

Power Analyse: Schwereinschätzung von Cannabis Delikten

Methoden

Power Analyse mit herkömmlichen Methoden

Da herkömmliche Methoden in der Regel nur für einfachere Modelle existieren, wurde für diese Analyse ein logistisches Regressionsmodell angenommen. Dabei wird eine binäre Zielvariable angenommen (0/1: keine Strafe/schwere Strafe) und die Wahrscheinlichkeit, Teil von Gruppe 1 zu sein (P_1), modelliert. Die Basiswahrscheinlichkeit (P_0) ist die Wahrscheinlichkeit, Teil von Gruppe 1 zu sein, ohne den Einfluss der Kovariablen des Modells. Die Modellgleichung in einem solchen Fall lautet:

$$\ln\left(\frac{P_1}{1-P_1}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$$

mit:

$$\beta_0 = \ln\left(\frac{P_0}{1-P_0}\right)$$

Das Paket `pwrss` verwendet den Wald-Test (oder z-Test), um die Signifikanz von Variablen zu schätzen und basierend darauf die benötigte Stichprobengröße zu ermitteln. Für diese Analyse wurde ein Signifikanzniveau von $\alpha = 0.05$ sowie eine Power von 0.8 verwendet. Die Funktion nutzt neben P_0 , P_1 (was in diesem Fall die vorhergesagte Wahrscheinlichkeit ist, wenn man X_1 (bzw. die Kovariable von Interesse) von 0 auf 1 erhöht), α und der Power noch ein Argument für die multiple Korrelation zwischen den Kovariablen in Form des adjustierten R^2 . Für die Zwecke dieser Analyse wurde dieses Argument auf 0.5 gesetzt.

In dieser Analyse wurden verschiedene Kombinationen von P_0 und P_1 verwendet, um die benötigte Stichprobengröße für verschiedene Datensätze abzuschätzen. Dabei wurde als minimale Veränderung zwischen den beiden eine Veränderung von 0.03 betrachtet.

Im Folgenden ist eine Tabelle mit verschiedenen Kombinationen von P_0 und P_1 , sowie der benötigten Stichprobengröße, um für diese Datenlage einen signifikanten Effekt feststellen zu können, zu sehen:

	1	2	3	4	5	6	7	8	9
p0	0.10	0.10	0.20	0.20	0.30	0.30	0.40	0.40	0.50
p1	0.07	0.13	0.17	0.23	0.27	0.33	0.37	0.43	0.47
sample_size	1132.00	1967.00	2485.00	3112.00	3471.00	3889.00	4102.00	4312.00	4382.00

	10	11	12	13	14	15	16	17	18
p0	0.50	0.60	0.60	0.70	0.70	0.80	0.80	0.90	0.90
p1	0.53	0.57	0.63	0.67	0.73	0.77	0.83	0.87	0.93
sample_size	4382.00	4312.00	4102.00	3889.00	3471.00	3112.00	2485.00	1967.00	1132.00

Für eine minimale Veränderung von 0.03 zwischen P_0 und P_1 bewegen sich die benötigten Stichprobengrößen zwischen ca. 1100 und 4400.

Power Analyse mit simulierten Daten

Da das angestrebte Modell etwas komplexer ist als das oben skizzierte Modell, wurde zusätzlich eine Power Analyse mithilfe von simulierten Daten durchgeführt. Dabei wurde ein kumulatives Logit-Modell für ordinale Zielvariablen geschätzt. Die Daten wurden für verschiedene Effektgrößen (zwischen -0.5 und 2.5) und Stichprobengrößen (zwischen 10 Personen und 3000 Personen) simuliert, und anschließend wurde ein Modell mit den Daten geschätzt. Zusätzlich zu den Effekten der Kovariablen wurden bei der Simulation der Daten auch Personen-Effekte μ_j , $j \in 1, \dots, 10$ im Intervall $[-3, 3]$ einbezogen, sowie verschiedene Intercept-Termine.

Das geschätzte Modell verwendet jeweils eine ordinale Variable Y mit 8 Kategorien als Zielvariable. Der lineare Prädiktor des Modells lautet $Y \sim age + treatment \times setup$

Die folgende Tabelle zeigt den jeweiligen Effekt, die Effektgröße, die Power und die benötigte Stichprobengröße.

effect	power	persons	log_odds	odds
setupRobbery	1.00	50	3.0	20.09
setupCocain	1.00	50	2.5	12.18
treatment	0.84	50	1.5	4.48
treatment:setupCocain	0.82	100	1.5	4.48
treatment:setupRobbery	0.82	200	1.0	2.72
setupAlcohol	0.84	450	-0.5	0.61
treatment:setupStealth	0.82	1250	0.4	1.49
treatment:setupAlcohol	0.84	1500	-0.4	0.67
setupStealth	0.90	3000	0.2	1.22

Die Ergebnisse zeigen, dass für alle simulierten Effekte bei einem verwendeten Signifikanzniveau von $\alpha = 0.05$ ab einer Stichprobengröße von 3000 eine Power größer als 0.8 erreicht werden kann.

Empfehlung

Basierend auf den Ergebnissen der Power Analyse empfehlen wir eine Stichprobengröße von etwa 3000-4000 Personen.