

INPower Package

October 27, 2020

Introduction

An R package to estimate the number of susceptibility loci and the distribution of their effect sizes for a trait on the basis of discoveries from existing genome-wide association studies (GWASs).

```
> library(INPower)
```

Height example

Get the path to the data which contains a known set of susceptibility SNPs for height.

```
> datafile <- system.file("sampleData", "data.rda", package="INPower")
```

Load and print the data frame.

```
> load(datafile)
> data
```

	SNP	MAF	Beta	Power
1	rs314277	0.13	0.03538462	0.01005508
2	rs4713858	0.14	-0.03900000	0.02774654
3	rs12198986	0.50	0.03300000	0.12273508
4	rs1492820	0.48	-0.03353846	0.13659137
5	rs10946808	0.28	-0.04830769	0.57978559
6	rs2274432	0.37	0.04500000	0.58359100
7	rs678962	0.22	0.05400000	0.65003547
8	rs8007661	0.30	-0.05015385	0.70992649
9	rs12735613	0.24	-0.05500000	0.75566530
10	rs724016	0.48	0.04723077	0.76483843
11	rs1812175	0.14	-0.06900000	0.79323028
12	rs4800148	0.21	-0.05900000	0.80022419
13	rs4896582	0.27	-0.05615385	0.86294616
14	rs11205277	0.44	0.05400000	0.94813202
15	rs4549631	0.50	0.05500000	0.96621958
16	rs7153027	0.48	-0.05700000	0.98270897
17	rs3760318	0.37	-0.06000000	0.98802127
18	rs2844479	0.33	-0.06300000	0.99289585

```

19 rs6060369 0.36 0.06307692 0.99593942
20 rs3791675 0.23 0.07500000 0.99879522
21 rs6830062 0.11 -0.10500000 0.99969644
22 rs2282978 0.31 0.07150000 0.99976245
23 rs2814993 0.15 0.09500000 0.99991429
24 rs1042725 0.50 0.06792308 0.99991841
25 rs798544 0.28 -0.07700000 0.99996252
26 rs8756 0.48 -0.07200000 0.99999452
27 rs6060373 0.38 -0.07500000 0.99999710
28 rs3748069 0.26 -0.08300000 0.99999711
29 rs6440003 0.44 0.09500000 1.00000000
30 rs6763931 0.45 0.13500000 1.00000000

```

For each known susceptibility SNP, we will need the MAF, effect size and power of detection.

```

> MAFs <- data[, "MAF"]
> betas <- data[, "Beta"]
> pow <- data[, "Power"]

```

Suppose that a one-stage design with a genome-wide significant level of $1e-7$ is considered for a future study. It is known that the heritability of height is 0.8. It is of interest to predict the expected number of discoveries with sample sizes from 25,000 to 125,000 with an increment of 25,000. In addition, for the given sample sizes, it is also of interest to find power to detect at least k loci, with k ranging from 25 to 125 with an increment of 25.

```

> INPower(MAFs, betas, pow, span=0.5, binary.outcome=FALSE,
+         sample.size=seq(25000,125000,by=25000),
+         signif.lvl=10^(-7), tgv=0.8, k=seq(25,125,by=25))

```

```

[1] FALSE
$esdist.summary
$esdist.summary$t.n.loci
[1] 200.8

```

```

$esdist.summary$gve
[1] "16.39% of the total genetic variance"

```

```

$esdist.summary$es.dist
      MAFs      betas      pow effect.size est.num.loci
1  0.13  0.03538462 0.01005508 0.0002832185         54.0
2  0.14 -0.03900000 0.02774654 0.0003662568         47.3
3  0.50  0.03300000 0.12273508 0.0005445000         33.4
4  0.48 -0.03353846 0.13659137 0.0005615143         32.2
5  0.28 -0.04830769 0.57978559 0.0009409209          3.1
6  0.37  0.04500000 0.58359100 0.0009440550          3.0
7  0.22  0.05400000 0.65003547 0.0010007712          2.3
8  0.30 -0.05015385 0.70992649 0.0010564715          1.7
9  0.24 -0.05500000 0.75566530 0.0011035200          1.4
10 0.48  0.04723077 0.76483843 0.0011135882          1.4

