

# 統計套用程式分析報告

## 關注的變數

1. 網站交貨量(連續變數)

2. 訂單量(連續變數)

## 影響的變數

1. 過程變更(類別變數)

2. 年(類別變數)

3. 季(類別變數)

## 目標問題

A. 網站交貨量是否受到過程變更、年、季的影響?

B. 訂單量是否受到過程變更、年、季的影響?

C. 網站交貨量和訂單量是否有相關性?

\*設定顯著水準=0.05

## 一. 分析類別變數:

### 1. 敘述統計

#### 過程變更:

N	平均值	標準差	最小值	最大值
73	0.3561644	0.4821786	0	1

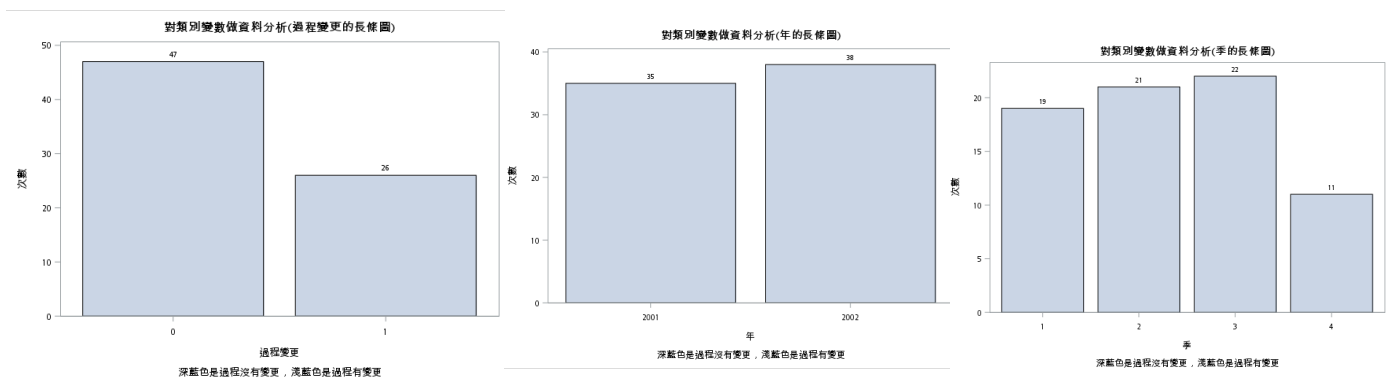
#### 年:

N	平均值	標準差	最小值	最大值
73	2001.52	0.5030349	2001	2002

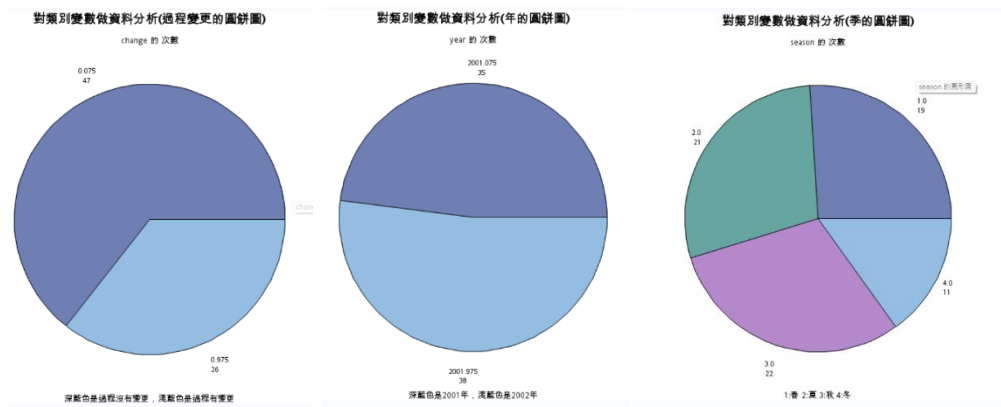
#### 季:

N	平均值	標準差	最小值	最大值
73	2.3424658	1.0303610	1	4

## 2. 長條圖



### 3. 圓餅圖



### 4. Binomial Test(過程變更&年)

#### 過程變更:

$H_0$  : 過程變更和不變更的機率都是相同的(0.5)

