رگرسیون وزنی

گاهی در عمل ثابت بودن واریانس خطاها برقرار نیست و خطاها در جهت y متناسب با افزایش x (با سرعت کمتر) افزایش می‌یابند در این موارد بهتر است از رگرسیون وزنی استفاده شود.

$$(x_1, y_1) \rightarrow s_1$$
$$(x_2, y_2) \quad s_2$$
$$\vdots$$
$$x_n \quad y_n \quad s_n$$

$$w_i = \frac{s_i^{-2}}{\sum_i s_i^{-2}/n} \qquad \frac{\frac{1}{s_i^2}}{\sum \frac{1}{n s_i^2}}$$

$$b_w = \frac{\sum w_i x_i y_i - n\bar{x}_w \bar{y}_w}{\sum w_i x_i^2 - n\bar{x}_w^2}$$

$$\bar{x}_w = \sum w_i x_i / n$$

$$\bar{y}_w = \sum w_i y_i / n \qquad a_w = \bar{y}_w - b\bar{x}_w$$

مثال ۲۸: رگرسیون معمولی و وزنی را برای داده‌های زیر بدست آورید (فقط خط)

x علف	۰	۲	۴	۶	۸	۱۰
انحراف معیار	۰/۰۰۱	۰/۰۰۴	۰/۰۱۰	۰/۰۱۳	۰/۰۱۷	۰/۰۲۲
y جذب	۰/۰۰۹	۰/۱۵۸	۰/۳۰۱	۰/۴۷۲	۰/۵۷۷	۰/۷۳۹

$\bar{x} = ۵$ $\quad \bar{y} = ۰.۳۷۶$ $\quad \sum x_i^2 = ۲۲۰$ $\quad \sum y_i^2 = ۱,۲۱۷۴۸$

$\sum x_i y_i = ۱۶,۳۵۸$ $\quad \sum y_i = ۲,۲۵۶$ $\quad \sum x_i = ۳۰$

Subject:
Date

$$\hat{b} = \frac{S_{xy}}{S_{xx}} = \frac{\sum x_i y_i - \sum x_i \sum y_i / n}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}} = \frac{۱۶٫۳۵۸ - \frac{(۲٫۲۵۶)(۳۰)}{۶}}{۲۲۰ - (۳۰)^۲/۶}$$

$$= \frac{۵٫۰۷۸}{۷۰} = ۰٫۰۷۲۵$$

$$\hat{a} = \bar{y} - \hat{b}\bar{x} = ۰٫۳۷۶ - ۰٫۰۷۲۵ \times ۵ = ۰٫۰۱۳۵$$

خط $= \hat{y} = ۰٫۰۱۳۵ + ۰٫۰۷۲۵\, x$

وزنی
برازش رگرسیون موزون

x_i	y_i	s_i	s_i^{-2}	w_i	$w_i x_i$	$w_i y_i$
۰	۰٫۰۰۹	۰٫۰۰۱	۱۰۰۰۰۰۰	۵٫۵۳۵	۰	۰٫۰۴۹۸
۲	۰٫۱۵۸	۰٫۰۰۴	۶۲۵۰۰	۰٫۳۴۶	۰٫۶۹۲	۰٫۰۵۴۷
۴	۰٫۳۰۱	۰٫۰۱	۱۰۰۰۰	۰٫۰۵۵	۰٫۲۲۰	۰٫۰۱۶۶
۶	۰٫۴۷۲	۰٫۰۱۳	۵۹۱۷	۰٫۰۳۳	۰٫۱۹۸	۰٫۰۱۵۶
۸	۰٫۵۷۷	۰٫۰۱۷	۳۴۶۰	۰٫۰۱۹	۰٫۱۵۲	۰٫۰۱۱۰
۱۰	۰٫۷۳۹	۰٫۰۲۲	۲۰۶۶	۰٫۰۱۱	۰٫۱۱۰	۰٫۰۰۸۱
مجموع				۵٫۹۹۹ $\simeq$ ۶ = n	۱٫۳۷۲	۰٫۱۵۵۸

$w_i x_i y_i$	$w_i x_i^2$
۰	۰
۰٫۱۰۹۳	۱٫۳۸۴
۰٫۰۶۶۲	۰٫۸۸۰
۰٫۰۹۳۵	۱٫۸۸۸
۰٫۰۸۷۷	۱٫۲۱۶
۰٫۰۸۱۳	۱٫۱۰۰
۰٫۴۳۸	۵٫۷۶۸

$$\bar{x}_w = \frac{۱٫۳۷۶}{۶} = ۰٫۲۲۹$$

$$\bar{y}_w = \frac{۰٫۱۵۵۸}{۶} = ۰٫۰۲۶$$

$$\hat{a}_w = ۰٫۰۲۶ - ۰٫۰۷۳۸ \times ۰٫۲۲۹ = ۰٫۰۰۹۱$$

$$\hat{b}_w = \frac{۰٫۴۳۸ - (۶ \times ۰٫۲۲۹ \times ۰٫۰۲۶)}{۵٫۷۶۸ - ۶(۰٫۲۲۹)^۲} = ۰٫۰۷۳۸$$

P4PCO

خط رگرسیون موزون $\hat{y}_w = 0.0091 + 0.0738\, x_w$

برآورد σ_w : $\hat{\sigma}_w = \sqrt{\sum_i w_i (y_i - \hat{y}_i)^2 / n-2}$

انحراف معیار سیگنال

$$S_{x_{0w}} = \frac{\hat{\sigma}_w}{\hat{b}} \left\{ \frac{1}{w_0} + \frac{1}{n} + \frac{(y_0 - \bar{y}_w)^2}{\hat{b}^2 (\sum w_i x_i^2 - n \bar{x}_w^2)} \right\}^{1/2}$$

فاصله اطمینان $x_{0w} : \hat{x}_{0w} \pm t_{n-2, 1-\alpha/2} S_{x_{0w}}$

تمرین تحویلی ۳۹: فلئورسانس هر یک از سری محلولهای استاندارد کینین پنج بار تعیین شده است نتایج زیر ارائه شده اند

غلظت	۰	۱۰	۲۰	۳۰	۴۰	۵۰
شدت فلئورسانس	۴	۲۲	۴۴	۶۰	۷۵	۱۰۴
	۳	۲۰	۴۶	۶۳	۸۱	۱۰۹
	۴	۲۱	۴۵	۶۰	۷۹	۱۰۷
	۵	۲۲	۴۴	۶۳	۷۸	۱۰۱
	۴	۲۱	۴۴	۶۳	۷۷	۱۰۵
	S_1	S_2	S_3	S_4	S_5	S_6

$n = 30$

الف) خط رگرسیون موزون ب) خط رگرسیون معمولی (ناموزون)

ج) حدود اطمینان برای غلظت محلولهایی با شدت فلئورسانس ۱۵ و ۹۰ را بدست آورید.

