sisg-2015-mod23-practical2

sisg23

July 23, 2015

```
setwd("C:/Users/vonholle/Documents/grad_school/misc/training/sisg-2015/module23/sisg-2015-module23") #

data <- read.table("practical_2/data/data.txt", header = T)
dim(data)
```

```
## [1] 11214     59
```

```
data[1:2, 1:24]
```

```
##      pairID chr1_pi chr2_pi chr3_pi chr4_pi chr5_pi chr6_pi chr7_pi
## 1 800301112  0.3025  0.4413  0.4603  0.3322  0.4801  0.2045  0.6479
## 2 801101112  0.5184  0.3173  0.4297  0.6732  0.5803  0.4266  0.3382
##   chr8_pi chr9_pi chr10_pi chr11_pi chr12_pi chr13_pi chr14_pi chr15_pi
## 1  0.1674  0.3484   0.4224   0.4550   0.2969   0.5484   0.2646   0.6484
## 2  0.6611  0.4706   0.3917   0.4542   0.4342   0.3589   0.4113   0.4557
##   chr16_pi chr17_pi chr18_pi chr19_pi chr20_pi chr21_pi chr22_pi chrx_pi
## 1   0.7575   0.5294   0.7745   0.6430   0.6818   0.8810   0.5376  0.4681
## 2   0.7145   0.6257   0.6315   0.8907   0.6510   0.3921   0.4344  0.7597
```

```
# Investigate the pi_hat_d values for the autosomes and X chromosome
data[1:2, 25:47]
```

```
##   chr1_ibd chr2_ibd chr3_ibd chr4_ibd chr5_ibd chr6_ibd chr7_ibd chr8_ibd
## 1   0.0146   0.1576   0.1428   0.0201   0.2047   0.0102   0.4009   0.0318
## 2   0.1987   0.0300   0.1771   0.5068   0.3305   0.1260   0.0731   0.4637
##   chr9_ibd chr10_ibd chr11_ibd chr12_ibd chr13_ibd chr14_ibd chr15_ibd
## 1   0.0458    0.1916    0.1324    0.1090    0.4508    0.0107    0.3754
## 2   0.2977    0.1373    0.2333    0.1679    0.0037    0.0597    0.3768
##   chr16_ibd chr17_ibd chr18_ibd chr19_ibd chr20_ibd chr21_ibd chr22_ibd
## 1    0.5247    0.2466    0.5560    0.3028    0.3735    0.7689    0.1238
## 2    0.4518    0.3782    0.3355    0.7852    0.3151    0.0176    0.0309
##   chrx_ibd
## 1   0.4681
## 2   0.7597
```

```
# Take a look at the rest of the data matrix
data[1:2, 48:59]
```

```
##        gw_pi    gw_ibd sex_sib1 age_sib1 ht_sib1    zht_sib1 sex_sib2
## 1 0.4920500 0.2361227        1       16   178.0  0.47048173        1
## 2 0.5118773 0.2498455        1       16   174.5 -0.05328332        1
##   age_sib2 ht_sib2  zht_sib2 sex_pair pop
## 1       16     166 -1.325284        1   2
## 2       16     185  1.518012        1   2
```

Exercise 1

Calculate the sample mean and SD of $\hat{\pi}_a$ and $\hat{\pi}_d$:

- for each autosome

- genome-wide

- Do a histogram of the genome-wide $\hat{\pi}_a$ and $\hat{\pi}_d$ (row means) and compare the mean and standard to that in the lectures.

- plot genome-wide $\hat{\pi}_a$ against $\hat{\pi}_d$ for each sibling pair.

