

Old Quiz 8  
Short 20 minutes

Stat Learning

1) Tremearne (1911) presents a data set of about 17 measurements on 112 people of Hausa nationality. We used  $Y = height$ . Along with a constant  $x_{i,1} \equiv 1$ , the five additional predictor variables used were  $x_{i,2} = height \text{ when sitting}$ ,  $x_{i,3} = height \text{ when kneeling}$ ,  $x_{i,4} = head \text{ length}$ ,  $x_{i,5} = nasal \text{ breadth}$ , and  $x_{i,6} = span$  (perhaps from left hand to right hand). The output below is for the OLS full model.

	Estimate	Std.Err	95% shorth CI
Intercept	-77.0042	65.2956	[-208.864, 55.051]
X2	0.0156	0.0992	[-0.177, 0.217] ←
→ X3	1.1553	0.0832	[ 0.983, 1.312]
X4	0.2186	0.3180	[-0.378, 0.805]
X5	0.2660	0.6615	[-1.038, 1.637]
→ X6	0.1396	0.0385	[0.0575, 0.217]

a) Give the shorth 95% CI for  $\beta_2$ .

17

$$[-0.177, 0.217]$$

b) Compute the standard 95% CI for  $\beta_2$ .

17

$$.0156 \pm 1.96(0.0992) = .0156 \pm .1944$$

$$= [-0.1788, 0.2100]$$

c) Which variables, if any, are needed in the MLR model given that the other variables are in the model?

16

x3 and x6  
height when kneeling span

3.11 a-c

Variables with CI that does not contain 0

50

3.11 def

SL Q4

2) Now we use forward selection and  $I_{min}$  is the minimum  $C_p$  model for the data from problem 1). Hence  $Y = height$ , the constant  $x_{i,1} \equiv 1$ ,  $x_{i,2} = height \text{ when sitting}$ ,  $x_{i,3} = height \text{ when kneeling}$ ,  $x_{i,4} = head \text{ length}$ ,  $x_{i,5} = nasal \text{ breadth}$ , and  $x_{i,6} = span$ .

	Estimate	Std.Err	95% shorth CI
Intercept	-42.4846	51.2863	[-192.281, 52.492]
X2	0		[ 0.000, 0.268]
→ X3	1.1707	0.0598	[ 0.992, 1.289]
X4	0		[ 0.000, 0.840]
X5	0		[ 0.000, 1.916]
→ X6	0.1467	0.0368	[ 0.0747, 0.215]

(19 min) E3d19

	(Intercept)	a	b	c	d	e
1	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE
→ 2	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE
3	TRUE	FALSE	TRUE	TRUE	FALSE	TRUE
4	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE
5	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE

TRUE = \*  
FALSE = blank

> tem2\$cp

[1] 14.389492 0.792566 2.189839 4.024738 6.000000

a) What is the value of  $C_p(I_{min})$  and what is  $\hat{\beta}_{I_{min},0}$ ?

16

3 →  $C_p(I_{min}) = 0.792566$

13 +  $\hat{\beta}_{I_{min},0} = (-42.8486, 0, 1.1707, 0, 0, 0.1467)^T$

-7 to -10  
if no numbers

b) Which variables, if any, are needed in the MLR model given that the other variables are in the model?

17

X3 and X6  
height while kneeling span

variables with CI that does not contain 0

c) List the variables, including a constant, that model 3 contains.

17

constant, X3, X4, X6  
height while kneeling head length span

b c e -2

90