1. Data Description

Data source: 衛生署新竹醫院

Case client: Dr. 黃

Case analyst: 吳漢銘 (交大統計所)

#observations: data set-1: 103 subjects (including Duke's "A")

data set-2: 96 subjects(not involve Duke's "A")

Categorical Variables:

Variable	Pre-CRP	Post-CRP	Sex	Differentiation	Duke.s	LN	Liver mets
Levels	L,H	L,H	M,F	M,W,P,MW	A,B,C,D	N,Y	N,Y

Numerical Variables:

Variable	Age	CEA	Survival

Study period: 84/9~88/9, total 49 months.

Part of formatted data:

No.	Survival(months)	Event(death)	Pre-CRP	Differentiation	Duke.s
1	18.00	1 (death)	L	M	B
2	17.00	0 (censored)	L	W	C
3	11.00	0	L	M	C
4	10.00	0	L	M	C
5	17.00	0	L	P	B
….					

2. Independence Testing

2.1 Test for independence of Pre-CRP(L,H) with Duke(A,B,C,D)

a. S-Plus output: Contingency table with Chi-square test

```
    ***  Summary Statistics for data in:  myCRPd ***

 Pre.CRP   Duke.s
 H:34     A: 7
 L:69     B:48
          C:36
          D:12


    *** Crosstabulations ***
Call:
crosstabs(formula =  ~ Pre.CRP + Duke.s, data = myCRPd, na.action = na.fail,
    drop.unused.levels = T)
103 cases in table
+----------+
|N         |
|N/RowTotal|
|N/ColTotal|
|N/Total   |
+----------+
Pre.CRP|Duke.s
      |A     |B     |C     |D     |RowTotl|
------+------+------+------+------+-------+
H     | 0    |17    |10    | 7    |34     |
      |0.000 |0.500 |0.294 |0.206 |0.33   |
```

```
      |0.000  |0.354  |0.278  |0.583  |       |
      |0.000  |0.165  |0.097  |0.068  |       |
-------+-------+-------+-------+-------+-------+
L     | 7     |31     |26     | 5     |69     |
      |0.101  |0.449  |0.377  |0.072  |0.67   |
      |1.000  |0.646  |0.722  |0.417  |       |
      |0.068  |0.301  |0.252  |0.049  |       |
-------+-------+-------+-------+-------+-------+
ColTotl|7     |48     |36     |12     |103    |
       |0.068  |0.466  |0.350  |0.117  |       |
-------+-------+-------+-------+-------+-------+
Test for independence of all factors
     Chi^2 = 7.50066 d.f.= 3 (p=0.05754149)
     Yates' correction not used
     Some expected values are less than 5, don't trust stated p-value
```

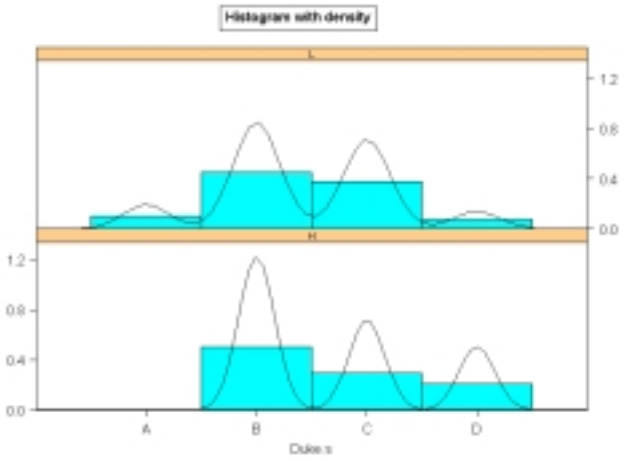
b. A p-value of 0.05754 indeicates that we have no enough evidence to reject the null hypothesis of independence. But, some expected values are less than 5, we consider Fisher's exact test.

```
     Fisher's exact test

data:  Pre.CRP and Duke.s from data set myCRPd

p-value = 0.0544
alternative hypothesis: two.sided
```

c. We conclude that the Pre.CRP is uncorrelated with Duke's. A histogram was shown below:



2.2 Test for independence of Pre-CRP(L,H) with Differentiation(M,W,MW,P)

a. S-Plus output: Contingency table with Chi-square test

```
     *** Crosstabulations ***
Call:
crosstabs(formula =  ~ Pre.CRP + Differentiation, data = myCRPd, na.action =
na.fail,
     drop.unused.levels = T)
103 cases in table
```

```
+----------+
|N         |
|N/RowTotal|
|N/ColTotal|
|N/Total   |
+----------+
Pre.CRP|Differentiation
       |M      |P      |W      |WM     |RowTotl|
-------+-------+-------+-------+-------+-------+
H      |23     | 6     | 5     | 0     |34     |
       |0.676  |0.176  |0.147  |0.000  |0.33   |
       |0.324  |0.600  |0.294  |0.000  |       |
       |0.223  |0.058  |0.049  |0.000  |       |
-------+-------+-------+-------+-------+-------+
L      |48     | 4     |12     | 5     |69     |
       |0.696  |0.058  |0.174  |0.072  |0.67   |
       |0.676  |0.400  |0.706  |1.000  |       |
       |0.466  |0.039  |0.117  |0.049  |       |
-------+-------+-------+-------+-------+-------+
ColTotl|71     |10     |17     |5      |103    |
       |0.689  |0.097  |0.165  |0.049  |       |
-------+-------+-------+-------+-------+-------+
Test for independence of all factors
     Chi^2 = 5.869732 d.f.= 3 (p=0.1181229)
     Yates' correction not used
     Some expected values are less than 5, don't trust stated p-value
```

```
     Fisher's exact test

data:  Pre.CRP and Differentiation from data set myCRPd
p-value = 0.138
alternative hypothesis: two.sided
```

b. The p-value is 0.138, which is not significant.

2.3 Adjusted

We doubt that Pre.CRP are related to Duke's "A" and "D" .

We then perform a Chi-Square test for this 2 by 2 contingency table.

(In fact, we should do the test for all combination of any two levels from Duke's).

```
     *** Crosstabulations ***
Call:
crosstabs(formula =  ~ Pre.CRP + Duke.s, data = atod, na.action = na.fail,
     drop.unused.levels = T)
19 cases in table
+----------+
|N         |
|N/RowTotal|
|N/ColTotal|
|N/Total   |
+----------+
Pre.CRP|Duke.s
       |A      |D      |RowTotl|
-------+-------+-------+-------+
H      |0      |7      |7      |
       |0.00   |1.00   |0.37   |
       |0.00   |0.58   |       |
       |0.00   |0.37   |       |
-------+-------+-------+-------+
L      |7      |5      |12     |
       |0.58   |0.42   |0.63   |
```

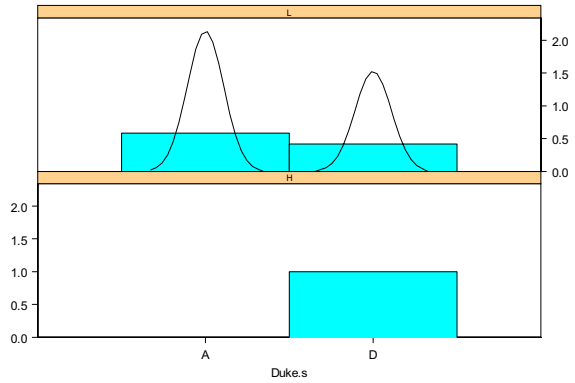
```
      |1.00   |0.42   |       |
      |0.37   |0.26   |       |
-------+-------+-------+-------+
ColTotl|7      |12     |19     |
      |0.37   |0.63   |       |
-------+-------+-------+-------+
Test for independence of all factors
     Chi^2 = 6.465278 d.f.= 1 (p=0.01100024)
     Yates' correction not used
     Some expected values are less than 5, don't trust stated p-value


     Fisher's exact test

data:  Pre.CRP and Duke.s from data set atod
p-value = 0.0174
alternative hypothesis: two.sided
```

The p-value is 0.0174 which is significant, and hence we believe that Pre.CRP is related to Duke's "A" and "D".

