

پروژه چند متغیره پیوسته

موارد مورد بحث در این پروژه

آزمون فرض مقایسه بردار میانگین های بیش از دو گروه مستقل

آزمون چند متغیره تحلیل واریانس یک طرفه MANOVA

آزمون Wilks

آزمون Pillai

آزمون Roy

آزمون Lawley_Hoteling

کاهش بعد

بررسی مقابله ها

1. معرفی داده ها

ما در این پروژه به بررسی 6 شرکت بزرگ از بورس ایران مربوط به صنعت خودرو شامل

۱. ایران خودرو (با نماد معاملاتی خودرو)

۲. سایپا (با نماد معاملاتی خساپا)

۳. گسترش سرمایه گذاری ایران خودرو (خگستر)

۴. خودروسازی بهمن (خبهمن)

۵. قطعات اتومبیل ایران (ختوقا)

۶. نیرو محرکه (خمحرکه)

در اردیبهشت ماه سال ۱۴۰۰ پرداخته ایم. در واقع ما سعی داریم ۶ گروه (نماد) از صنعت خودروسازی را بررسی کنیم.

متغیرهایی که ما در این بخش مورد بررسی قرار میدهیم شامل موارد زیر است :

۱. تعداد معاملات

تعداد معاملات در بورس تعداد دفعات انجام معامله در یک نماد را نشان می دهد. در واقع تعداد معاملات، نشان می -

دهد که یک نماد چندبار تعداد سهامش خرید و فروش شده است.

۲. حجم معاملات

در واقع حجم معاملات در سهم تعداد سهام یا اوراق بهادار خرید و فروش شده در یک بازه زمانی خاص (در اینجا یک روزه کاری) است.

۳. درصد تغییر قیمت پایانی

هر سهم در طول یک روز معاملاتی ممکن است در قیمت های متفاوتی معامله شود. اگر از این قیمت ها میانگین وزنی بگیریم، قیمت پایانی سهم تا همان لحظه بدست می آید پس در درصد تغییر قیمت پایانی می شود میزان تغییر قیمت پایانی یک سهم نسبت به روز کاری قبلی.

برای اجرای دستورات و انجام تحلیل ها ابتدا باید داده های خود را تعریف و فراخوانی کنیم

#تمام داده ها باهم بصورت زیر تعریف و بارگزاری میکنیم

```
#all  
data.all<-data.frame(read.table("C:/Users/12345/Desktop/mm.txt",header= T))
```

#هرگروه را بصورت جداگانه تعریف میکنیم

```
#group1:khesapa  
#group2:khodro  
#group3:khegostar  
#group4:khebahman  
#group5:khetogha  
#group6:khmohareke
```

#متغیرها را نام گذاری میکنیم

```
y1<-data.all$Volume  
y2<-data.all$`Number of transactions`  
y3<-data.all$`Final price percentage`  
group<-as.factor(data.all$group)
```

۲. آمار توصیفی

```
> (mean<-aggregate(data.all[,2:4], list(Group), mean))
```

	Group.1	Volume	Number.of.transactions	Final.price.percentage
1	1	373769150	9250.90	-1.005
2	2	476443698	12185.00	-0.885
3	3	144812570	5126.80	-1.105
4	4	281736639	5488.15	-1.150
5	5	10452519	512.05	-1.074
6	6	6153339	458.85	-0.830

از خروجی بالا نتایج زیر حاصل می‌شود

اگر به میانگین درصد تغییرات قیمت پایانی هر نماد را نگاه کنیم می‌بینیم که بسیار به یکدیگر نزدیک هستند یعنی اختلاف زیادی بین میانگین درصدها نیست که این نشأت گرفته از این است که این نمادها از یک نوع هستند. درواقع تفاوت معناداری بین میانگین‌های متغیرسوم هر گروه نیست.

گروه دوم (خودرو) بیشترین حجم معامله در روز یعنی چیزی در حدود ۴۷۰ میلیون و بیشترین تعداد معامله یعنی ۱۲ هزار سهم در روز است یعنی بطور میانگین در هر بار معامله ۴۰ هزار سهم رد و بدل شده است.

گروه سوم (خگستر) و گروه چهارم (خبهمن) حدوداً در هر روز تعداد معاملات برابری دارند ولی نماد خبهمن دارای حجم معاملات بیشتری است یعنی قدرت معامله‌گران نماد خبهمن بیشتر از قدرت معامله‌گران نماد خگستر است. این موضوع برای گروه پنجم (ختوقا) و ششم (خمحرکه) نیز صادق است یعنی هر معامله‌گر نماد ختوقا به طور میانگین در هر بار معامله ۲۰ هزار سهم معامله کرده ولی هر معامله‌گر نماد خمحرکه به طور میانگین در هر بار معامله ۱۳ هزار سهم معامله کرده که یعنی معامله‌گران نماد ختوقا از خمحرکه قوی‌تر هستند.

