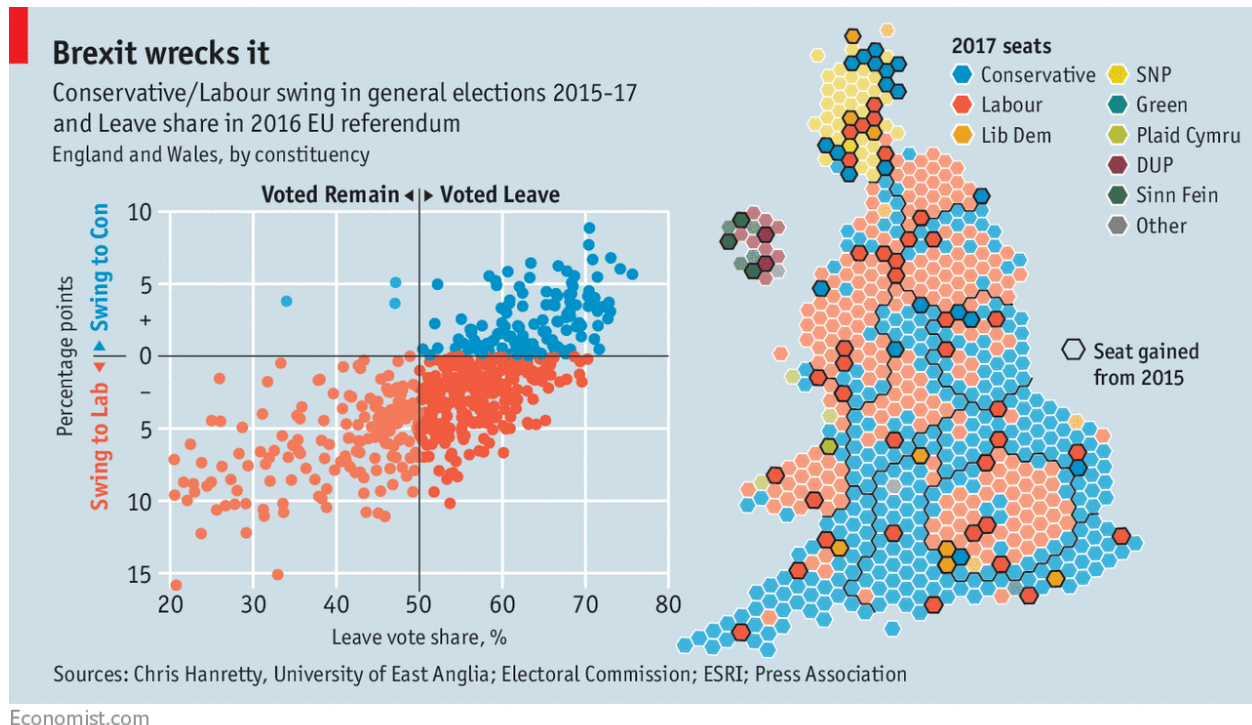


Gabarito - Lista 9

Correlação e Regressão Bivariada

Gabarito

Exercício 1 [4 pontos]. O gráfico abaixo, publicado pela revista *The Economist*, apresenta a relação entre o plebiscito do Brexit e o desempenho dos partidos Trabalhista (Lab) e Conservador (Con) nas eleições britânicas.



- (a) Resuma em um parágrafo a associação encontrada pela revista.
- (b) Explique por que seria equivocado concluir, somente a partir desses dados, que a ascensão do Partido Conservador é uma consequência do Brexit.
- (c) Dê uma explicação alternativa que seria plausível para explicar os resultados.
- (a) O gráfico de dispersão mostra uma associação positiva entre os distritos que votaram a favor do Brexit e aqueles que tendem a votar no Partido Conservador. De interesse particular é o fato de que apenas 3 distritos que votaram contra o Brexit fortaleceram o apoio aos Conservadores.

- (b) O primeiro ponto a ser notado, e também o mais óbvio, é que o gráfico mostra apenas uma associação entre as duas variáveis, e não mostra causalidade. É possível que uma variável omitida no gráfico, como crise econômica ou aversão à presença de imigrantes, tenha influência tanto sobre o voto no Brexit quanto naquele sobre o Partido Conservador.

