

End of Chapter 8 and Chapter 9

Administrivia

1. By end of this lecture you should be able to do HW from chapters 7, 8, 9
2. Go to lab UH112 this Thursday
A-M 8:30-9:05 N-Z 9:05-9:40

Let's go through Team 1's regression analysis trying to predict, forecast, determine

Y=predicted=response=MILES on a car by knowing its

X=predictor=explanatory=AGE

WHAT MUST HAPPEN?

1. We assume Y (miles) (and sometimes X) comes from a normal distribution
2. We think there is a linear relationship – plot your data and look – with few outliers.....view their plot

5. Now we can find predictions \hat{Y} (Y-hat) by plugging an X into our line equation...see spreadsheet

and

6. We can talk about the error we make when predicting: $Y - \hat{Y} = e$

age	miles	zage	zmiles	prod
1.00	10.00	-1.29	-1.34	1.73
6.00	117.00	-0.30	0.60	-0.18
9.00	150.00	0.30	1.20	0.36
7.00	84.00	-0.10	0.00	0.00
14.00	150.00	1.29	1.20	1.55
15.00	98.00	1.49	0.26	0.38
6.00	35.00	-0.30	-0.89	0.26
<u>2.00</u>	<u>26.00</u>	<u>-1.09</u>	<u>-1.05</u>	<u>1.14</u>
7.50	83.75			5.25
5.04	55.08			
				0.75

▷ Regression: age by miles

Dependent variable is: **miles**

No Selector

R squared = 56.2% R squared (adjusted) = 48.8%

s = 39.39 with 8 - 2 = 6 degrees of freedom

Source	Sum of Squares	df	Mean Square	F-ratio
Regression	11926.1	1	11926.1	7.68
Residual	9311.38	6	1551.9	

Variable	Coefficient	s.e. of Coeff	t-ratio	prob
Constant	22.3596	26.16	0.855	0.4255
age	8.18539	2.953	2.77	0.0323

age	miles	zage	zmiles	prod	pred	error	e-square
1.00	10.00	-1.29	-1.34	1.73	30.5	-20.5	422.0
6.00	117.00	-0.30	0.60	-0.18	71.4	45.5	2072.8
9.00	150.00	0.30	1.20	0.36	96.0	53.9	2912.9
7.00	84.00	-0.10	0.00	0.00	79.6	4.3	18.8
14.00	150.00	1.29	1.20	1.55	136.9	13.0	170.1
15.00	98.00	1.49	0.26	0.38	145.1	-47.1	2222.2
6.00	35.00	-0.30	-0.89	0.26	71.4	-36.4	1330.2
<u>2.00</u>	<u>26.00</u>	<u>-1.09</u>	<u>-1.05</u>	<u>1.14</u>	<u>38.7</u>	<u>-12.7</u>	<u>162.0</u>
				5.25			9311.3
7.50	83.75			0.75			
5.04	55.08						1551.90

Q. By looking at the handout for all teams, which lines have points most tightly clustered?

Ans:

HOW DO WE JUDGE THE MODEL?

1. Is r far from 0?
2. Plot e =errors on a y-axis versus their \hat{Y} =predicted values and look for:
 - a) good scatter; no pattern <GOOD>
3. Do the e =errors seem to be from a $\frac{(e - 0)}{\text{stdev of } e}$ distribution? Any of these z-scores equal to 3 or -3 ?