- regress genome-wide $\hat{\pi}_d$ values on genome-wide $\hat{\pi}_a$.

```
# Row means over a matrix can be calculated with the below function
rowMeans(data[1:2, 2:23])
```

```
##           1         2
## 0.4920500 0.5118773
```

```
# A similar function does not exist for row standard deviations and thus we
# must appeal to the apply function
head(apply(data[, 2:23], 1, sd))
```

```
## [1] 0.1934229 0.1461863 0.1777696 0.1620989 0.1846666 0.1840230
```

```
pihat.mean <- rowMeans(data[, 2:23])
pihat.sd <- apply(data[, 2:23], 1, sd)
ibd.mean <- rowMeans(data[, 25:46])
ibd.sd <- apply(data[, 25:46], 1, sd)
plot(pihat.mean, ibd.mean, pch = 20,
col = 4, ylab = "Dominance relationship (mean IBD2 sharing)", xlab = "Additive relationship (mean IBD
         sharing)")
# Regress IBD2 on pi-hat. Use the lm function un R. Try ?lm if you are interested
reg <- lm(ibd.mean ~ pihat.mean)
abline(reg, lwd = 2.5)
```
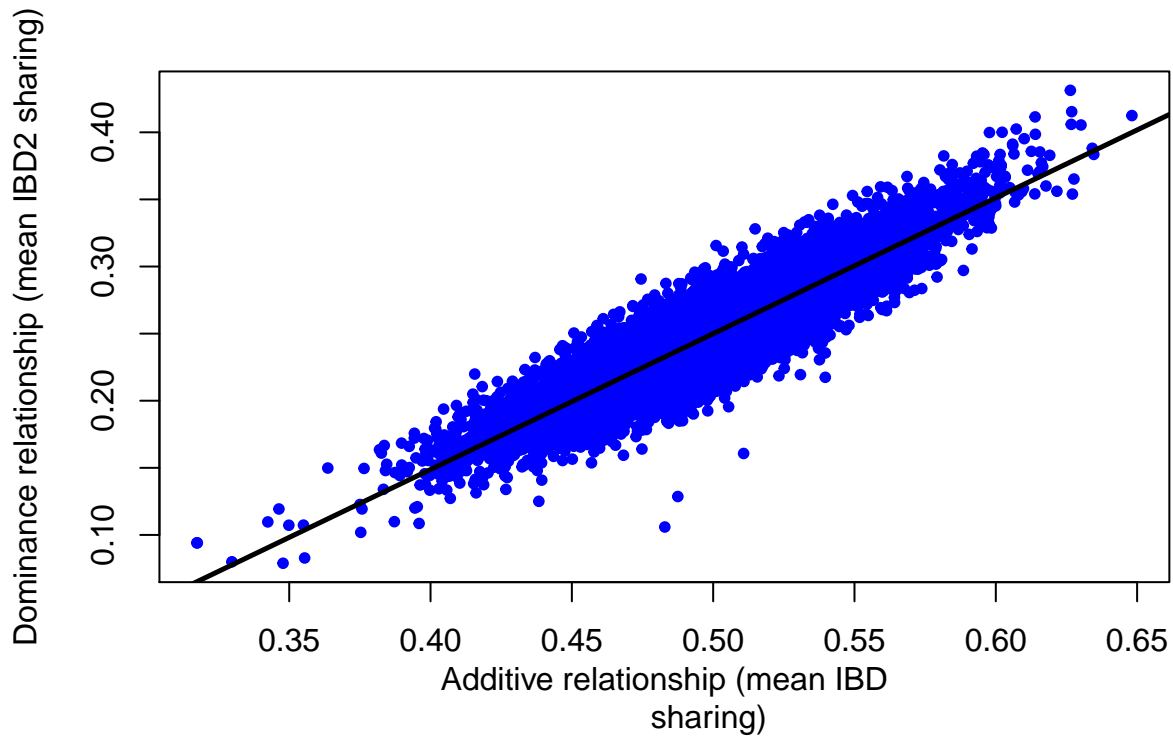
```
# Regress IBD on pi-hat. Use the lm function un R. Try ?lm if you are interested
summary(lm(ibd.mean ~ pihat.mean))
```

```
##
## Call:
## lm(formula = ibd.mean ~ pihat.mean)
##
## Residuals:
##       Min       1Q    Median       3Q      Max
## -0.126745 -0.011290  0.000044  0.011251  0.066568
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.256023   0.002184   -117.2   <2e-16 ***
## pihat.mean   1.011945   0.004362    232.0   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.01709 on 11212 degrees of freedom
## Multiple R-squared:  0.8276, Adjusted R-squared:  0.8276
## F-statistic: 5.382e+04 on 1 and 11212 DF,  p-value: < 2.2e-16
```

```
# Read in .ped and .dat files
setwd("C:/Users/vonholle/Documents/grad_school/misc/training/sisg-2015/module23/sisg-2015-module23") #
ped <- read.table("practical_2/data/qtdt.ped")
```

```r
dat <- read.table("practical_2/data/qtdt.dat")
ibd <- read.table("practical_2/data/qtdt.ibd")

# The head of the .ped file. Two columns have been cut to fit in the listing
head(ped)
```

```
##    V1 V2 V3 V4 V5      V6 V7  V8  V9 V10 V11 V12 V13 V14 V15 V16 V17 V18
## 1  1  1  0  0  1       X  X   X 1/2 1/2 1/2 1/2 1/2 1/2 1/2 1/2 1/2 1/2
## 2  1  2  0  0  2       X  X   X 1/2 1/2 1/2 1/2 1/2 1/2 1/2 1/2 1/2 1/2
## 3  1  3  1  2  1  0.4705  1 16. 1/1 1/1 1/1 1/1 1/1 1/1 1/1 1/1 1/1 1/1
## 4  1  4  1  2  1 -1.3253  1 16. 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2
## 5  2  1  0  0  1       X  X   X 1/2 1/2 1/2 1/2 1/2 1/2 1/2 1/2 1/2 1/2
## 6  2  2  0  0  2       X  X   X 1/2 1/2 1/2 1/2 1/2 1/2 1/2 1/2 1/2 1/2
##   V19 V20 V21 V22 V23 V24 V25 V26 V27 V28 V29 V30 V31
## 1 1/2 1/2 1/2 1/2 1/2 1/2 1/2 1/2 1/2 1/2 1/2 1/2 1/2
## 2 1/2 1/2 1/2 1/2 1/2 1/2 1/2 1/2 1/2 1/2 1/2 1/2 1/2
## 3 1/1 1/1 1/1 1/1 1/1 1/1 1/1 1/1 1/1 1/1 1/1 1/1 1/1
## 4 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2 2/2
## 5 1/2 1/2 1/2 1/2 1/2 1/2 1/2 1/2 1/2 1/2 1/2 1/2 1/2
## 6 1/2 1/2 1/2 1/2 1/2 1/2 1/2 1/2 1/2 1/2 1/2 1/2 1/2
```