```

11	0.14	-0.06900000	0.79323028	0.0011464488	1.3
12	0.21	-0.05900000	0.80022419	0.0011549958	1.3
13	0.27	-0.05615385	0.86294616	0.0012430129	1.2
14	0.44	0.05400000	0.94813202	0.0014370048	1.1
15	0.50	0.05500000	0.96621958	0.0015125000	1.1
16	0.48	-0.05700000	0.98270897	0.0016219008	1.0
17	0.37	-0.06000000	0.98802127	0.0016783200	1.0
18	0.33	-0.06300000	0.99289585	0.0017550918	1.0
19	0.36	0.06307692	0.99593942	0.0018333841	1.0
20	0.23	0.07500000	0.99879522	0.0019923750	1.0
21	0.11	-0.10500000	0.99969644	0.0021586950	1.0
22	0.31	0.07150000	0.99976245	0.0021870205	1.0
23	0.15	0.09500000	0.99991429	0.0023013750	1.0
24	0.50	0.06792308	0.99991841	0.0023067722	1.0
25	0.28	-0.07700000	0.99996252	0.0023905728	1.0
26	0.48	-0.07200000	0.99999452	0.0025878528	1.0
27	0.38	-0.07500000	0.99999710	0.0026505000	1.0
28	0.26	-0.08300000	0.99999711	0.0026508872	1.0
29	0.44	0.09500000	1.00000000	0.0044475200	1.0
30	0.45	0.13500000	1.00000000	0.0090213750	1.0

```
$future.study.summary
$future.study.summary$e.discov
  sample.size e.discov
1      25000      27.3
2      50000      74.4
3      75000     125.6
4     100000     161.5
5     125000     182.8
```

```
$future.study.summary$e.gve
  sample.size          e.gve
1      25000 6.58% of the total genetic variance
2      50000 10.28% of the total genetic variance
3      75000 13.17% of the total genetic variance
4     100000 14.85% of the total genetic variance
5     125000 15.71% of the total genetic variance
```

```
$future.study.summary$prob.k
      k Pr(X>=k).n=25000      k Pr(X>=k).n=50000      k Pr(X>=k).n=75000      k
1    24              0.90  25              1.00  25              1.00  25
2    25              0.82  50              1.00  50              1.00  50
3    25              0.82  68              0.91  75              1.00  75
4    26              0.72  70              0.83  100             1.00  100
5    27              0.59  72              0.71  119             0.90  125
6    28              0.46  74              0.57  121             0.82  155
7    28              0.46  75              0.49  123             0.71  158
8    29              0.34  75              0.49  125             0.57  159
9    30              0.23  76              0.42  125             0.57  161
```

10	32	0.09	78	0.28	126	0.50	162
11	50	0.00	79	0.22	127	0.43	163
12	75	0.00	82	0.09	129	0.30	165
13	100	0.00	100	0.00	131	0.19	166
14	125	0.00	125	0.00	133	0.11	168

	Pr(X>=k) .n=1e+05	k	Pr(X>=k) .n=125000
1	1.00	25	1.00
2	1.00	50	1.00
3	1.00	75	1.00
4	1.00	100	1.00
5	1.00	125	1.00
6	0.92	178	0.90
7	0.78	180	0.79
8	0.72	181	0.72
9	0.58	182	0.63
10	0.50	183	0.53
11	0.42	184	0.43
12	0.27	185	0.34
13	0.21	187	0.18
14	0.11	188	0.12

The function call below shows the same results as the above, but without the `tg` (total genetic variance) argument. As a result, the genetic variance explained is expressed as a percentage of the total variance of the outcome, not of the total genetic variance.

```
> INPower(MAFs, betas, pow, span=0.5, binary.outcome=FALSE,
+         sample.size=seq(25000,125000,by=25000),
+         signif.lvl=10^(-7), k=seq(25,125,by=25))
```

```
[1] FALSE
```

```
$esdist.summary
```

```
$esdist.summary$t.n.loci
```

```
[1] 200.8
```

```
$esdist.summary$gve
```

```
[1] "13.11% of the total variance of the outcome"
```

```
$esdist.summary$es.dist
```

	MAFs	betas	pow	effect.size	est.num.loci
1	0.13	0.03538462	0.01005508	0.0002832185	54.0
2	0.14	-0.03900000	0.02774654	0.0003662568	47.3
3	0.50	0.03300000	0.12273508	0.0005445000	33.4
4	0.48	-0.03353846	0.13659137	0.0005615143	32.2
5	0.28	-0.04830769	0.57978559	0.0009409209	3.1
6	0.37	0.04500000	0.58359100	0.0009440550	3.0
7	0.22	0.05400000	0.65003547	0.0010007712	2.3
8	0.30	-0.05015385	0.70992649	0.0010564715	1.7
9	0.24	-0.05500000	0.75566530	0.0011035200	1.4
10	0.48	0.04723077	0.76483843	0.0011135882	1.4
11	0.14	-0.06900000	0.79323028	0.0011464488	1.3

