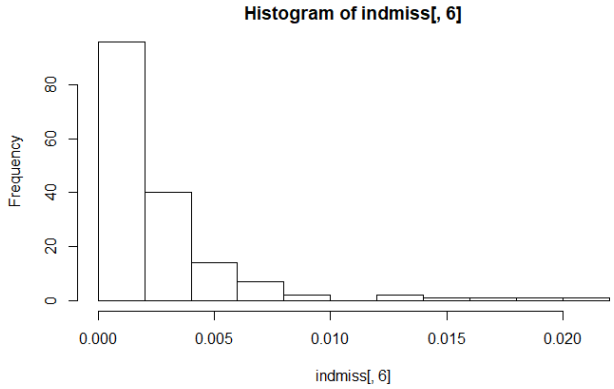
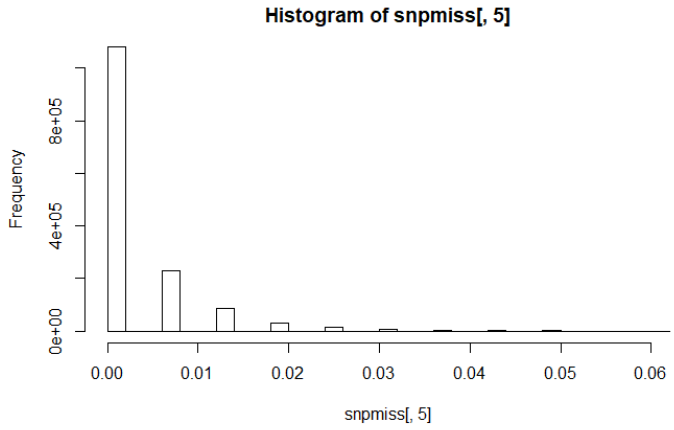
Ćwiczenia- Biologia molekularna z genetyką II

Zad 1

Histogram z kolumny 6, która zawiera informacje o tym jakiego odsetka SNP brakuje u danej osoby



Histogram z kolumny 5 zawierającej informacje o tym u jakiego odsetka ludzi brakuje danego SNP



Po odrzuceniu SNP nieoznaczonych u >20% osób

```
Reading map (extended format) from [ HapMap.bim ]
1457897 markers to be included from [ HapMap.bim ]
Reading pedigree information from [ HapMap.fam ]
165 individuals read from [ HapMap.fam ]
112 individuals with nonmissing phenotypes
Assuming a disease phenotype (1=unaff, 2=aff, 0=miss)
Missing phenotype value is also -9
56 cases, 56 controls and 53 missing
80 males, 85 females, and 0 of unspecified sex
Reading genotype bitfile from [ HapMap.bed ]
Detected that binary PED file is v1.00 SNP-major mode
Before frequency and genotyping pruning, there are 1457897 SNPs
112 founders and 53 non-founders found
225 heterozygous haploid genotypes; set to missing
Writing list of heterozygous haploid genotypes to [ HapMap_2.hh ]
Total genotyping rate in remaining individuals is 0.997378
0 SNPs failed missingness test ( GENO > 0.2 )
0 SNPs failed frequency test ( MAF < 0 )
After frequency and genotyping pruning, there are 1457897 SNPs
After filtering, 56 cases, 56 controls and 53 missing
After filtering, 80 males, 85 females, and 0 of unspecified sex
Writing pedigree information to [ HapMap_2.fam ]
Writing map (extended format) information to [ HapMap_2.bim ]
Writing genotype bitfile to [ HapMap_2.bed ]
Using (default) SNP-major mode

Analysis finished: Mon Jun 08 22:24:00 2020
```

Żadne dane nie zostały odfiltrowane

Najpierw odrzucam SNP, których brakuje u więcej, niż 2% osób:

```
Reading map (extended format) from [ HapMap.bim ]
1457897 markers to be included from [ HapMap.bim ]
Reading pedigree information from [ HapMap.fam ]
165 individuals read from [ HapMap.fam ]
112 individuals with nonmissing phenotypes
Assuming a disease phenotype (1=unaff, 2=aff, 0=miss)
Missing phenotype value is also -9
56 cases, 56 controls and 53 missing
80 males, 85 females, and 0 of unspecified sex
Reading genotype bitfile from [ HapMap.bed ]
Detected that binary PED file is v1.00 SNP-major mode
Before frequency and genotyping pruning, there are 1457897 SNPs
112 founders and 53 non-founders found
225 heterozygous haploid genotypes; set to missing
Writing list of heterozygous haploid genotypes to [ HapMap_2_2.hh ]
Total genotyping rate in remaining individuals is 0.997378
27454 SNPs failed missingness test ( GENO > 0.02 )
0 SNPs failed frequency test ( MAF < 0 )
After frequency and genotyping pruning, there are 1430443 SNPs
After filtering, 56 cases, 56 controls and 53 missing
After filtering, 80 males, 85 females, and 0 of unspecified sex
Writing pedigree information to [ HapMap_2_2.fam ]
Writing map (extended format) information to [ HapMap_2_2.bim ]
Writing genotype bitfile to [ HapMap_2_2.bed ]
Using (default) SNP-major mode

Analysis finished: Mon Jun 08 22:40:57 2020
```