```r
dat
```

```
##     V1  V2
## 1    T   Y
## 2    C SEX
## 3    C AGE
## 4   S2  C1
## 5   S2  C2
## 6   S2  C3
## 7   S2  C4
## 8   S2  C5
## 9   S2  C6
## 10  S2  C7
## 11  S2  C8
## 12  S2  C9
## 13  S2 C10
## 14  S2 C11
## 15  S2 C12
## 16  S2 C13
## 17  S2 C14
## 18  S2 C15
## 19  S2 C16
## 20  S2 C17
## 21  S2 C18
## 22  S2 C19
## 23  S2 C20
## 24  S2 C21
## 25  S2 C22
## 26   M   G
```

```r
head(ibd)
```

```
##   V1 V2 V3 V4     V5     V6     V7
## 1  1  3  4 C1 0.4096 0.5758 0.0146
## 2  1  3  4 C2 0.2750 0.5674 0.1576
## 3  1  3  4 C3 0.2222 0.6350 0.1428
## 4  1  3  4 C4 0.3557 0.6242 0.0201
## 5  1  3  4 C5 0.2445 0.5508 0.2047
## 6  1  3  4 C6 0.6012 0.3886 0.0102
```

```r
ibd[1:23,1:ncol(ibd)]
```

```
##    V1 V2 V3  V4     V5     V6     V7
## 1   1  3  4  C1 0.4096 0.5758 0.0146
## 2   1  3  4  C2 0.2750 0.5674 0.1576
## 3   1  3  4  C3 0.2222 0.6350 0.1428
## 4   1  3  4  C4 0.3557 0.6242 0.0201
## 5   1  3  4  C5 0.2445 0.5508 0.2047
## 6   1  3  4  C6 0.6012 0.3886 0.0102
## 7   1  3  4  C7 0.1051 0.4940 0.4009
## 8   1  3  4  C8 0.6970 0.2712 0.0318
## 9   1  3  4  C9 0.3490 0.6052 0.0458
## 10  1  3  4 C10 0.3468 0.4616 0.1916
## 11  1  3  4 C11 0.2224 0.6452 0.1324
## 12  1  3  4 C12 0.5152 0.3758 0.1090
## 13  1  3  4 C13 0.3540 0.1952 0.4508
## 14  1  3  4 C14 0.4815 0.5078 0.0107
## 15  1  3  4 C15 0.0786 0.5460 0.3754
## 16  1  3  4 C16 0.0097 0.4656 0.5247
## 17  1  3  4 C17 0.1878 0.5656 0.2466
## 18  1  3  4 C18 0.0070 0.4370 0.5560
## 19  1  3  4 C19 0.0168 0.6804 0.3028
## 20  1  3  4 C20 0.0099 0.6166 0.3735
## 21  1  3  4 C21 0.0069 0.2242 0.7689
## 22  1  3  4 C22 0.0486 0.8276 0.1238
## 23  1  3  4   G 0.2520 0.5119 0.2361
```

```r
# following lines from a document from ibg.colorado.edu/~howrigan/howrigan/.../Howrigan_WCPG_2010.pdf
# THE 'system()' COMMAND:
# #The system command is a way to use your computer terminal from within R. Within the system() functio
# whatever you put in quotes will be read as if that's what you typed in your terminal window. I'm doing
# my analyses using a UNIX shell (bash), and so in order to use the qtdt program, I have to put a './'
# before I write 'qtdt'. Windows users only need to write 'qtdt'. In addition, I use 'intern=TRUE' so t
# the output will be displayed in the R console.
# #NOTE: this won't look as good in R as in terminal unless you have your width setup accordingly
options(width=100) #works well
```