برای تمام گروه‌ها میانگین سهم‌های معامله شده در هر بار معامله بصورت زیر است.

```
> mean$Volume/mean$Number.of.transactions
```

[1]	40403.54	39100.84	28246.19	51335.45	20413.08	13410.35
-----	----------	----------	----------	----------	----------	----------

3. آزمون فرض مقایسه بردار میانگین های بیش از دو گروه مستقل

ایتدا فرض آزمون را تشکیل می دهیم

$$\begin{cases} H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6 \\ H_1: \text{حداقل میانگین دو گروه متفاوت است} \end{cases}$$

که در اینجا μ_1 یک بردار شامل میانگین ۳ متغیر (حجم و تعداد معامله و درصد تغییرات) برای گروه اول (خسایا) است.

آزمون چندمتغیره (MANOVA)

```
> result<-manova(cbind(y1,y2,y3)~Group,data=data.all)

> summary(result,test = "Pillai")
              Df  Pillai approx F num Df den Df    Pr(>F)
Group           5 0.63551      6.128    15   342 2.032e-11 ***
Residuals 114
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> summary(result,test = "Hotelling-Lawley")
              Df Hotelling-Lawley approx F num Df den Df    Pr(>F)
Group           5           1.0114    7.4616    15   332 2.753e-14 ***
Residuals 114
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> summary(result,test = "Wilks")
              Df  Wilks approx F num Df den Df    Pr(>F)
Group           5 0.45319    6.8525    15 309.58 7.772e-13 ***
Residuals 114
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> summary(result,test = "Roy")
              Df      Roy approx F num Df den Df    Pr(>F)
Group           5 0.75398   17.191      5   114 1.202e-12 ***
Residuals 114
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

چون در تمام آزمون ها p-مقدار ما کمتر از ۰.۰۵ است لذا فرض H_0 را رد میکنیم. یعنی حداقل میانگین دو گروه متفاوت است.

4. آزمون تک متغیره برای هر متغیر

```
> summary(aov(y1~group))
      Df      Sum Sq   Mean Sq F value    Pr(>F)
group    5 3.768e+18 7.536e+17   15.43 1.43e-11 ***
Residuals 114 5.568e+18 4.884e+16
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> summary(aov(y2~group))
      Df      Sum Sq   Mean Sq F value    Pr(>F)
group    5 2.184e+09 436762617   13.51 2.39e-10 ***
Residuals 114 3.685e+09 32326556
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> summary(aov(y3~group))
      Df Sum Sq Mean Sq F value Pr(>F)
group    5     2.8   0.570   0.113 0.989
Residuals 114  574.3   5.037
```

بنا به p -مقدار (بزرگتر از ۰.۰۵) مشخص می‌شود که متغیر سوم ما فقط فرض H_0 را رد نمی‌کند یعنی درصد تغییرات قیمت پایانی در هر ۶ گروه دارای اختلاف معنا داری نمی‌باشد به عبارتی دیگر ۶ گروه دارای درصد تغییرات قیمت پایانی برابری با هم هستند .

بنا به p -مقدار (کوچکتر از ۰.۰۵) مشخص می‌شود که متغیرها اول و دوم ما یعنی حجم معاملات و تعداد معاملات هر گروه فرض H_0 ما را رد می‌کند یعنی بین تعداد دفعات معامله شده و حجم معاملات ۶ گروه اختلاف معناداری وجود دارد که این نتایج را در بخش آمار توصیفی نیز مشاهده کردیم.

5. کاهش بعد

برای کاهش بعد ابتدا باید مقادیر ویژه بدست آوریم، برای بدست آوردن مقادیر ویژه نیاز به ماتریس‌های E و H داریم که از دستورات زیر بدست می‌آید

```
> summary(result)$SS
```

#ماتریس H

	y1	y2	y3
y1	3.767904e+18	8.883538e+13	1.258290e+08
y2	8.883538e+13	2.183813e+09	6.065488e+03
y3	1.258290e+08	6.065488e+03	1.615017e+00

#ماتریس E

	y1	y2	y3
y1	5.567861e+18	1.361820e+14	1.459957e+10
y2	1.361820e+14	3.685227e+09	2.694894e+05
y3	1.459957e+10	2.694894e+05	5.754974e+02

اکنون باید مقادیر ویژه ماتریس $E^{-1}H$ را بدست آوریم

