Chapter 46
Probit Models for Estimating Effective Pharmacological Treatment Dosages (14 Tests)

General Purpose

Probit regression is, just like logistic regression, for estimating the effect of predictors on yes/no outcomes. If your predictor is multiple pharmacological treatment dosages, then probit regression may be more convenient than logistic regression, because your results will be reported in the form of response rates instead of odds ratios. The dependent variable of the two methods log odds (otherwise called logit) and log prob (otherwise called probit) are closely related to one another. Log prob (probability), is the z-value corresponding to its area under the curve value of the normal distribution. It can be shown that the log odds of responding $\approx (\pi/\sqrt{3})$ x log prob of responding (see Chap. 7, Machine learning in medicine part three, Probit regression, pp 63–68, 2013, Springer Heidelberg Germany, from the same authors).

Primary Scientific Question

This chapter will assess whether probit regression is able to find response rates of different dosages of mosquito repellents.

Example

Simple Probit Regression

repellent nonchem	repellent chem	mosquitos gone	n mosquitos
1	,02	1000	18000
1	,03	1000	18500

(continued)

T.J. Cleophas, A.H. Zwinderman, *Machine Learning in Medicine - a Complete Overview*, DOI 10.1007/978-3-319-15195-3_46

repellent nonchem	repellent chem	mosquitos gone	n mosquitos
1	,03	3500	19500
1	,04	4500	18000
1	,07	9500	16500
1	,09	17000	22500
1	,10	20500	24000

In 14 test sessions the effect measured as the numbers of mosquitos gone after administration of different dosages of a chemical repellent was assessed. The first 7 sessions are in the above table. The entire data file is entitled probit.sav, and is in extras.springer.com. Start by opening the data file in SPSS statistical software.

Command:

Analyze....Regression....Probit Regression....Response Frequency: enter "mosquitos gone"....Total Observed: enter "n mosquitos"....Covariate(s): enter "chemical".... Transform: select "natural log"....click OK.

Chi-Square tests				
		Chi-Square	df[a]	Sig.
PROBIT	Pearson Goodness-of-Fit Test	7706,816	12	,000[b]

[a]Statistics based on individual cases differ from statistics based on aggregated cases
[b]Since the significance level is less than, 150, a heterogeneity factor is used in the calculation of confidence limits

In the output sheets the above table shows that the goodness of fit tests of the data is significant, and, thus, the data do not fit the probit model very well. However, SPSS is going to produce a heterogeneity correction factor and we can proceed. The underneath shows that chemical dilution levels are a very significant predictor of proportions of mosquitos gone.

Parameter estimates							
						95 % Confidence interval	
Parameter		Estimate	Std. Error	Z	Sig.	Lower bound	Upper bound
PROBIT[a]	Chemical (dilution)	1,649	,006	286,098	,000	1,638	1,660
	Intercept	4,489	,017	267,094	,000	4,472	4,506

[a]PROBIT model: PROBIT(p) = Intercept + BX (Covariates X are transformed using the base 2.718 logarithm.)

Cell counts and residuals							
Number		Chemical (dilution)	Number of subjects	Observed responses	Expected responses	Residual	Probability
PROBIT	1	−3,912	18000	1000	448,194	551,806	,025
	2	−3,624	18500	1000	1266,672	−266,672	,068
	3	−3,401	19500	3500	2564,259	935,741	,132
	4	−3,124	18000	4500	4574,575	−74,575	,254
	5	−2,708	16500	9500	8405,866	1094,134	,509
	6	−2,430	22500	17000	15410,676	1589,324	,685
	7	−2,303	24000	20500	18134,992	2365,008	,756
	8	−3,912	22500	500	560,243	−60,243	,025
	9	−3,624	18500	1500	1266,672	233,328	,068
	10	−3,401	19000	1000	2498,508	−1498,508	,132
	11	−3,124	20000	5000	5082,861	−82,861	,254
	12	−2,708	22000	10000	11207,821	−1207,821	,509
	13	−2,430	16500	8000	11301,162	−3301,162	,685
	14	−2,303	18500	13500	13979,056	−479,056	,756

The above table shows that according to chi-square tests the differences between observed and expected proportions of mosquitos gone is several times statistically significant.