Original run from the assignment

```r
setwd("C:/Users/vonholle/Documents/grad_school/misc/training/sisg-2015/module23/sisg-2015-module23") #
# Let's see if these files will work in QTDT
# Once you've set up your binary directory for this analysis and have set R to that directory, do this:
system("qtdt -d practical_2/data/qtdt.dat -p practical_2/data/qtdt.ped -a- -we- -veg-", intern=TRUE)
```

```
##  [1] "QTDT - Quantitative TDT 2.6.1"
##  [2] "(c) 1998-2007 Goncalo Abecasis (goncalo@umich.edu)"
##  [3] ""
##  [4] "This program implements tests described by"
##  [5] "Abecasis et al, Am J Hum Genet 66:279-292 (2000)"
##  [6] "Abecasis et al, Eur J Hum Genet 8:545-551 (2000)"
##  [7] "and others"
##  [8] ""
##  [9] "The following parameters are in effect:"
## [10] "           QTDT Data File : practical_2/data/qtdt.dat (-dname)"
## [11] "       QTDT Pedigree File : practical_2/data/qtdt.ped (-pname)"
## [12] "     QTDT IBD Status File :              qtdt.ibd (-iname)"
## [13] "       Missing Value Code :               -99.999 (-xname)"
## [14] "               Covariates :     USER SPECIFIED (-c{p|s|u|-})"
## [15] "        Association Model :          NONE (-a[a|d|f|m|o|p|r|t|w|-])"
## [16] "      Full Model Variances :      NON SHARED (-v{e|c|g|n|t|a|d|-})"
## [17] "                        &          POLYGENIC"
## [18] "      Null Model Variances :      NON SHARED (-w{e|c|g|n|t|a|d|-})"
## [19] " Parent of Origin Effects :          NONE (-o[f|t|m|p|-])"
## [20] " Monte-Carlo Permutations :             0 (-m9999)"
## [21] "              Random Seed :         123456 (-r9999)"
## [22] "         Numeric Minimizer :     NELDER AND MEAD (-n[f|n|p])"
## [23] "      Transmission Scoring :       FULL PEDIGREE (-t[n|p])"
## [24] ""
## [25] "Additional Options"
## [26] " --dominance, --snp, --multi-allelic, --deviates, --references,"
## [27] " --exclude-founder-phenotypes, --p-values, --no-regress-tbl"
## [28] ""
## [29] "Online documentation http://www.sph.umich.edu/csg/abecasis/QTDT"
## [30] "Comments, bugs: goncalo@umich.edu"
## [31] ""
## [32] "The following models will be evaluated..."
## [33] "  NULL MODEL"
## [34] "     Means = Mu + SEX + AGE"
## [35] " Variances = Ve"
## [36] ""
## [37] "  FULL MODEL"
## [38] "     Means = Mu + SEX + AGE"
## [39] " Variances = Ve + Vg"
## [40] ""
## [41] "Testing trait:                              Y"
## [42] "============================================="
## [43] ""
## [44] " Allele   df(0) -LnLk(0)   df(V) -LnLk(V)   ChiSq         p"
## [45] "    N/A   22424 31521.42   22423 30196.57 2649.70  0e+000  (22428 probands)"
## [46] ""
## [47] ""
## [48] "Run completed on Thu Jul 23 16:00:30 2015"
```

```
## [49] "1 tests carried out"
## [50] ""
## [51] "The most significant result refers to:"
## [52] "     Trait: Y"
## [53] "     ChiSq: 2649.700"
## [54] "   p-value: 0e+000"
## [55] ""
## [56] ""
```

```r
system("qtdt -d practical_2/data/qtdt.dat -p practical_2/data/qtdt.ped -i practical_2/data/qtdt.ibd -a-
```

```
