

Multi Linear

Hossein Vatani

April 1, 2016

بنام خدا

پسرفت چندگانه

الف-پسرفت چندگانه

یک پسرفت با بیش از یک متغیر تشریح را ، پسرفت چندگانه گویند. سوای ساخت یک خط راست برای میانگین متغیر پاسخ توسط نمونه پسرفت، اینبار یک تابع از چند متغیر تشریح برای متغیر پاسخ خواهیم داشت. دستور lm() برای این کاربرد نیز قابل استفاده است. دستور کلی آن بدین صورت است:

```
lm(response ~ explanatory_1 + explanatory_2 + ... + explanatory_p)
```

متأسفانه فایل های مربوط به این سری از مقاله ها یافت نشد، لذا داده هایی دیگر جهت مثال ها مورد استفاده قرار گرفته است. (پس بدنبال یکی بودن اعداد و ارقام مثال زیر با متن اصلی نباشید!)

نکته: ازین پس هرگاه در مثالی از داده هایی استفاده شد که منبع آن در دسترس نباشد ، ابتدای آن ذکر می کنم که: فایل موجود نبود.

برای این مثال از داده های مربوط به خودرو که با نام mtcars در نرم افزار وجود دارد استفاده می کنیم.

```
Car=mtcars[,c("mpg", "disp", "hp", "wt", "am")]
head(Car)
```

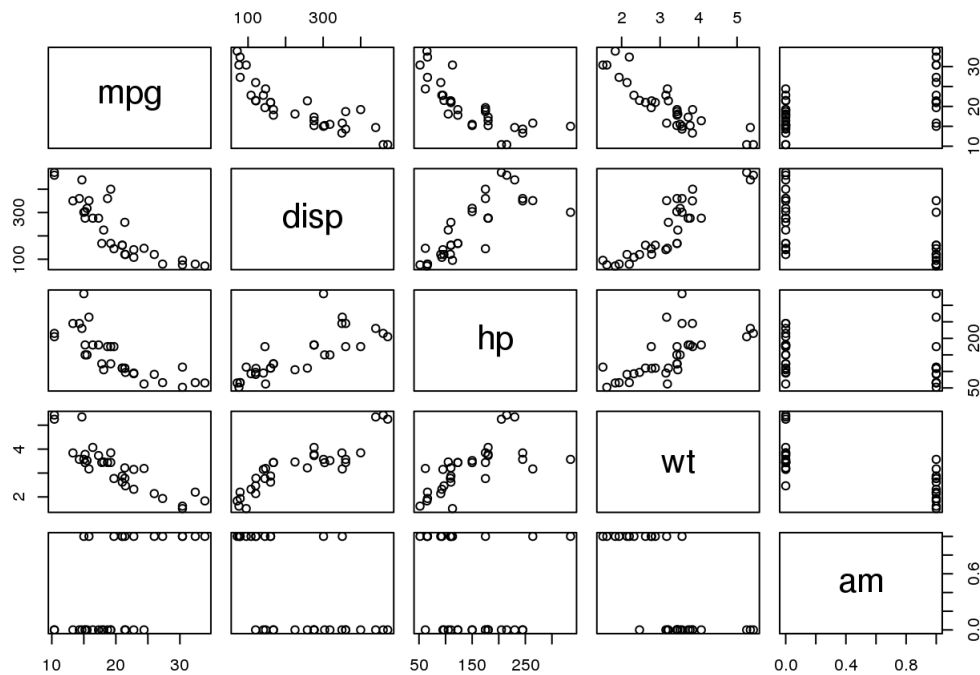
```
##           mpg  disp  hp   wt  am
## Mazda RX4    21.0  160  110 2.620  1
## Mazda RX4 Wag 21.0  160  110 2.875  1
## Datsun 710    22.8  108   93 2.320  1
## Hornet 4 Drive 21.4  258  110 3.215  0
## Hornet Sportabout 18.7  360  175 3.440  0
## Valiant      18.1  225  105 3.460  0
```

```
summary(Car)
```

```
##           mpg           disp           hp           wt
## Min.      :10.40   Min.    : 71.1   Min.    : 52.0   Min.    :1.513
## 1st Qu.:15.43   1st Qu.:120.8   1st Qu.: 96.5   1st Qu.:2.581
## Median :19.20   Median :196.3   Median :123.0   Median :3.325
## Mean     :20.09   Mean     :230.7   Mean     :146.7   Mean     :3.217
## 3rd Qu.:22.80   3rd Qu.:326.0   3rd Qu.:180.0   3rd Qu.:3.610
## Max.     :33.90   Max.     :472.0   Max.     :335.0   Max.     :5.424
##           am
## Min.      :0.0000
## 1st Qu.:0.0000
## Median :0.0000
## Mean     :0.4062
## 3rd Qu.:1.0000
## Max.     :1.0000
```

قبل از برآزش و نمونه سازی تصویرسازی ای از اطلاعات را لازم است انجام دهیم تا دریافتی کلی از ارتباط اطلاعات را بدست آوریم.

```
plot(Car)
```



از ظاهر نمودار مشخص است که mpg با disp و hp و wt رابطه ای خطی دارد. ما جهت آزمایش نظر خودمان عامل am را که ظاهراً بطور کامل بهم نوابسته اند! را نیز در نمونه وارد می کنیم.