۰	۰	۰	۰	۰	۱۰	۱۰	...	۵۰
۴	۳	۴	۵	۴	۲۲	۲۰	...	۱۰۵

```
> x<-c(0,2,4,6,8,10)
> x
[1]  0  2  4  6  8 10
> y<-c(0.009,0.158,0.301,0.472,0.577,0.739)
>
> M<-lm(y~x)
> M

Call:
lm(formula = y ~ x)

Coefficients:
(Intercept)            x
    0.01329      0.07254

> anova(M)
Analysis of Variance Table

Response: y
          Df  Sum Sq Mean Sq F value    Pr(>F)
x          1 0.36837 0.36837  1730.7 1.995e-06 ***
Residuals  4 0.00085 0.00021
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> summary(M)

Call:
lm(formula = y ~ x)

Residuals:
         1          2          3          4          5          6
-0.0042857 -0.0003714 -0.0024571  0.0234571 -0.0166286  0.0002857

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.013286   0.010559   1.258    0.277
x           0.072543   0.001744  41.602    2e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.01459 on 4 degrees of freedom
Multiple R-squared:  0.9977,	Adjusted R-squared:  0.9971
F-statistic:  1731 on 1 and 4 DF,  p-value: 1.995e-06
```

```
>
> Si<-c(0.001,0.004,0.01,0.013,0.017,0.022)
> Si
[1] 0.001 0.004 0.010 0.013 0.017 0.022
> cbind(x,y,Si)
      x     y    Si
[1,]  0 0.009 0.001
[2,]  2 0.158 0.004
[3,]  4 0.301 0.010
[4,]  6 0.472 0.013
[5,]  8 0.577 0.017
[6,] 10 0.739 0.022
>
> wi<-Si^(-2)/mean(Si^(-2))
> wi
[1] 5.53534395 0.34595900 0.05535344 0.03275351 0.01915344 0.01143666
> sum(wi)
[1] 6
>
> M1<-lm(y~x, weights=wi)
> M1

Call:
lm(formula = y ~ x, weights = wi)

Coefficients:
(Intercept)            x
   0.009084     0.073760

>
>
```

```
> ### Example 39
> x<-rep(c(0,10,20,30,40,50),each=5)
> x
 [1]  0  0  0  0  0 10 10 10 10 10 20 20 20 20 20 30 30 30 30 30 40 40 40 40 40 50
[27] 50 50 50 50
> y<-c(4,3,4,5,4,22,20,21,22,21,44,46,45,44,44,60,63,60,63,63,
+    75,81,79,78,77,104,109,107,101,105)
>
> M<-lm(y~x)
> M

Call:
lm(formula = y ~ x)

Coefficients:
(Intercept)            x
      2.924        1.982

> anova(M)
Analysis of Variance Table

Response: y
          Df Sum Sq Mean Sq F value    Pr(>F)
x          1  34363   34363    3780 < 2.2e-16 ***
Residuals 28    255       9
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> summary(M)

Call:
lm(formula = y ~ x)

Residuals:
    Min      1Q  Median      3Q     Max
-7.1924 -1.7410  0.6248  1.4419  6.9905

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.92381    0.97589   2.996  0.00567 **
x            1.98171    0.03223  61.482  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.015 on 28 degrees of freedom
Multiple R-squared:  0.9926,	Adjusted R-squared:  0.9924
F-statistic:  3780 on 1 and 28 DF,  p-value: < 2.2e-16
```

```
> ################################
> ### weighted regression ###
> #############################
> S<-aggregate(y~x,FUN=sd)
> S
   x         y
1  0 0.7071068
2 10 0.8366600
3 20 0.8944272
4 30 1.6431677
5 40 2.2360680
6 50 3.0331502
> Si<-rep(S$y,each=5)
> Si
 [1] 0.7071068 0.7071068 0.7071068 0.7071068 0.7071068 0.8366600 0.8366600
 [8] 0.8366600 0.8366600 0.8366600 0.8944272 0.8944272 0.8944272 0.8944272
[15] 0.8944272 1.6431677 1.6431677 1.6431677 1.6431677 1.6431677 2.2360680
[22] 2.2360680 2.2360680 2.2360680 2.2360680 3.0331502 3.0331502 3.0331502
[29] 3.0331502 3.0331502
> cbind(x,y,Si)
       x   y        Si
 [1,]  0   4 0.7071068
 [2,]  0   3 0.7071068
 [3,]  0   4 0.7071068
 [4,]  0   5 0.7071068
 [5,]  0   4 0.7071068
 [6,] 10  22 0.8366600
 [7,] 10  20 0.8366600
 [8,] 10  21 0.8366600
 [9,] 10  22 0.8366600
[10,] 10  21 0.8366600
[11,] 20  44 0.8944272
[12,] 20  46 0.8944272
[13,] 20  45 0.8944272
[14,] 20  44 0.8944272
[15,] 20  44 0.8944272
[16,] 30  60 1.6431677
[17,] 30  63 1.6431677
[18,] 30  60 1.6431677
[19,] 30  63 1.6431677
[20,] 30  63 1.6431677
[21,] 40  75 2.2360680
[22,] 40  81 2.2360680
[23,] 40  79 2.2360680
[24,] 40  78 2.2360680
[25,] 40  77 2.2360680
[26,] 50 104 3.0331502
[27,] 50 109 3.0331502
[28,] 50 107 3.0331502
[29,] 50 101 3.0331502
[30,] 50 105 3.0331502
```

```
> wi<-Si^(-2)/mean(Si^(-2))
> wi
 [1] 2.2397932 2.2397932 2.2397932 2.2397932 2.2397932 1.5998523 1.5998523
 [8] 1.5998523 1.5998523 1.5998523 1.3998708 1.3998708 1.3998708 1.3998708
[15] 1.3998708 0.4147765 0.4147765 0.4147765 0.4147765 0.4147765 0.2239793
[22] 0.2239793 0.2239793 0.2239793 0.2239793 0.1217279 0.1217279 0.1217279
[29] 0.1217279 0.1217279
> sum(wi)
[1] 30
> M1<-lm(y~x, weights=wi)
> M1

Call:
lm(formula = y ~ x, weights = wi)

Coefficients:
(Intercept)            x
      3.481        1.963

>
```

Subject:
Date

مثال

۱- داده های زیر بازیابی برومور را از نمونه های گیاهی آغشته که با استفاده از یک روش کروماتوگرافی گاز-مایع اندازه گیری شده اند به دست می دهد. مقدار برابری از برومور پتاسیم به هر نمونه افزوده شده است

گوجه فرنگی	۷۶۴	۷۵۸	۷۷۰	۷۹۰	۷۵۹	۷۹۰	۷۷۷
خیار	۷۸۳	۷۹۷	۷۸۹	۷۶۵	۷۷۸	۷۷۳	۷۸۲

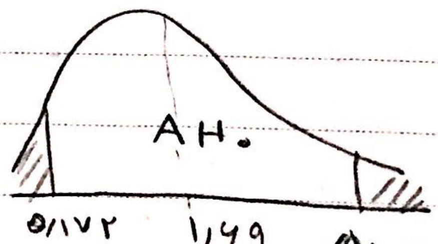
الف) بررسی کنید آیا بازیابی ها از دو گروه از واریانس های با تفاوت های معنی دار برخوردارند؟

ب) بررسی کنید آیا میانگین بازیابی ها به طور معنی دار از هم متفاوتند؟

$\bar{X} = 772.57 \quad \bar{Y} = 780.86 \quad n_1 = 7$

$S_1 = 13.563 \quad S_2^2 = 10.42 \quad n_2 = 7$

$$F_0 = \frac{S_1^2}{S_2^2} = 1.69$$



$$F_{6,6,0.975} = 5.82$$

$$F_{6,6,0.025} = \frac{1}{F_{6,6,0.975}} = \frac{1}{5.82} = 0.172$$

$1.69 < 5.82 \rightarrow AH_0$

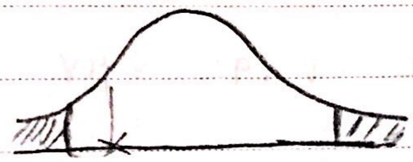
Subject:
Date

$$T_0 = \frac{\bar{X} - \bar{Y}}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} =$$

$$S_p = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{6(13.563)^2 + 6(10.42)^2}{7 + 7 - 2}}$$

$$\cong 12.1$$

$$T_0 = \frac{772.57 - 780.86}{9.86\sqrt{\frac{1}{7}+\frac{1}{7}}} \cong -1.28$$



$-t_{7+7-2}$ $\quad$ t_{7+7-2}

-2.18 $\quad$ 2.18

$$|1.28| \ngtr 2.18$$

$$\Rightarrow AH_0$$

$$\mu_1 - \mu_2 : (772.57 - 780.86) \pm 2.18 \cdot 12.1 \sqrt{1/7 + 1/7}$$

$$\cong (-22.37 \,,\, 5.79)$$

با اطمینان ۹۵٪ تفاضل میانگین غلظت در این بازه

قرار می‌گیرد که شامل صفر است نشان می‌دهد

تفاوت معنی داری بین میانگین‌ها وجود ندارد

PAPCO

Scanned with CamScanner

مثال ۴۱

۲- جدول زیر غلظت نوراپی نفرین (میکرومول بر گرم کراتینین) را در ادرار داوطلبان سالم در شروع ۲۰ سالگی نشان می دهد آیا نشانه ای از متفاوت بودن بین غلظت نوراپی نفرین در دو جنس وجود دارد.

