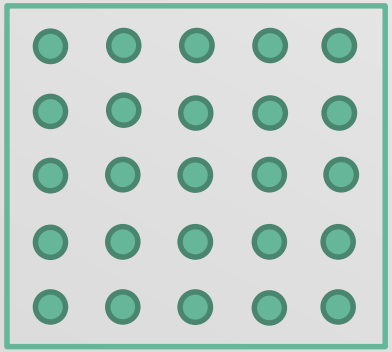
豆象資料分析

7109018032謝綺珊
7109018037李蕙君

Data介紹

實驗設置

在5x5的綠豆上記錄豆象的路徑及產卵情況(8隻豆象)

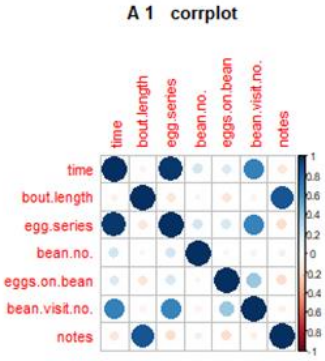


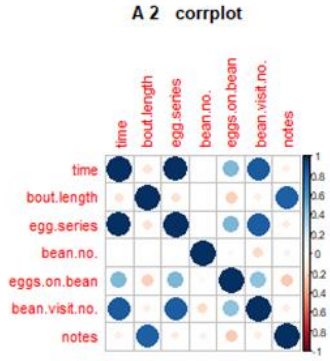
變數

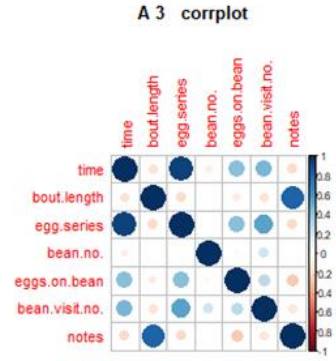
up series : 第幾次爬上綠豆
time : 在第幾秒爬上綠豆
behavior : 爬上綠豆
bout length : 待在綠豆上的時間
egg series : 已產下的卵數量
bean no. : 綠豆編號
egg on bean : 在綠豆上的卵數量
bean visit no. : 豆象到此綠豆次數
notes : 有無產卵(0和1)

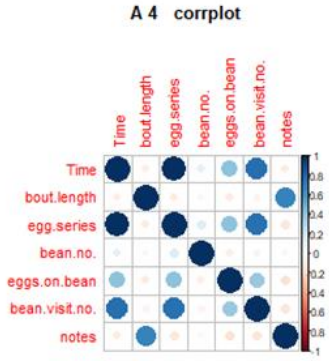
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	up series	time	behavior	bout length	egg series	bean no.	eggs on bean	bean visit no.	notes	patch value	bean quality	average encountered bean quality within every ovp bout(EBQ)					
2	1	29.6	up	3.23	1	21	0	1		1	1						
3	2	73.77	up	74.43	1	31	0	1	ovp	1	1	1					
4	3	300.23	up	1.97	2	22	0	1		0.98	1						
5	4	306.33	up	2.9	2	23	0	1		0.98	1						
6	5	309.33	up	2.57	2	24	0	1		0.98	1						

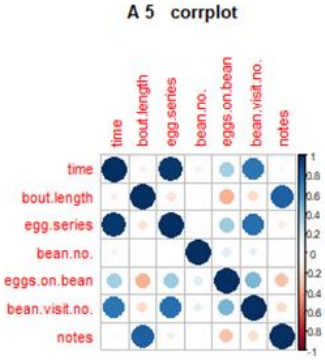
corrplot

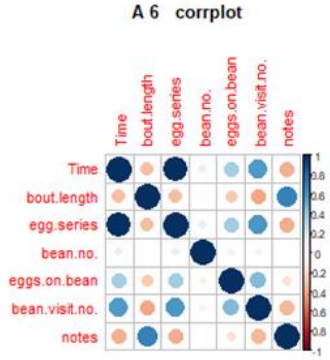

A 1 corrplot
time
bout.length
egg.series
bean.no.
eggs.on.bean
bean.visit.no.
notes