توضیح متغیر ها را در رهنمای نرم افزار می توانید بیابید.

```
Fit1=lm(mpg ~ disp+ hp+ wt+ am, data = Car)
Fit1
```

```
##
## Call:
## lm(formula = mpg ~ disp + hp + wt + am, data = Car)
##
## Coefficients:
## (Intercept)      disp          hp          wt          am
##  34.209443    0.002489   -0.039323   -3.046747    2.159271
```

$y = 34.20943 + 0.002489 \text{disp} - 0.039323 \text{hp} - 3.046747 \text{wt} + 2.159271 \text{am}$ رابطه به خروجی، با توجه به $P\text{-Value}$ حاصله مقدار $P\text{-Value}$ ها را برای متغیرهای تشریح در نمونه را بررسی می کنیم.

```
summary(Fit1)
```

```
##
## Call:
## lm(formula = mpg ~ disp + hp + wt + am, data = Car)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.4590 -1.6900 -0.3708  1.1301  5.5011
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 34.209443   2.822826  12.119 1.98e-12 ***
## disp         0.002489   0.010377   0.240  0.81222
## hp          -0.039323   0.012434  -3.163  0.00384 **
## wt          -3.046747   1.157119  -2.633  0.01383 *
## am           2.159271   1.435176   1.505  0.14405
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.581 on 27 degrees of freedom
## Multiple R-squared:  0.8402, Adjusted R-squared:  0.8166
## F-statistic: 35.5 on 4 and 27 DF,  p-value: 2.181e-10
```

با توجه به مقادیر P-Value متوجه می شویم که disp و am عاملهای مناسبی نیستند یا عبارتی دیگر میانگین این دو عامل با میانگین mpg تفاوت معنی داری ندارد. لذا نمونه را دوباره بدون این دو بازسازی می کنیم.

```
Fit2=lm(mpg ~ hp+ wt, data = Car)
Fit2
```

```
##
## Call:
## lm(formula = mpg ~ hp + wt, data = Car)
##
## Coefficients:
## (Intercept)          hp          wt
##    37.22727    -0.03177    -3.87783
```

```
summary(Fit2)
```

```
##
## Call:
## lm(formula = mpg ~ hp + wt, data = Car)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.941 -1.600 -0.182  1.050  5.854
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 37.22727   1.59879  23.285 < 2e-16 ***
## hp          -0.03177   0.00903  -3.519  0.00145 ***
## wt          -3.87783   0.63273  -6.129 1.12e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.593 on 29 degrees of freedom
## Multiple R-squared:  0.8268, Adjusted R-squared:  0.8148
## F-statistic: 69.21 on 2 and 29 DF,  p-value: 9.109e-12
```

ب-آزمایش زیر مجموعه ای عامل های تشریحی بوسیله آزمون

فیشتر-قسمتی

برخی مواقع ، می خواهیم بدانیم که کدامیک از عامل های تشریح استفاده شده در نمونه پسرقت ، تاثیر معنی داری بر روی نمونه ندارند. برای اینگونه از آزمایش ها می توان از آزمون فیشتر-قسمتی استفاده کرد؛ بدین شرح که دو نمونه خطی که یکی دارای تمام عامل ها و دیگر فقط با عامل های مورد نظر می باشد ایجاد می کنیم و در نهایت از روش تحلیل پراکنش جهت تحلیل و نتیجه گیری استفاده می کنیم.

مثال- در مثال قبل می خواهیم بررسی کنیم که آیا واقعا تاثیر عاملهای disp و am بدون اهمیت می باشد یا خیر؟ از نتایج بدست آمده قبلی استفاده می کنیم.

```
anova(Fit2,Fit1)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ hp + wt
## Model 2: mpg ~ disp + hp + wt + am
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1      29 195.05
## 2      27 179.91  2      15.14 1.1361 0.3359
```

با توجه به مقدار P-Value خروجی که برابر با ۰.۳۳ و بیشتر از فاصله اطمینان مورد نظر که ۰.۰۵ است، می باشد، بوضوح می بینیم که دو عامل disp و am اختلاف معنی داری در نمونه ایجاد نمی کنند ولذا قابل صرفنظر میباشند(کاری که ما در همان مثال قبل انجام دادیم).

ج-بازه های اطمینان و پیش بینی

مانند آنچه در متن نمونه ساده خطی گذشت، در این نوع از نمونه سازی نیز در نهایت بدنبال یافتن بازه اطمینان جهت میانگین متغیر پاسخ و همچنین پیش بینی وضعیت آینده هستیم که هر دوی آنها با تابع lm() مانند قبل قابل انجام است.

مثال- می خواهیم مقدار mpg را برای خودرویی با hp برابر ۴۰۰ و wt برابر ۶ محاسبه کنیم.

```
predict(Fit2,data.frame(hp=400,wt=6),interval="confidence") # جهت پیش بینی میانگین
```

```
##           fit           lwr           upr
## 1 1.251107 -2.428152 4.930366
```

```
predict(Fit2,data.frame(hp=400,wt=6),interval="predict") # جهت پیش بینی مقدار
```

```
##           fit           lwr           upr
## 1 1.251107 -5.204175 7.706389
```

```
predict(Fit2,data.frame(hp=110,wt=4),interval="confidence") # جهت پیش بینی میانگین
```

```
##           fit           lwr           upr
## 1 18.22092 16.41302 20.02882
```

```
predict(Fit2,data.frame(hp=110,wt=4),interval="predict") # جهت پیش بینی مقدار
```

```
##           fit           lwr           upr
## 1 18.22092 12.61715 23.82469
```