##  [1] "QTDT - Quantitative TDT 2.6.1"
##  [2] "(c) 1998-2007 Goncalo Abecasis (goncalo@umich.edu)"
##  [3] ""
##  [4] "This program implements tests described by"
##  [5] "Abecasis et al, Am J Hum Genet 66:279-292 (2000)"
##  [6] "Abecasis et al, Eur J Hum Genet 8:545-551 (2000)"
##  [7] "and others"
##  [8] ""
##  [9] "The following parameters are in effect:"
## [10] "          QTDT Data File : practical_2/data/qtdt.dat (-dname)"
## [11] "      QTDT Pedigree File : practical_2/data/qtdt.ped (-pname)"
## [12] "    QTDT IBD Status File : practical_2/data/qtdt.ibd (-iname)"
## [13] "      Missing Value Code :            -99.999 (-xname)"
## [14] "              Covariates :      USER SPECIFIED (-c{p|s|u|-})"
## [15] "       Association Model :               NONE (-a[a|d|f|m|o|p|r|t|w|-])"
## [16] "     Full Model Variances :        NON SHARED (-v{e|c|g|n|t|a|d|-})"
## [17] "                        &           POLYGENIC"
## [18] "                        &            ADDITIVE"
## [19] "     Null Model Variances :         NON SHARED (-w{e|c|g|n|t|a|d|-})"
## [20] "                        &           POLYGENIC"
## [21] " Parent of Origin Effects :             NONE (-o[f|t|m|p|-])"
## [22] " Monte-Carlo Permutations :                0 (-m9999)"
## [23] "             Random Seed :           123456 (-r9999)"
## [24] "        Numeric Minimizer :    NELDER AND MEAD (-n[f|n|p])"
## [25] "     Transmission Scoring :      FULL PEDIGREE (-t[n|p])"
## [26] ""
## [27] "Additional Options"
## [28] " --dominance, --snp, --multi-allelic, --deviates, --references,"
## [29] " --exclude-founder-phenotypes, --p-values, --no-regress-tbl"
## [30] ""
## [31] "Online documentation http://www.sph.umich.edu/csg/abecasis/QTDT"
## [32] "Comments, bugs: goncalo@umich.edu"
## [33] ""
## [34] "The following models will be evaluated..."
## [35] "  NULL MODEL"
## [36] "     Means = Mu + SEX + AGE"
## [37] " Variances = Ve + Vg"
## [38] ""
## [39] "  FULL MODEL"
## [40] "     Means = Mu + SEX + AGE"
## [41] " Variances = Ve + Vg + Va"
## [42] ""
## [43] "Testing trait:                        Y"
```

```
## [44] "============================================"
## [45] ""
## [46] "Testing marker:                          G"
## [47] "--------------------------------------------"
## [48] ""
## [49] " Allele   df(0) -LnLk(0)   df(V) -LnLk(V)   ChiSq         p"
## [50] "    All   22423 30196.57   22422 30186.27   20.61   6e-006  (22428 probands)"
## [51] ""
## [52] ""
## [53] "Run completed on Thu Jul 23 16:01:06 2015"
## [54] "1 tests carried out"
## [55] ""
## [56] "The most significant result refers to:"
## [57] "     Trait: Y"
## [58] "    Marker: G"
## [59] "     ChiSq: 20.609"
## [60] "   p-value: 6e-006"
## [61] ""
## [62] "Overall Bonferroni significance level: 6e-006"
## [63] ""
```

```r
# print off variance info
cat(readChar("regress.tbl", 1e5))
```

```
## Trait: Y                Marker: G
## ================================================================
## Total Probands: 22428
## Family #1 Phenotypes - scores :   0.470  -1.325
##
## NULL HYPOTHESIS
## ---------------
##
## Family #1 var-covar matrix terms [2]...[[Ve]][[Vg]]
## Family #1 regression matrix...
##        [linear] =
##         [2 x 3]      Mu      SEX      AGE
##            1.3   1.000    1.000   16.000
##            1.4   1.000    1.000   16.000
##
## Some useful information...
##             df : 22423
##   log(likelihood) : 30196.57
##        variances :   0.080    0.894
##            means :   0.079    0.019   -0.002
##
## FULL HYPOTHESIS
## ---------------
##
## Family #1 var-covar matrix terms [3]...[[Ve]][[Vg]][[Va]]
## Family #1 regression matrix...
##        [linear] =
##         [2 x 3]      Mu      SEX      AGE
##            1.3   1.000    1.000   16.000
##            1.4   1.000    1.000   16.000
```

```
##
## Some useful information...
##             df : 22422
##   log(likelihood) : 30186.27
##        variances :   0.079   0.056   0.839
##            means :   0.079   0.019  -0.002
##
##
```

