7주차 3차시 : 다중회귀분석(예측)

1.예측

1.예측

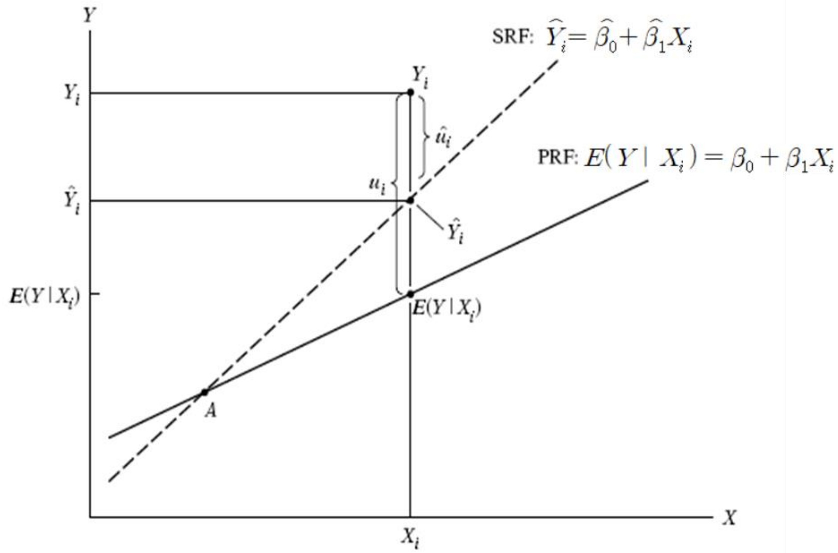
-모형설정→모형추정→가설검정→예측이라는 회귀분석의 절차에 따라 주어진 독립변수의 값에 대한 종속변수의 값을 구하는 것을 예측(forecasting, prediction)이라고 함

-모집단회귀선 위의 한 점을 예측하는 평균예측(mean prediction)과 독립변수의 값에 대응하는 개별 Y의 값을 예측하는 개별예측(individual prediction)이 있음

-예측에는 주어진 X에 대해 하나의 Y 값을 구하는 점예측(point forecasting)과 점예측에 대한 신뢰구간을 구하는 구간예측(interval forecasting)이 있음