12	0.21	-0.05900000	0.80022419	0.0011549958	1.3
13	0.27	-0.05615385	0.86294616	0.0012430129	1.2
14	0.44	0.05400000	0.94813202	0.0014370048	1.1
15	0.50	0.05500000	0.96621958	0.0015125000	1.1
16	0.48	-0.05700000	0.98270897	0.0016219008	1.0
17	0.37	-0.06000000	0.98802127	0.0016783200	1.0
18	0.33	-0.06300000	0.99289585	0.0017550918	1.0
19	0.36	0.06307692	0.99593942	0.0018333841	1.0
20	0.23	0.07500000	0.99879522	0.0019923750	1.0
21	0.11	-0.10500000	0.99969644	0.0021586950	1.0
22	0.31	0.07150000	0.99976245	0.0021870205	1.0
23	0.15	0.09500000	0.99991429	0.0023013750	1.0
24	0.50	0.06792308	0.99991841	0.0023067722	1.0
25	0.28	-0.07700000	0.99996252	0.0023905728	1.0
26	0.48	-0.07200000	0.99999452	0.0025878528	1.0
27	0.38	-0.07500000	0.99999710	0.0026505000	1.0
28	0.26	-0.08300000	0.99999711	0.0026508872	1.0
29	0.44	0.09500000	1.00000000	0.0044475200	1.0
30	0.45	0.13500000	1.00000000	0.0090213750	1.0

```
$future.study.summary
$future.study.summary$e.discov
```

	sample.size	e.discov
1	25000	27.3
2	50000	74.4
3	75000	125.6
4	100000	161.5
5	125000	182.8

```
$future.study.summary$e.gve
```

	sample.size	e.gve
1	25000	5.26% of the total variance of the outcome
2	50000	8.23% of the total variance of the outcome
3	75000	10.54% of the total variance of the outcome
4	100000	11.88% of the total variance of the outcome
5	125000	12.57% of the total variance of the outcome

```
$future.study.summary$prob.k
```

	k	Pr(X>=k).n=25000	k	Pr(X>=k).n=50000	k	Pr(X>=k).n=75000	k
1	24	0.90	25	1.00	25	1.00	25
2	25	0.82	50	1.00	50	1.00	50
3	25	0.82	68	0.91	75	1.00	75
4	26	0.72	70	0.83	100	1.00	100
5	27	0.59	72	0.71	119	0.90	125
6	28	0.46	74	0.57	121	0.82	155
7	28	0.46	75	0.49	123	0.71	158
8	29	0.34	75	0.49	125	0.57	159
9	30	0.23	76	0.42	125	0.57	161
10	32	0.09	78	0.28	126	0.50	162

11	50	0.00	79	0.22	127	0.43	163
12	75	0.00	82	0.09	129	0.30	165
13	100	0.00	100	0.00	131	0.19	166
14	125	0.00	125	0.00	133	0.11	168

	Pr(X>=k) .n=1e+05	k	Pr(X>=k) .n=125000
1	1.00	25	1.00
2	1.00	50	1.00
3	1.00	75	1.00
4	1.00	100	1.00
5	1.00	125	1.00
6	0.92	178	0.90
7	0.78	180	0.79
8	0.72	181	0.72
9	0.58	182	0.63
10	0.50	183	0.53
11	0.42	184	0.43
12	0.27	185	0.34
13	0.21	187	0.18
14	0.11	188	0.12

Now a two-stage study is considered with all the other conditions remaining the same (including the estimate of total heritability). In addition, the selection criterion for SNPs taken toward the second stage is $5e-5$ and 30 assigned to the first stage (and hence 70)

```
> INPower(MAFs, betas, pow, span=0.5, binary.outcome=FALSE,
+         sample.size=seq(25000,125000,by=25000),
+         signif.lvl=10^(-7), multi.stage.option=list(al=5*10^(-5), pi=0.3),
+         tgv=0.8 , k=seq(25,125,by=25))
```

```
$esdist.summary
$esdist.summary$t.n.loci
[1] 200.8
```

```
$esdist.summary$gve
[1] "16.39% of the total genetic variance"
```

```
$esdist.summary$es.dist
```