Odfiltrowano 27454 SNP, a pozostało 1430443

Z pozostałego pliku eliminuje te osoby, u których brakuje więcej, niż 2% SNP:

```
Reading map (extended format) from [ HapMap_2_2.bim ]
1430443 markers to be included from [ HapMap_2_2.bim ]
Reading pedigree information from [ HapMap_2_2.fam ]
165 individuals read from [ HapMap_2_2.fam ]
112 individuals with nonmissing phenotypes
Assuming a disease phenotype (1=unaff, 2=aff, 0=miss)
Missing phenotype value is also -9
56 cases, 56 controls and 53 missing
80 males, 85 females, and 0 of unspecified sex
Reading genotype bitfile from [ HapMap_2_2.bed ]
Detected that binary PED file is v1.00 SNP-major mode
Before frequency and genotyping pruning, there are 1430443 SNPs
112 founders and 53 non-founders found
0 of 165 individuals removed for low genotyping ( MIND > 0.02 )
179 heterozygous haploid genotypes; set to missing
Writing list of heterozygous haploid genotypes to [ HapMap_3.hh ]
Total genotyping rate in remaining individuals is 0.997899
0 SNPs failed missingness test ( GENO > 1 )
0 SNPs failed frequency test ( MAF < 0 )
After frequency and genotyping pruning, there are 1430443 SNPs
After filtering, 56 cases, 56 controls and 53 missing
After filtering, 80 males, 85 females, and 0 of unspecified sex
Writing pedigree information to [ HapMap_3.fam ]
Writing map (extended format) information to [ HapMap_3.bim ]
Writing genotype bitfile to [ HapMap_3.bed ]
Using (default) SNP-major mode

Analysis finished: Mon Jun 08 22:43:12 2020
```

Ta komenda nie usunęła żadnej osoby z danych

Filtrowanie przeprowadza się w tej kolejności, ponieważ lepiej jest najpierw usunąć rzadkie SNP, a dopiero potem ludzi, u których nie ma więcej niż 2% SNP. W innym wypadku usuwana byłaby większa liczba osób, bo mogliby nie mieć tych nie występują nieczęsto.

Zad 2

FID	IID	PEDSEX	SNPSEX	STATUS	F
1328	NA06989	2	2	OK	-0.01184
1377	NA11891	1	1	OK	1
1349	NA10854	2	1	PROBLEM	0.99

W tym pliku kolejne kolumny odpowiadają identyfikatorowi rodziny, identyfikatorowi osoby, płci przypisanej w pliku z rodowodem (.ped) i płci zdeterminowanej przez chromosom X. Kolejna kolumna precyzuje czy przypisane płcie się zgadzają (OK), czy są różne (PROBLEM). Ostatnia kolumna odpowiada estymowanej homozygotyczności chromosomu X.

Podjęłam decyzje o usunięciu osoby z problematycznie przypisaną płcią

Zad 3

Wybieram tylko autosomalne:

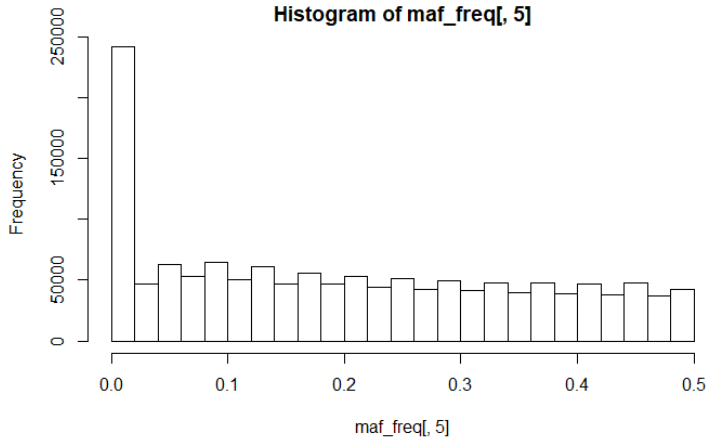
```
Reading map (extended format) from [ HapMap_4.bim ]
1430443 markers to be included from [ HapMap_4.bim ]
Reading pedigree information from [ HapMap_4.fam ]
164 individuals read from [ HapMap_4.fam ]
112 individuals with nonmissing phenotypes
Assuming a disease phenotype (1=unaff, 2=aff, 0=miss)
Missing phenotype value is also -9
56 cases, 56 controls and 52 missing
80 males, 84 females, and 0 of unspecified sex
Reading genotype bitfile from [ HapMap_4.bed ]
Detected that binary PED file is v1.00 SNP-major mode
Reading list of SNPs to extract [ autosomal.txt ] ... 1398544 read
Before frequency and genotyping pruning, there are 1398544 SNPs
112 founders and 52 non-founders found
Total genotyping rate in remaining individuals is 0.998052
0 SNPs failed missingness test ( GENO > 1 )
0 SNPs failed frequency test ( MAF < 0 )
After frequency and genotyping pruning, there are 1398544 SNPs
After filtering, 56 cases, 56 controls and 52 missing
After filtering, 80 males, 84 females, and 0 of unspecified sex
Writing pedigree information to [ HapMap_5.fam ]
Writing map (extended format) information to [ HapMap_5.bim ]
Writing genotype bitfile to [ HapMap_5.bed ]
Using (default) SNP-major mode

Analysis finished: Tue Jun 09 00:22:12 2020
```

