1. 該資料源自於1978年Cameron與Pauling所收集之資料，該資料集涵蓋不同器官部位(胃、支氣管、結腸、卵巢及乳房)診斷出癌症後，接受補充抗壞血酸治療後的存活時間。其中表1列出此資料集一共有兩個變數，並說明變數意義。而以器官分組畫出存活時間的盒型圖。

圖1為未將存活時間做轉換，看出因不同器官的存活時間彼此差異很大，對於後續分析可能會有所影響，因此對存活時間做log轉換，圖2為轉換後之盒型圖，可以看出較圖1而言其數值差異變小，方便後續分析。

|  |  |
| --- | --- |
| Variable | 說明 |
| Surv.time | 患者從接受治療後的存活時間 |
| organ | 癌症發生之器官處 |

表1 變數說明

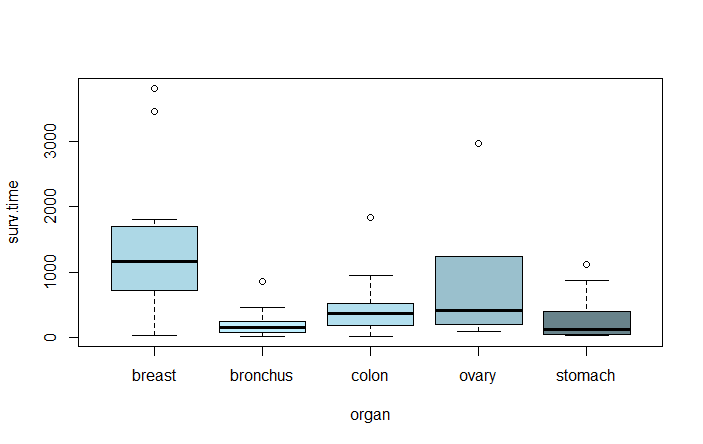


圖1 轉換前盒型圖

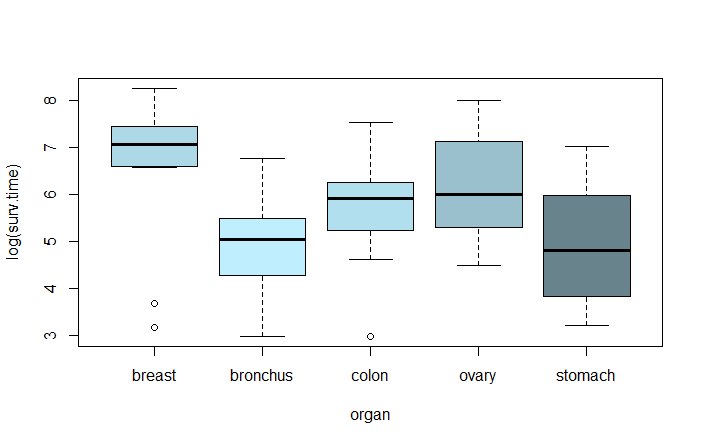


圖2 轉換後盒型圖

將存活時間做log轉換後，想了解不同器官的癌症對於存活時間是否有所影響，這邊使用變異數分析(ANOVA)，比較每組平均存活時間是否相等。其模型如下:

一開始將先檢定同質性，檢查此因素是否有隨機效應的影響，其虛無假設與對立假說如下

而得到的檢定統計量與p-value分別為

表示不拒絕虛無假設，認為該筆資料通過同質性檢定。因此認為其沒有隨機效應的影響，進一步使用固定效應的模型。其虛無假設與對立假說如下

檢定方式為建立ANOVA表，檢定統計量

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 平方和(SS) | 自由度(df) | 均方和(MS) | F-值 | p-value |
| 組內 |  |  |  |  |  |
| 組間 |  |  |  |  |  |
| Total |  |  |  |  |  |

表2 ANOVA表

其中*k*表示組數，*N*表示總樣本數

最後根據此資料計算出檢定統計量=4.29，而將信心水準設定為95%的情況下，可得到檢定量的臨界值，因故我們拒絕虛無假設，**至少存在一組。**此外可算出p-value如下:

另外檢查殘差是否有服從常態性，其虛無假設與對立假說如下

而得到的檢定統計量W=0.96508與p-value=0.067，表示不拒絕虛無假設，認為殘差通過常態性之檢定。

1. 該資料源自於1978年Box所收集之資料，該資料涵蓋四種不同飲食方式(分別為A、B、C、D)，從動物身上抽取血液的凝固時間(秒)。該研究想比較這些飲食效果對於血液凝固的時間是否有所影響，並且進一步使用Tukey多重比較，考慮是否有任何飲食有相同的效果。

從圖3利用不同飲食方式畫出血液凝固時間分組盒型圖，可以看出C方式的血液凝固時間為最大，而D方式的變異性最大。

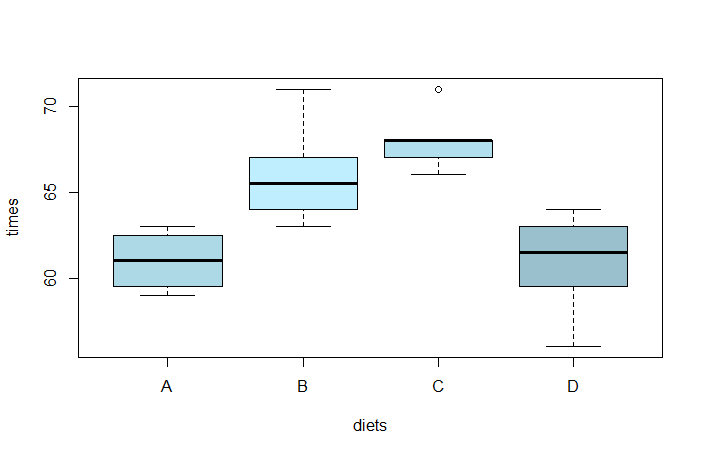


圖3 血液凝固盒型圖

進一步想利用統計檢定不同飲食方式對於血液凝固時間是否有所影響，這邊使用變異數分析(ANOVA)，比較每組平均血液凝固時間是否相等。其模型如下:

