simulation to study strategies to handle small strata

Wei Zou

2023-12-19 10:13:59

this program compare stratified vs unstratified analyses

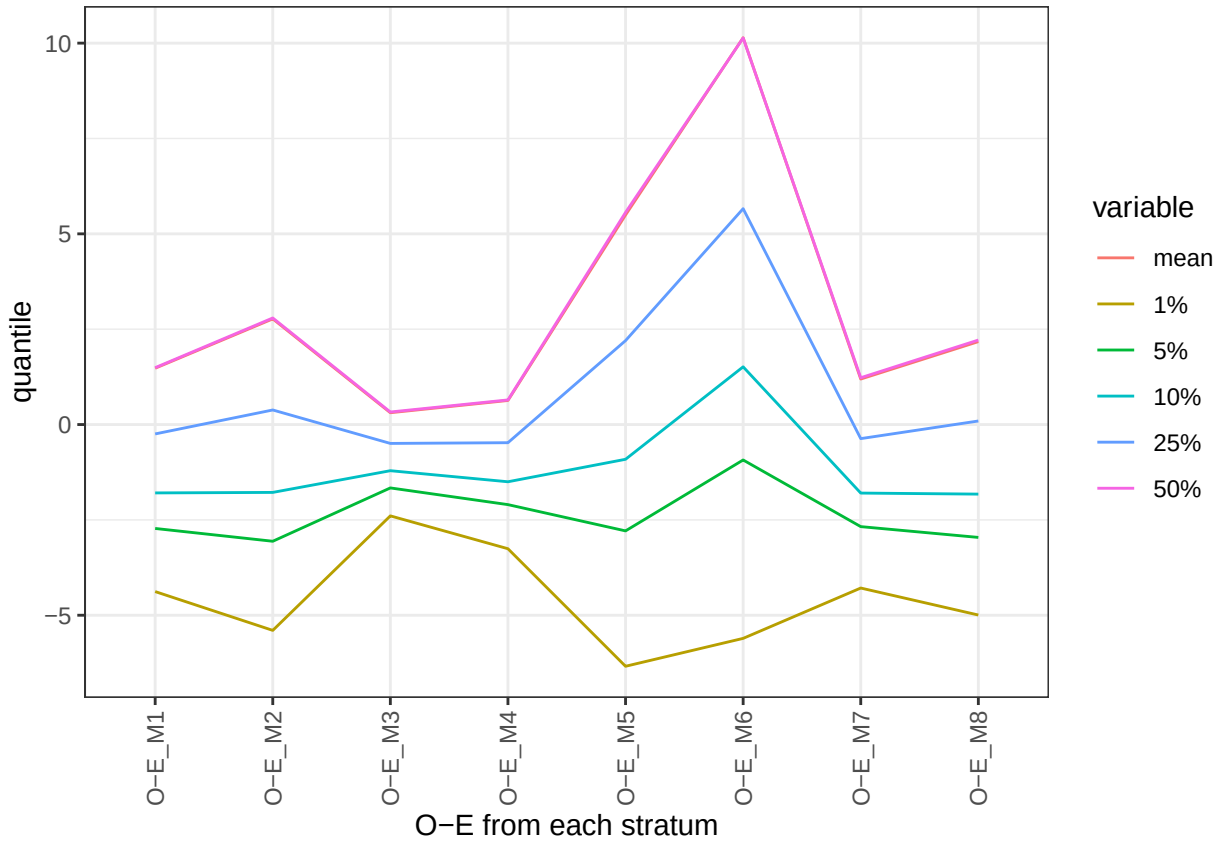
simulate imp150 data

```
## [1] "finding 972 input files when looking for r1"
## [1] "modification interval: 9.1 min"
## [1] "will add file name to the returned data"
```