Um outro equívoco de análise possível neste gráfico não tem a ver com causalidade, e é puramente descritivo. O fato de haver uma correlação positiva entre o voto no Brexit e o partido conservador não significa que o último está em “ascensão”. De fato, se analisarmos o gráfico com cuidado, observamos que há muito mais distritos com *swing* para os trabalhistas depois do Brexit do que distritos com *swing* para os conservadores.

- (c) Serão aceitas quaisquer respostas que indiquem uma terceira variável que tenha uma conexão plausível com o Brexit e com o voto no Partido Conservador, como, por exemplo, crise econômica e percepção em relação à imigração. Respostas que indicarem uma terceira variável que só tem relação plausível com um dos fenômenos receberão crédito parcial.

Exercício 2 [6 pontos]. Com base nos resultados da regressão abaixo, responda:

. reg babymort lifeexpf						
Source	SS	df	MS	Number of obs = 102		
Model	140230.169	1	140230.169	F(1, 100) = 1249.96		
Residual	11218.8011	100	112.188011	Prob > F = 0.0000		
Total	151448.97	101	1499.49476	R-squared = 0.9259		
				Adj R-squared = 0.9252		
				Root MSE = 10.592		
babymort	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lifeexpf	-3.445937	.0974675	-35.35	0.000	-3.63931	-3.252564
_cons	284.315	6.895303	41.23	0.000	270.6349	297.9951

- (a) Quais são as variáveis independente e dependente?
- (b) Qual é o valor de α ? Interprete-o.
- (c) Qual é o valor de β ? Interprete-o.
- (d) As colunas “t”, “P>|t|”, e “95% Conf Interval” correspondem a três critérios de decisão. Identifique-os e aponte, em cada um deles, quais valores nos permitiriam rejeitar H_0 a 95% de significância.

- (e) Formalize as hipóteses sendo testadas para α e β , e decida em cada caso se H_0 pode ser rejeitada.
- (f) Identifique e interprete as duas medidas de ajuste do modelo apresentadas na tabela.

(a) VI: expectativa de vida feminina; VD: mortalidade infantil

(b) $\alpha = 284,315$.

A interpretação do intercepto depende de uma extrapolação que não esperamos ver na realidade: o modelo prevê que, em um país em que a expectativa de vida média das mulheres é zero, a mortalidade infantil seria de 284 mortes a cada 1000 nascimentos vivos.

(c) $\beta = -3,446$.

O modelo prevê que o aumento de um ano de vida na expectativa de vida média das mulheres está associado a uma queda de 3,446 mortes a cada 1000 nascimentos vivos. Trata-se de um efeito alto, mas devemos nos lembrar de que aumentar a expectativa de vida *média* das mulheres de um país em um ano não é tarefa trivial. Ademais, esse modelo não está controlado por nenhuma outra variável, o que enviesaria o estimador (como veremos a partir da próxima aula).

- (d)
- “t” é a estatística-teste. Assumindo um teste bilateral com n alto, podemos rejeitar H_0 a 95% de significância se $t > 1,96$
 - “ $P > |t|$ ” é o p-valor. Podemos rejeitar H_0 a 95% de significância se $p < 0,05$
 - “95% Conf Interval” é a estimativa intervalar para β . Podemos rejeitar H_0 a 95% de significância se o intervalo **não** passar por 0.
- (e)
- Teste para α
 - $H_0 : \alpha = 0$
 - $H_a : \alpha \neq 0$
 - Rejeitamos a hipótese nula
 - Teste para β
 - $H_0 : \beta = 0$
 - $H_a : \beta \neq 0$
 - Rejeitamos a hipótese nula
- (f)
- R^2 : A expectativa de vida feminina explica 92,59% da variação da mortalidade infantil
 - Root MSE: A diferença média entre a mortalidade infantil observada e a mortalidade infantil prevista pelo modelo é de 10.592 mortes a cada 1000 nascimentos vivos (lembre-se de que esta medida é lida na escala da variável dependente).