

تحلیل رگرسیون با R و مقدمات لازم برای فراهم کردن برازش مدل

تذکر: این یک نمونه است، سایر معیارهای برازش مدل را شما بررسی کنید، مانند ضریب تعیین، ...

۱- ابتدا بسته‌هایی را که در تحلیل رگرسیون لازم است نصب و باگذاری می‌کنیم:

```
install.packages("ggplot2")
install.packages("dplyr")
install.packages("broom") # this support the leaps package
install.packages("ggpubr")

library(ggplot2)
library(dplyr)
library(broom)
library(ggpubr)
```

- ۲- فراخوانی داده‌ها
- ۳- بررسی شرایط مجموعه داده‌ها (نرمال بودن متغیر پاسخ و خطی بودن رابطه بین متغیرها)
- ۴- اجرای مدل رگرسیون خطی (برازش مدل رگرسیون خطی) بر داده‌ها
- ۵- بررسی هم‌وارانسی خطاها (باقیمانده‌ها)
- ۶- بررسی نموداری نتایج (باقیمانده‌ها)
- ۷- بررسی boxcox برای مدل برازش شده
- ۸- گزارش نتایج برازش مدل رگرسیون (معنی‌داری برآورد پارامترها)
- ۹- انتخاب مدل نهایی رگرسیون برازش شده.

خواندن داده‌ها از پوشه مورد نظر و انجام مراحل فوق روی داده‌ها:

```
Data=read.table("income_data.csv",header=TRUE)
```

فایل داده را در پوشه تعریف شده در مسیر R قرار دهید و سپس داده‌ها را فراخوانی کنید (این فایل داده‌ها به فرمت اکسل است):

```
Data=read.csv("income_data.csv",header=TRUE)
```

با دستور زیر می‌توانید ابتدای فایل داده‌ها (متغیرهای موجود) را مشاهده بکنید:

```
head(Data)
```

رگرسیون خطی ساده :Simple regression

در صورت لزوم، خلاصه آمار توصیفی متغیرهای موجود فایل داده را مشاهده کنید:

```
summary(Data)
```

با دستور attach فایل داده‌ها را در مسیر جاری (روی میزکار) قرار داده و سپس متغیرهای فراخوانی شده از این فایل در نظر می‌گیرد:
`Data=attach(Data)`

هیستوگرام متغیر پاسخ و یا باقیمانده‌ها را رسم کنید (بررسی چشمی نرمال بودن):

```
hist(happiness)
```

مدل رگرسیون "happiness" را بر حسب "income" برازش دهید و داخل متغیر mod بگذارید:
`mod=lm(happiness~income)`

خلاصه آنچه برازش شده است را مشاهده کنید:

```
summary(mod)
```

نمودار پراکندگی متغیرها را ببینید و خط برازش‌شده مدل رگرسیون را روی نمودار پراکندگی داده‌ها رسم کنید:

```
plot(income,happiness)
```

با دستور زیر می‌توان خط رگرسیون برازش‌شده را روی نمودار پراکنش داده‌ها رسم کرد.

```
abline(mod)
```

نمودار پراکندگی باقیمانده‌ها را رسم و مشاهده کنید:

```
plot(mod$residuals)
```

نمودار باکس-پلات را برای باقیمانده‌ها رسم و نقاط پرت آنرا مشاهده کنید:

```
boxplot(mod$residuals)
```

آزمون نرمالیتی را برای باقیمانده‌ها انجام دهید:

```
shapiro.test(mod$residuals)
```

برای بررسی اینکه آیا متغیر پاسخ نیاز به تبدیل دارد یا خیر، از دستور زیر استفاده کنید:

```
boxcox(mod) #
```

۱۰- در صورتی که متغیر پاسخ نیاز به تبدیل ندارد نوشتن مدل نهایی رگرسیون برآزش شده به داده ها

```
happiness=0.2043+0.7138income
```

برای برآزش مدل رگرسیون بدون β_0 از دستور زیر می توان استفاده کرد
`Mod1=lm(happiness~income-1)`

برای اینکه بدانیم شرایط اولیه رگرسیون رعایت شده است، از دستورات زیر استفاده می کنیم. چهار نمودار حاصل از مدل را در کنار هم رسم می کند که قابل تفسیر است.

```
par(mfrow=c(2,2))  
Plot(mod)
```

رگرسیون استاندارد:

```
mod=lm(scale(happiness)~scale(income))  
summary(mod)
```

رگرسیون چندگانه Multiple regression

ابتدا فایل داده ها بصورت زیر خوانده می شود:

```
heart=read.csv("heart_data.csv",header=TRUE)
```

خلاصه (آمار توصیفی) متغیرها را مشاهده می کنید (در صورت لزوم):
`summary(heart)`

```
summary(heart[,2:4])
```

باکس-پلات متغیرهای مورد نظر را رسم می کنیم:

```
boxplot(heart[,2:4])
```

فایل داده را روی میز کار قرار می دهیم:

```
heart =attach(heart)
```

ابتدا ضرایب همبستگی بین متغیرهای آزاد را بررسی می کنیم:
`cor(biking, smoking)`

نرمال بودن متغیر پاسخ را بررسی می کنیم:

```
hist(heart.disease)
```

مدل رگرسیون چندگانه را بر داده ها برازش می دهیم:

```
mod=lm(heart.disease~biking+smoking)
```

خلاصه آنچه برازش شده را مشاهده می کنیم:

```
summary(mod)
```

پراکندگی باقیمانده ها را مشاهده می کنیم:

```
plot(mod$residuals)
```

آزمون نرمالیتی را برای باقیمانده ها انجام می دهیم:

```
shapiro.test(mod$residuals)
```

با دستور زیر نیاز به تبدیل متغیر پاسخ را بررسی می کنیم:

```
boxcox(mod,lambda = seq(0.7, 1.2, 1/10))#
```

نوشتن مدل نهایی رگرسیون برازش شده (چندمتغیره) به داده ها را می نویسیم .

```
Calories=0.7558+3.9763sugars+4.0426carbo+3.9271protein+8.4369fat
```

در فایل داده Bodyfat متغیرهای زیر موجود و مورد بررسی قرار می گیرند.

Density

Pct.BF درصد چربی

Age

Weight

Height

Neck

Chest

Abdomen شکم

Waist کمر

Hip لڱن

Thigh ران

Knee زانو

Ankle مچ پا

Bicep دوپا

Forearm پيش بازو

Wrist مچ دست