$$\begin{cases} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 \neq \mu_2 \end{cases}$$

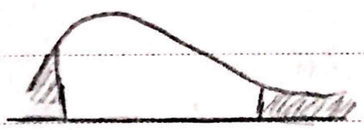
X	مذکر	۰٫۴۸	۰٫۳۶	۰٫۲	۰٫۵۵	۰٫۴۵	۰٫۴۶	۰٫۴۷	۰٫۲۳
Y	مونث	۰٫۳۵	۰٫۳۷	۰٫۲۷	۰٫۲۹				

$$\begin{cases} H_0: \sigma_1^2 = \sigma_2^2 \\ H_1: \sigma_1^2 \neq \sigma_2^2 \end{cases}$$ ابتدا این فرضیه آزمون می شود

$\bar{X} = 0.4$ $\qquad S_1 = 0.126$ $\qquad n_1 = 8$

$\bar{Y} = 0.32$ $\qquad S_2 = 0.0476$ $\qquad n_2 = 4$

$$F_0 = \frac{S_1^2}{S_2^2} = \frac{0.126^2}{0.0476^2} \cong 7 \qquad F_{7,3,0.025} = ?$$



۰٫۱۶۷ $\qquad$ ۱۴٫۶

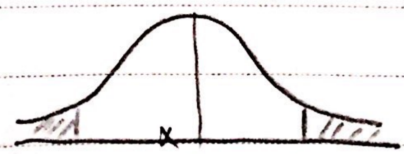
$$F_{7,3,0.975} = 14.6$$

$$F_{3,7,0.975} = 5.99$$

$$F_{7,3,0.025} = \frac{1}{5.99} = 0.167$$

فرض برابری واریانس پذیرفته می شود

$$T_0 = \frac{0.4 - 0.32}{\underbrace{s_p\sqrt{\frac{1}{8}+\frac{1}{4}}}_{0.114}} \simeq 1.2 \qquad s_p = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}} \simeq 0.11$$



-2.22 $\qquad 0 \qquad$ 2.22

$|0.11| < 2.22$

$\Rightarrow A\,H_0$

$$\mu_1 - \mu_2 : \bar{X} - \bar{Y} \pm S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\; t_{n_1+n_2-2,\,\alpha/2} =$$

$$= (-0.098\,,\ 0.23)$$

مثال ۴۲

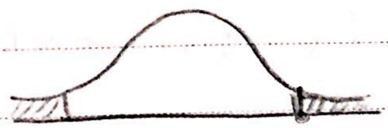
۳- به منظور ارزیابی یک روش اسپکتروفوتومتری برای تعیین مقدار تیتان نمونه‌های آلیاژ حاوی مقادیر متفاوت تأیید شده از تیتان بکار گرفته شده است

نتایج برای هر آلیاژ براساس ۸ تکرار به شرح زیر است

نمونه آلیاژ	مقدار تأیید شده	میانگین نتیجه اسپکتروفوتومتری	انحراف معیار
1	۰/۴۹۶	۰/۴۸۲	۰/۰۲۵۸
2	۰/۹۹۵	۱/۰۰۹	۰/۰۲۴۸
3	۱/۴۹۳	۱/۵۰۵	۰/۰۲۸۷
4	۱/۹۹۰	۲/۰۰۲	۰/۰۲۱۲

برای هر آلیاژ تفاوت معنی‌دار میانگین از مقدار تأیید شده را بیابید. برای آلیاژ ۲ فاصله اطمینان ۹۵٪ به دست آورید.

$$\begin{cases} H_0: \mu_1 = 0.496 \\ H_1: \mu_1 \neq 0.496 \end{cases} \qquad T_0 = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}} = \frac{0.48 - 0.49}{\frac{0.0258}{\sqrt{8}}} \simeq -1.53$$



$-t_{7, 0.975}$ $\quad$ $t_{7, 0.975}$

-2.36 $\quad$ 2.36

$$|T_0| = 1.53 < 2.36 \Rightarrow AH_0$$

$$\begin{cases} H_0: \mu_2 = 0.995 \\ H_1: \mu_2 \neq 0.995 \end{cases} \qquad T_0 = \frac{1.009 - 0.995}{\frac{0.0248}{\sqrt{8}}} = 1.597$$

$$1.597 < 2.36 \rightarrow AH_0$$

$$\begin{cases} H_0: \mu_3 = 1.493 \\ H_1: \mu_3 \neq 1.493 \end{cases} \qquad T_0 = \frac{1.493 - 1.505}{\frac{0.0287}{\sqrt{8}}} = -1.182 \rightarrow AH_0$$

$$\begin{cases} H_0: \mu_4 = 1.990 \\ H_1: \mu_4 \neq 1.990 \end{cases} \qquad T_0 = \frac{2.002 - 1.990}{0.0212/\sqrt{8}} = 1.6 \rightarrow AH_0$$

$$\mu: \bar{X} \pm t_{n-1} \frac{S}{\sqrt{n}} = 1.009 \mp 2.36 \frac{0.0248}{\sqrt{8}} = (0.988 , 1.0296)$$

مثال ۴۲. یک روش جدید طیف سنجی جذب اتمی شعله ای برای تعیین مقدار آنتیموان در اتمسفر با روش رنگ سنجی مقایسه شده است برای نمونه هایی از یک اتمسفر شهری نتایج زیر بدست آمده است

نمونه	روش جدید	روش استاندارد	d_i
1	22.2	25	-2.8
2	19.2	19.5	-0.3
3	15.7	16.6	-0.9
4	20.4	21.3	-0.9
5	19.6	20.7	-1.1
6	15.7	16.8	-1.1

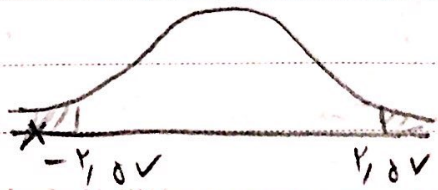
sum = -7.1

$$\begin{cases} H_0: \overset{\delta}{\overline{\mu_d}} = 0 \\ H_1: \mu_d \neq 0 \end{cases}$$

$\bar{D} = -1.18$

$$S_d = \sqrt{\frac{1}{5}\left[(-1.18-1.017)^2 + \dots (-1.1-1.017)^2\right.} = 0.84$$

$$T_0 = \frac{\overset{\bar{D}}{-1.18}}{\underset{S_d/\sqrt{n}}{\frac{0.84}{\sqrt{6}}}} = -3.43$$



$|-3.43| > 2.57 \Rightarrow RH_0$

تفاوت معنی داری بین دو روش وجود دارد

$$\delta = \mu_d : \bar{d} \pm \frac{S_d}{\sqrt{n}} t_{n-1} = (-2.07, -0.296)$$

نکته: برای آزمون کردن واریانس جامعه برابر مقدار خاصی هست یا خیر نیز

$$\begin{cases} H_0: \sigma^2 = \sigma_0^2 \\ H_1: \sigma^2 \neq \sigma_0^2 \end{cases} \quad \chi_0^2 = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2_{n-1}$$

می توانید از آماره ی

$$\text{if } \chi_0^2 > \chi^2_{n-1, 1-\alpha/2} \Rightarrow RH_0$$

$$S^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$$

استفاده کنید اگر میانگین جامعه معلوم بود (μ) آن گاه آماره ی آزمون

$$\begin{cases} H_0: \sigma^2 = \sigma_0^2 \\ H_1: \sigma^2 \neq \sigma_0^2 \end{cases} \quad \chi_0^2 = \frac{n S_n^2}{\sigma_0^2} \sim \chi^2_n \quad \text{if } \chi_0^2 > \chi^2_{n-1, 1-\alpha/2} \Rightarrow RH_0$$

$$S_n^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2$$

است.