單側檢驗中， $p\text{-value} = 0.0070 < \alpha = 0.05$ ，雙側檢驗中， $p\text{-value} = 0.0140 < \alpha = 0.05$ ，因此可以拒絕原假設。

結論：過程變更和不變更的機率不像預期的那樣平均分佈

#### 手算報表值

(1)二項式比例

Change = 0

比例：

$47/73 \approx 0.6438$

ASE (標準誤)：

$\sqrt{0.6438 \times (1 - 0.6438)/73} \approx 0.0560$

95% 信賴區間：

信賴下限：

$0.6438 - 1.96 \times 0.0560 \approx 0.5340$

信賴上限：

$0.6438 + 1.96 \times 0.0560 \approx 0.7537$

(2)精準信賴界限

95% 信賴區間：

信賴下限：

$0.6438 - 1.96 \sqrt{(0.6438 \times (1 - 0.6438) + 0.5^2/(4 \times 73))}/(73 - 1) \approx 0.5231$

信賴上限：

$0.6438 + 1.96 \sqrt{(0.6438 \times (1 - 0.6438) + 0.5^2/(4 \times 73))}/(73 - 1) \approx 0.7525$

(3) $H_0$  的檢定：比例 = 0.5

ASE：

$$\sqrt{(0.5 \times (1 - 0.5))/73} \approx 0.0585$$

Z :

$$(0.6438 - 0.5)/0.0585 \approx 2.4579$$

單邊  $\Pr > Z$  :

$$1 - \text{CDF}(2.4579) \approx 0.0070$$

雙邊  $\Pr > |Z|$  :

$$2 \times (1 - \text{CDF}(|2.4579|)) \approx 0.0140$$

年:

虛無假設 ( $H_0$ ): 比例等於 0.5 的檢定結果

單邊 p-value ( $\Pr < Z$ ) 為 0.3627, 雙邊 p-value ( $\Pr > |Z|$ ) 為 0.7255。

由於 p-value 大於通常的顯著性水平, 不能拒絕虛無假設, 即沒有足夠的證據支持比例不等於 0.5。

結論: 這一年份的比例與 0.5 沒有顯著差異, 平均分布

## 5. Multinomial Test(季:卡方檢定)

卡方值 (Chi-Square): 4.0959

$$\chi^2 = ((19 - 17.25)/17.25)^2 + ((21 - 20.25)/20.25)^2 + ((22 - 21.75)/21.75)^2 + ((11 - 13.75)/13.75)^2$$

自由度 (DF):  $(4 - 1) \times (1 - 1) = 3$

p-value ( $\Pr > \text{ChiSq}$ ): 0.2513

卡方檢定的結果顯示, 觀察到的季節分佈與期望的多項式分佈之間的差異在統計上並不顯著。p-value (0.2513) 大於通常的顯著性水平 (例如 0.05), 因此不拒絕虛無假設, 即無法拒絕觀察到的頻率分佈符合期望的多項式分佈的假設。

相對偏差長條圖:

通常, 相對偏差長條圖可以用來比較觀察到的頻率和期望頻率之間的差異。如果柱狀圖的條形高度大致相等, 則表示觀察到的頻率分佈與期望分佈接近。

結論:

由於 p-value 大於顯著性水平, 無法拒絕虛無假設, 即頻率分佈符合期望的多項式分佈的假設。這意味著, 根據這次 Multinomial Test 的結果, 沒有足夠的證據認為觀察到的季節分佈和等比例分佈之間存在統計上的顯著差異。

## 6. 無兩兩獨立性檢定

過程變更對年：

過程變更對年的無兩兩獨立性檢定(chisq)				
FREQ 程序				
次數 百分比 列百分比 欄百分比	change * year的表格			
	change(過程變更)	year(年)		總計
		2001	2002	
0		35	12	47
		47.95	16.44	64.38
		74.47	25.53	
		100.00	31.58	
1		0	26	26
		0.00	35.62	35.62
		0.00	100.00	
		0.00	68.42	
總計		35	38	73
		47.95	52.05	100.00

change * year 之表格的統計值			
統計值	DF	值	機率
卡方	1	37.1948	<.0001
概化比卡方	1	47.6744	<.0001
連續性調整卡方	1	34.2709	<.0001
Mantel-Haenszel 卡方	1	36.6853	<.0001
Phi 係數		0.7138	
列聯係數		0.5810	
Cramer V		0.7138	

