

应用广义估计方程进行多应变量回归分析

在很多医学研究中，结果变量和/或危险因素是从多个渠道获得的，或由多个测量指标组成的。如呼吸道症状包括咳嗽、咳痰、喘鸣、气短。对儿童行为的评价可以从家长和教师两方面获得。对于这种多个应变量的数据，一种分析方法是分别对每个应变量与危险因素的关系进行回归分析。也可以使用 GEE（广义估计方程）同时对多个应变量与危险因素的关系进行回归。这样可以提高统计效率。每个变量单独分析可能都没有显著性，但综合起来就有可能显著了。比如，吸烟可能导致咳嗽、咳痰、喘鸣、气短，每个人表现的症状不一样，综合起来就是分析吸烟是否导致呼吸道症状的增加。

例： DEMO 数据四个呼吸道症状与单核苷酸多态性标记 SNP1 的关系分析，输入界面如下：

广义估计方程多应变量回归 ?

标题：

选择分析对象：

应变量(必须是同一种分布类型)

变量名	分布类型	联系函数
COUGH	Binomial	Logit
PHLEGM	Binomial	Logit
WHEEZE	Binomial	Logit
Shortness of breath	Binomial	Logit

暴露变量：

应变量之间相关类型：

其它自(协)变量：

分层变量：

输出结果：

回归方程 1：不同回归系数拟合危险因素对每个应变量的作用

	Estimate	Std.err	Wald	Pr(> W)	exp(coef)	95%区间下限	95%区间上限
--	----------	---------	------	----------	-----------	---------	---------

(Intercept)	-2.1203	0.9101	5.4279	0.0198	0.1200	0.0202	0.7142
factor(SNP1.COUGH) 1	-0.3782	0.2487	2.3126	0.1283	0.6851	0.4208	1.1154
factor(SNP1.COUGH) 2	0.4892	0.4493	1.1854	0.2763	1.6310	0.6761	3.9345
factor(SNP1.PHLEGM) 1	0.1457	0.2041	0.5100	0.4752	1.1569	0.7755	1.7259
factor(SNP1.PHLEGM) 2	0.2274	0.4230	0.2891	0.5908	1.2554	0.5479	2.8761
factor(SNP1.WHEEZE) 1	-0.1984	0.2489	0.6355	0.4253	0.8201	0.5035	1.3356
factor(SNP1.WHEEZE) 2	-0.2824	0.5820	0.2355	0.6275	0.7540	0.2409	2.3592
factor(SNP1.SOB) 1	-0.0535	0.1814	0.0869	0.7681	0.9479	0.6643	1.3526
factor(SNP1.SOB) 2	-0.1527	0.3903	0.1529	0.6957	0.8584	0.3994	1.8448
factor(tmp.yidx) 2	0.2617	0.0979	7.1427	0.0075	1.2992	1.0723	1.5741
factor(tmp.yidx) 3	-0.0916	0.1503	0.3714	0.5422	0.9125	0.6796	1.2251
factor(tmp.yidx) 4	0.9549	0.1406	46.1291	0.0000	2.5983	1.9725	3.4227
AGE	0.0322	0.0052	38.8722	0.0000	1.0327	1.0223	1.0432
SEX	-0.1172	0.2041	0.3296	0.5659	0.8894	0.5961	1.3270
BMI	-0.0227	0.0351	0.4161	0.5189	0.9776	0.9125	1.0473
SMOKE	0.3420	0.1789	3.6547	0.0559	1.4078	0.9914	1.9990
ALH	-0.2864	0.2158	1.7609	0.1845	0.7510	0.4919	1.1464
OCCU.NEW	-0.2709	0.1452	3.4797	0.0621	0.7627	0.5738	1.0138
factor(EDU.NEW) 2	0.0996	0.1829	0.2966	0.5860	1.1047	0.7719	1.5810
factor(EDU.NEW) 3	-0.2612	0.2412	1.1736	0.2787	0.7701	0.4800	1.2354

组数:	771
最大的组内观察数:	4
组内相关性类型:	independence
使用的总观察数:	3084

对数似然比检验比较两回归模型(未使用 GEE, 使用 GLM):

Model 1 用不同回归系数拟合危险因素对各应变量的作用:

tmp.y ~

factor(SNP1.COUGH)+factor(SNP1.PHLEGM)+factor(SNP1.WHEEZE)+factor(SNP1.SOB)+factor(tmp.yidx)+AGE+SEX+BMI+SMOKE+ALH+OCCU.NEW+factor(EDU.NEW)

Model 2 用同一回归系数拟合危险因素对各应变量的作用:

tmp.y ~ factor(SNP1)+factor(tmp.yidx)+AGE+SEX+BMI+SMOKE+ALH+OCCU.NEW+factor(EDU.NEW)

	Log likelihood	df.	X-square	P.value
Model 1	-1340.83471218796	20		
Model 2	-1343.31221690092	14	4.95500942591116	0.549596351589834

广义估计方程多应变量回归 (同一回归系数拟合危险因素对每个应变量的作用)

	Estimate	Std.err	Wald	Pr(> W)	exp(coef)	95%区间下限	95%区间上限
(Intercept)	-2.1767	0.9090	5.7343	0.0166	0.1134	0.0191	0.6736
factor(SNP1) 1	-0.0888	0.1503	0.3490	0.5547	0.9150	0.6815	1.2285
factor(SNP1) 2	0.0764	0.3123	0.0599	0.8067	1.0794	0.5853	1.9908
factor(tmp.yidx) 2	0.4030	0.0873	21.3349	0.0000	1.4963	1.2611	1.7754
factor(tmp.yidx) 3	-0.0850	0.1259	0.4560	0.4995	0.9185	0.7176	1.1756

factor(tmp.yidx) 4	1.0116	0.1147	77.8279	0.0000	2.7501	2.1965	3.4432
AGE	0.0321	0.0052	38.8914	0.0000	1.0327	1.0223	1.0431
SEX	-0.1168	0.2038	0.3285	0.5665	0.8898	0.5968	1.3265
BMI	-0.0226	0.0351	0.4172	0.5183	0.9776	0.9127	1.0472
SMOKE	0.3420	0.1784	3.6743	0.0553	1.4078	0.9923	1.9971
ALH	-0.2865	0.2155	1.7680	0.1836	0.7509	0.4922	1.1455
OCCU.NEW	-0.2711	0.1450	3.4939	0.0616	0.7626	0.5739	1.0133
factor(EDU.NEW) 2	0.0998	0.1826	0.2988	0.5846	1.1050	0.7725	1.5806
factor(EDU.NEW) 3	-0.2602	0.2408	1.1682	0.2798	0.7709	0.4809	1.2357

组数:	771
最大的组内观察数:	4
组内相关性类型:	independence
使用的总观察数:	3084