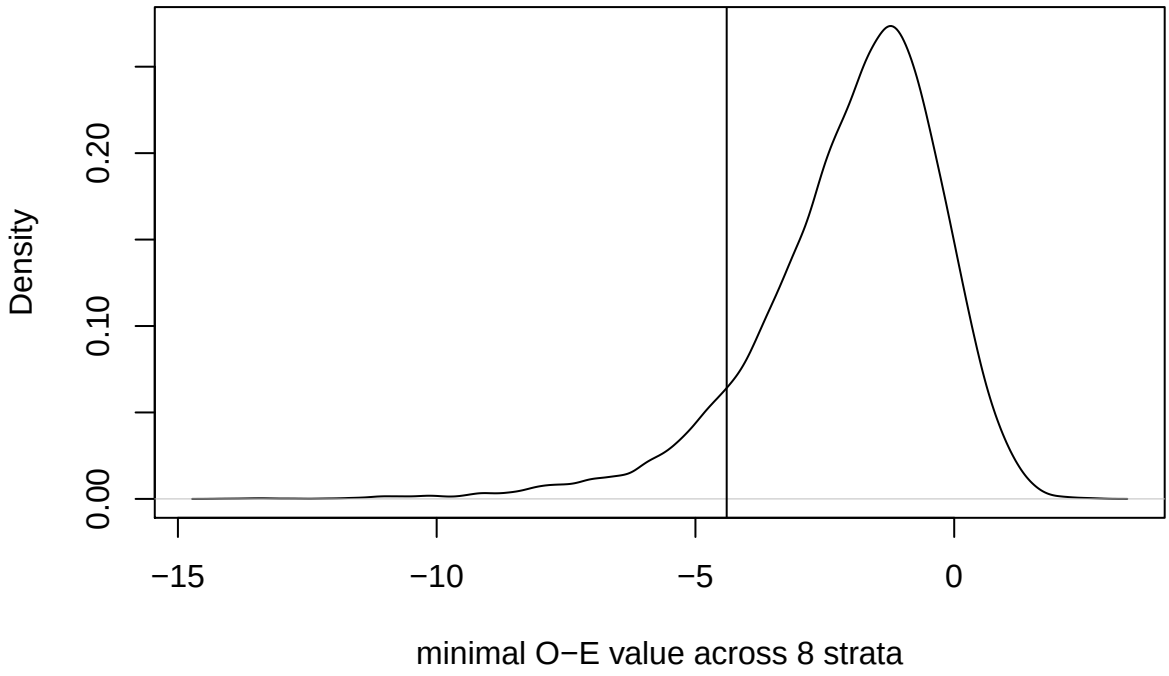
var	n	mean	sd	min	25%	50%	75%	max
n	19440	702.000	0.000	702.000	702.000	702.00	702.000	702.000
nevent	19440	438.669	12.639	392.000	430.000	439.00	447.000	487.000
p_s	19440	0.095	0.173	0.000	0.003	0.02	0.097	1.000
delta_p	19440	0.005	0.042	-0.431	-0.002	0.00	0.009	0.409

var	n	mean	sd	1%	5%	10%	25%	50%
O-E_M1	19440	1.476	2.545	-4.379	-2.725	-1.793	-0.246	1.491
O-E_M2	19440	2.770	3.522	-5.397	-3.060	-1.779	0.382	2.794
O-E_M3	19440	0.310	1.161	-2.395	-1.662	-1.210	-0.495	0.329
O-E_M4	19440	0.631	1.642	-3.256	-2.100	-1.500	-0.476	0.645
O-E_M5	19440	5.494	4.962	-6.338	-2.788	-0.911	2.201	5.562
O-E_M6	19440	10.143	6.710	-5.607	-0.929	1.513	5.660	10.135
O-E_M7	19440	1.193	2.311	-4.286	-2.676	-1.796	-0.369	1.227
O-E_M8	19440	2.173	3.069	-4.994	-2.959	-1.824	0.093	2.215

```
## Using var as id variables
```