Zad 4

Sprawdzam, jak wygląda rozkład częstości alleli:

```
zosia@DESKTOP-UG4OILP: /mnt/c/Users/Zosia/Downloads/plink_1.07-dos$ head MAF_check.frq
 CHR         SNP   A1   A2          MAF  NCHROBS
   1   rs2185539    T    C            0      224
   1  rs11240767    T    C            0      224
   1   rs3131972    A    G       0.1652      224
   1   rs3131969    A    G       0.1339      224
   1   rs1048488    C    T       0.1667      222
   1  rs12562034    A    G       0.1027      224
   1  rs12124819    G    A       0.2902      224
   1   rs4040617    G    A       0.1295      224
   1   rs2905036    C    T            0      224
```

Histogram rozkładu częstości allelu rzadkiego:

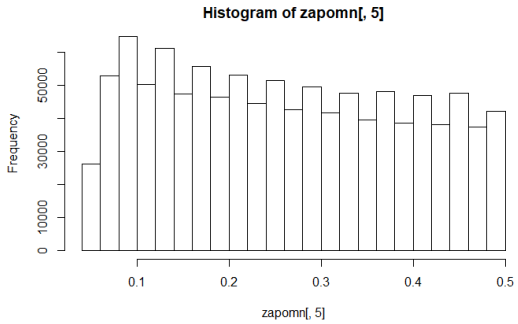


Usuwam wszystkie SNP, dla których MAF (częstość allelu rzadkiego) wynosi mniej niż 5%:

```
Reading map (extended format) from [ HapMap_5.bim ]
1398544 markers to be included from [ HapMap_5.bim ]
Reading pedigree information from [ HapMap_5.fam ]
164 individuals read from [ HapMap_5.fam ]
112 individuals with nonmissing phenotypes
Assuming a disease phenotype (1=unaff, 2=aff, 0=miss)
Missing phenotype value is also -9
56 cases, 56 controls and 52 missing
80 males, 84 females, and 0 of unspecified sex
Reading genotype bitfile from [ HapMap_5.bed ]
Detected that binary PED file is v1.00 SNP-major mode
Before frequency and genotyping pruning, there are 1398544 SNPs
112 founders and 52 non-founders found
Total genotyping rate in remaining individuals is 0.998052
0 SNPs failed missingness test ( GENO > 1 )
325318 SNPs failed frequency test ( MAF < 0.05 )
After frequency and genotyping pruning, there are 1073226 SNPs
After filtering, 56 cases, 56 controls and 52 missing
After filtering, 80 males, 84 females, and 0 of unspecified sex
Writing pedigree information to [ HapMap_6.fam ]
Writing map (extended format) information to [ HapMap_6.bim ]
Writing genotype bitfile to [ HapMap_6.bed ]
Using (default) SNP-major mode

Analysis finished: Tue Jun 09 00:29:30 2020
```

Usunięto 325318 SNP, a zostało ich 1073226

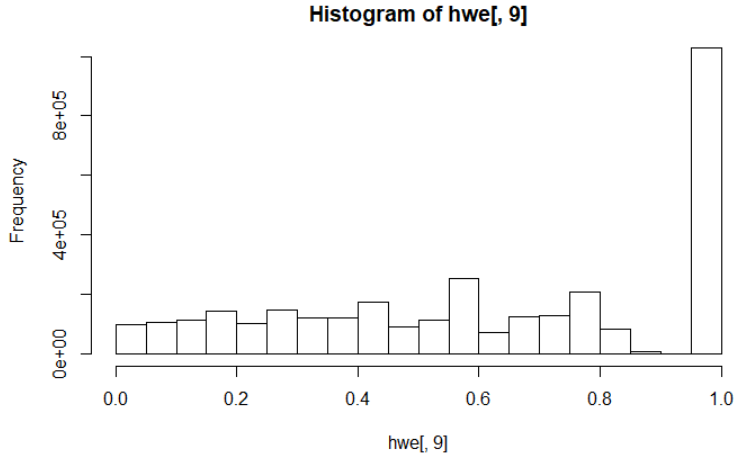
Histogram rozkładu częstości:



Zad 5

Sprawdzam równowagę Hardy'ego-Weinberga i robię histogram wartości p

```
zosia@DESKTOP-UG4OILP:/mnt/c/Users/zosia/Downloads/plink-1.07-dos$ head plink.hwe
 CHR         SNP       TEST   A1   A2                 GENO   O(HET)   E(HET)            P
   1   rs3131972        ALL    A    G              2/33/77   0.2946   0.2758       0.7324
   1   rs3131972        AFF    A    G              1/19/36   0.3393   0.3047        0.667
   1   rs3131972      UNAFF    A    G              1/14/41     0.25   0.2449            1
   1   rs3131969        ALL    A    G              2/26/84   0.2321    0.232            1
   1   rs3131969        AFF    A    G              1/17/38   0.3036   0.2817            1
   1   rs3131969      UNAFF    A    G               1/9/46   0.1607   0.1771       0.4189
   1   rs1048488        ALL    C    T              2/33/76   0.2973   0.2778       0.7324
   1   rs1048488        AFF    C    T              1/19/35   0.3455   0.3089       0.6661
   1   rs1048488      UNAFF    C    T              1/14/41     0.25   0.2449            1
```



