

Stats HW2

Mateusz Grobelny

Q1

```
plink --file hapmap1 --mind 0.06 --maf 0.05 --geno 0.1 --hwe 0.0001 \
--noweb --recode --out qcd
```

How many samples were removed?

No samples were removed

How many SNPs are left?

58737 SNPs

Q2

```
plink --file qcd --logistic --pheno pop.phe --adjust --noweb --out nopc
```

What is the genomic inflation factor?

Genomic inflation factor (based on median chi-squared) is 1.63725

Report the OR of the SNP rs2222162.

```
grep "rs2222162" nopc.assoc.logistic
2   rs2222162      10602    1      ADD      89      1.673      1.713
0.08668
```

OR = 1.673

Is having more minor alleles of this SNP associated with higher or lower risk?

Increasing Odds ratio means there is a higher probability of something happening. Since this SNP has an OR of 1.673 that means that it is associated with higher risk

Q3

```
~/Downloads/EIG-6.1.4/bin/convertf -p par.PED.EIGENSTRAT
```

Top five lines of qcd.pca

```
$ head -n 5 qcd.pca
3
1.6190
1.1200
1.1120
0.1068 -0.0439 -0.0735
```

Q4

```
evvec <- read.table("qcd.pca.evec");
fid <- sapply(as.character(evvec$V1),function(x){ strsplit(x,":")[[1]][1]; });
iid <- sapply(as.character(evvec$V1),function(x){ strsplit(x,":")[[1]][2]; });
out <- data.frame(FID=fid,IID=iid,
                  PC1=evvec$V2,
                  PC2=evvec$V3,
                  PC3=evvec$V4);
write.table(out,file="pcs.txt",row.names=FALSE,quote=FALSE);
```

Rerun GWAS with PCA

```
plink --file qcd --logistic --pheno pop.phe --adjust --covar pcs.txt --noweb \
--covar-name PC1 --out pc1
```

```
$ grep "rs2222162" pc1.assoc.logistic | grep "ADD"
2    rs2222162    10602    1          ADD          89    0.01695    -4.512
6.407e-06
```

OR = 0.01695

Odds ratio decreased after controlling for pop stratification. The rs2222162 is actually associated with lower risk.

Q5

Sort by Unadjusted

```
pc1 <- read.table("pc1.assoc.logistic.adjust",header=TRUE);

> head(pc1[order(pc1$UNADJ),]);
1    2    rs2222162 6.407e-06 8.243e-06 0.3763 0.3763    0.3136    0.3136 0.3763
```

Sort by BONF

```
> head(pcl[order(pcl$BONF),]);
```

	CHR	SNP	UNADJ	GC	BONF	HOLM	SIDAK_SS	SIDAK_SD	FDR_BH
1	2	rs2222162	6.407e-06	8.243e-06	0.3763	0.3763	0.3136	0.3136	0.3763
2	9	rs7046471	1.747e-04	2.086e-04	1.0000	1.0000	1.0000	1.0000	0.9884
3	13	rs9585021	1.927e-04	2.296e-04	1.0000	1.0000	1.0000	1.0000	0.9884
4	2	rs4675607	1.938e-04	2.309e-04	1.0000	1.0000	1.0000	1.0000	0.9884
5	8	rs2395989	2.985e-04	3.522e-04	1.0000	1.0000	1.0000	1.0000	0.9884
6	8	rs10505549	3.099e-04	3.653e-04	1.0000	1.0000	1.0000	1.0000	0.9884

Bonferroni threshold = $0.05/58737 = \sim 8.512 \times 10^{-7}$

No SNP was found to be pass the Bonferroni threshold.

Q6

On Grobelny_manhattan_plot.pdf