prob of seeing O−E < −4.396: 0.09



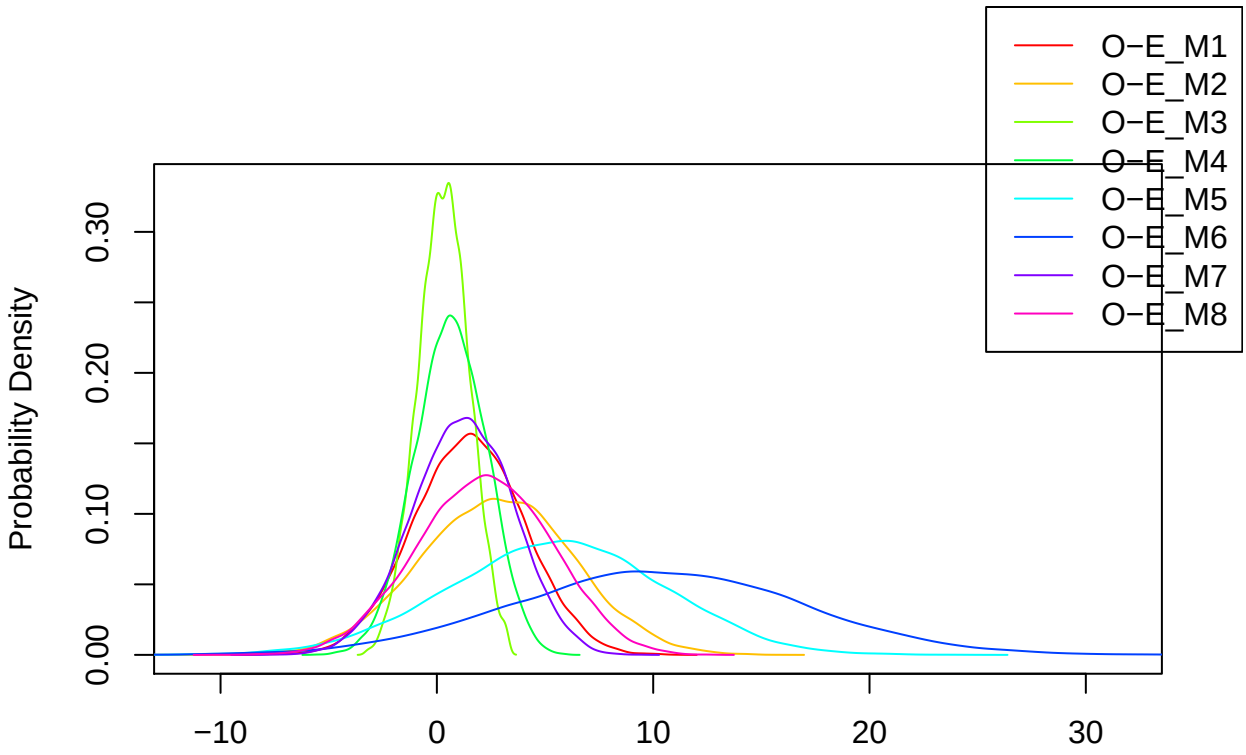
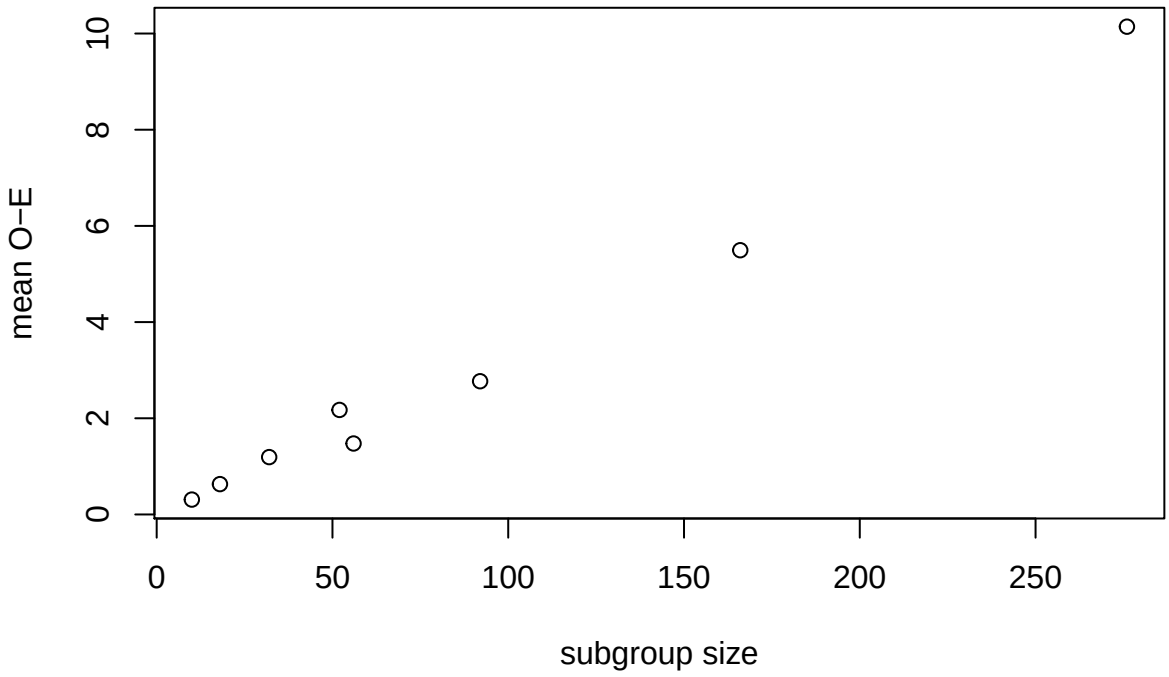