Używam komendy awk by znaleźć SNP o wartości poniżej 10 ^-5

```
zosia@DESKTOP-UG4OILP:/mnt/c/Users/zosia/Downloads/plink-1.07-dos$ awk '{ if ($9 <1e-5) print $0 }' plink.hwe
   3    rs7623291        ALL    T    C            22/28/62     0.25   0.4362   8.938e-006
   7   rs34238522        ALL    C    T             0/64/48   0.5714   0.4082   3.515e-006
   8    rs3102841        ALL    C    A             8/78/23   0.7156   0.4905   1.899e-006
   9     rs354831        ALL    C    T            12/18/82   0.1607   0.3047   6.339e-006
   9   rs10990625        ALL    C    T            23/28/61     0.25   0.4424   9.391e-006
   9   rs10990625        AFF    C    T              15/8/33   0.1429   0.4483   3.574e-007
  10    rs2918624        ALL    C    T             0/62/50   0.5536   0.4004    8.54e-006
  10    rs4934139        ALL    C    A             0/65/47   0.5804   0.4119   1.722e-006
  12    rs2303632      UNAFF    T    G            15/10/31   0.1786   0.4592   4.934e-006
  12    rs7963063      UNAFF    C    T            15/10/31   0.1786   0.4592   4.934e-006
  13   rs17080881        ALL    T    G             5/75/32   0.6696   0.4709   6.863e-006
  13   rs10507731        ALL    G    A             5/75/32   0.6696   0.4709   6.863e-006
  19   rs12608717        ALL    G    A             0/66/45   0.5946   0.4178    1.41e-006
```

Powinno się usuwać SNP dla których P wynosi poniżej $1 \cdot 10^{-6}$ w grupie kontrolnej i poniżej $1 \cdot 10^{-10}$ w grupie chorych.

```
Options in effect:
        --bfile HapMap_6
        --hwe 1e-6
        --make-bed
        --out HapMap_7

Reading map (extended format) from [ HapMap_6.bim ]
1073226 markers to be included from [ HapMap_6.bim ]
Reading pedigree information from [ HapMap_6.fam ]
164 individuals read from [ HapMap_6.fam ]
112 individuals with nonmissing phenotypes
Assuming a disease phenotype (1=unaff, 2=aff, 0=miss)
Missing phenotype value is also -9
56 cases, 56 controls and 52 missing
80 males, 84 females, and 0 of unspecified sex
Reading genotype bitfile from [ HapMap_6.bed ]
Detected that binary PED file is v1.00 SNP-major mode
Before frequency and genotyping pruning, there are 1073226 SNPs
112 founders and 52 non-founders found
0 markers to be excluded based on HWE test ( p <= 1e-006 )
        1 markers failed HWE test in cases
        0 markers failed HWE test in controls
Total genotyping rate in remaining individuals is 0.998039
0 SNPs failed missingness test ( GENO > 1 )
0 SNPs failed frequency test ( MAF < 0 )
After frequency and genotyping pruning, there are 1073226 SNPs
After filtering, 56 cases, 56 controls and 52 missing
After filtering, 80 males, 84 females, and 0 of unspecified sex
Writing pedigree information to [ HapMap_7.fam ]
Writing map (extended format) information to [ HapMap_7.bim ]
Writing genotype bitfile to [ HapMap_7.bed ]
Using (default) SNP-major mode

Analysis finished: Tue Jun 09 00:39:07 2020
```

```
Options in effect:
        --bfile HapMap_7
        --hwe 1e-10
        --hwe-all
        --make-bed
        --out HapMap_8

Reading map (extended format) from [ HapMap_7.bim ]
1073226 markers to be included from [ HapMap_7.bim ]
Reading pedigree information from [ HapMap_7.fam ]
164 individuals read from [ HapMap_7.fam ]
112 individuals with nonmissing phenotypes
Assuming a disease phenotype (1=unaff, 2=aff, 0=miss)
Missing phenotype value is also -9
56 cases, 56 controls and 52 missing
80 males, 84 females, and 0 of unspecified sex
Reading genotype bitfile from [ HapMap_7.bed ]
Detected that binary PED file is v1.00 SNP-major mode
Before frequency and genotyping pruning, there are 1073226 SNPs
112 founders and 52 non-founders found
0 markers to be excluded based on HWE test ( p <= 1e-010 )
        0 markers failed HWE test in cases
        0 markers failed HWE test in controls
Total genotyping rate in remaining individuals is 0.998039
0 SNPs failed missingness test ( GENO > 1 )
0 SNPs failed frequency test ( MAF < 0 )
After frequency and genotyping pruning, there are 1073226 SNPs
After filtering, 56 cases, 56 controls and 52 missing
After filtering, 80 males, 84 females, and 0 of unspecified sex
Writing pedigree information to [ HapMap_8.fam ]
Writing map (extended format) information to [ HapMap_8.bim ]
Writing genotype bitfile to [ HapMap_8.bed ]
Using (default) SNP-major mode

Analysis finished: Tue Jun 09 00:41:20 2020
```