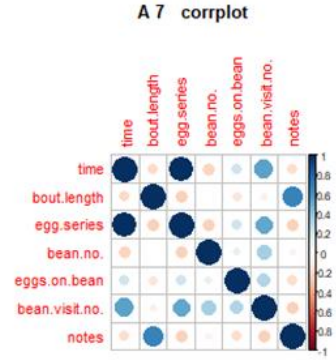

A 2 corrplot
time
bout.length
egg.series
bean.no.
eggs.on.bean
bean.visit.no.
notes

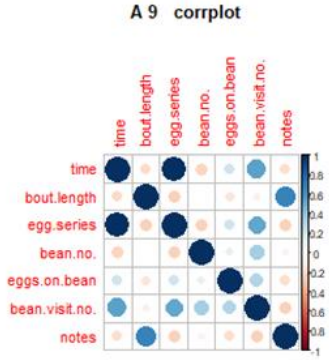

A 3 corrplot
time
bout.length
egg.series
bean.no.
eggs.on.bean
bean.visit.no.
notes


A 4 corrplot
Time
bout.length
egg.series
bean.no.
eggs.on.bean
bean.visit.no.
notes

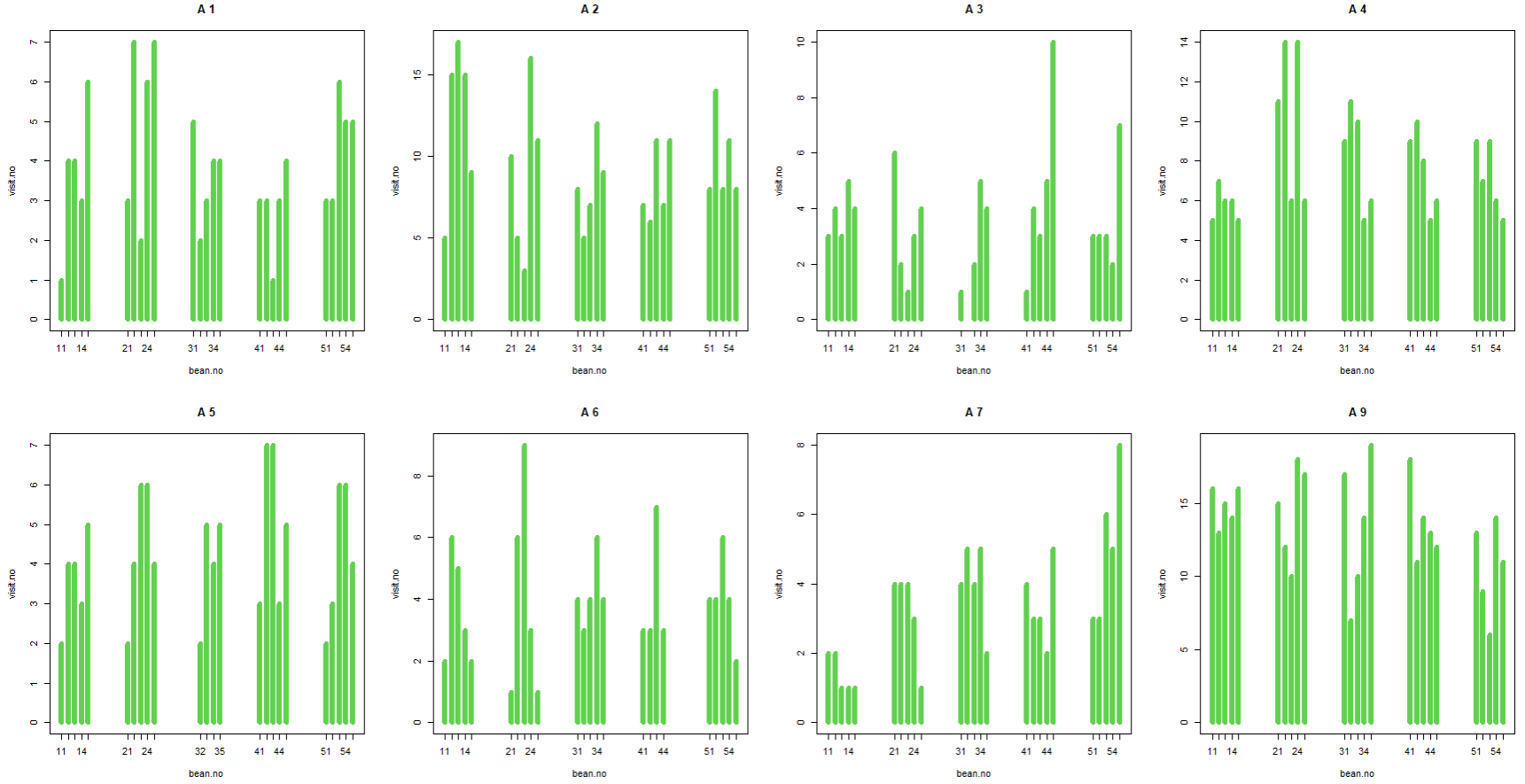

A 5 corrplot
time
bout.length
egg.series
bean.no.
eggs.on.bean
bean.visit.no.
notes


A 6 corrplot
Time
bout.length
egg.series
bean.no.
eggs.on.bean
bean.visit.no.
notes


A 7 corrplot
time
bout.length
egg.series
bean.no.
eggs.on.bean
bean.visit.no.
notes


A 9 corrplot
time
bout.length
egg.series
bean.no.
eggs.on.bean
bean.visit.no.
notes

Bar Chart



將八隻豆象的數據合併起來，80%訓練組，20%測試組

```
> head(data)
  bout.time ovp bean.visit.no. eggs.on.bean position
1      3.23   0              1            0        2
2     74.43   1              1            0        2