```
> E<-summary(result)$SS$Group
> H<-summary(result)$SS$Residuals
> A<-t(E)%*%H
> e <- eigen(A)
> e$values
[1] 2.097917e+37 3.166235e+16 4.538450e+04
```

می‌دانیم که مقادیر ویژه شامل اطلاعات واریانس است. در اینجا می‌توان گفت تقریباً تمام واریانس مشاهدات در مقدار ویژه اول نمایان می‌شود لذا میتوان کاهش بعد از ۳ به ۱ داشته باشیم.

6. بررسی مقابله‌ها

برای مثال ما می‌خواهیم مقابله‌های زیر را آزمون کنیم:

$$1: 2 \mu_1 - \mu_2 = -1 \mu_3 = 2 \mu_4 = \mu_5 = -2 \mu_6$$

$$2: -2 \mu_1 = 1 \mu_2 = -1 \mu_3 = 1 \mu_5 = -2 \mu_6$$

آزمون فرض برای مقابله‌های بالا بصورت زیر است:

$$H_{01}: 2\mu_1 = \mu_2 = -1\mu_3 = 2\mu_4 = \mu_5 = -2\mu_6 \Rightarrow H_{01}: 2\mu_1 + \mu_2 + 2\mu_4 + \mu_5 = 2\mu_6 + 1\mu_3$$

$$H_{02}: -2\mu_1 = 1\mu_2 = -1\mu_3 = 1\mu_5 = -2\mu_6 \Rightarrow H_{02}: \mu_2 + \mu_5 = 2\mu_1 + 2\mu_6 + 1\mu_3$$

قبل از انجام آزمون باید به این نکته توجه داشت که نرم افزار R به بردارهایی که به صفر نزدیک هستند در اجرای بعضی از دستورات حساسیت نشان می‌دهد مثل دستورات پایین، از این رو چون متغیر سوم ما یعنی درصد تغییرات قیمت پایانی نزدیک به صفر است لذا ما این بردار را در ۱۰ ضرب می‌کنیم تا در اجرای دستور به مشکل نخوریم لازم به گفتن است این تغییر در مقادیر تغییری در نتایج ایجاد نمی‌کند (چون ما تمام داده‌های متغیر سوم را تغییر می‌دهیم) و این تغییر صرفاً برای رفع مشکل نرم‌افزار است.

اکنون فرضیات بالا در آزمون می‌کنیم

```
> data.all$y3<-(data.all$y3)*10
> Modell<-lm(cbind(data.all$y1,data.all$y2,data.all$y3)~Group,data = data.a
11)
> h01<- matrix(c(2 , 1 , -1 , 2 , 1 , -2),1 ,6)
> linearHypothesis(Modell, h01)

Sum of squares and products for the hypothesis:
      [,1]      [,2]      [,3]
[1,] 2.675564e+18 5.676194e+13 -5.277136e+12
[2,] 5.676194e+13 1.204202e+09 -1.119542e+08
[3,] -5.277136e+12 -1.119542e+08 1.040834e+07

Sum of squares and products for error:
      [,1]      [,2]      [,3]
[1,] 5.567861e+18 1.361820e+14 1.459957e+13
[2,] 1.361820e+14 3.685227e+09 2.694894e+08
[3,] 1.459957e+13 2.694894e+08 5.754974e+08

Multivariate Tests:
      Df test stat approx F num Df den Df      Pr(>F)
Pillai      1 0.4155732 26.54693      3    112 4.7964e-13 ***
Wilks      1 0.5844268 26.54693      3    112 4.7964e-13 ***
Hotelling-Lawley 1 0.7110783 26.54693      3    112 4.7964e-13 ***
Roy      1 0.7110783 26.54693      3    112 4.7964e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

مشخص است فرض اول طبق p-مقادیرهای بدست آمده رد می‌شود یعنی مقابله اول برقرار نیست.

حال برای مقابله دوم داریم

```
> h02<- matrix(c(-2,1,-1,0,1,-2),1 ,6)
> linearHypothesis(Moell1, h02)
```

Sum of squares and products for the hypothesis:

	[,1]	[,2]	[,3]
[1,]	4.838280e+15	285767706155	-1.991746e+10
[2,]	2.857677e+11	16878557	-1.176403e+06
[3,]	-1.991746e+10	-1176403	8.199303e+04

Sum of squares and products for error:

	[,1]	[,2]	[,3]
[1,]	5.567861e+18	1.361820e+14	1.459957e+12
[2,]	1.361820e+14	3.685227e+09	2.694894e+07
[3,]	1.459957e+12	2.694894e+07	5.754974e+06

Multivariate Tests:

	Df	test	stat	approx	F	num	Df	den	Df	Pr(>F)
Pillai	1	0.0280584	1.077754			3		112		0.36157
Wilks	1	0.9719416	1.077754			3		112		0.36157
Hotelling-Lawley	1	0.0288684	1.077754			3		112		0.36157
Roy	1	0.0288684	1.077754			3		112		0.36157

لذا بنا به p -مقدار فرض H_0 ما پذیرفته می‌شود یعنی مقابله دوم برقرار است. یعنی حاصل جمع میانگین گروه دوم و پنج برابر است با دو برابر میانگین‌های گروه اول و ششم به علاوه میانگین گروه سوم.

منبع داده‌ها

<https://www.tse.ir/archive.html>