

TABLE 1. Summary Statistics

Variable	Min	Max	Mean	Std dev
(Subdistrict) Muslim Brotherhood branch, 1937	0	1	0.05	0.21
(Subdistrict) Muslim population, logged	0	10.99	7.75	0.98
(Subdistrict) Literate %	0	100	15.82	10.88
(Subdistrict) Employed in agriculture %	0	83.96	32.69	13.67
(Subdistrict) Unemployed males %	0	70	10.30	4.98
(Subdistrict) Non-Muslim %	0	100	6.48	14.40
(Subdistrict) Egyptian state railway	0	1	0.09	0.29
(District) Population change since 1917 %	-71.38	2466.13	25.49	88.06
(District) European %	0	59.98	0.61	3
(District) Missionaries, per 10,000	0	11.10	0.16	0.62
(District) Employed in state administration %	0.19	7.83	0.87	0.76
(District) Distance from Cairo HQ, sqrt km	0	26.49	11.51	5.18
(District) Administrative center	0	1	0.03	0.16

Each observation comprises one subdistrict in one district

TABLE 2. Predicting the Probability of a Muslim Brotherhood Branch (1937)

	Model 1	Model 2	Model 3
Subdistrict: Literate (%)	1.048*** (0.013)	1.065*** (0.015)	1.069*** (0.016)
Subdistrict: Employed in agriculture (%)	0.962*** (0.011)	0.959*** (0.012)	0.970* (0.012)
Subdistrict: Unemployed males (%)	1.050* (0.022)	1.054* (0.023)	1.042 (0.024)
District: Population change since 1917 (%)	0.990* (0.004)	0.994 (0.005)	0.999 (0.003)
Subdistrict: Non-Muslim (%)		1.016 (0.009)	1.015 (0.009)
District: Europeans (%)		0.721*** (0.065)	0.742*** (0.057)
District: Missionaries (per 10,000)		0.872 (0.161)	0.930 (0.182)
District: Employed in state administration (%)			0.636** (0.108)
Subdistrict: Egyptian state railway			3.734*** (0.735)
District: Distance from Cairo HQ (sqrt km)	1.095*** (0.029)	1.084** (0.029)	1.050 (0.027)
District: Administrative center	0.572 (0.295)	0.820 (0.467)	1.077 (0.593)
Random intercept: district (std dev)	1.120*** (0.183)	1.115*** (0.170)	0.972*** (0.151)
Reduction in AIC compared to Model 1		-28	-76
Branches	202	202	202
Subdistricts	4178	4178	4178
Districts	134	134	134

Multilevel logistic regression, with $\ln(\text{Muslim population})$ as offset
Standard errors in parentheses; *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Question 1 (30 percent). Brooke and Ketchley (2018) use multilevel logistic regression to model where Muslim Brotherhood branches were founded in 1930s Egypt. Coefficients are expressed as odds ratios (the exponent of β). Several hypotheses are tested:

H1: Muslim Brotherhood branches were more likely in subdistricts with higher literacy

H2: Muslim Brotherhood branches were less likely in subdistricts with greater employment in agriculture

H3: Muslim Brotherhood branches were more likely in areas where Christian missionaries were active

H4: Muslim Brotherhood branches were less likely in areas where state administration was more extensive

Questions:

- a) Which of Brooke and Ketchley's hypotheses (H1-H4) are supported?
- b) Interpret the association between having a train station in a subdistrict and the likelihood of having a Muslim Brotherhood branch
- c) The model includes random intercepts at the district level. Explain the rationale for this modeling choice
- d) Moving from Model 1 to Model 3, interpret the AIC statistic

Question 2 (20 percent)

Peace and conflict data is often noisy and subject to missingness.

- a) Why is measurement error on one or more independent variables potentially more serious than measurement error on the outcome variable?
- b) What are some strategies for dealing with missingness?

Question 3 (25 percent).

You are asked to design a study on the effects of democracy on economic performance using the V-Dem dataset. The outcome measure is GDP per capita. To give an illustration: for the year 2019, this variable ranges from \$275 to \$113,197. V-Dem contains measures for electoral democracy for 201 countries captured yearly from 1789 to 2017. The unit of analysis is thus the country year. Explain how you would model this, with particular emphasis on the estimator and data structure.

Question 4 (25 percent)

Under a number of assumptions, the Ordinary Least Squares model is the Best Linear Unbiased Estimator (BLUE). Choose four of these assumptions and:

- 1) explain them;
- 2) explain the consequences of violating them;
- 3) describe how we can test whether they hold or not.