	MAFs	betas	pow	effect.size	est.num.loci
1	0.13	0.03538462	0.01005508	0.0002832185	54.0
2	0.14	-0.03900000	0.02774654	0.0003662568	47.3
3	0.50	0.03300000	0.12273508	0.0005445000	33.4
4	0.48	-0.03353846	0.13659137	0.0005615143	32.2
5	0.28	-0.04830769	0.57978559	0.0009409209	3.1
6	0.37	0.04500000	0.58359100	0.0009440550	3.0
7	0.22	0.05400000	0.65003547	0.0010007712	2.3
8	0.30	-0.05015385	0.70992649	0.0010564715	1.7
9	0.24	-0.05500000	0.75566530	0.0011035200	1.4
10	0.48	0.04723077	0.76483843	0.0011135882	1.4
11	0.14	-0.06900000	0.79323028	0.0011464488	1.3
12	0.21	-0.05900000	0.80022419	0.0011549958	1.3

13	0.27	-0.05615385	0.86294616	0.0012430129	1.2
14	0.44	0.05400000	0.94813202	0.0014370048	1.1
15	0.50	0.05500000	0.96621958	0.0015125000	1.1
16	0.48	-0.05700000	0.98270897	0.0016219008	1.0
17	0.37	-0.06000000	0.98802127	0.0016783200	1.0
18	0.33	-0.06300000	0.99289585	0.0017550918	1.0
19	0.36	0.06307692	0.99593942	0.0018333841	1.0
20	0.23	0.07500000	0.99879522	0.0019923750	1.0
21	0.11	-0.10500000	0.99969644	0.0021586950	1.0
22	0.31	0.07150000	0.99976245	0.0021870205	1.0
23	0.15	0.09500000	0.99991429	0.0023013750	1.0
24	0.50	0.06792308	0.99991841	0.0023067722	1.0
25	0.28	-0.07700000	0.99996252	0.0023905728	1.0
26	0.48	-0.07200000	0.99999452	0.0025878528	1.0
27	0.38	-0.07500000	0.99999710	0.0026505000	1.0
28	0.26	-0.08300000	0.99999711	0.0026508872	1.0
29	0.44	0.09500000	1.00000000	0.0044475200	1.0
30	0.45	0.13500000	1.00000000	0.0090213750	1.0

```
$future.study.summary
$future.study.summary$e.discov
```

	sample.size	e.discov
1	25000	10.8
2	50000	31.5
3	75000	56.0
4	100000	82.1
5	125000	106.7

```
$future.study.summary$e.gve
```

	sample.size	e.gve
1	25000	3.75% of the total genetic variance
2	50000	6.81% of the total genetic variance
3	75000	8.89% of the total genetic variance
4	100000	10.61% of the total genetic variance
5	125000	12.04% of the total genetic variance

```
$future.study.summary$prob.k
```

	k	Pr(X>=k).n=25000	k	Pr(X>=k).n=50000	k	Pr(X>=k).n=75000	k
1	8	0.93	25	0.98	25	1.00	25
2	9	0.84	27	0.92	50	0.91	50
3	10	0.70	29	0.79	50	0.91	75
4	11	0.53	30	0.70	52	0.82	75
5	11	0.53	31	0.60	54	0.69	78
6	12	0.36	32	0.49	55	0.62	80
7	12	0.36	33	0.38	56	0.54	81
8	13	0.22	34	0.29	58	0.38	83
9	14	0.12	35	0.20	59	0.30	84
10	25	0.00	37	0.09	61	0.18	86
11	50	0.00	50	0.00	63	0.10	87

12	75	0.00	75	0.00	75	0.00	90
13	100	0.00	100	0.00	100	0.00	100
14	125	0.00	125	0.00	125	0.00	125

	Pr(X>=k) .n=1e+05	k	Pr(X>=k) .n=125000
1	1.00	25	1.00
2	1.00	50	1.00
3	0.91	75	1.00
4	0.91	100	0.89
5	0.79	100	0.89
6	0.67	102	0.81
7	0.60	104	0.70
8	0.47	106	0.58
9	0.40	107	0.51
10	0.27	109	0.38
11	0.22	110	0.32
12	0.10	112	0.21
13	0.00	115	0.10
14	0.00	125	0.00

Session Information

```
> sessionInfo()
```

```
R version 4.0.3 (2020-10-10)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: Windows Server 2012 R2 x64 (build 9600)
```

```
Matrix products: default
```

```
locale:
[1] LC_COLLATE=C
[2] LC_CTYPE=English_United States.1252
[3] LC_MONETARY=English_United States.1252
[4] LC_NUMERIC=C
[5] LC_TIME=English_United States.1252
```

```
attached base packages:
[1] stats      graphics  grDevices  utils      datasets  methods   base
```

```
other attached packages:
[1] INPower_1.26.0 mvtnorm_1.1-1
```

```
loaded via a namespace (and not attached):
[1] compiler_4.0.3 tools_4.0.3
```