-구간예측을 위해서는 실제치와 예측치의 차이인 예측오차의 분산을 알아야 함

-일반적으로 개별예측에서 점예측치 및 예측구간대를 구함



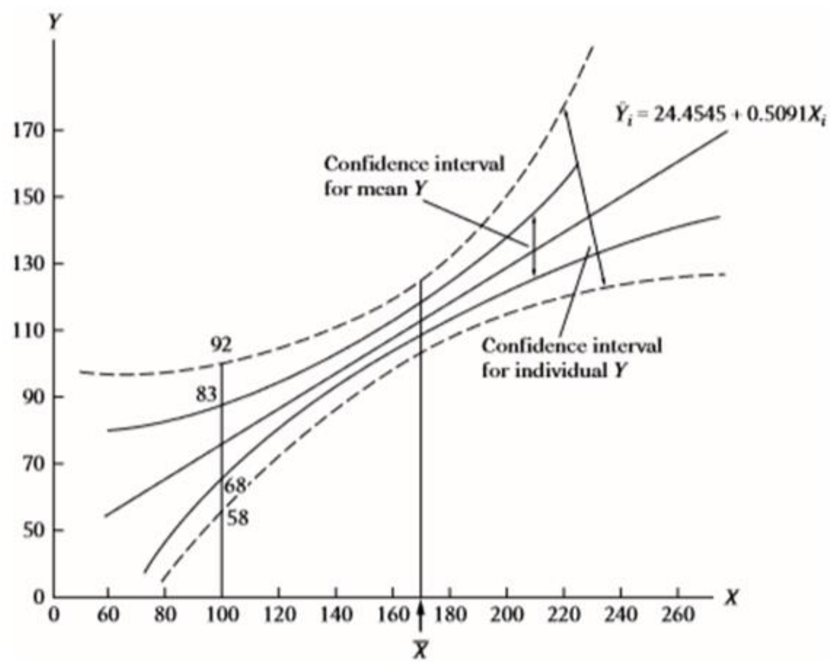
(점예측) $\hat{Y}_0 = x'_0\hat{B}$

(평균예측의 예측오차 분산) $\sigma_\epsilon^2 = \hat{\sigma}_u^2 x'_0(X'X)^{-1}x_0 \longleftrightarrow \sigma_u^2\left(\frac{1}{n} + \frac{(X_0-\bar{X})^2}{\sum_{i=1}^n x_i^2}\right)$

(개별예측의 예측오차 분산) $\sigma_\epsilon^2 = \hat{\sigma}_u^2(1 + x'_0(X'X)^{-1}x_0) \longleftrightarrow \sigma_u^2\left(1 + \frac{1}{n} + \frac{(X_0-\bar{X})^2}{\sum_{i=1}^n x_i^2}\right)$

(평균예측의 예측구간) $\hat{Y}_0 \pm t_{\frac{\alpha}{2},n-k}\sqrt{\hat{\sigma}_u^2 x'_0(X'X)^{-1}x_0}$

(개별예측의 예측구간) $\hat{Y}_0 \pm t_{\frac{\alpha}{2},n-k}\sqrt{\hat{\sigma}_u^2(1 + x'_0(X'X)^{-1}x_0)}$



(예제-계속)

$X_2 = 3, X_3 = 2$일 때 점예측치는?

$$\hat{Y}_0 = x'_0\hat{B} = \begin{bmatrix}1 & 3 & 2\end{bmatrix}\begin{bmatrix}-4.25\\1.5\\2\end{bmatrix} = 4.25$$

```
> x0<-matrix(c(1,3,2))
> yhat<-t(x0)%*%beta
> x0
     [,1]
[1,]    1
[2,]    3
[3,]    2
> yhat
     [,1]
[1,] 4.25
>
```

```
Value of x0 for prediction is :
[[1]
 [3]
 [2]]
Point Prediction of y given x0 is :
[[4.25]]
```

(평균예측의 예측오차 분산) $\boldsymbol{\sigma_\epsilon^2 = \hat{\sigma}_u^2 x'_0(X'X)^{-1}x_0} = (0.25)\begin{bmatrix}1 & 3 & 2\end{bmatrix}\begin{bmatrix}14.75 & -3.5 & -5\\-3.5 & 1 & 1\\-5 & 1 & 2\end{bmatrix}\begin{bmatrix}1\\2\\3\end{bmatrix} = 0.6875$

(개별예측의 예측오차 분산) $\boldsymbol{\sigma_\epsilon^2 = \hat{\sigma}_u^2(1 + x'_0(X'X)^{-1}x_0)}$

$$= (0.25)(1 + \begin{bmatrix}1 & 3 & 2\end{bmatrix}\begin{bmatrix}14.75 & -3.5 & -5\\-3.5 & 1 & 1\\-5 & 1 & 2\end{bmatrix}\begin{bmatrix}1\\2\\3\end{bmatrix}) = 0.9375$$

(평균예측의 95% 예측구간) $\boldsymbol{\hat{Y}_0 \pm t_{\frac{\alpha}{2},n-k}\sqrt{\hat{\sigma}_u^2 x'_0(X'X)^{-1}x_0}} = 4.25 \pm (12.706)(0.829) = [-6.285, 14.785]$