The Histogram for this data set is shown below:



3. Survival Analysis

3.1 Question: Is the survival time of Duke's with levels B,C,D different?

a. Kaplan-Meier estimate

```
Call: survfit(formula = Surv(myCRP$survival, myCRP$event) ~ Duke.s, data = myCRP,
type =
    "kaplan-meier")

        n events mean se(mean) median 0.95LCL 0.95UCL
Duke.s=B 48     8 39.9    1.94     NA      NA      NA
Duke.s=C 36     5 41.7    2.51     NA      NA      NA
Duke.s=D 12     6 18.0    3.61     16       9      NA
```

b. Print the survival curve estimate, standard errors, and confidence intervals.

```
Call: survfit(formula = Surv(myCRP$survival, myCRP$event) ~ Duke.s, data = myCRP,
type =
    "kaplan-meier")

             Duke.s=B
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
    2     48       1    0.979  0.0206       0.940       1.000
    4     42       1    0.956  0.0306       0.898       1.000
   18     25       1    0.918  0.0476       0.829       1.000
   29     20       1    0.872  0.0636       0.756       1.000
   30     19       1    0.826  0.0750       0.691       0.987
   34     17       1    0.777  0.0849       0.628       0.963
   35     16       1    0.729  0.0924       0.568       0.934
   36     14       1    0.677  0.0994       0.507       0.902

             Duke.s=C
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
   10     30       1    0.967  0.0328       0.905       1.000
   16     22       1    0.923  0.0531       0.824       1.000
   22     19       1    0.874  0.0690       0.749       1.000
   25     15       1    0.816  0.0856       0.664       1.000
   34     11       1    0.742  0.1051       0.562       0.979

             Duke.s=D
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
    1     12       1    0.917  0.0798       0.773          1
    8      9       1    0.815  0.1194       0.611          1
    9      8       1    0.713  0.1414       0.483          1
   12      5       1    0.570  0.1705       0.318          1
   16      3       1    0.380  0.1924       0.141          1
   30      1       1    0.000      NA          NA         NA
```

c. Comparing Kaplan-Meier Survival Curves: Log-Rank test

```
Call:
survdiff(formula = Surv(myCRP$survival, myCRP$event) ~ Duke.s, data = myCRP, rho
= 0)

        N Observed Expected (O-E)^2/E (O-E)^2/V
Duke.s=B 48       8    10.24     0.490     1.08
Duke.s=C 36       5     7.69     0.939     1.59
Duke.s=D 12       6     1.07    22.633    24.97
```

```
 Chisq= 25  on 2 degrees of freedom, p= 3.68e-006
```

d. Comparing Kaplan-Meier Survival Curves: Peto-Wilcoxon test

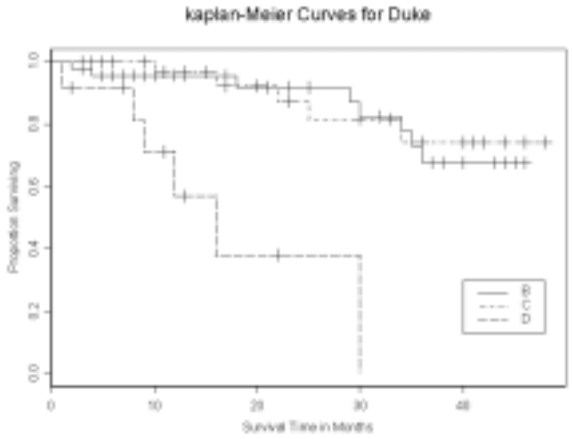
```
Call:
survdiff(formula = Surv(myCRP$survival, myCRP$event) ~ Duke.s, data = myCRP,
rho = 1)

        N Observed Expected (O-E)^2/E (O-E)^2/V
Duke.s=B 48     6.56     8.77     0.554      1.38
Duke.s=C 36     4.30     6.65     0.830      1.61
Duke.s=D 12     5.56     1.01    20.602     24.24

 Chisq= 24.3  on 2 degrees of freedom, p= 5.38e-006
```

Conclusion: Duke's B.C,D groups have significantly different KM survival curves.