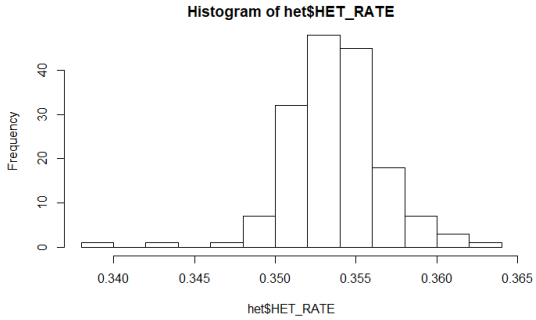
Obie komendy nie usunęły nic z danych, ma to związek z różnicą w punkcie odcięcia którą zastosowałam w komendzie awk (która znalazła dane) i przy filtrowaniu w programie plink

Zad 6

Robimy plik tylko z SNP niezależnymi od sb-wyłączymy obszary genomu, w których wiadomo o występowaniu nierównowagi sprzężeń.

```
For chromosome 16, 30132 SNPs pruned out, 3687 remaining
Scan region on chromosome 17 from [ rs8069278 ] to [ rs9897769 ]
Pruning SNPs 27791 to 27837 of 27837
For chromosome 17, 24467 SNPs pruned out, 3370 remaining
Scan region on chromosome 18 from [ rs8096071 ] to [ rs12960632 ]
Pruning SNPs 31606 to 31655 of 31655
For chromosome 18, 28237 SNPs pruned out, 3418 remaining
Scan region on chromosome 19 from [ rs8110113 ] to [ rs7910 ]
Pruning SNPs 19616 to 19662 of 19662
For chromosome 19, 16852 SNPs pruned out, 2810 remaining
Scan region on chromosome 20 from [ rs4814683 ] to [ rs6062363 ]
Pruning SNPs 27901 to 27948 of 27948
For chromosome 20, 24898 SNPs pruned out, 3050 remaining
Scan region on chromosome 21 from [ rs1736316 ] to [ rs35829851 ]
Pruning SNPs 15436 to 15482 of 15482
For chromosome 21, 13773 SNPs pruned out, 1709 remaining
Scan region on chromosome 22 from [ rs11089128 ] to [ rs28729663 ]
Pruning SNPs 15421 to 15468 of 15468
For chromosome 22, 13511 SNPs pruned out, 1957 remaining

Analysis finished: Tue Jun 09 00:50:30 2020
```

Zad 7 Analiza heterozygotyczności



W histogramie widać osoby o skrajnych wartościach heterozygotyczności. Maksymalna wartość wynosi 0.3621799, a minimalna 0.3387942

```
"FID" "IID" "O.HOM." "E.HOM." "N.NM." "F" "HET_RATE" "HET_DST"
1330 "NA12342" 68107 67290 103636 0.0226 0.342824887104867 -3.7033837588451
1459 "NA12874" 68854 67610 104134 0.034 0.338794245875507 -5.06222644998476
```

Plik z usuniętymi, wykorzystuję do odfiltrowania danych:

Zad 8 Analiza pokrewieństwa

Sprawdzam pokrewieństwo szacowane na podstawie alleli SNP dla wszystkich par osób w zbiorze danych:

```
** For gPLINK compatibility, do not use '.' in --out **
Reading map (extended format) from [ HapMap_9.bim ]
1073226 markers to be included from [ HapMap_9.bim ]
Reading pedigree information from [ HapMap_9.fam ]
164 individuals read from [ HapMap_9.fam ]
112 individuals with nonmissing phenotypes
Assuming a disease phenotype (1=unaff, 2=aff, 0=miss)
Missing phenotype value is also -9
56 cases, 56 controls and 52 missing
80 males, 84 females, and 0 of unspecified sex
Reading genotype bitfile from [ HapMap_9.bed ]
Detected that binary PED file is v1.00 SNP-major mode
Before frequency and genotyping pruning, there are 1073226 SNPs
112 founders and 52 non-founders found
Total genotyping rate in remaining individuals is 0.998039
0 SNPs failed missingness test ( GENO > 1 )
0 SNPs failed frequency test ( MAF < 0 )
After frequency and genotyping pruning, there are 1073226 SNPs
After filtering, 56 cases, 56 controls and 52 missing
After filtering, 80 males, 84 females, and 0 of unspecified sex
62 nuclear families, 10 founder singletons found
43 non-founders with 2 parents in 43 nuclear families
17 non-founders without 2 parents in 9 nuclear families
0 affected offspring trios
19 phenotypically discordant parent pairs found
Added 5 dummy parents
Converting data to Individual-major format
Writing whole genome IBS/IBD information to [ pihat_min0.2.genome ]
Filtering output to include pairs with ( 0.2 <= PI-HAT <= 1 )
IBD(g) calculation: 13300 of 13366

Analysis finished: Tue Jun 09 01:29:26 2020
```