It does, therefore, make sense to make some inferences using the underneath confidence limits table.

Confidence limits							
		95 % Confidence limits for chemical (dilution)			95 % Confidence limits for log (chemical (dilution)[b]		
Probability		Estimate	Lower bound	Upper bound	Estimate	Lower bound	Upper bound
PROBIT[a]	,010	,016	,012	,020	−4,133	−4,453	−3,911
	,020	,019	,014	,023	−3,968	−4,250	−3,770
	,030	,021	,016	,025	−3,863	−4,122	−3,680
	,040	,023	,018	,027	−3,784	−4,026	−3,612
	,050	,024	,019	,029	−3,720	−3,949	−3,557
	,060	,026	,021	,030	−3,665	−3,882	−3,509
	,070	,027	,022	,031	−3,617	−3,825	−3,468
	,080	,028	,023	,032	−3,574	−3,773	−3,430
	,090	,029	,024	,034	−3,535	−3,726	−3,396
	,100	,030	,025	,035	−3,500	−3,683	−3,365
	,150	,035	,030	,039	−3,351	−3,506	−3,232
	,200	,039	,034	,044	−3,233	−3,368	−3,125
	,250	,044	,039	,048	−3,131	−3,252	−3,031

(continued)

Confidence limits							
		95 % Confidence limits for chemical (dilution)			95 % Confidence limits for log (chemical (dilution)[b]		
Probability		Estimate	Lower bound	Upper bound	Estimate	Lower bound	Upper bound
	,300	,048	,043	,053	−3,040	−3,150	−2,943
	,350	,052	,047	,057	−2,956	−3,059	−2,860
	,400	,056	,051	,062	−2,876	−2,974	−2,778
	,450	,061	,055	,067	−2,799	−2,895	−2,697
	,500	,066	,060	,073	−2,722	−2,819	−2,614
	,550	,071	,064	,080	−2,646	−2,745	−2,529
	,600	,077	,069	,087	−2,569	−2,672	−2/442
	,650	,083	,074	,095	−2,489	−2,598	−2,349
	,700	,090	,080	,105	−2,404	−2,522	−2,251
	,750	,099	,087	,117	−2,313	−2/441	−2,143
	,800	,109	,095	,132	−2,212	−2,351	−2,022
	,850	,123	,106	,153	−2,094	−2,248	−1,879
	,900	,143	,120	,183	−1,945	−2,120	−1,699
	,910	,148	,124	,191	−1,909	−2,089	−1,655
	,920	,154	,128	,200	−1,870	−2,055	−1,608
	,930	,161	,133	,211	−1,827	−2,018	−1,556
	,940	,169	,138	,224	−1,780	−1,977	−1/497
	,950	,178	,145	,239	−1,725	−1,931	−1/430
	,960	,190	,153	,259	−1,661	−1,876	−1,352
	,970	,206	,164	,285	−1,582	−1,809	−1,255
	,980	,228	,179	,324	−1,477	−1,719	−1,126
	,990	,269	,206	,397	−1,312	−1,579	−,923

[a]A heterogeneity factor is used
[b]Logarithm base = 2.718

E.g., one might conclude that a 0,143 dilution of the chemical repellent causes 0,900 (=90 %) of the mosquitos to have gone. And 0,066 dilution would mean that 0,500 (=50 %) of the mosquitos disappeared.

Multiple Probit Regression

Like multiple logistic regression using multiple predictors, probit regression can also be applied with multiple predictors. We will add as second predictor to the above example the nonchemical repellents ultrasound (=1) and burning candles (=2) (see uppermost table of this chapter).