e. Kaplan-Meier Survival Curves



3.2 Question: Is the survival time of Pre.CRP with levels L, H different?

a. Kaplan-Meier estimate

```
Call: survfit(formula = Surv(myCRP$survival, myCRP$event) ~ Pre.CRP, data =
myCRP, type =
    "kaplan-meier")

          n events mean se(mean) median 0.95LCL 0.95UCL
Pre.CRP=H 34      8 34.2     3.18     NA      30      NA
Pre.CRP=L 62     11 40.6     1.94     NA      NA      NA
```

b. Print the survival curve estimate, standard errors, and confidence intervals.

```
Call: survfit(formula = Surv(myCRP$survival, myCRP$event) ~ Pre.CRP, data =
myCRP, type =
    "kaplan-meier")

              Pre.CRP=H
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
    2     34       1    0.971  0.0290       0.915       1.000
    8     27       1    0.935  0.0450       0.851       1.000
    9     25       1    0.897  0.0566       0.793       1.000
   10     21       1    0.855  0.0682       0.731       0.999
   12     18       1    0.807  0.0792       0.666       0.978
   16     14       1    0.749  0.0922       0.589       0.954
   30      9       1    0.666  0.1135       0.477       0.930
   34      8       1    0.583  0.1262       0.381       0.891
```

```
            Pre.CRP=L
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
    1     62       1    0.984  0.0160        0.953        1.000
    4     56       1    0.966  0.0235        0.921        1.000
   16     39       1    0.942  0.0335        0.878        1.000
   18     35       1    0.915  0.0420        0.836        1.000
   22     33       1    0.887  0.0490        0.796        0.988
   25     27       1    0.854  0.0571        0.749        0.974
   29     25       1    0.820  0.0643        0.703        0.956
   30     24       1    0.786  0.0701        0.660        0.936
   34     20       1    0.746  0.0768        0.610        0.913
   35     19       1    0.707  0.0822        0.563        0.888
   36     17       1    0.666  0.0873        0.515        0.861
```

c. Comparing Kaplan-Meier Survival Curves: Log-Rank test

```
Call:
survdiff(formula = Surv(myCRP$survival, myCRP$event) ~ Pre.CRP, data = myCRP,
rho = 0)

          N Observed Expected (O-E)^2/E (O-E)^2/V
Pre.CRP=H 34       8     5.64    0.987      1.43
Pre.CRP=L 62      11    13.36    0.417      1.43

 Chisq= 1.4  on 1 degrees of freedom, p= 0.233
```

d. Comparing Kaplan-Meier Survival Curves: Peto-Wilcoxon test

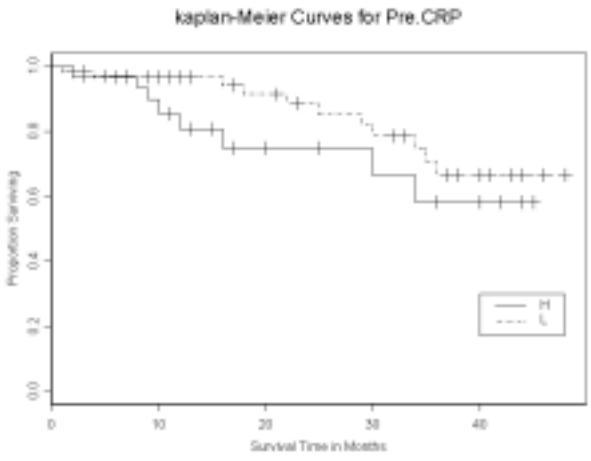
```
Call:
survdiff(formula = Surv(myCRP$survival, myCRP$event) ~ Pre.CRP, data = myCRP,
rho = 1)

          N Observed Expected (O-E)^2/E (O-E)^2/V
Pre.CRP=H 34     7.24     4.92     1.092      1.8
Pre.CRP=L 62     9.19    11.50     0.467      1.8

 Chisq= 1.8  on 1 degrees of freedom, p= 0.18
```

Conclusion: Pre.CRP's L, H groups have significantly different KM survival curves.