فاصله اطمینان $\sigma^2 : \left(\frac{(n-1)S^2}{\chi^2_{n-1, 1-\alpha/2}} , \frac{(n-1)S^2}{\chi^2_{n-1, \alpha/2}} \right)$

فاصله اطمینان وقتی μ معلوم $\sigma^2 : \left(\frac{n S_n^2}{\chi^2_{n, 1-\alpha/2}} , \frac{n S_n^2}{\chi^2_{n, \alpha/2}} \right)$

مثال ٤٤: میزان جذب یک طیف سنج در طول موج خاصی با یک محلول استاندارد که دارای جذب 0.47 و انحراف معیار 0.002 است آزمایش شده است با ده بار تکرار میانگین جذب 0.461 و انحراف معیار 0.003 بدست آمد یک فاصله اطمینان ٪۹۵ برای میانگین و انحراف معیار جذب بدست آورید و $\begin{cases} H_0: \sigma^2 = 0.002^2 \\ H_1: \sigma^2 \neq 0.002^2 \end{cases}$ را آزمون کنید

$$\mu: \bar{x} \pm t_{n-1, 1-\alpha/2} \frac{S}{\sqrt{n}} = 0.461 \pm 2.26 \cdot \frac{0.003}{\sqrt{10}} = (0.459, 0.463)$$

$$\sqrt{\sigma^2}: \left(\sqrt{\frac{9(0.003)^2}{19.02}} , \sqrt{\frac{9(0.003)^2}{2.7}} \right) = (0.0021, 0.0055)$$

$$\chi_0^2 = \frac{9(0.003)^2}{0.002^2} = 20.25 > 19.02 \Rightarrow RH_0$$

```
> #######################example40######################################
> x=c(777,790,759,790,770,758,764)
> y=c(782,773,778,765,789,797,782)
> mean(x)
[1] 772.5714
> mean(y)
[1] 780.8571
> sd(x)
[1] 13.5629
> sd(y)
[1] 10.41519
> t.test(x,y,alternative="two.sided",conf.level=0.95,var.equal = TRUE)

	Two Sample t-test

data:  x and y
t = -1.2819, df = 12, p-value = 0.2241
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -22.36825   5.79682
sample estimates:
mean of x mean of y 
 772.5714  780.8571 

> var.test(x, y, alternative = "two.sided")

	F test to compare two variances

data:  x and y
F = 1.6958, num df = 6, denom df = 6, p-value = 0.5371
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.2913843 9.8690604
sample estimates:
ratio of variances 
          1.695786 

> qt(0.975,12)
[1] 2.178813
> 
```

```
> ########################example41########################################
> x=c(0.48,0.36,0.2,0.55,0.45,0.46,0.47,0.23)
> y=c(0.35,0.37,0.27,0.29)
> mean(x)
[1] 0.4
> mean(y)
[1] 0.32
> sd(x)
[1] 0.1255843
> sd(y)
[1] 0.04760952
> t.test(x,y,alternative="two.sided",conf.level=0.95,var.equal = TRUE)

	Two Sample t-test

data:  x and y
t = 1.2067, df = 10, p-value = 0.2553
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.06771401  0.22771401
sample estimates:
mean of x mean of y 
     0.40      0.32 

> var.test(x, y, alternative = "two.sided")

	F test to compare two variances

data:  x and y
F = 6.958, num df = 7, denom df = 3, p-value = 0.139
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
  0.4757792 40.9812628
sample estimates:
ratio of variances 
          6.957983 

>
```

```
> #######################example43#######################################
> x=c(22.2,19.2,15.7,20.4,19.6,15.7)
> y=c(25,19.5,16.6,21.3,20.7,16.8)
> d=x-y
> d
[1] -2.8 -0.3 -0.9 -0.9 -1.1 -1.1
> sum(d)
[1] -7.1
> mean(x-y)
[1] -1.183333
> sd(x-y)
[1] 0.8447879
>
> t.test(x,y,paired=TRUE, alternative="two.sided",conf.level=0.95,var.equal =
FALSE)

    Paired t-test

data:  x and y
t = -3.4311, df = 5, p-value = 0.01861
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2.0698839 -0.2967828
sample estimates:
mean of the differences
              -1.183333

>
```

رگرسیون چندگانه Multiple Regression

مثال ۴۵

در بسیاری از تحقیقات متغیر وابسته‌ی Y تحت تاثیر چند متغیر مستقل هستند

X_1	X_2	X_3	Y
۹۱٫۵	۵۶٫۱	۷۳٫۶	۰٫۸۸۸
۹۳٫۸	۵۶٫۳	۷۴٫۱	۰٫۴۶۱
۹۳٫۴	۵۶٫۴	۷۴٫۵	۰٫۴۵۳
۹۲٫۵	۵۶٫۷	۷۳٫۷	۰٫۵۶
۹۴٫۸	۵۶٫۵	۷۳٫۶	۰٫۴۱۴
۹۳٫۲	۵۶٫۸	۷۳٫۹	۰٫۴۳۸
۹۳٫۷	۵۷	۷۴٫۴	۰٫۳۴۲
۹۱٫۵	۵۶٫۸	۷۳٫۹	۰٫۷۴۳
۹۲٫۷	۵۵٫۷	۷۳٫۹	۰٫۷۵۱
۹۲٫۷	۵۷٫۷	۷۳٫۸	۰٫۴۷۷

و به دنبال $Y = a + b_1X_1 + b_2X_2 + b_3X_3 + \varepsilon$ هستیم که به صورت

$$\hat{Y} = \hat{a} + \hat{b}_1X_1 + \hat{b}_2X_2 + \hat{b}_3X_3$$

برآورد کنیم — یا به طور کلی

$$\hat{Y} = \hat{a} + \hat{b}_1X_1 + \cdots + \hat{b}_pX_p$$

برای رگرسیون چندگانه

$$\underset{\sim}{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \qquad \underset{\sim}{X} = \begin{pmatrix} 1 & X_{11} & \cdots & X_{1p} \\ \vdots & \vdots & & \\ 1 & X_{n1} & \cdots & X_{np} \end{pmatrix}_{n \times (p+1)}$$

$$\underset{\sim}{\beta} = \begin{pmatrix} a \\ b_1 \\ \vdots \\ b_p \end{pmatrix}_{(p+1) \times 1} \qquad \underset{\sim}{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

و مدل ماتریسی رگرسیون چندگانه به صورت

$$\underset{\sim}{Y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon} \quad , \quad \hat{\underset{\sim}{Y}} = X\hat{\underset{\sim}{\beta}}$$

تعریف می شود و فرضیات $E(\underset{\sim}{\varepsilon}) = \underset{\sim}{0}$ میانگین خطا صفر ، $var(\underset{\sim}{\varepsilon}) = \sigma^2 I = \begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ 0 & \cdots & 0 & \sigma^2 \end{pmatrix}$

از روش حداقل مربعات خطا $\sum e_i^2 \to min$ ثابت می شود

$$\hat{\underset{\sim}{\beta}} = (X'X)^{-1}X'\underset{\sim}{Y}$$

مثال ۴۶

در مثال ۸.ب $\hat{\underset{\sim}{\beta}}$ را بدست آورید (با نرم افزار) M = lm(Y ~ X1 + X2 + X3)