Command:

Analyze....Regression....Probit Regression....Response Frequency: enter "mosquitos gone"....Total Observed: enter "n mosquitos"....Covariate(s): enter "chemical".... Transform: select "natural log"....click OK.

Chi-Square tests				
		Chi-Square	df[a]	Sig.
PROBIT	Pearson Goodness-of-Fit Test	3863,489	11	,000[b]

[a]Statistics based on individual cases differ from statistics based on aggregated cases
[b]Since the significance level is less than, 150, a heterogeneity factor is used in the calculation of confidence limits

Again, the goodness of fit is not what it should be, but SPSS adds a correction factor for heterogeneity. The underneath shows the regression coefficients for the multiple model. The no chemical repellents have significantly different effects on the outcome.

Parameter estimates								
							95 % Confidence interval	
Parameter			Estimate	Std. Error	Z	Sig.	Lower bound	Upper bound
PROBIT[a]	Chemical (dilution)		1,654	,006	284,386	,000	1,643	1,665
	Intercept[b]	Ultrasound	4,678	,017	269,650	,000	4,661	4,696
		Burning candles	4,321	,017	253,076	,000	4,304	4,338

[a]PROBIT model: PROBIT(p) = Intercept + BX(Covariates X are transformed using the base 2.718 logarithm.)
[b]Corresponds to the grouping variable repellentnonchemical

Cell counts and residuals								
Number		Repellent non chemical	chemical (dilution)	Number of subjects	Observed responses	Expected responses	Residual	Probability
PROBIT	1	1	−3,912	18000	1000	658,233	341,767	,037
	2	1	−3,624	18500	1000	1740,139	−740,139	,094
	3	1	−3,401	19500	3500	3350,108	149,892	,172
	4	1	−3,124	18000	4500	5630,750	−1130,750	,313
	5	1	−2,708	16500	9500	9553,811	−53,811	,579
	6	1	−2,430	22500	17000	16760,668	239,332	,745
	7	1	−2,303	24000	20500	19388,521	1111,479	,808
	8	2	−3,912	22500	500	355,534	144,466	,016
	9	2	−3,624	18500	1500	871,485	628,515	,047
	10	2	−3,401	19000	1000	1824,614	−824,614	,096
	11	2	−3,124	20000	5000	3979,458	1020,542	,199
	12	2	−2,708	22000	10000	9618,701	381,299	,437
	13	2	−2,430	16500	8000	10202,854	−2202,654	,618
	14	2	−2,303	18500	13500	12873,848	626,152	,696

In the above Cell Counts table, it is shown that according to the chi-square tests the differences of observed and expected proportions of mosquitos gone were statistically significant several times. The next page table gives interesting results. E.g., a 0,128 dilution of the chemical repellent causes 0,900 (=90 %) of the mosquitos to have gone in the ultrasound tests. And 0,059 dilution would mean that 0,500 (=50 %) of the mosquitos disappeared. The results of burning candles were less impressive. 0,159 dilution caused 90 % of the mosquitos to disappear, 0,073 dilution 50 %.

