	4.27762016	-1.52499116	-0.264823271
30	2.130415904	0.605086096	0.105076595
31	8.280658971	-5.891619971	-1.023112862
32	6.210915766	-3.876144766	-0.673114285
33	4.72856894	-2.82527994	-0.490625712
34	6.253692383	-4.438338383	-0.770742361
35	1.436650517	0.310563483	0.053931091
36	2.547825319	-1.230572319	-0.213695787
37	5.4940189	-4.1945189	-0.72840174
38	8.273711857	-7.010871857	-1.217477231
39	3.280906549	-2.200274549	-0.38209002
40	-0.770685357	1.697550357	0.294789143
41	5.749559099	-4.919195099	-0.854245827
42	-0.196012569	0.966895569	0.16790684
43	3.689078712	-3.033643712	-0.52680925
44	-2.524756746	3.159122746	0.548599388
45	1.489019247	-0.847625247	-0.14719488
46	10.26019197	-9.706668967	-1.685617523
47	-0.085012152	0.591541152	0.10272444

PROBABILITY OUTPUT

Percentile	*Y*
1.063829787	0.506529
3.191489362	0.553523
5.319148936	0.621394
7.446808511	0.634366
9.574468085	0.655435
11.70212766	0.770883
13.82978723	0.830364
15.95744681	0.926865
18.08510638	1.080632
20.21276596	1.26284
22.34042553	1.2995
24.46808511	1.317253
26.59574468	1.747214
28.72340426	1.815354
30.85106383	1.903289
32.9787234	2.334771
35.10638298	2.389039
37.23404255	2.735502
39.36170213	2.752629
41.4893617	2.902966
43.61702128	3.503604
45.74468085	3.523553
47.87234043	3.594586
50	4.145922
52.12765957	4.198068
54.25531915	4.51577
56.38297872	4.530182
58.5106383	4.601403
60.63829787	5.100958
62.76595745	5.509026
64.89361702	5.558058
67.0212766	5.743834
69.14893617	5.754618
71.27659574	5.900962
73.40425532	6.203788
75.53191489	6.416505
77.65957447	7.459827
79.78723404	8.541221
81.91489362	8.698879
84.04255319	8.829383
86.17021277	10.11262
88.29787234	11.459011
90.42553191	12.406292
92.55319149	17.397161
94.68085106	19.227088
96.80851064	22.490022
98.93617021	35.893799

(b) Assuming normal distribution,

$$\overline{X} - Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \eta < \overline{X} + Z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

, where η = true mean, $\overline{X}$ = sample mean, σ^2 = true variance, n = sample size

Suppose that $\gamma = 1 - \alpha = 0.95$.

$$\alpha = 0.05$$

$$1 - \frac{\alpha}{2} = 0.975 \Rightarrow Z_{0.975} = 1.967$$

$$\overline{X} - 1.967 \frac{\sigma}{\sqrt{n}} < \eta < \overline{X} + 1.967 \frac{\sigma}{\sqrt{n}}$$

A : $3.842435 < \eta < 7.662096$

B : $6.387729 < \eta < 6.854824$

C : $22.76749 < \eta < 24.80698$

$$\frac{(n-1)\,s^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)} < \sigma^2 < \frac{(n-1)\,s^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}$$

$$\left.\begin{aligned} \chi^2_{1-\frac{\alpha}{2}}(n-1) &= \chi^2_{0.975}(46) = 66.6 \\ \chi^2_{\frac{\alpha}{2}}(n-1) &= \chi^2_{0.025}(46) = 29.2 \end{aligned}\right\}$$

by Interpolation from Chi-square percentiles table

A : $30.60289 < \sigma^2 < 69.79975$

B : $0.457638 < \sigma^2 < 1.043792$

C : $8.72481 < \sigma^2 < 19.89974$

(C) C can be categorized into 9 bins with width 2

Bin #	1	2	3	4	5	6	7	8	9
Range	15-17	17-19	19-21	21-23	23-25	25-27	27-29	29-31	31-33
Frequency	2	3	3	10	13	8	4	2	2

Mean = 5.22 variance = 16.69.

From the normal distribution $\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\eta)^2}{2\sigma^2}\right)$,

we can get $q = \Sigma \cdot \frac{(k_i^\circ - np_i^\circ)^2}{np_i^\circ}$

Bin #	Freq.	Pi	$(Ki-nPi)^2/nPi$
1	2	0.057245773	0.177235477
2	3	0.071544207	0.039095783
3	3	0.084215341	0.231927195
4	10	0.09336705	7.176371712
5	13	0.097494867	15.46363301
6	8	0.095886078	2.707895305
7	4	0.088820874	0.007300986
8	2	0.077492598	0.740403889
9	2	0.063678246	0.329383949
Mean	5.22		q
Variance	16.69		26.87

$q = 26.87$

$\chi^2_{0.95}(8) = 15.51$

$\therefore \quad q > \chi^2_{0.95}$

↳ normal distribution is <u>not</u> acceptable.

From Poisson distribution $P_i^\circ = \frac{e^{-a} a^k}{k!}$

Bin #	Freq.	Pi	$(Ki-nPi)^2/nPi$
1	2	0.028175592	0.344824051
2	3	0.073569602	0.060603939
3	3	0.128065604	1.514327652
4	10	0.16719676	0.583730997
5	13	0.174627727	2.798418129
6	8	0.1519908	0.102676494
7	4	0.113389962	0.331582781
8	2	0.074018447	0.628666646
9	2	0.042948976	0.00017142
Mean	5.22		q
Variance	16.69		6.37

$q = 6.37$

$q < \chi^2_{0.95} = 15.51$

→ The hypothesis that the data follows Poisson distribution is accepted.