$\hat{a} = 31.6879 \quad \hat{b}_1 = -0.1289 \quad \hat{b}_2 = -0.1526$

$\hat{b}_3 = -0.1421$

$$\hat{y} = 31.6879 - 0.1289X_1 - 0.1526X_2 - 0.1421X_3$$

جدول آنالیز واریانس

منبع	SS مجموع مربعات	df	MS	F_0
رگرسیون	SSR	p	$MSR = SSR/p$	$\frac{MSR}{MSE}$
خطا	SSE	n − p − 1	$MSE = SSE/(n-p-1)$	
کل	SST	n − 1		

معنی داری رگرسیون if $F_0 > F_{p, n-p-1, 1-\alpha} \Rightarrow RH_0$

$$\begin{cases} H_0: \underset{\sim}{\beta} = 0 \\ H_1: \underset{\sim}{\beta} \neq 0 \end{cases}$$

$$SSR = \hat{\underset{\sim}{\beta}}' X' \underset{\sim}{y} - n\bar{y}^2$$

$$SST = \underset{\sim}{y}'\underset{\sim}{y} - n\bar{y}^2 = \sum y_i^2 - n\bar{y}^2$$

$$SSE = SST - SSR$$

$$\hat{\sigma}^2 = MSE = \frac{SSE}{n-p-1} \qquad var(\hat{\underset{\sim}{\beta}}) = \hat{\sigma}^2 (X'X)^{-1}$$

$$R^2 = \frac{SSR}{SST}$$ ضریب تعیین

مثال ۴۷ با نرم افزار جدول آنالیز واریانس مثال ۴۵ : anova(M)

منبع	SS	df	MS	F_0
رگرسیون	۰٫۲۷۵۸۳۳	۳	0.0919	42.27
خطا	۰٫۰۱۳۰۵۱	4	0.002175	
کل	۰٫۲۸۸۸۸۴	9		

