STAT:2010/4200
Statistical Methods and Computing

Regression confidence intervals and prediction intervals

Lecture 23
April 17, 2017

Kate Cowles
374 SH, 335-0727
kate-cowles@uiowa.edu

Inferences concerning the regression line

- Estimating the mean of the Y's for a particular value of X, say X_0
  - Example: what is the average per capita health spending for a country with per capita gross domestic product 10 PPP

$$E[Y|X_0] = \hat{Y}_{X_0} = a + bX_0$$

- estimated standard error of $E[Y|X_0]$

$$\hat{\sigma}_{\hat{Y}_{X_0}} = \hat{\sigma}\sqrt{\frac{1}{n} + \frac{(X_0 - \bar{X})^2)}{\Sigma\,(X_i - \bar{X})^2}}$$

- A (1 - α)% confidence interval for $E[Y|X_0]$ is given by:

$$\hat{Y}_{X_0} \pm (t_{1-\alpha/2,df=n-2})(\hat{\sigma}_{\hat{Y}_{X_0}})$$


```
options linesize = 72 ;

data manatee ;
infile 'manatee.dat' ;
input obsno year powerbt killed ;
run ;

proc reg data = manatee ;
model killed = powerbt / clm ;
run ;
```


Model: MODEL1
Dependent Variable: killed

Output Statistics

Obs	Dep Var killed	Predicted Value	Std Error Mean Predict	95% CL Mean		Residual
1	13.0000	14.3827	1.9299	10.1779	18.5876	-1.3827
2	21.0000	16.0059	1.7974	12.0896	19.9222	4.9941
3	24.0000	18.6280	1.5976	15.1472	22.1089	5.3720
4	16.0000	20.7507	1.4528	17.5853	23.9161	-4.7507
5	24.0000	22.6236	1.3420	19.6997	25.5475	1.3764
6	20.0000	22.4987	1.3488	19.5600	25.4375	-2.4987
7	15.0000	24.2468	1.2622	21.4968	26.9968	-9.2468
8	34.0000	28.3672	1.1482	25.8656	30.8689	5.6328
9	33.0000	31.6137	1.1650	29.0753	34.1520	1.3863
10	33.0000	35.2346	1.2909	32.4221	38.0472	-2.2346
11	39.0000	39.1054	1.5187	35.7963	42.4144	-0.1054
12	43.0000	42.8512	1.7974	38.9349	46.7675	0.1488
13	50.0000	47.3462	2.1762	42.6048	52.0877	2.6538
14	47.0000	48.3451	2.2647	43.4109	53.2794	-1.3451

Sum of Residuals	0
Sum of Squared Residuals	219.44991
Predicted Residual SS (PRESS)	281.76275

Prediction interval for a new individual's Y given that we know their value of X, say X_0

- Point estimate is the group mean

$$\hat{Y}_{new} = a + bX_0$$

- Would you expect an individual's response to be more or less variable than the group's mean response?
- two sources of variability in prediction of individual Y
  - uncertainty of the group mean $E[Y|X_0]$
  - variability of individual responses around the group mean
    - Example: The Netherlands has per capita PCGDP of 13, but is unlikely to have PCH of exactly the mean of all possible countries with the same GDP.

- estimated standard error of $\hat{Y}_{new}$

$$\hat{\sigma}_{\hat{Y}_{new}} = \hat{\sigma}\sqrt{1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2)}{\Sigma\,(X_i - \bar{X})^2}}$$

- A (1 - α)% prediction interval for Y_{new} is given by:

$$\hat{Y}_{new} \pm (t_{1-\alpha/2,df=n-2})(\hat{\sigma}_{\hat{Y}_{new}})$$


```
proc reg data = manatee ;
model killed = powerbt / cli ;
run ;
```

The REG Procedure
Model: MODEL1
Dependent Variable: killed

Output Statistics

Obs	Dep Var killed	Predicted Value	Std Error Mean Predict	95% CL Predict		Residual
1	13.0000	14.3827	1.9299	4.1604	24.6050	-1.3827
2	21.0000	16.0059	1.7974	5.8989	26.1130	4.9941
3	24.0000	18.6280	1.5976	8.6816	28.5745	5.3720
4	16.0000	20.7507	1.4528	10.9102	30.5911	-4.7507
5	24.0000	22.6236	1.3420	12.8582	32.3891	1.3764
6	20.0000	22.4987	1.3488	12.7288	32.2687	-2.4987
7	15.0000	24.2468	1.2622	14.5320	33.9616	-9.2468
8	34.0000	28.3672	1.1482	18.7198	38.0147	5.6328
9	33.0000	31.6137	1.1650	21.9566	41.2707	1.3863
10	33.0000	35.2346	1.2909	25.5019	44.9673	-2.2346
11	39.0000	39.1054	1.5187	29.2178	48.9929	-0.1054
12	43.0000	42.8512	1.7974	32.7442	52.9582	0.1488
13	50.0000	47.3462	2.1762	36.8917	57.8007	2.6538
14	47.0000	48.3451	2.2647	37.8018	58.8884	-1.3451

Sum of Residuals	0
Sum of Squared Residuals	219.44991
Predicted Residual SS (PRESS)	281.76275