Fisher 精準檢定	
儲存格 (1,1) 次數 (F)	35
左邊 Pr ≤ F	1.0000
右邊 Pr ≥ F	<.0001
表格機率 (P)	<.0001
雙邊 Pr ≤ P	<.0001

卡方  $p\text{-value} < .0001$  表示在過程變更和年之間存在著非常顯著的統計差異，拒絕零假設，支持過程變更和年之間的關係不是隨機的。

在這個情況下，相關性係數 Phi 係數和 Cramer V 皆為 0.7138 蠻接近 1，過程變更和年之間的相關性蠻強的。

### 手算報表值

卡方檢定統計值：

$$E_{\{11\}} = (35 + 0) * (47 + 26) / 73 \approx 24.5479$$

$$E_{\{12\}} = (35 + 0) * (38 + 26) / 73 \approx 26.4521$$

$$E_{\{21\}} = (12 + 26) * (47 + 26) / 73 \approx 11.4521$$

$$E_{\{22\}} = (12 + 26) * (38 + 26) / 73 \approx 12.5479$$

$$\chi^2 = (35 - 24.5479)^2 / 24.5479 + (0 - 26.4521)^2 / 26.4521 + (12 - 11.4521)^2 / 11.4521 + (26 - 12.5479)^2 / 12.5479 \approx 37.1948$$

卡方檢定的統計值  $\chi^2$ ，這是 37.1948，而總樣本數  $n$  是 73。將值帶入公式，計算 Phi 係數：

$$\phi = \sqrt{37.1948 / 73} \approx \sqrt{0.5097} \approx 0.7138$$

(Phi 係數的值介於 0 和 1 之間，越接近 1 表示相關性越強。)

## 過程變更對季：

$$\chi^2 = 1.24 + 0.32 + 0.11 + 1.43 + 3.87 + 0.25 + 0.76 + 2.99 \approx 10.07$$

卡方檢定的統計值 (Chi-square statistic) 為 28.2113。

對應的自由度 (DF) 為 3。

檢定的 p-value 小於

0.0001，顯示在給定的顯著性水準下，拒絕虛無假設，即認為過程變更和季節之間存在顯著的相關性。

概度比卡方 (Likelihood Ratio Chi-square) 統計值為 37.3948。

對應的自由度為 3。

檢定的 p-value 小於

0.0001，這再次強調了過程變更和季節之間的相關性。

Phi 係數為 0.6217，列聯係數為 0.5280，Cramer V 也為 0.6217。

這些係數提供了相關性的度量，值越接近 1 表示相關性越強。

## 過程變更對年跟季：

表格 1：過程變更等於 0 的控制

卡方檢定結果 (觀察到的卡方統計值為 23.7474) 顯示年份和季節之間的相關性極高。概度比卡方、Mantel-Haenszel 卡方、Phi 係數、列聯係數 Cramer's V 都表明了相當強的相關性。

表格 2：過程變更等於 1 的控制

在這個控制組中，年份和季節之間呈現獨立。樣本較小，結果可靠性有限。

總的來說，過程變更等於 0 時，年份

過程變更對季的無兩兩獨立性檢定(chisq)

FREQ 程序

次數 百分比 列百分比 欄百分比	change * season 的表格					
	change(過程變更)	season(季)				總計
		1	2	3	4	
0	19	8	9	11	47	
	26.03	10.96	12.33	15.07	64.38	
	40.43	17.02	19.15	23.40		
	100.00	38.10	40.91	100.00		
1	0	13	13	0	26	
	0.00	17.81	17.81	0.00	35.62	
	0.00	50.00	50.00	0.00		
	0.00	61.90	59.09	0.00		
總計	19	21	22	11	73	
	26.03	28.77	30.14	15.07	100.00	

change \* season 之表格的統計值

過程變更對年跟季的無兩兩獨立性檢定(chisq)