e. Kaplan-Meier Survival Curves



3.2 Question: Is the survival time of Pre.CRP with levels L, H different?

a. Kaplan-Meier estimate

```
Call: survfit(formula = Surv(myCRP$survival, myCRP$event) ~ Pre.CRP + Duke.s,
data = myCRP
,
    type = "kaplan-meier")

                 n events mean se(mean) median 0.95LCL 0.95UCL
Pre.CRP=H, Duke.s=B 17      2 40.4    2.98     NA      34      NA
Pre.CRP=H, Duke.s=C 10      1 39.1    4.50     NA      NA      NA
Pre.CRP=H, Duke.s=D  7      5 15.7    3.63     12       9      NA
Pre.CRP=L, Duke.s=B 31      6 39.5    2.37     NA      36      NA
Pre.CRP=L, Duke.s=C 26      4 41.6    2.79     NA      34      NA
Pre.CRP=L, Duke.s=D  5      1 17.8    3.76     NA      NA      NA
```

b. Print the survival curve estimate, standard errors, and confidence intervals.

```
Call: survfit(formula = Surv(myCRP$survival, myCRP$event) ~ Pre.CRP + Duke.s,
data = myCRP
,
    type = "kaplan-meier")

             Pre.CRP=H, Duke.s=B
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
    2     17       1    0.941  0.0571       0.836          1
   34      5       1    0.753  0.1744       0.478          1

             Pre.CRP=H, Duke.s=C
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
   10      7       1    0.857   0.132       0.633          1

             Pre.CRP=H, Duke.s=D
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
    8      6       1    0.833   0.152      0.5827          1
    9      5       1    0.667   0.192      0.3786          1
   12      3       1    0.444   0.222      0.1668          1
   16      2       1    0.222   0.192      0.0407          1
   30      1       1    0.000      NA          NA         NA

             Pre.CRP=L, Duke.s=B
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
    4     27       1    0.963  0.0363       0.894       1.000
   18     18       1    0.909  0.0623       0.795       1.000
   29     15       1    0.849  0.0825       0.702       1.000
   30     14       1    0.788  0.0964       0.620       1.000
   35     12       1    0.723  0.1084       0.538       0.970
   36     10       1    0.650  0.1193       0.454       0.932

             Pre.CRP=L, Duke.s=C
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
   16     18       1    0.944  0.0540       0.844          1
   22     16       1    0.885  0.0763       0.748          1
   25     12       1    0.812  0.0994       0.638          1
   34      8       1    0.710  0.1288       0.498          1

             Pre.CRP=L, Duke.s=D
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
    1      5       1      0.8   0.179       0.516          1
```

c. Comparing Kaplan-Meier Survival Curves: Log-Rank test

```
Call:
survdiff(formula = Surv(myCRP$survival, myCRP$event) ~ Pre.CRP + Duke.s, data
= myCRP,
    rho = 0)

                  N Observed Expected (O-E)^2/E (O-E)^2/V
Pre.CRP=H, Duke.s=B 17       2    3.083    0.380    0.459
Pre.CRP=H, Duke.s=C 10       1    1.839    0.383    0.428
Pre.CRP=H, Duke.s=D  7       5    0.719   25.485   27.168
Pre.CRP=L, Duke.s=B 31       6    7.158    0.187    0.307
Pre.CRP=L, Duke.s=C 26       4    5.848    0.584    0.855
Pre.CRP=L, Duke.s=D  5       1    0.354    1.182    1.228

 Chisq= 29.1  on 5 degrees of freedom, p= 0.0000221
```

d. Comparing Kaplan-Meier Survival Curves: Peto-Wilcoxon test

```
Call:
survdiff(formula = Surv(myCRP$survival, myCRP$event) ~ Pre.CRP + Duke.s, data
= myCRP,
    rho = 1)

                  N Observed Expected (O-E)^2/E (O-E)^2/V
Pre.CRP=H, Duke.s=B 17    1.735    2.666    0.325    0.446
Pre.CRP=H, Duke.s=C 10    0.942    1.583    0.259    0.331
Pre.CRP=H, Duke.s=D  7    4.558    0.670   22.578   25.773
Pre.CRP=L, Duke.s=B 31    4.828    6.100    0.265    0.494
Pre.CRP=L, Duke.s=C 26    3.358    5.067    0.576    0.961
Pre.CRP=L, Duke.s=D  5    1.000    0.336    1.311    1.431

 Chisq= 27.9  on 5 degrees of freedom, p= 0.0000382
```

Conclusion: Pre.CRP's L, H and Duke's B,C,D groups have significantly different KM survival curves.

e. Kaplan-Meier Survival Curves