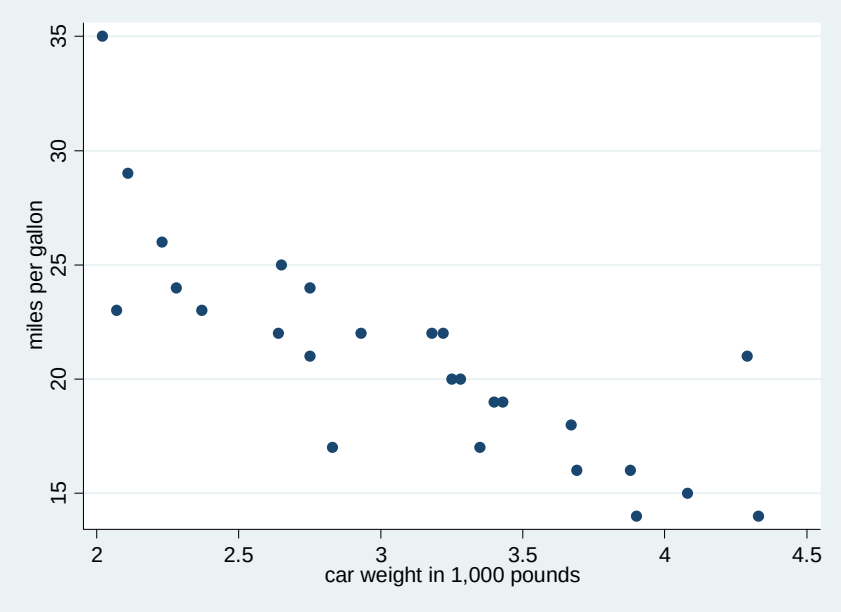
Linear Regression Example

Ani Katchova

Linear Regression Example

- Data set: auto
- Dependent variable: mpg (miles per gallon)
- Independent variable for the simple linear regression: weight
- Independent variables for the multiple linear regression: weight, price, and foreign

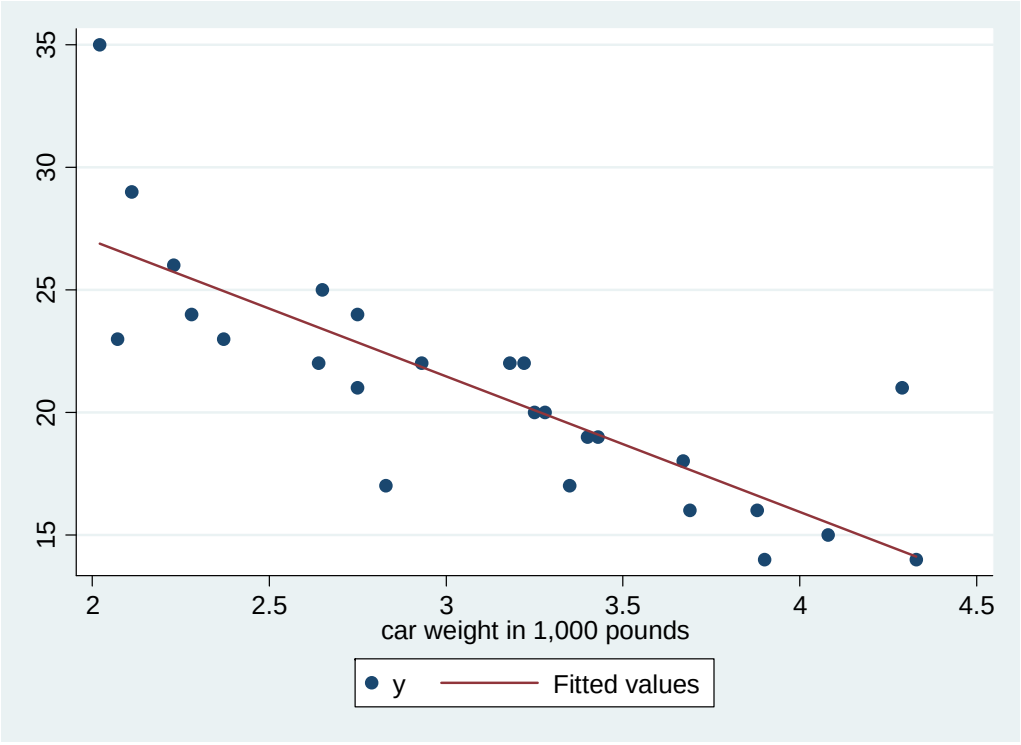
- Estimate a simple and multiple linear regression
- Plot regression line
- Interpret coefficients and their significance
- Estimate the predicted values for the dependent variable
- Calculate residuals



Simple linear regression: $y = \beta_0 + \beta_1 x_1 + u$

mpg (y)	Coefficient	Standard error	t-statistic	p-value	[95% Confidence	Interval]
weight (x)	-5.53	0.82	-6.72	0.00	-7.23	-3.83
intercept	38.07	2.61	14.58	0.00	32.68	43.46

- Regression line equation: $\hat{y} = b_0 + b_1 x_1 = 38.07 - 5.53 x_1$
- One unit increase in weight (corresponding to a 1,000 pounds) leads to a reduction of 5.53 miles per gallon.



Multiple linear regression

Source	SS	df	MS			Number of obs =26
						F(3, 22)=15.25
Model	382.1	3	127.4			Prob > F =0
Residual	183.8	22	8.4			R-squared =0.6752
						Adj R-squared =0.6309
Total	565.8	25	22.6			Root MSE =2.8902

mpg (y)	Coefficient	Standard error	t-statistic	p-value	[95% Confidence	Interval]
weight	-7.12	1.60	-4.44	0.00	-10.45	-3.79
price	0.0002	0.0003	0.85	0.40	-0.0003	0.0008
foreign	-2.51	2.06	-1.22	0.24	-6.77	1.76
intercept	42.17	4.26	9.89	0.00	33.32	51.01

- From the ANOVA table, the F-statistic is 15.25 and the p-value is less than 0.05. Therefore all coefficients are jointly significantly different from zero.
- The R-square is 0.67 and the adjusted R-squared is 0.63. The regression line has a good fit.
- If p-value<0.05 then the coefficient is significantly different from zero. Also if the 95% confidence interval does not contain zero, the coefficient is significantly different from zero. Here, only the coefficients on weight and the intercept are significantly different than zero.
- If weight is higher by 1 unit (corresponding to 1,000 pounds) then the mpg is 7.12 units lower in terms of miles per gallon.

Regression models (summary tables)

y=mpg	Simple linear regression	Multiple linear regression
Weight	-5.5*	-7.1*
Price		0.0002
Foreign		-2.5
Intercept	38.1*	42.1*

*: Coefficient is significantly different from zero at the 5% significance level.

- Interpretation of the coefficients in the simple linear regression model: for cars with 1 unit higher weight (corresponding to 1,000 pounds more), this is associated with a reduction of 5.5 miles per gallon.
- Interpretation of the coefficients in the multiple linear regression model: for cars with 1 unit higher weight (corresponding to 1,000 pounds more), this is associated with a reduction of 7.1miles per gallon, holding all else constant.
- Other coefficients are not significantly different from zero.