一開始將先檢定同質性，檢查此因素是否有隨機效應的影響，其虛無假設與對立假說如下

而得到的檢定統計量與p-value分別為

表示不拒絕虛無假設，認為該筆資料通過同質性檢定。因此認為其沒有隨機效應的影響，進一步使用固定效應的模型。其虛無假設與對立假說如下

檢定方式為建立ANOVA表，檢定統計量

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 平方和(SS) | 自由度(df) | 均方和(MS) | F-值 | p-value |
| 組內 |  |  |  |  |  |
| 組間 |  |  |  |  |  |
| Total |  |  |  |  |  |

表4 ANOVA表

其中*k*表示組數，*N*表示總樣本數

最後根據此資料計算出檢定統計量=13.57，而將信心水準設定為95%的情況下，可得到檢定量的臨界值，因故我們拒絕虛無假設，**至少存在一組。**此外可算出p-value如下:

另外檢查殘差是否有服從常態性，其虛無假設與對立假說如下

而得到的檢定統計量W=0.9783與p-value=0.8629，表示不拒絕虛無假設，認為殘差通過常態性之檢定。

進一步利用tukey多重比較分析為哪幾組平均效果相等，表5中diff表兩組平均差異，lwr.ci表信賴區間下界，upr.ci表信賴區間下界，若信賴區間範圍有包含0，則認為這兩組之間並無差異；反之若都小於0或大於0，則認為某組較好。而從結果來看，A與B、A與D、B與C並無顯著差異。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | diff | lwr.ci | upr.ci | p-value |
| B-A | 5 | -0.54 | 10.54 | 0.02 |
| C-A | 7 | 1.46 | 12.54 | 0.00 |
| D-A | 0 | -5.26 | 5.26 | 1.00 |
| C-B | 2 | -2.96 | 6.96 | 0.48 |
| D-B | -5 | -9.64 | -0.36 | 0.00 |
| D-C | -7 | -11.64 | -2.36 | 0.00 |

表5 tukey多重比較

1. 該資料源自於1942年Shaw而後1977年由Mosteller、Tukey完善此資料集，該資料集涵蓋Lake Victoria Nyanza相對於標準水平的水位，以及連續20年的太陽黑子數，想將兩變數建立迴歸模型。

從圖4可以看出，將水位作為X軸，太陽中子數作為Y軸，整體似乎有正相關，太陽中子數越多時，該年的水位有越高的趨勢，進一步建立迴歸模型檢定其效果是否存在。

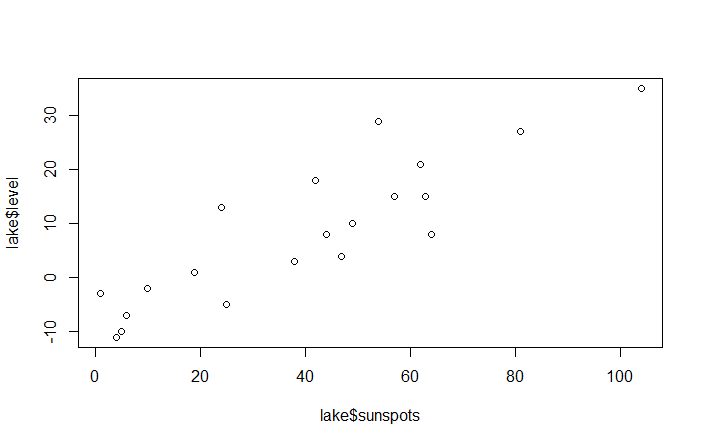


圖4 散佈圖

其模型如下:

其中Y表示水位，為截距項，為太陽中子數對水位的影響，x表太陽中子數，表殘差項，且殘差項必須服從平均數為0，變異數為的常態分配。

檢定太陽中子數是否對水位有影響，其虛無假設與對立假說如下

而得到的檢定統計量與p-value分別為

表示拒絕虛無假設，認為太陽中子數對於水位是有影響的。

另外進一步，迴歸模型需滿足常態性、獨立性、同質性三個假設，因此逐一檢查其假設是否有通過

1. 常態性

利用Shapiro-Wilk檢定殘差是否服從常態分配，其虛無假設與對立假設如下

而得到的檢定統計量與p-value分別為

表示不拒絕虛無假設，認為殘差通過常態性檢定。

1. 獨立性

利用durbinWatson檢查殘差之間是否獨立，其虛無假設與對立假設如下

而得到的檢定統計量與p-value分別為

表示不拒絕虛無假設，認為殘差通過獨立性檢定。

1. 同質性

檢查殘差是否符合同質性，其虛無假設與對立假設如下

而得到的檢定統計量與p-value分別為

表示不拒絕虛無假設，認為殘差通過同質性檢定。

總結上述的迴歸三大前提假設皆有通過，表示此配飾模型是合理可用來推論的，模型如下

當太陽中子數每增加一單位時，水位平均會上升0.41單位。