Table 3: parameters derived from imp150, used in simulation

	subgroup size	outcome medians	mean O-E
M1+M2+M3+	56	23.2	1.5
M1-M2+M3+	92	19.0	2.8
M1+M2-M3+	10	16.2	0.3
M1-M2-M3+	18	13.3	0.6
M1+M2+M3-	166	16.2	5.5
M1-M2+M3-	276	13.3	10.1
M1+M2-M3-	32	11.4	1.2
M1-M2-M3-	52	9.3	2.2



requiring 10 events per stratum

collapse small groups

```
## [1] "finding 100 input files when looking for r1"
## [1] "modification interval: 0.8 min"
## [1] "will add file name to the returned data"
```

var	n	mean	sd	min	25%	50%	75%	max
n	2000	702.000	0.000	702.000	702.000	702.000	702.000	702.000
nevent	2000	390.250	12.896	347.000	381.000	390.000	399.000	431.000
p_s	2000	0.114	0.192	0.000	0.004	0.027	0.129	0.996
p_c	2000	0.113	0.191	0.000	0.004	0.027	0.127	0.998
n_strata_c	2000	6.990	0.118	6.000	7.000	7.000	7.000	8.000
delta_p	2000	0.006	0.045	-0.385	-0.002	0.001	0.012	0.308
delta_p2	2000	0.000	0.010	-0.140	-0.001	0.000	0.001	0.108

remove the smallest stratum

```
## [1] "finding 100 input files when looking for r1"
## [1] "modification interval: 0.7 min"
## [1] "will add file name to the returned data"
```

var	n	mean	sd	min	25%	50%	75%	max
n	2000	702.000	0.000	702.000	702.000	702.000	702.000	702.000
nevent	2000	390.250	12.896	347.000	381.000	390.000	399.000	431.000
p_s	2000	0.114	0.192	0.000	0.004	0.027	0.129	0.996
p_c	2000	0.116	0.193	0.000	0.004	0.029	0.134	1.000

var	n	mean	sd	min	25%	50%	75%	max
n_strata_c	2000	4.000	0.000	4.000	4.000	4.000	4.000	4.000
delta_p	2000	0.006	0.045	-0.385	-0.002	0.001	0.012	0.308
delta_p2	2000	0.002	0.032	-0.253	-0.003	0.000	0.007	0.270

requiring 30 events per stratum

```
## [1] "finding 100 input files when looking for r1"
## [1] "modification interval: 0.7 min"
## [1] "will add file name to the returned data"
```

var	n	mean	sd	min	25%	50%	75%	max
n	2000	702.000	0.000	702.000	702.000	702.000	702.000	702.000
nevent	2000	390.250	12.896	347.000	381.000	390.000	399.000	431.000
p_s	2000	0.114	0.192	0.000	0.004	0.027	0.129	0.996
p_c	2000	0.115	0.193	0.000	0.004	0.028	0.129	1.000
n_strata_c	2000	5.029	0.208	4.000	5.000	5.000	5.000	6.000
delta_p	2000	0.006	0.045	-0.385	-0.002	0.001	0.012	0.308
delta_p2	2000	0.001	0.023	-0.236	-0.002	0.000	0.004	0.233