3      1.97   0              1            0        3
4      2.90   0              1            0        3
5      2.57   0              1            0        3
6      4.50   0              1            0        2
```

Logistic Regression(不考慮豆子位置)

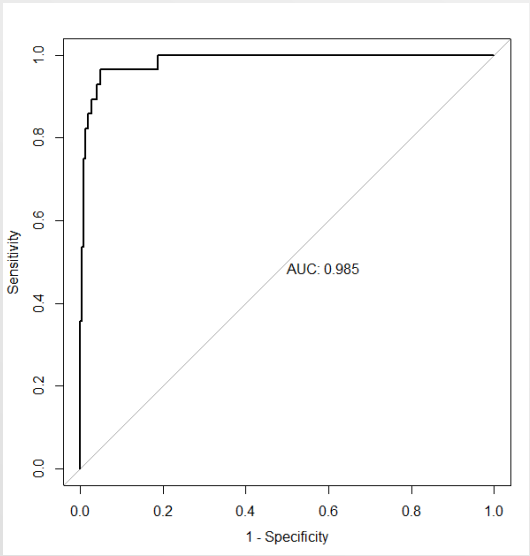
Trainset

```
Call:
glm(formula = ovp ~ bout.time + bean.visit.no. + eggs.on.bean,
    family = binomial(logit), data = train)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.5932  -0.2690  -0.2057  -0.1278   2.3838

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  -3.67673    0.31118 -11.815  < 2e-16 ***
bout.time     0.17131    0.01429  11.987  < 2e-16 ***
bean.visit.no. -0.06176    0.05175  -1.193  0.23270
eggs.on.bean  -1.06195    0.40215  -2.641  0.00827 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 754.17  on 987  degrees of freedom
Residual deviance: 321.17  on 984  degrees of freedom
AIC: 329.17

Number of Fisher Scoring iterations: 7
```