FID1	IID1	FID2	IID2	RT	EZ	Z0	Z1	Z2	PI_HAT	PHE	DST	PPC	RATIO
1377	NA11891	1377	NA10865	PO	0.5	0.0015	0.9970	0.0015	0.5000	0	0.823332	1.0000	244.9444
1349	NA11843	1349	NA10853	PO	0.5	0.0019	0.9960	0.0021	0.5001	-1	0.823398	1.0000	265.0588
1330	NA12341	1330	NA12335	PO	0.5	0.0000	1.0000	0.0000	0.5000	0	0.822095	1.0000	638.5714
1444	NA12739	1444	NA12749	PO	0.5	0.0016	0.9905	0.0079	0.5032	-1	0.824460	1.0000	373.0833
1444	NA12739	1444	NA12748	PO	0.5	0.0014	0.9906	0.0080	0.5033	0	0.824491	1.0000	297.2000
1463	NA12877	1463	NA12890	PO	0.5	0.0000	1.0000	0.0000	0.5000	0	0.821923	1.0000	296.9333
1463	NA12877	1463	NA12889	PO	0.5	0.0119	0.9881	0.0000	0.4941	0	0.821810	1.0000	117.0789
1418	NA12275	1418	NA10836	PO	0.5	0.0021	0.9957	0.0021	0.5000	-1	0.823380	1.0000	322.6429
13291	NA06986	13291	NA06997	PO	0.5	0.0020	0.9929	0.0050	0.5015	-1	0.823909	1.0000	249.5000
1418	NA12272	1418	NA10837	PO	0.5	0.0014	0.9901	0.0086	0.5036	-1	0.824599	1.0000	748.3333
1424	NA10845	1424	NA11930	PO	0.5	0.0000	1.0000	0.0000	0.5000	0	0.822622	1.0000	750.1667
1424	NA10845	1424	NA11931	PO	0.5	0.0011	0.9989	0.0000	0.4994	-1	0.822994	1.0000	451.4000
1346	NA10852	1346	NA12045	PO	0.5	0.0017	0.9897	0.0086	0.5035	-1	0.824581	1.0000	233.3158
13292	NA07051	13292	NA07014	PO	0.5	0.0021	0.9853	0.0127	0.5053	0	0.825247	1.0000	247.8889
1354	NA12400	1354	NA12386	PO	0.5	0.0012	0.9927	0.0060	0.5024	0	0.824169	1.0000	341.9231
13281	NA12344	13281	NA12348	PO	0.5	0.0000	1.0000	0.0000	0.5000	0	0.820917	1.0000	444.3000
13281	NA12344	13281	NA12347	PO	0.5	0.0011	0.9915	0.0074	0.5032	0	0.824426	1.0000	747.3333
1451	NA12777	1451	NA12767	PO	0.5	0.0011	0.9937	0.0051	0.5020	0	0.824017	1.0000	341.7692
1421	NA12287	1421	NA10840	PO	0.5	0.0013	0.9940	0.0047	0.5017	-1	0.823934	1.0000	377.0833
1418	NA10837	1418	NA12273	PO	0.5	0.0000	1.0000	0.0000	0.5000	-1	0.822661	1.0000	639.2857
1353	NA12383	1353	NA12375	PO	0.5	0.0017	0.9893	0.0091	0.5037	0	0.824654	1.0000	279.6250
1330	NA12340	1330	NA12335	PO	0.5	0.0014	0.9962	0.0024	0.5005	0	0.823512	1.0000	375.0833
1358	NA12708	1358	NA12718	PO	0.5	0.0000	1.0000	0.0000	0.5000	-1	0.822156	1.0000	372.5833
1377	NA11892	1377	NA10865	PO	0.5	0.0000	1.0000	0.0000	0.5000	0	0.822967	1.0000	743.3333
1353	NA12546	1353	NA12376	PO	0.5	0.0011	0.9859	0.0130	0.5060	0	0.825418	1.0000	636.2857
1458	NA12843	1458	NA12832	PO	0.5	0.0031	0.9910	0.0059	0.5014	0	0.823943	1.0000	199.2927

Zostawiam tylko założycieli i powtarzam analizę pi-hat:

FID1	IID1	FID2	IID2	RT	EZ	Z0	Z1	Z2	PI_HAT	PHE	DST	PPC	RATIO
13291	NA07045	1454	NA12813	UN	NA	0.2745	0.5006	0.2250	0.4753	0	0.833661	1.0000	9.8218

Została jedna para o wartości pi-hat 0,4755- być może rodzeństwo

```
zosia@DESKTOP-UG4OILP:/mnt/c/Users/zosia/Downloads/plink-1.07-dos$ head plink.imiss
    FID       IID MISS_PHENO   N_MISS   N_GENO   F_MISS
   1328   NA06989          N     4203  1457897 0.002883
   1377   NA11891          N    20787  1457897  0.01426
   1349   NA11843          N     1564  1457897 0.001073
   1330   NA12341          N     6218  1457897 0.004265
   1444   NA12739          Y    29584  1457897  0.02029
   1344   NA10850          Y     2631  1457897 0.001805
   1328   NA06984          N     9639  1457897 0.006612
   1463   NA12877          Y     3789  1457897 0.002599
   1418   NA12275          N     5349  1457897 0.003669
```

```
   1418   NA12275          N     5349  1457897 0.003669
  13291   NA06986          N     1758  1457897 0.001206
   1454   NA12812          N     2563  1457897 0.001758
```