FREQ 程序

次數 百分比 列百分比 欄百分比	year * season 的表格 1					
	year(年)	season(季)				總計
		1	2	3	4	
2001	7	8	9	11	35	
	14.89	17.02	19.15	23.40	74.47	
	20.00	22.86	25.71	31.43		
	36.84	100.00	100.00	100.00		
2002	12	0	0	0	12	
	25.53	0.00	0.00	0.00	25.53	
	100.00	0.00	0.00	0.00		
	63.16	0.00	0.00	0.00		
總計	19	8	9	11	47	
	40.43	17.02	19.15	23.40	100.00	

year \* season 之表格 1 的統計值  
對照 change=0

統計值	DF	值	機率
卡方	3	23.7474	<.0001
概度比卡方	3	28.3936	<.0001
Mantel-Haenszel 卡方	1	16.9446	<.0001
Phi 係數		0.7108	
列聯係數		0.5794	
Cramer V		0.7108	
WARNING: 50% of the cells have expected counts less than 5. Chi-Square may not be a valid test.			

和季節相關性高；而在過程變更等於 1 時，它們呈現獨立。

樣本大小 = 47

次數 百分比 列百分比 欄百分比	year * season的表格 2					
	change=1的控制					
	year(年)	season(季)				總計
		1	2	3	4	
2001	0	0	0	0	0	0
	0.00	0.00	0.00	0.00	0.00	0.00
	-	-	-	-	-	-
	-	0.00	0.00	-	-	-
2002	0	13	13	0	26	26
	0.00	50.00	50.00	0.00	100.00	100.00
	0.00	50.00	50.00	0.00	-	-
	-	100.00	100.00	-	-	-
總計	0	13	13	0	26	26
	0.00	50.00	50.00	0.00	100.00	100.00

year \* season 之表格 2 的統計值  
對照 change=1

Row or column sum zero. No statistics computed for this table.

樣本大小 = 26

## 二. 分析連續資料：

### 網站交易量

#### 1. 敘述統計

N	平均值	標準差	最小值	最大值
73	9.0410959	7.0837026	0	30.0000000

#### 2. 盒鬚圖

中位數：7

第一四分位數：3

第三四分位數：13.5

最小值：0

最大值：30

圖表顯示了網站交易量分佈的情況。網站交易量大部分都集中在 3 到 13.5 之間，有少部分交易量超過 25。

#### 3. 直方圖(含機率密度函數)

從圖表中可以看到，大部分交易量集中在 0 至 10 個之間，佔總交易量的 75%。其中，交易量在 5 至 10 個之間的佔比最高，達 32.5%。

交易量在 10 至 20 個之間的佔比為 25%。交易量在 20 個以上佔比為 20%。

#### 4. 檢定其是否為 normal

QQPlot:數據點並沒有完全沿著直線分佈，不確定是否符合正態分佈

Shapiro-Wilk 檢驗: P 值為 $<0.0001$ 。P 值小於 0.05，因此我們拒絕原假設，即網站交易量不服從正態分佈

### 訂單量

#### 1. 敘述統計

N	平均值	標準差	最小值	最大值
73	27.8219178	7.9762538	3	45

#### 2. 盒鬚圖

中位數：27

第一四分位數：22

第三四分位數：34

最小值：3

最大值：45

圖表顯示了網站交易量分佈的情況。網站交易量大部分都集中在 22 到 34 之間，有少部分交易量低於 10 或超過 40。

#### 3. 直方圖(含機率密度函數)

從圖表中可以看到，大部分交易量集中在 18 至 34 個之間，佔總交易量的 65%。其中，交易量在 22 至 29 個之間的佔比最高，達 38%。

#### 4. 檢定其是否為 normal

QQPlot:數據點並沒有完全沿著直線分佈，不確定是否符合正態分佈

Shapiro-Wilk 檢驗: P 值為 0.3643。P 值大於 0.05，因此我們無法拒絕原假設，即訂單量服從正態分佈

### 三. 研究在哪些類別變數下，哪些連續變數平均值是會有

#### 差異？

#### 過程變更對網站交易量

##### 1. proc ttest

(1)變異數相等性的 p 值= $0.0004 < 0.05$  拒絕虛無假設，變異數不均等。

因此看 Satterthwaite (不均等)，p 值 $<0.0001$  拒絕虛無假設，過程變更對於網站交易量的平均值存在統計上的顯著影響。

(2)過程不變更時的平均值為 5.4468，過程變更時的平均值為 15.5385，差異的平均值為 -10.0917

根據(1)(2):