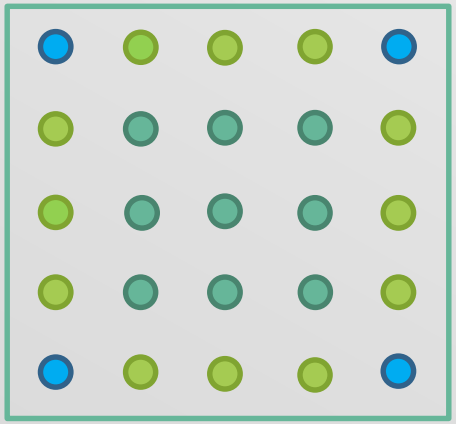
```
> cm
    預測
實際   0   1
   0 814  48
   1   8 118
```

Testset

```
Confusion Matrix and Statistics

          Reference
Prediction   0   1
         0 215   5
         1   4  23

               Accuracy : 0.9636
                 95% CI : (0.932, 0.9832)
    No Information Rate : 0.8866
    P-Value [Acc > NIR] : 1.283e-05

                  Kappa : 0.8159

 Mcnemar's Test P-Value : 1

            Sensitivity : 0.9817
            Specificity : 0.8214
         Pos Pred Value : 0.9773
         Neg Pred Value : 0.8519
             Prevalence : 0.8866
         Detection Rate : 0.8704
   Detection Prevalence : 0.8907
      Balanced Accuracy : 0.9016

       'Positive' Class : 0
```



考慮豆子位置



：第一種綠豆

：第二種綠豆

：第三種綠豆

上三種豆的次數及比例

```
> visit
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
[1,]   15   30   17   24   13   10   14   56
[2,]   51  136   46   89   45   42   38  169
[3,]   31   72   25   83   44   43   33  109
```

```
> ratio.visi
          [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]      [,8]
[1,] 0.1546392 0.1260504 0.1931818 0.1224490 0.1274510 0.1052632 0.1647059 0.1676647
[2,] 0.5257732 0.5714286 0.5227273 0.4540816 0.4411765 0.4421053 0.4470588 0.5059880
[3,] 0.3195876 0.3025210 0.2840909 0.4234694 0.4313725 0.4526316 0.3882353 0.3263473
```

上三種豆產卵的次數及比例

```
> bout
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8]
[1,]    0    4    5    3    1    5    1    4
[2,]    7   13    8   15    9    6    7   12
[3,]    3    8    7    6    7    6    8    9
```

```
> ratio.bout
     [,1] [,2] [,3]  [,4]       [,5]      [,6]   [,7] [,8]
[1,]  0.0 0.16 0.25 0.125 0.05882353 0.2941176 0.0625 0.16
[2,]  0.7 0.52 0.40 0.625 0.52941176 0.3529412 0.4375 0.48
[3,]  0.3 0.32 0.35 0.250 0.41176471 0.3529412 0.5000 0.36
```

三種豆上產卵的機率

```
> p.bout = round(bout/visit,3)
> p.bout
      [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]  [,8]
[1,] 0.000 0.133 0.294 0.125 0.077 0.500 0.071 0.071
[2,] 0.137 0.096 0.174 0.169 0.200 0.143 0.184 0.071
[3,] 0.097 0.111 0.280 0.072 0.159 0.140 0.242 0.083
```

Logistic Regression(考慮豆子位置)

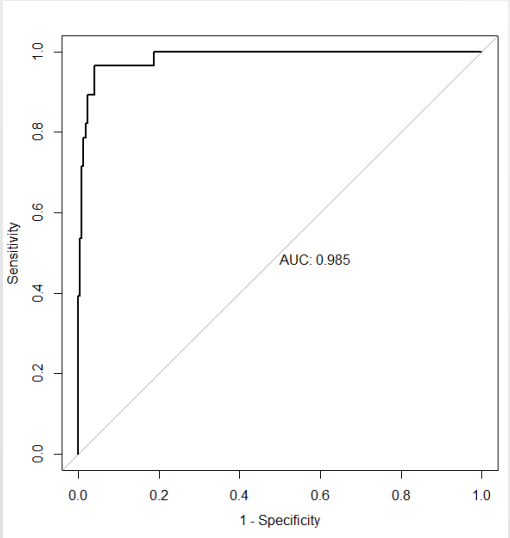
Trainset

```
Call:
glm(formula = ovp ~ ., family = binomial(logit), data = train)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.5945  -0.2693  -0.2080  -0.1284   2.3892

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    -3.86881    0.58502  -6.613 3.76e-11 ***
bout.time       0.17159    0.01434  11.963  < 2e-16 ***
bean.visit.no. -0.06208    0.05175  -1.200  0.23030
eggs.on.bean   -1.06072    0.40205  -2.638  0.00833 **
position        0.08751    0.22340   0.392  0.69528
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 754.17  on 987  degrees of freedom
Residual deviance: 321.02  on 983  degrees of freedom
AIC: 331.02

Number of Fisher Scoring iterations: 7
```

```
> cm
    預測
實際   0   1
   0 816  46
   1   7 119
```