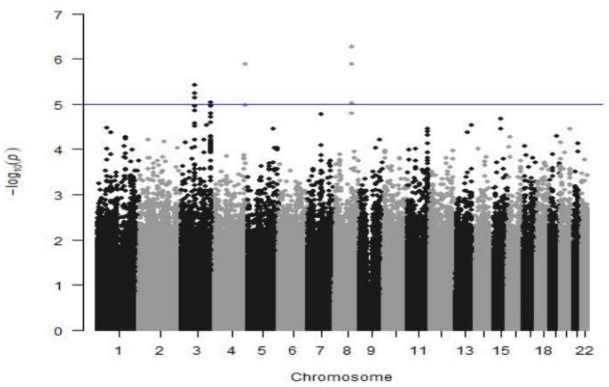
Lepiej usunąć osobę z rodziny o identyfikatorze 13291, co robię i ponownie filtruję dane

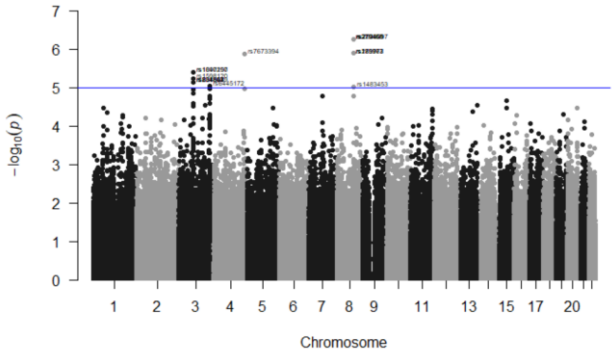
```
Options in effect:
        --bfile HapMap_10
        --remove usun.txt
        --make-bed
        --out HapMap_11

Reading map (extended format) from [ HapMap_10.bim ]
1073226 markers to be included from [ HapMap_10.bim ]
Reading pedigree information from [ HapMap_10.fam ]
112 individuals read from [ HapMap_10.fam ]
112 individuals with nonmissing phenotypes
Assuming a disease phenotype (1=unaff, 2=aff, 0=miss)
Missing phenotype value is also -9
56 cases, 56 controls and 0 missing
57 males, 55 females, and 0 of unspecified sex
Reading genotype bitfile from [ HapMap_10.bed ]
Detected that binary PED file is v1.00 SNP-major mode
Reading individuals to remove [ usun.txt ] ... 0 read
0 individuals removed with --remove option
Before frequency and genotyping pruning, there are 1073226 SNPs
112 founders and 0 non-founders found
Total genotyping rate in remaining individuals is 0.997991
0 SNPs failed missingness test ( GENO > 1 )
0 SNPs failed frequency test ( MAF < 0 )
After frequency and genotyping pruning, there are 1073226 SNPs
After filtering, 56 cases, 56 controls and 0 missing
After filtering, 57 males, 55 females, and 0 of unspecified sex
Writing pedigree information to [ HapMap_11.fam ]
Writing map (extended format) information to [ HapMap_11.bim ]
Writing genotype bitfile to [ HapMap_11.bed ]
Using (default) SNP-major mode

Analysis finished: Tue Jun 09 01:58:36 2020
```

Po zakończeniu filtrowania, w próbie zostało 112 osób

Zad 9

Analizy asocjacji wykorzystująca test *chi*$_2$





Pierwsze 25 wyników dla analizy z poprawkami:

```
$ head -n 25 wynik_ostat_popraw.assoc.adjusted
 CHR         SNP         UNADJ         GC         BONF         HOLM    SIDAK_SS    SIDAK_SD      FDR_BH      FDR_BY
   8    rs279466 5.549e-007 6.154e-007       0.5955       0.5955      0.4487      0.4487      0.1489           1
   8    rs279460 5.549e-007 6.154e-007       0.5955       0.5955      0.4487      0.4487      0.1489           1
   8    rs279453 5.549e-007 6.154e-007       0.5955       0.5955      0.4487      0.4487      0.1489           1
   8  rs2623297 5.549e-007 6.154e-007       0.5955       0.5955      0.4487      0.4487      0.1489           1
   8    rs279973 1.289e-006  1.42e-006            1            1      0.7491      0.7491      0.1777           1
   8    rs185773 1.289e-006  1.42e-006            1            1      0.7491      0.7491      0.1777           1
   8    rs183967 1.289e-006  1.42e-006            1            1      0.7491      0.7491      0.1777           1
   4  rs7673394 1.325e-006 1.459e-006            1            1      0.7587      0.7587      0.1777           1
   3  rs1840290 3.881e-006  4.24e-006            1            1      0.9845      0.9845      0.4165           1
   3  rs1097157 3.881e-006  4.24e-006            1            1      0.9845      0.9845      0.4165           1
   3  rs1598120 5.813e-006 6.331e-006            1            1       0.998       0.998       0.511           1
   3   rs834843 7.141e-006 7.766e-006            1            1      0.9995      0.9995       0.511           1
   3  rs2085079 7.141e-006 7.766e-006            1            1      0.9995      0.9995       0.511           1
   3   rs834858 7.141e-006 7.766e-006            1            1      0.9995      0.9995       0.511           1
   3   rs834864 7.141e-006 7.766e-006            1            1      0.9995      0.9995       0.511           1
   3  rs6445172 9.065e-006 9.839e-006            1            1      0.9999      0.9999      0.6037           1
   8  rs1483453 9.684e-006 1.051e-005            1            1           1           1      0.6037           1
   4 rs12505348 1.038e-005 1.147e-005            1            1           1           1      0.6037           1
   3  rs9855684 1.082e-005 1.172e-005            1            1           1           1      0.6037           1
   3   rs834856 1.125e-005 1.219e-005            1            1           1           1      0.6037           1
   3  rs1600058 1.386e-005   1.5e-005            1            1           1           1      0.6762           1
   3  rs1580295 1.386e-005   1.5e-005            1            1           1           1      0.6762           1
   3  rs9817983 1.625e-005 1.756e-005            1            1           1           1      0.6919           1
   8  rs7015011 1.636e-005 1.768e-005            1            1           1           1      0.6919           1
```