過程變更時的平均值明顯高於過程不變更時，平均增加了約 10.0917。

95% 信賴區間排除了零，這進一步強調了這個平均值差異的統計顯著性。

T-Test 統計結果也確認了這個平均值差異是高度統計上顯著的。

綜合這些結果，我們可以得出結論：過程變更對網站交易量有顯著的正向影響，即過程變更伴隨著平均交易量的增加。

### (3)手算報表值

給定數據：

Mean1 = 5.4468, Mean2 = 15.5385

Std Dev1 = 3.8381, Std Dev2 = 7.0156

N1 = 47, N2 = 26

計算均等變異數的 t 檢定：

Std Err1 = 3.8381 / sqrt(47)  $\approx$  0.5552

Std Err2 = 7.0156 / sqrt(26)  $\approx$  1.3740

DF = 47 + 26 - 2 = 71

(集區) t Value = (5.4468 - 15.5385) / sqrt((0.5552<sup>2</sup> / 47) + (1.3740<sup>2</sup> / 26))  $\approx$  -7.96

計算不均等變異數的 t 檢定：

Std Err Diff = sqrt(((47-1)\*3.8381<sup>2</sup> + (26-1)\*7.0156<sup>2</sup>) / (47 + 26 - 2)) \* sqrt(1/47 + 1/26)  $\approx$  5.1840

t Value = (5.4468 - 15.5385) / 5.1840  $\approx$  -6.79

計算變異數相等性的 F 檢定：

F Value = 3.8381<sup>2</sup> / 7.0156<sup>2</sup>  $\approx$  0.2893

## 2. proc anova

在過程變更的 ANOVA 分析中 F 值= 63.44, p 值=<.0001 小於 0.05，拒絕虛無假設，代表對網站交易量的影響是統計上顯著的

手算報表值：

Model MS = 1704.798153 / 1 = 1704.798153

Error MS = 1908.078560 / 71  $\approx$  26.874346

F Value = 1704.798153 / 26.874346  $\approx$  63.435

## 3. 盒鬚圖比較

從盒鬚圖的比較中，我們可以看到，變化後的盒鬚圖向上移動了，中位數和 Q3 都增加了。這表明，變化導致了網站交易量的增加。

## 過程變更對訂單量

### 1.proc ttest



(1)變異數相等性的 p 值=0.3650>0.05 無法拒絕虛無假設，變異數均等。  
因此看集區（均等），p 值=0.0001 拒絕虛無假設，過程變更對於訂單量的平均值存在統計上的顯著影響。

(2)過程不變更時的平均值為 25.2553，過程變更時的平均值為 32.4615，差異的平均值為 -7.2062

根據(1)(2):

過程變更時的平均值明顯高於過程不變更時，平均增加了約 7.2062。

95% 信賴區間排除了零，這進一步強調了這個平均值差異的統計顯著性。

T-Test 統計結果也確認了這個平均值差異是高度統計上顯著的。

綜合這些結果，我們可以得出結論：過程變更對訂單量有顯著的正向影響，即過程變更伴隨著平均訂單量的增加。

### (3)手算報表值

給定數據：

Mean1 = 25.2553, Mean2 = 32.4615

Std Dev1 = 6.8158, Std Dev2 = 7.9359

N1 = 47, N2 = 26

計算差異的標準誤差 (Satterthwaite)：

Std Err Diff =  $\sqrt{((47-1)*6.8158^2 + (26-1)*7.9359^2) / (47 + 26 - 2)} * \sqrt{1/47 + 1/26} \approx 1.8468$

計算 t 值 (均等變異數)：

t Value =  $(25.2553 - 32.4615) / \sqrt{(6.8158^2 / 47) + (7.9359^2 / 26)} \approx -4.08$