Confidence limits

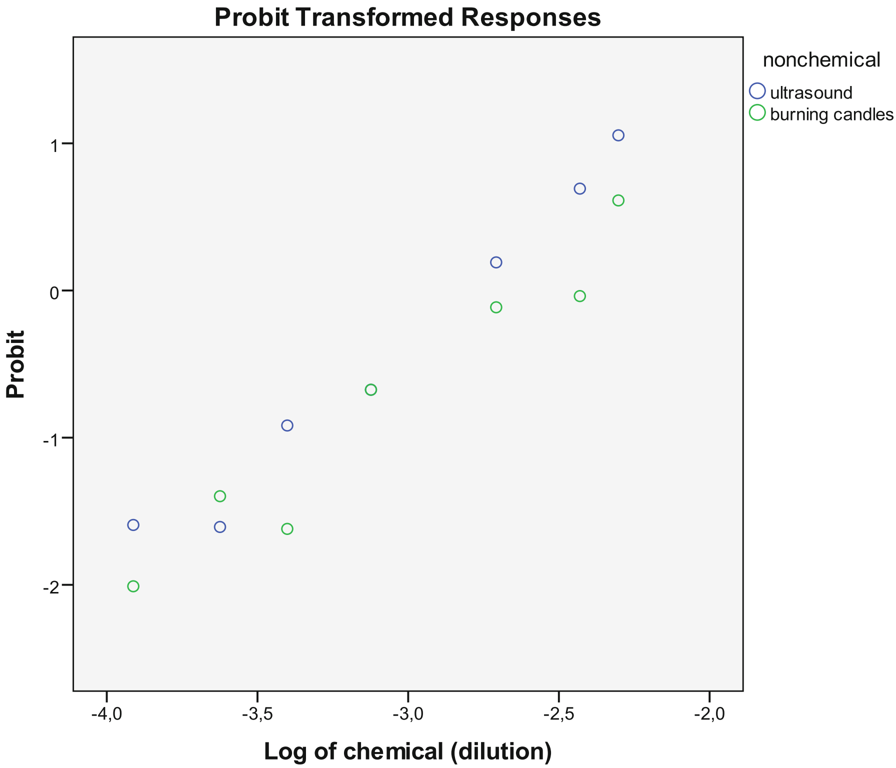
Nonchemical		Probability	95 % Confidence limits for chemical (dilution) Estimate	Lower bound	Upper bound	95 % Confidence limits for log (chemical (dilution)[b] Estimate	Lower bound	Upper bound
PROBIT[a]	Ultrasound	,010	,014	,011	,018	−4,235	−4,486	−4,042
		,020	,017	,014	,020	−4,070	−4,296	−3,895
		,030	,019	,015	,022	−3,966	−4,176	−3,801
		,040	,021	,017	,024	−3,887	−4,086	−3,731
		,050	,022	,018	,025	−3,823	−4,013	−3,673
		,060	,023	,019	,027	−3,769	−3,951	−3,624
		,070	,024	,020	,028	−3,721	−3,896	−3,581
		,080	,025	,021	,029	−3,678	−3,848	−3,542
		,090	,026	,022	,030	−3,639	−3,804	−3,506
		,100	,027	,023	,031	−3,603	−3,763	−3,473
		,150	,032	,027	,036	−3,455	−3,597	−3,337
		,200	,036	,031	,040	−3,337	−3,467	−3,227
		,250	,039	,035	,044	−3,236	−3,356	−3,131
		,300	,043	,038	,048	−3,146	−3,258	−3,043
		,350	,047	,042	,052	−3,062	−3,169	−2,961
		,400	,051	,046	,056	−2,982	−3,085	−2,882
		,450	,055	,049	,061	−2,905	−3,006	−2,803
		,500	,059	,053	,066	−2,829	−2,929	−2,725
		,550	,064	,058	,071	−2,753	−2,853	−2,646
		,600	,069	,062	,077	−2,675	−2,777	−2,564
		,650	,075	,067	,084	−2,596	−2,700	−2,478
		,700	,081	,073	,092	−2,512	−2,620	−2,387
		,750	,089	,079	,102	−2,421	−2,534	−2,287
		,800	,098	,087	,114	−2,320	−2,440	−2,174
		,850	,111	,097	,130	−2,202	−2,332	−2,042
		,900	,128	,111	,153	−2,054	−2,197	−1,874
		,910	,133	,115	,160	−2,018	−2,165	−1,833
		,920	,138	,119	,167	−1,979	−2,129	−1,789
		,930	,144	,124	,175	−1,936	−2,091	−1,740
		,940	,151	,129	,185	−1,889	−2,048	−1,686
		,950	,160	,135	,197	−1,834	−1,999	−1,623
		,960	,170	,143	,212	−1,770	−1,942	−1,550
		,970	,184	,154	,232	−1,691	−1,871	−1,459
		,980	,205	,169	,262	−1,587	−1,778	−1,339
		,990	,241	,196	,317	−1,422	−1,632	−1,149