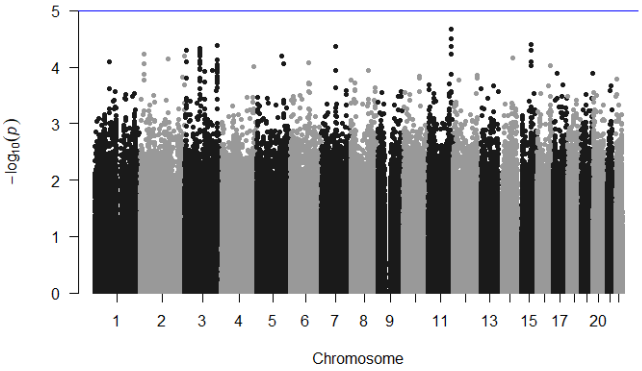
Zad 10

Uwzględniam MDS w analizie i prowadzę analizę korelacji dla niezależnych SNP

```
Writing this text to log file [ HapMap_11.log ]
Analysis started: Tue Jun 09 11:38:16 2020

Options in effect:
        --bfile HapMap_11
        --genome
        --extract indepSNP.prune.in
        --out HapMap_11

Reading map (extended format) from [ HapMap_11.bim ]
1073226 markers to be included from [ HapMap_11.bim ]
Reading pedigree information from [ HapMap_11.fam ]
112 individuals read from [ HapMap_11.fam ]
112 individuals with nonmissing phenotypes
Assuming a disease phenotype (1=unaff, 2=aff, 0=miss)
Missing phenotype value is also -9
56 cases, 56 controls and 0 missing
57 males, 55 females, and 0 of unspecified sex
Reading genotype bitfile from [ HapMap_11.bed ]
Detected that binary PED file is v1.00 SNP-major mode
Reading list of SNPs to extract [ indepSNP.prune.in ] ... 104210 read
Before frequency and genotyping pruning, there are 104210 SNPs
112 founders and 0 non-founders found
Total genotyping rate in remaining individuals is 0.997952
0 SNPs failed missingness test ( GENO > 1 )
0 SNPs failed frequency test ( MAF < 0 )
After frequency and genotyping pruning, there are 104210 SNPs
After filtering, 56 cases, 56 controls and 0 missing
After filtering, 57 males, 55 females, and 0 of unspecified sex
Converting data to Individual-major format
Writing whole genome IBS/IBD information to [ HapMap_11.genome ]
Filtering output to include pairs with ( 0 <= PI-HAT <= 1 )
IBD(g) calculation: 6200 of 6216

Analysis finished: Tue Jun 09 11:40:47 2020
```

Uzyskuję kowarianty dzięki analizie MDS dla 10 wymiarów i używam ich do analizy korelacji z regresją logistyczną:

```
Options in effect:
        --bfile HapMap_11
        --covar covar_mds.txt
        --logistic
        --hide-covar
        --out wyniki_mds_covar

Reading map (extended format) from [ HapMap_11.bim ]
1073226 markers to be included from [ HapMap_11.bim ]
Reading pedigree information from [ HapMap_11.fam ]
112 individuals read from [ HapMap_11.fam ]
112 individuals with nonmissing phenotypes
Assuming a disease phenotype (1=unaff, 2=aff, 0=miss)
Missing phenotype value is also -9
56 cases, 56 controls and 0 missing
57 males, 55 females, and 0 of unspecified sex
Reading genotype bitfile from [ HapMap_11.bed ]
Detected that binary PED file is v1.00 SNP-major mode
Reading 10 covariates from [ covar_mds.txt ] with nonmissing values for 112 individuals
Before frequency and genotyping pruning, there are 1073226 SNPs
112 founders and 0 non-founders found
Total genotyping rate in remaining individuals is 0.997991
0 SNPs failed missingness test ( GENO > 1 )
0 SNPs failed frequency test ( MAF < 0 )
After frequency and genotyping pruning, there are 1073226 SNPs
After filtering, 56 cases, 56 controls and 0 missing
After filtering, 57 males, 55 females, and 0 of unspecified sex
Converting data to Individual-major format
Writing logistic model association results to [ wyniki_mds_covar.assoc.logistic ]

Analysis finished: Tue Jun 09 12:26:58 2020
```



Uzyskane wartości – log(P) są mniejsze, czyli wartości P są większe