計算 t 值 (不均等變異數 - Satterthwaite)：

t Value =  $(25.2553 - 32.4615) / 1.8468 \approx -3.90$

計算變異數相等性的 F 值：

F Value =  $6.8158^2 / 7.9359^2 \approx 1.36$

## 2. proc anova

在過程變更的 ANOVA 分析中 F 值= 16.63, p 值= 0.0001 小於 0.05，拒絕虛無假設，代表對網站交易量的影響是統計上顯著的。

手算報表值：

Model MS =  $869.287223 / 1 = 869.287223$

Error MS =  $3711.397709 / 71 \approx 52.273207$

F Value =  $869.287223 / 52.273207 \approx 16.629$

## 3. 盒鬚圖比較

綜合來看，變化後的 Q1 和 Q3 都比變化前增加了。平均值和中位數都增加了，也就是說，變化導致了訂單量的分佈更加向上移動。

## 年對網站交易量

### 1. proc ttest

(1)變異數相等性的 p 值 $<.0001<0.05$  拒絕虛無假設，變異數不均等。

因此看 Satterthwaite (不均等)，p 值 $<.0001$  拒絕虛無假設，過程變更對於網站交易量的平均值存在統計上的顯著影響。

(2)2001 年時的平均值為 5，2002 年時的平均值為 12.7632，差異的平均值為 -7.7632

根據(1)(2):

2002 年時的平均值明顯高於 2001 年時，平均增加了約 7.7632。

95% 信賴區間排除了零，這進一步強調了這個平均值差異的統計顯著性。

T-Test 統計結果也確認了這個平均值差異是高度統計上顯著的。

綜合這些結果，我們可以得出結論：年對網站交易量有顯著的正向影響。

### 2. proc anova

在年的 ANOVA 分析中 F 值= 31，p 值 $=<.0001$  小於 0.05，拒絕虛無假設，代表對網站交易量的影響是統計上顯著的

### 3. 盒鬚圖比較

從盒鬚圖的比較中，我們可以看到，變化後的盒鬚圖向上移動了，中位數和 Q3 都增加了。這表明，變化導致了網站交易量的增加。

## 年對訂單量

### 1. proc ttest

(1)變異數相等性的 p 值=0.0195  $< 0.05$  拒絕虛無假設，變異數不均等。

因此看 Satterthwaite (不均等)，p 值 $<.0001$  拒絕虛無假設，年對於訂單量的平均值存在統計上的顯著影響。

(2) 2001 年時的平均值為 22.2571，2002 年時的平均值為 32.9474，差異的平均值為 -10.6902

根據(1)(2):