(continued)

Confidence limits

Nonchemical		Probability	95 % Confidence limits for chemical (dilution)			95 % Confidence limits for log (chemical (dilution)[b]		
			Estimate	Lower bound	Upper bound	Estimate	Lower bound	Upper bound
	Burning candles	,010	,018	,014	,021	−4,019	−4,247	−3,841
		,020	,021	,017	,025	−3,854	−4,058	−3,693
		,030	,024	,019	,027	−3,750	−3,939	−3,599
		,040	,025	,021	,029	−3,671	−3,850	−3,528
		,050	,027	,023	,031	−3,607	−3,777	−3,469
		,060	,029	,024	,033	−3,553	−3,716	−3,420
		,070	,030	,026	,034	−3,505	−3,662	−3,376
		,080	,031	,027	,036	−3,462	−3,614	−3,336
		,090	,033	,028	,037	−3,423	−3,571	−3,300
		,100	,034	,029	,038	−3,387	−3,531	−3,267
		,150	,039	,034	,044	−3,239	−3,367	−3,128
		,200	,044	,039	,049	−3,121	−3,240	−3,015
		,250	,049	,044	,054	−3,020	−3,132	−2,916
		,300	,053	,048	,059	−2,930	−3,037	−2,826
		,350	,058	,052	,065	−2,845	−2,950	−2,741
		,400	,063	,057	,070	−2,766	−2,869	−2,658
		,450	,068	,061	,076	−2,688	−2,793	−2,578
		,500	,073	,066	,082	−2,613	−2,718	−2,497
		,550	,079	,071	,089	−2,537	−2,644	−2,415
		,600	,085	,076	,097	−2,459	−2,571	−2,331
		,650	,093	,082	,106	−2,380	−2,495	−2,244
		,700	,101	,089	,116	−2,295	−2,417	−2,151
		,750	,110	,097	,129	−2,205	−2,333	−2,049
		,800	,122	,106	,144	−2,104	−2,240	−1,936
		,850	,137	,119	,165	−1,986	−2,133	−1,802
		,900	,159	,136	,195	−1,838	−1,999	−1,633
		,910	,165	,140	,203	−1,802	−1,966	−1,592
		,920	,172	,145	,213	−1,763	−1,932	−1,548
		,930	,179	,151	,223	−1,720	−1,893	−1,499
		,940	,188	,157	,236	−1,672	−1,850	−1,444
		,950	,198	,165	,251	−1,618	−1,802	−1,381
		,960	,211	,175	,270	−1,554	−1,745	−1,308
		,970	,229	,187	,296	−1,475	−1,675	−1,217
		,980	,254	,206	,334	−1,371	−1,582	−1,096
		,990	,299	,238	,404	−1,206	−1,436	−,906

[a]A heterogenelty factor is used
[b]Logarithm base = 2.718



The above figure supports the adequacy of the multiple variables probit model, with two similarly sloped linear patterns (the blue and the green one) of "chemical repellent levels" versus "mosquitos gone levels" regressions.

Conclusion

Probit regression is, just like logistic regression, for estimating the effect of predictors on yes/no outcomes. If your predictor is multiple pharmacological treatment dosages, then probit regression may be more convenient than logistic regression, because your results will be reported in the form of response rates instead of odds ratios.

This chapter shows that probit regression is able to find response rates of different dosages of mosquito repellents.

Note

More background, theoretical and mathematical information of probit regression is given in the Chap. 7, Machine learning in medicine part three, Probit regression, pp 63–68, 2013, Springer Heidelberg Germany, (from the same authors).