$$F_{3,6,0.95} = 4.757$$

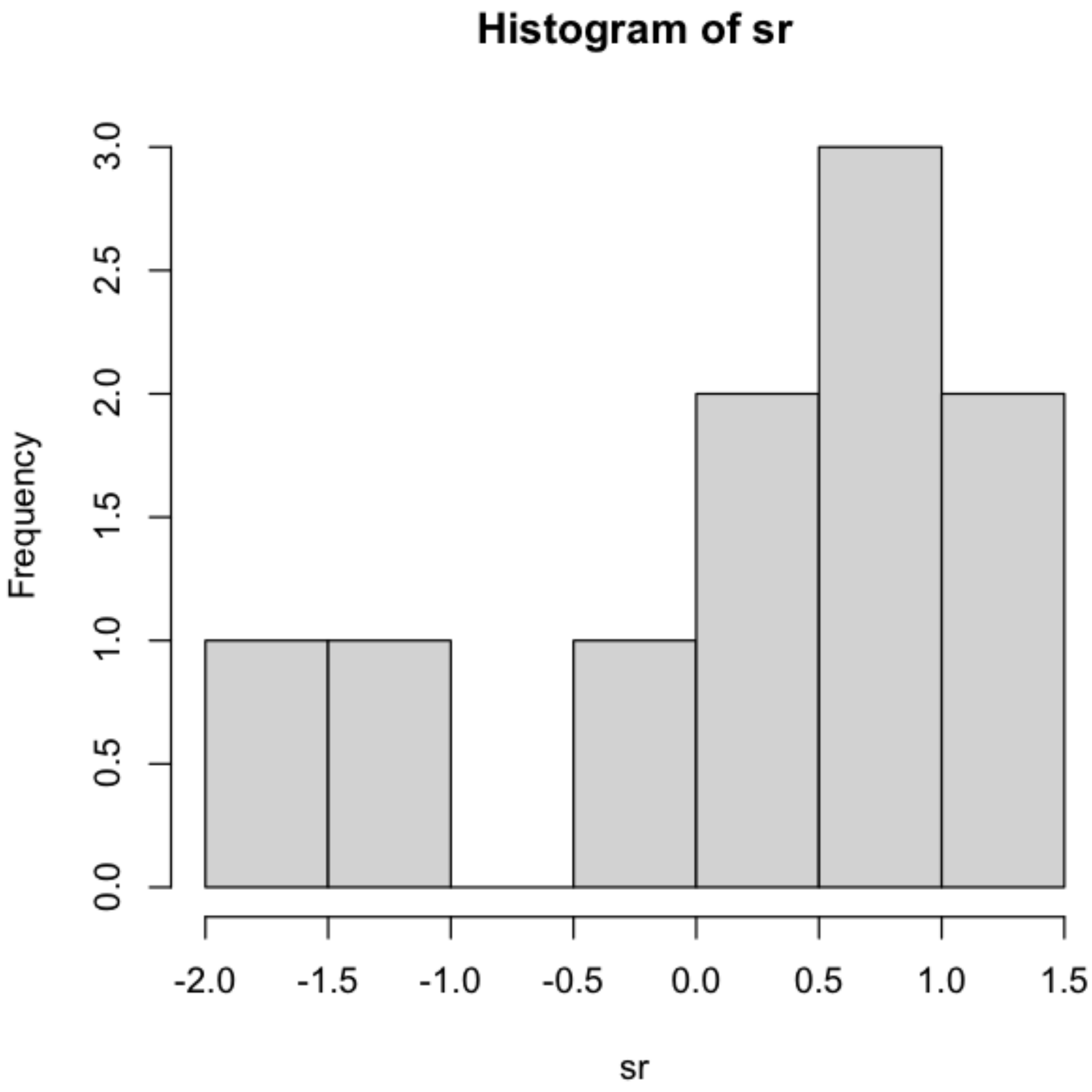
$$42.27 > 4.757 \Rightarrow RH_0$$

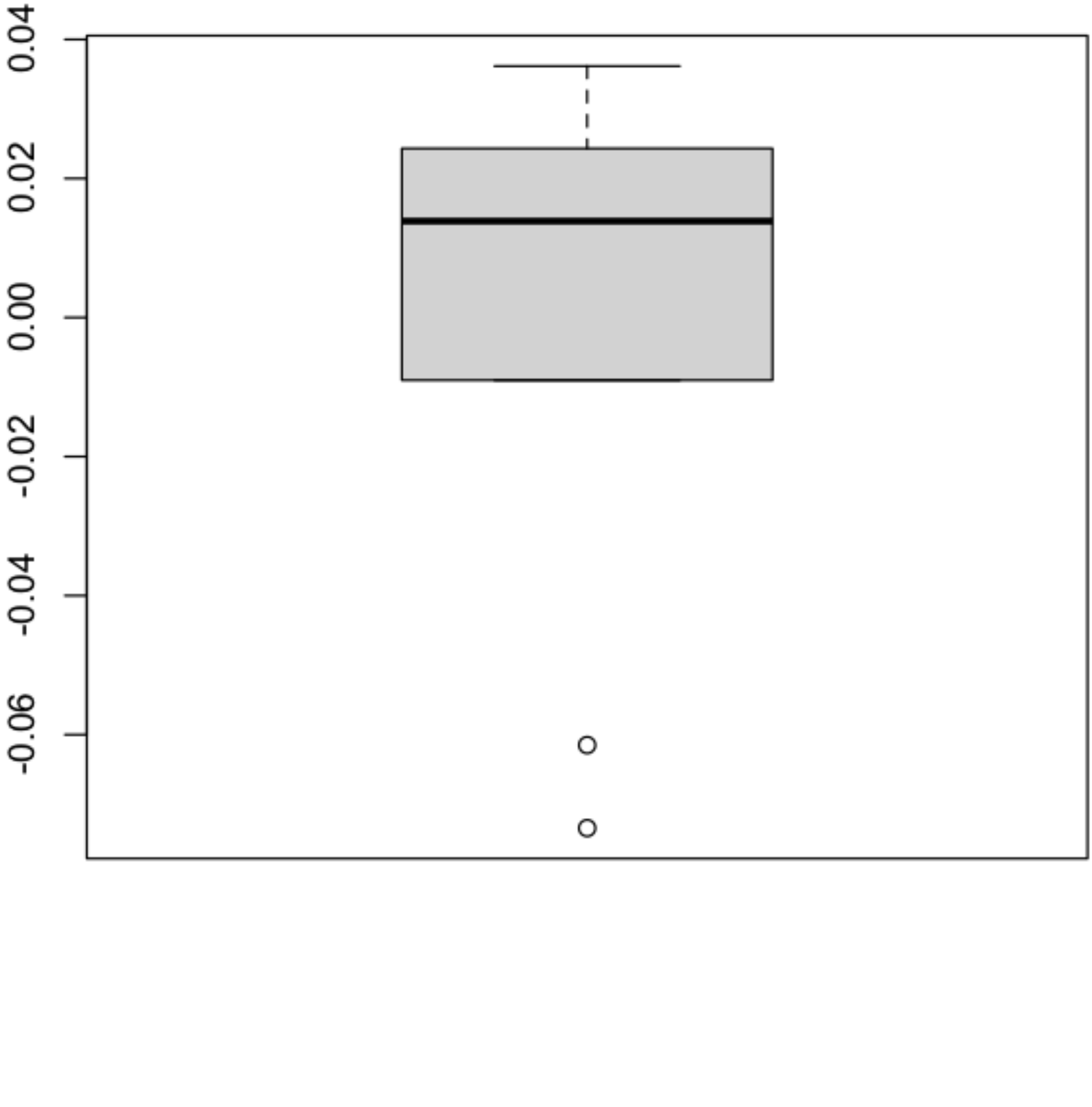
رگرسیون معنی دار است

$R^2 = 0.9548$ بیش از ۹۵٪ تغییرات y توسط x_1 و x_2 و x_3 بیان می شود

```
> #######################example45-46-47######################################
> x1=c(91.5,93.8,93.4,92.5,94.8,93.2,93.7,91.5,92.7,92.7)
> x2=c(56.1,56.3,56.4,56.7,56.5,56.8,57,56.8,55.7,57.7)
> x3=c(73.6,74.1,74.5,73.7,73.6,73.9,74.4,73.9,73.9,73.8)
> y=c(0.888,0.461,0.453,0.56,0.414,0.438,0.342,0.743,0.751,0.477)
> m=lm(y~x1+x2+x3)
> anova(m)
Analysis of Variance Table

Response: y
          Df   Sum Sq  Mean Sq F value    Pr(>F)    
x1         1 0.195396 0.195396 89.8303 7.857e-05 ***
x2         1 0.064359 0.064359 29.5882  0.001603 ** 
x3         1 0.016078 0.016078  7.3915  0.034701 *  
Residuals  6 0.013051 0.002175                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> summary(m)

Call:
lm(formula = y ~ x1 + x2 + x3)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.073440 -0.005879  0.013864  0.023203  0.036162 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 31.68793    3.99911   7.924 0.000215 ***
x1          -0.12893    0.01576  -8.180 0.000180 ***
x2          -0.15260    0.02863  -5.330 0.001779 ** 
x3          -0.14214    0.05228  -2.719 0.034701 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.04664 on 6 degrees of freedom
Multiple R-squared:  0.9548,	Adjusted R-squared:  0.9322 
F-statistic: 42.27 on 3 and 6 DF,  p-value: 0.0001983
```

```
> qf(0.95,3,6)
[1] 4.757063
> betahat<-coef(m)
> r=resid(m)
> r
           1            2            3            4            5            6 
 0.019854709 -0.009015530  0.003530461 -0.073440397  0.032362287 -0.061500906 
           7            8            9           10 
 0.008556061  0.024319233  0.019172262  0.036161820 
> sr=rstandard(m)
> sr
         1          2          3          4          5          6          7          8 
 0.5840325 -0.2188591  0.1056406 -1.7388426  1.4081560 -1.4093373  0.2375454  0.6506133 
         9         10 
 0.5351348  1.2140885 
> yhat=fitted(m)
> yhat
        1         2         3         4         5         6         7         8         9 
0.8681453 0.4700155 0.4494695 0.6334404 0.3816377 0.4995009 0.3334439 0.7186808 0.7318277 
       10 
0.4408382 
> 
> confint(m)
                 2.5 %      97.5 %
(Intercept) 21.9024595 41.47340919
x1          -0.1674970 -0.09036166
x2          -0.2226623 -0.08254162
x3          -0.2700761 -0.01421152
> shapiro.test(m$res)

	Shapiro-Wilk normality test

data:  m$res
W = 0.81817, p-value = 0.02408
```


Histogram of sr
Frequency
0.0
0.5
1.0
1.5
2.0
2.5
3.0
-2.0
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
sr


0.04
0.02
0.00
-0.02
-0.04
-0.06

Subject:
Date

مثال ۴۸ برای داده‌های زیر یک رگرسیون چندگانه کامل اجرا کنید (بدون عرض از مبدأ)

y	x_1	x_2	x_3
8	2	1	4
10	-1	2	1
9	1	-3	4
6	2	1	2
12	1	4	6

$n = ۵$

$p = ۳$

۵-۳-۱

(بدون عرض از مبدأ یعنی ستون اول را در X در نظر نگیرید.)

$$X = \begin{pmatrix} 2 & 1 & 4 \\ -1 & 2 & 1 \\ 1 & -3 & 4 \\ 2 & 1 & 2 \\ 1 & 4 & 6 \end{pmatrix} \qquad X' = \begin{pmatrix} 2 & -1 & 1 & 2 & 1 \\ 1 & 2 & -3 & 1 & 4 \\ 4 & 1 & 4 & 2 & 6 \end{pmatrix}$$

$$X'X = \begin{pmatrix} 2 & -1 & 1 & 2 & 1 \\ 1 & 2 & -3 & 1 & 4 \\ 4 & 1 & 4 & 2 & 6 \end{pmatrix} \begin{pmatrix} 2 & 1 & 4 \\ -1 & 2 & 1 \\ 1 & -3 & 4 \\ 2 & 1 & 2 \\ 1 & 4 & 6 \end{pmatrix}$$

$$= \begin{pmatrix} 11 & 3 & 21 \\ 3 & 31 & 20 \\ 21 & 20 & 73 \end{pmatrix}$$

$$(X'X)^{-1} = \begin{pmatrix} 0.2145 & 0.0231 & -0.068 \\ 0.0231 & 0.0417 & -0.0181 \\ -0.068 & -0.0181 & +0.0382 \end{pmatrix}$$

$$X' \begin{pmatrix} 8 \\ 10 \\ 9 \\ 6 \\ 12 \end{pmatrix} = \begin{pmatrix} 39 \\ 55 \\ 162 \end{pmatrix} \qquad \hat{\underset{\sim}{\beta}} = (X'X)^{-1} X'y = \begin{pmatrix} -1.39 \\ 0.27 \\ 2.54 \end{pmatrix}$$

$SSR = \hat{\underset{\sim}{\beta}}' X' \underset{\sim}{y} = 372.9$ بدون عرض از مبدأ: $df = p$