2002 年時的平均值明顯高於 2001 年，平均增加了約 10.6902。

95% 信賴區間排除了零，這進一步強調了這個平均值差異的統計顯著性。

T-Test 統計結果也確認了這個平均值差異是高度統計上顯著的。

綜合這些結果，我們可以得出結論：年對訂單量有顯著的正向影響。

### 2. proc anova

在過程變更的 ANOVA 分析中 F 值= 59.17，p 值 $=<.0001$  小於 0.05，拒絕虛無假設，代表對訂單量的影響是統計上顯著的

### 3. 盒鬚圖比較

從盒鬚圖的比較中，我們可以看到，變化後的盒鬚圖向上移動了，中位數和 Q3 都增加了。這表明，變化導致了訂單量的增加。

## 季對網站交易量

### 1. proc ttest

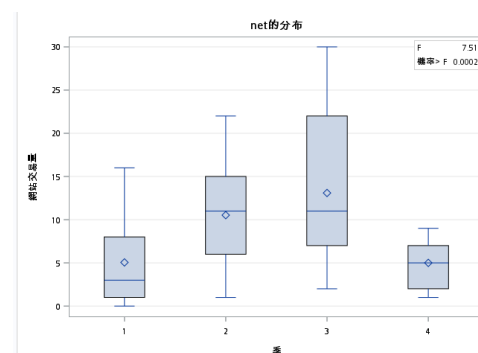
季有四個變數所以無法用 ttest

### 2. proc anova

在過程變更的 ANOVA 分析中 F 值= 7.51, p 值= 0.0002 小於 0.05, 拒絕虛無假設, 代表對網站交易量的影響是統計上顯著的

### 3. 盒鬚圖比較

從盒鬚圖的比較中, 我們可以看到, 網站交易量最高的是秋天, 然後是夏天、春天, 冬天是最低的。這表明, 季節對網站交易量的影響是統計上顯著的。



## 季對訂單量

### 1. proc ttest

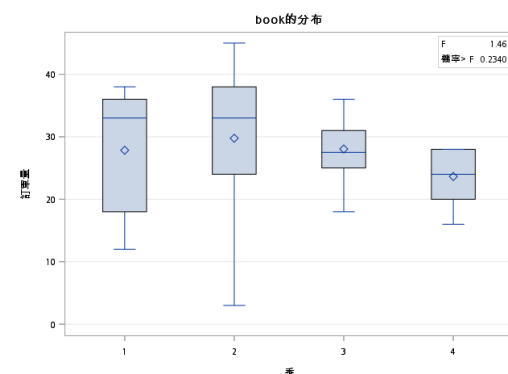
季有四個變數所以無法用 ttest

### 2. proc anova

在季的 ANOVA 分析中 F 值= 1.46, p 值=0.2340 大於 0.05, 無法拒絕虛無假設, 代表對訂單量的影響不顯著的

### 3. 盒鬚圖比較

從盒鬚圖的比較中, 我們可以看到, 夏天的訂單量是最多的, 再來是春天、秋天、冬天。但差別沒有很大。



#### 四. 研究兩兩連續間的關係：

##### 1. 網站交易與訂單量相關的相關性

網站交易量和訂單量之間的相關係數為 0.36517，表示它們之間存在正相關性，但不是非常強烈的相關。假設檢定的結果顯示，這種相關性是統計上顯著的，因為  $p\text{-value} = 0.0015$  低於顯著性水平 (0.05)，網站交易量和訂單量之間的相關性不是由於隨機變異引起的，而是真實存在的。

##### 2. 使用網站交易解釋訂單量的迴歸直線

(1)模型 F-檢定：模型的 F 值為 10.92，對應的  $p\text{-value}$  為 0.0015。這個結果顯示模型整體上是統計上顯著的，即至少有一個解釋變數對應變數有影響。

(2)參數估計值 net (網站交易量)：估計值為 0.41118，標準誤差為 0.12440，t 值為 3.31， $p\text{-value}$  為 0.0015。這顯示訂單量對網站交易量有統計上顯著的影響。

(3)配適圖：從圖表中可以看到，網站交易量與訂單量的百分比之間存在一定的正相關關係。也就是說，網站交易量越高，訂單量的百分比也越高。然而，這種關係並不完全線性，而是呈現出一定的散點。

根據圖表中提供的統計資訊，R 平方值為 0.1333，調整 R 平方值為 0.1211。這表明網站交易量與訂單量的百分比之間的關係並不強，但仍存在一定的相關性。

##### 3. 使用訂單量解釋網站交易量的迴歸直線

(1)模型 F-檢定：模型的 F 值為 10.92，對應的  $p\text{-value}$  為 0.0015。這個結果顯示模型整體上是統計上顯著的，即至少有一個解釋變數對應變數有影響。

(2)參數估計值 book (訂單量)：估計值為 0.32430，標準誤差為 0.09812，t 值為 3.31， $p\text{-value}$  為 0.0015。這顯示訂單量對網站交易量有統計上顯著的影響。

(3)配適圖：從圖表中可以看到，大部分的散點圖點位都位於一條傾斜的直線上，這表明網站交易量與訂單量之間存在正相關關係。也就是說，網站交易量越大，訂單量也越大。根據圖表中的 R 平方值，可以得出網站交易量與訂單量之間的相關性為 0.1333。R 平方值越大，表示相關性越強。因此，可以得出結論，網站交易量與訂單量之間的相關性不是很強，但也存在一定的相關性。

## 五. 回答目標問題

### A. 網站交貨量是否受到過程變更、年、季的影響？

\*進行了 t-test 和 ANOVA 分析，結果顯示過程變更對網站交易量有統計上顯著的正向影響，年對網站交易量有統計上顯著的正向影響，季對網站交易量有統計上顯著的影響。

總結：是的，網站交貨量受到過程變更、年、季的影響，並且這種影響是統計上顯著的。

### B. 訂單量是否受到過程變更、年、季的影響？

\*進行了 t-test 和 ANOVA 分析，結果顯示過程變更對訂單量有統計上顯著的正向影響，年對訂單量有統計上顯著的正向影響。

但是季對訂單量並沒有統計上顯著的影響。

總結：是的，訂單量受到過程變更、年的影響，並且這種影響是統計上顯著的。但是訂單量不會受到季的影響。

### C. 網站交貨量和訂單量是否有相關性？

\*進行了相關性分析和迴歸分析，結果顯示網站交易量和訂單量之間存在統計上顯著的正向相關性。

總結：是的，網站交貨量和訂單量之間存在統計上顯著的正向相關性。