Problem set 1: Linear regression

Silje Synnøve Lyder Hermansen

2025-02-17

We can think of linear regression as a comparison of means or as a way to extract the relationship (correlation) between different variables.

Conceptual points: the statistics

Upload the King et al PDF and dialogue with ChatTutor and/or ChatGPT and ask it for explainers

- Can you express in layman's terms what a "standard deviation" of a variable is?

- What are the "residuals" of the regression? How are they calculated?

- What is a variance-covariance matrix?

- What is the role of the variance-covariance matrix in the King et al article (pdf) I uploaded?

- Can you explain what the covariance matrix is good for in this example?

- What is the difference between fundamental and estimation uncertainty?

- What is the difference between expected and predicted values of Y and how does this relate to the difference between fundamental and estimation uncertainty? When am I interested in one rather than the other?

Exercises in R

Scenarios: point estimates and simulation using ggpredict

MEPs would have to compensate for lack of party funding with their own parliamentary allowance if their party lacks funding. State funding for parties is often proportinal to party size in the national party, so we'll add this as a variable to the regression.

- can you re-fit model 2 with each MEP's national party size in the national parliament as a predictor?

- what is the marginal effect of party size on MEP's local investment?

- create two scenarios, justify your choice and calculate the first difference between the two.

- visualize the effect of party size on MEP's local investment.

Fundamental variation

- can you calculate the residuals for model 1, then model 2 and store them as separate variables in R? (you can use my slides from Monday)

- can you describe the resituals of the two models in a histogram, then in numbers by calculating the mean and standard deviation?

- what is the difference between the two sets of residuals and why?

- can you extract the variance-covariance matrix for model 2? (you can use my slides)

- can you calculate the standard error for the regression coefficients (parameters) from the variance-covariance matrix?

- are there any predictors that correlate more than others?

- how does this relate to King et al's argument?

Table 1:

	Local staff size (y)	
	LocalAssistants	
	Model 1	Model 2
OpenList	0.684***	0.829***
	(0.068)	(0.228)
LaborCost		−0.070***
		(0.010)
Constant	2.177***	4.127***
	(0.046)	(0.286)
Observations	7,143	739
R^2	0.014	0.081
Adjusted R^2	0.014	0.079
Residual Std. Error	2.856 (df = 7141)	3.083 (df = 736)
F Statistic	101.946*** (df = 1; 7141)	32.612*** (df = 2; 736)

Note: *p<0.1; **p<0.05; ***p<0.01

The models

Literature