$SST = \underset{\sim}{y}'\underset{\sim}{y} = \sum y_i^2 = 425$ $df = n$

$SSE = 425 - 372.9 = 52.1$ $df = n - p$

منبع	SS	df	MS	F_0
رگرسیون	372.9	3	372.9/3	4.768
خطا	52.1	2	52.1/2	
کل	425	5		

$F_{3,2,0.95} = 19.16$

$4.768 < 19.16 \Rightarrow AH_0$

رگرسیون معنی‌دار نیست.

```
> #######################example48######################################
> x1=c(2,-1,1,2,1)
> x2=c(1,2,-3,1,4)
> x3=c(4,1,4,2,6)
> y=c(8,10,9,6,12)
>
> m=lm(y~-1+x1+x2+x3)
> anova(m)
Analysis of Variance Table

Response: y
          Df  Sum Sq Mean Sq F value Pr(>F)
x1         1 138.273 138.273  5.3047 0.1478
x2         1  65.209  65.209  2.5017 0.2545
x3         1 169.386 169.386  6.4983 0.1256
Residuals  2  52.132  26.066
> summary(m)

Call:
lm(formula = y ~ -1 + x1 + x2 + x3)

Residuals:
      1       2       3       4       5
 0.3253  5.5371  1.0063  3.4145 -2.9488

Coefficients:
   Estimate Std. Error t value Pr(>|t|)
x1  -1.3851     2.3646  -0.586    0.617
x2   0.2666     1.0423   0.256    0.822
x3   2.5446     0.9982   2.549    0.126

Residual standard error: 5.106 on 2 degrees of freedom
Multiple R-squared:  0.8773,	Adjusted R-squared:  0.6933
F-statistic: 4.768 on 3 and 2 DF,  p-value: 0.1782

> qf(0.95,3,2)
[1] 19.16429
>
```

Subject:
Date

تحلیل مؤلفه های اصلی

دیدیم در رگرسیون ممکن است p متغیر مورد بررسی قرار گیرد که p تعداد زیادی

هم باشد در این صورت می توان با استفاده از "تحلیل مؤلفه های اصلی"

با تعداد مؤلفه کمتر یعنی k مؤلفه ی اصلی کل سیستم را مورد مطالعه قرار داد

فرض کنید Σ ماتریس کوواریانس بردار تصادفی $\underline{X}' = (X_1 \dots X_p)$

باشد و Σ دارای زوج مقدار ویژه - بردار ویژه ی

$(\lambda_1 , \underline{e}_1) \dots\dots (\lambda_p , \underline{e}_p)$

باشد که $\underline{e}$ ها استاندارد شده ی بردار ای ویژه اند)

و $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ آن گاه مؤلفه ی اصلی i ام:

$$\underline{Y}_i = \underline{e}'_i \underline{X} = e_{1i} X_1 + \dots + e_{pi} X_p \qquad i = 1 \dots p$$

و ثابت می شود

$$\begin{cases} V(\underline{Y}_i) = \lambda_i \\ Cov(\underline{Y}_i , \underline{Y}_k) = 0 \end{cases}$$

سهم کل واریانس مربوط به مؤلفه ی اصلی i ام: $\dfrac{\lambda_i}{\lambda_1 + \dots + \lambda_p}$

مثلاً اگر ۸۰ تا ۹۰ درصد این سهم مربوط به مؤلفه اول و دوم اصلی باشد

می توان به جای p متغیر از این دو مؤلفه استفاده کرد.

Subject:
Date

مثال ٤٩ فرض کنید متغیرهای تصادفی X_1 و X_2 و X_3 دارای ماتریس

$$\Sigma = \begin{pmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

کوواریانس زیر باشند

مؤلفه‌های اصلی را بدست آورید و تفسیر کنید

$$|A - \lambda I| = 0$$

$$|\Sigma - \lambda I| = 0 \qquad \begin{vmatrix} 1-\lambda & -2 & 0 \\ -2 & 5-\lambda & 0 \\ 0 & 0 & 2-\lambda \end{vmatrix}$$

$$= (1-\lambda)\begin{vmatrix} 5-\lambda & 0 \\ 0 & 2-\lambda \end{vmatrix} + 2\begin{vmatrix} -2 & 0 \\ 0 & 2-\lambda \end{vmatrix} + 0$$

$$\stackrel{\text{یا}}{=} (2-\lambda)\begin{vmatrix} 1-\lambda & -2 \\ -2 & 5-\lambda \end{vmatrix} = 0$$

$$\underbrace{(2-\lambda)}\underbrace{\left((1-\lambda)(5-\lambda)-4\right)}_{\text{معادله درجه ۲}} = 0$$

$$\lambda = 2 \qquad \lambda = ۵/۸۳ \quad , \quad \lambda = ۰/۱۷$$

$$\lambda_1 = ۵/۸۳ \qquad \lambda_2 = ۲ \qquad \lambda_3 = ۰/۱۷$$

$$\tilde{e}'_1 = (0.383, -0.924, 0) \qquad \tilde{e}'_2 = (0, 0, 1)$$

$$\tilde{e}'_3 = (0.924, 0.383, 0)$$

Subject:
Date

پس مؤلفه‌های اصلی عبارتند از

$$y_1 = \underset{\sim}{e}_1' \underset{\sim}{X} = 0.383 X_1 - 0.924 X_2$$

$$y_2 = \underset{\sim}{e}_2' \underset{\sim}{X} = X_3$$

$$y_3 = \underset{\sim}{e}_3' \underset{\sim}{X} = 0.924 X_1 + 0.383 X_2$$

سهم هر مولفه:

$$\left.\begin{aligned} \frac{\lambda_1}{\lambda_1+\lambda_2+\lambda_3} &= \frac{5.83}{8} = 0.73 \\ \frac{\lambda_2}{\lambda_1+\lambda_2+\lambda_3} &= \frac{2}{8} = 0.25 \end{aligned}\right\} \rightarrow 0.98$$

یعنی مثلاً در یک رگرسیون چندگانه ۳ متغیر X_1 و X_2 و X_3 داریم بعد از اینکه مؤلفه‌های اصلی را حساب کردیم دیدیم ۹۸٪ واریانس مربوط به مؤلفه اول و دوم است ما در رگرسیون به جای X_1 و X_2 و X_3 از دو مؤلفه w_1 و w_2 استفاده می‌کنیم.

تمرین. فرض کنید $\Sigma_{X_1,X_2} = \begin{pmatrix} 70 & 6 \\ 6 & 1 \end{pmatrix}$ دو مؤلفه اصلی را بدست آورید و سهم واریانس هر مؤلفه را بدست آورید

$$|\Sigma - \lambda I| = \begin{vmatrix} 70-\lambda & 6 \\ 6 & 1-\lambda \end{vmatrix} = \lambda^2 - 71\lambda + 34 = 0$$

$$\Delta = b^2 - 4ac$$

$$\text{جواب} = \frac{-b \pm \sqrt{\Delta}}{2a}$$

$$\begin{cases} \lambda_1 = 70.518 \\ \lambda_2 = 0.482 \end{cases}$$

Subject:

Date

$$\Sigma_1 \underset{\sim}{x} = \lambda_1 \underset{\sim}{x} \Rightarrow \begin{pmatrix} 70 & 6 \\ 6 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 70.518 \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$\Rightarrow X_1 = 1 \qquad X_2 = 0.084$$

$$\underset{\sim}{e}'_1 = \left(\overbrace{\frac{1}{\sqrt{1^2 + 0.084^2}}}^{0.996} \qquad \overbrace{\frac{0.084}{\sqrt{1^2 + 0.084^2}}}^{0.084} \right)$$

$$\underset{\sim}{e}'_2 = (0.084 \qquad -0.996)$$

$$w_1 = (0.996 \quad 0.084) \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = 0.996 X_1 + 0.084 X_2$$

$$w_2 = 0.084 X_1 - 0.996 X_2$$

$$\frac{\lambda_1}{\lambda_1 + \lambda_2} = 0.993 \qquad \frac{\lambda_2}{\lambda_1 + \lambda_2} = 0.007$$