Testset

```
Confusion Matrix and Statistics

          Reference
Prediction   0   1
         0 215   5
         1   4  23

               Accuracy : 0.9636
                 95% CI : (0.932, 0.9832)
    No Information Rate : 0.8866
    P-Value [Acc > NIR] : 1.283e-05

                  Kappa : 0.8159

 Mcnemar's Test P-Value : 1

            Sensitivity : 0.9817
            Specificity : 0.8214
         Pos Pred Value : 0.9773
         Neg Pred Value : 0.8519
             Prevalence : 0.8866
         Detection Rate : 0.8704
   Detection Prevalence : 0.8907
      Balanced Accuracy : 0.9016

       'Positive' Class : 0
```



Logistic Regression(移除不顯著變數)

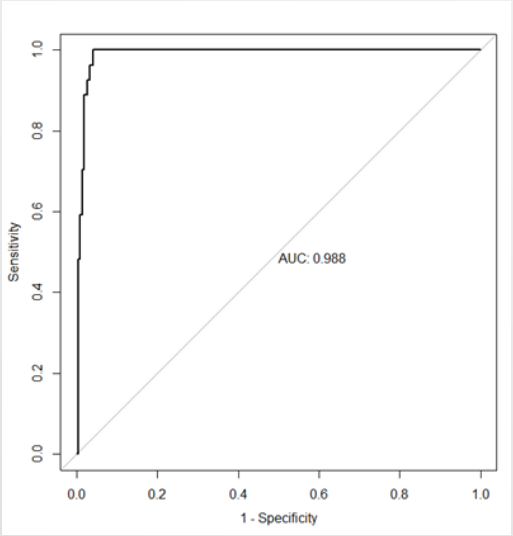
Trainset

```
Call:
glm(formula = ovp ~ bout.time + eggs.on.bean, family = binomial(logit),
    data = train)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.5661  -0.2644  -0.2224  -0.1271   2.4160

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  -3.89351    0.26315 -14.796  < 2e-16 ***
bout.time     0.17168    0.01425  12.047  < 2e-16 ***
eggs.on.bean -1.24077    0.37446  -3.313 0.000921 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 754.17  on 987  degrees of freedom
Residual deviance: 322.66  on 985  degrees of freedom
AIC: 328.66

Number of Fisher Scoring iterations: 7
```

```
> cm
    預測
實際   0   1
   0 817  45
   1   8 118
```

Testset

```
Confusion Matrix and Statistics

          Reference
Prediction   0   1
         0 212   1
         1   8  26

               Accuracy : 0.9636
                 95% CI : (0.932, 0.9832)
    No Information Rate : 0.8907
    P-Value [Acc > NIR] : 2.779e-05

                  Kappa : 0.832

 Mcnemar's Test P-Value : 0.0455

            Sensitivity : 0.9636
            Specificity : 0.9630
         Pos Pred Value : 0.9953
         Neg Pred Value : 0.7647
             Prevalence : 0.8907
         Detection Rate : 0.8583
   Detection Prevalence : 0.8623
      Balanced Accuracy : 0.9633

       'Positive' Class : 0
```



Thank You for Your Attention