(개별예측의 95% 예측구간) $\boldsymbol{\hat{Y}_0 \pm t_{\frac{\alpha}{2},n-k}\sqrt{\hat{\sigma}_u^2(1 + x'_0(X'X)^{-1}x_0)}} = 4.25 \pm (12.706)(0.968) = [-8.049, 16.549]$

```
> sigesq<-0.25*(1+t(x0)%*%xpxinv%*%x0)
> sigesq
       [,1]
[1,] 0.9375
> sige<-sqrt(sigesq)
> sige
          [,1]
[1,] 0.9682458
> yhat+c(-1,1)*qt(.975, 1)*sige
[1] -8.05273 16.55273
```

```
Variance of Individual Predicition Error is : [[0.9375]]
Stand Error of Individual Predicition is : [[0.96824584]]
t-critical value of t Distribution w. df= 1 is : 12.706204736432095
Lower Value of 95% Individual Predicition Interval is : [[-8.05272983]]
Upper Value of 95% Individual Predicition Interval is : [[16.55272983]]
```

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

Residuals:
    1     2     3     4 
-0.25  0.25 -0.25  0.25 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -4.2500     1.9203  -2.213    0.270
x2            1.5000     0.5000   3.000    0.205
x3            2.0000     0.7071   2.828    0.216

Residual standard error: 0.5 on 1 degrees of freedom
Multiple R-squared:  0.9091,	Adjusted R-squared:  0.7273 
F-statistic:     5 on 2 and 1 DF,  p-value: 0.3015


> confint(ols)
                 2.5 %    97.5 %
(Intercept) -28.649553 20.149553
x2           -4.853102  7.853102
x3           -6.984644 10.984644

> resid(ols)
    1     2     3     4 
-0.25  0.25 -0.25  0.25 

> predict(lm(y~x2+x3))
   1    2    3    4 
1.25 0.75 2.25 2.75 

> new<-data.frame(x2=3, x3=2)

> predict(lm(y~x2+x3), new, se.fit = TRUE)
$fit
   1 
4.25 
```

(점예측) →

```
$se.fit
[1] 0.8291562
```

(평균예측의 예측오차 표준오차) →

```
$df
[1] 1

$residual.scale
[1] 0.5
```

(교란항 표준오차) →

```
> pred.w.clim<-predict(lm(y~x2+x3), new, interval = "confidence")

> pred.w.clim
   fit       lwr      upr
1 4.25 -6.285428 14.78543
```

(평균예측의 95% 예측구간) →

```
> pred.w.plim<-predict(lm(y~x2+x3), new, interval = "prediction")

> pred.w.plim
   fit      lwr      upr
1 4.25 -8.05273 16.55273
```

(개별예측의 95% 예측구간) →

단순회귀 및 행렬표현 다중회귀의 비교

구분	단순회귀	다중회귀
$\hat{\beta}$	$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$	$\hat{B} = (X'X)^{-1}X'Y$
$\hat{\sigma}_u^2$	$\hat{\sigma}_u^2 = \frac{\sum_{i=1}^{n} \hat{e}_i^2}{n-2}$	$\frac{e'e}{n-k}$
$Var(\hat{\beta})$	$\hat{\sigma}_u^2 \frac{1}{\sum_{i=1}^{n} x_i^2}$	$\sigma_u^2 (X'X)^{-1}$
$\sum e_i^2$	$\sum_{i=1}^{n}(Y_i - \bar{Y})^2 - \hat{\beta}_1 \sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})$	$Y'Y - \hat{B}'X'Y$
R^2	$\frac{\hat{\beta}_1^2 \sum_{i=1}^{n}(X_i - \bar{X})^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$	$\frac{\hat{B}'X'X\hat{B} - n\bar{Y}^2}{Y'Y - n\bar{Y}^2}$
σ_ϵ^2 (평균예측)	$\sigma_u^2 (\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum_{i=1}^{n} x_i^2})$	$\hat{\sigma}_u^2 x'_0 (X'X)^{-1} x_0$
σ_ϵ^2 (개별예측)	$\sigma_u^2 (1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum_{i=1}^{n} x_i^2})$	$\hat{\sigma}_u^2 (1 + x'_0 (X'X)^{-1} x_0)$