مؤلفه‌ی اصلی اول ۹۹ ٪ واریانس کل را بیان می‌کند پس بدون آنکه اطلاعاتی از دست بدهیم به جای X_1 و X_2 در مطالعه می‌توان از w_1 استفاده کرد.

```
> x=matrix(scan(),byrow=T,ncol=3)
1: 1 -2 0
4: -2 5 0
7: 0 0 2
10:
Read 9 items
> e=eigen(x)
> e
eigen() decomposition
$values
[1] 5.8284271 2.0000000 0.1715729

$vectors
           [,1] [,2]      [,3]
[1,] -0.3826834    0 0.9238795
[2,]  0.9238795    0 0.3826834
[3,]  0.0000000    1 0.0000000

> e$values[1]/(e$values[1]+e$values[2]+e$values[3])
[1] 0.7285534
> e$values[2]/(e$values[1]+e$values[2]+e$values[3])
[1] 0.25
>
```

Subject:
Date

طرح بلوکی (تحلیل واریانس دوطرفه بدون اثر متقابل)

یک نوع طرح تک عاملی است که سطوح عامل در چند بلوک اندازه گیری می شود

عوامل شلاته کننده

سطوح عامل ←

بلوک ها (روز)	A	B	C	D
1	۸۴	۸۰	۸۳	۷۹
2	۷۹	۷۷	۸۰	۷۹
3	۸۳	۷۸	۸۰	۷۸

$j = 1, 2, 3, 4$

$i = 1, 2, 3$

برای مثال داده های بالا هدف مقایسه درصد کارائی عوامل شلاته کننده مختلف در استخراج یک یون فلزی از محلول آبی است. در هر روز، محلول تازه ای از یون فلزی (با غلظت مشخص) تهیه شده و استخراج با هر یک از عوامل شلاته کننده که به طور تصادفی برداشته می شوند انجام گرفته است. در این آزمایش استفاده از عوامل شلاته کننده یک عامل کنترل شده در نظر گرفته می شود چون توسط آزمایشگر انتخاب می شوند ولی روزی که آزمایش در آن انجام می گیرد تغییرات کنترل نشده که از تغییر در دمای آزمایشگاه، فشار و غیره است. پس روز به عنوان یک عامل تصادفی که به آن بلوک می گویند به حساب می آید.

فرض کنید $j = 1, \dots, a$ تعداد سطوح عامل (تیمارها)

$i = 1, \dots, b$ تعداد بلوک

مدل $$Y_{ij} = \mu + \tau_j + B_i + \varepsilon_{ij}$$

میانگین کل ، اثر عامل (تیمار j ام) ، اثر بلوک i ام ، خطای تصادفی

$\varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$ $\qquad \sum_{i=1}^{a} \tau_i = 0 \qquad \sum_{j=1}^{b} B_j = 0$

$$\begin{cases} H_0: \tau_1 = \dots = \tau_a = 0 \\ H_1: \text{Not } H_0 \end{cases} \equiv \begin{cases} H_0: \mu_1 = \dots = \mu_a \\ H_1: \text{Not } H_0 \end{cases}$$ فرضیه

i \ j	1	2	...	a	
1	Y_{11}	Y_{12}	...	Y_{1a}	$T_{1\cdot}$
2	⋮			⋮	$T_{2\cdot}$
⋮	⋮			⋮	⋮
b	Y_{b1}	Y_{b2}	...	Y_{ba}	$T_{b\cdot}$
$T_{\cdot j}$	$T_{\cdot 1}$	$T_{\cdot 2}$	...	$T_{\cdot a}$	جمع $T_{\cdot\cdot}$

$\left.\begin{matrix} T_{1\cdot} \\ \vdots \\ T_{b\cdot} \end{matrix}\right\} T_{i\cdot}$

کل $SST = \sum \sum Y_{ij}^2 - \frac{T_{\cdot\cdot}^2}{N}$ $\qquad df = \underbrace{N}_{ab} - 1 \qquad N = ab$

عامل $SSA = \sum_j \frac{T_{\cdot j}^2}{b} - \frac{T_{\cdot\cdot}^2}{N}$ $\qquad df = a - 1$

بلوک $SSB = \sum_i \frac{T_{i\cdot}^2}{a} - \frac{T_{\cdot\cdot}^2}{N}$ $\qquad df = b - 1$

$SSE = SST - SSA - SSB$ $\qquad df = (a-1)(b-1)$

$N = ab \qquad ab - \cancel{1} - a + \cancel{1} - b + 1 =$
$= ab - a - b + 1 = (a-1)(b-1)$

منبع	SS	df	MS	F
عامل	SSA	$a-1$	$SSA/a-1 = MSA$	$F_0 = \frac{MSA}{MSE}$
بلوک	SSB	$b-1$	$SSB/b-1 = MSB$	
خطا	SSE	$(a-1)(b-1)$	$SSE/(a-1)(b-1) = MSE$	
کل	SST	$\underbrace{ab}_{N}-1$		

$$\text{if } F_0 > F_{a-1,\,(a-1)(b-1),\,1-\alpha} \Rightarrow RH_0$$

$$\begin{cases} H_0: \mu_1 = \dots = \mu_a \\ H_1: \text{Not } H_0 \end{cases}$$

حل مثال .

برای سادگی اعداد از ۸۰ کم کنید .

	A	B	C	D
1	۸۴	۸۰	۸۳	۷۹
2	۷۹	۷۷	۸۰	۷۹
3	۸۳	۷۸	۸۰	۷۸

	A	B	C	D	
1	4	0	3	-1	6
2	-1	-3	0	-1	-5
3	3	-2	0	-2	-1
	6	-5	3	-4	$0 = T_{..}$

$$SST = (4^2 + \dots + (-2)^2) - 0 = ۵۴$$

Subject:
Date

$$SSA = \frac{6^2 + (-5)^2 + (3^2) + (-4)^2}{3} - 0 = 28.667$$

$$SSB = \frac{6^2 + (-5)^2 + (-1)^2}{4} - 0 = 15.5$$

$$SSE = 54 - 28.667 - 15.5 = 9.833$$

جدول آنالیز واریانس

منبع	SS	df	MS	F_0
عامل	۲۸٫۶۶۷	3	۹٫۵۵۶	$\frac{9.556}{1.6389} = 5.83$
بلوک	۱۵٫۵	2	۷٫۷۵	
خطا	۹٫۸۳۳	6	۱٫۶۳۸۹	
کل	۵۴	11		

$$F_{3,6,0.95} = 4.75$$

$$5.83 > 4.75 \Rightarrow R H_0$$

یعنی تیمارها روی پاسخ تاثیر گذارند.

```
> ###################### Randomized block design -50##################
>
> y<-c(84,79,83,80,77,78,83,80,80,79,79,78)
> A<-factor(rep(1:4,each=3))
> A
 [1] 1 1 1 2 2 2 3 3 3 4 4 4
Levels: 1 2 3 4
> Block<-factor(rep(1:3,4))
> Block
 [1] 1 2 3 1 2 3 1 2 3 1 2 3
Levels: 1 2 3
> y<-y-80
> y
 [1]  4 -1  3  0 -3 -2  3  0  0 -1 -1 -2
> aggregate(y~A,FUN=sum)
  A  y
1 1  6
2 2 -5
3 3  3
4 4 -4
> aggregate(y~Block,FUN=sum)
  Block  y
1     1  6
2     2 -5
3     3 -1
> sum(y)
[1] 0
>
> M<-aov(y~A+Block)
> anova(M)
Analysis of Variance Table

Response: y
          Df  Sum Sq Mean Sq F value  Pr(>F)
A          3 28.6667  9.5556  5.8305 0.03276 *
Block      2 15.5000  7.7500  4.7288 0.05848 .
Residuals  6  9.8333  1.6389
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> qf(0.95,3,6)
[1] 4.757063
>
```