Run on the 10th chromosome.

```
setwd("C:/Users/vonholle/Documents/grad_school/misc/training/sisg-2015/module23/sisg-2015-module23") #
# marker 10
system("qtdt -d practical_2/data/qtdt10.dat -p practical_2/data/qtdt.ped -i practical_2/data/qtdt.ibd -
```

Note: no variance showing up.

```
##  [1] "QTDT - Quantitative TDT 2.6.1"
##  [2] "(c) 1998-2007 Goncalo Abecasis (goncalo@umich.edu)"
##  [3] ""
##  [4] "This program implements tests described by"
##  [5] "Abecasis et al, Am J Hum Genet 66:279-292 (2000)"
##  [6] "Abecasis et al, Eur J Hum Genet 8:545-551 (2000)"
##  [7] "and others"
##  [8] ""
##  [9] "The following parameters are in effect:"
## [10] "        QTDT Data File : practical_2/data/qtdt10.dat (-dname)"
## [11] "    QTDT Pedigree File : practical_2/data/qtdt.ped (-pname)"
## [12] "  QTDT IBD Status File : practical_2/data/qtdt.ibd (-iname)"
## [13] "     Missing Value Code :           -99.999 (-xname)"
## [14] "            Covariates :     USER SPECIFIED (-c{p|s|u|-})"
## [15] "       Association Model :           NONE (-a[a|d|f|m|o|p|r|t|w|-])"
## [16] "    Full Model Variances :        NON SHARED (-v{e|c|g|n|t|a|d|-})"
## [17] "                      &            POLYGENIC"
## [18] "                      &            ADDITIVE"
## [19] "    Null Model Variances :        NON SHARED (-w{e|c|g|n|t|a|d|-})"
## [20] "                      &            POLYGENIC"
## [21] " Parent of Origin Effects :           NONE (-o[f|t|m|p|-])"
## [22] " Monte-Carlo Permutations :              0 (-m9999)"
## [23] "          Random Seed :          123456 (-r9999)"
## [24] "      Numeric Minimizer :     NELDER AND MEAD (-n[f|n|p])"
## [25] "    Transmission Scoring :     FULL PEDIGREE (-t[n|p])"
## [26] ""
## [27] "Additional Options"
## [28] " --dominance, --snp, --multi-allelic, --deviates, --references,"
## [29] " --exclude-founder-phenotypes, --p-values, --no-regress-tbl"
## [30] ""
## [31] "Online documentation http://www.sph.umich.edu/csg/abecasis/QTDT"
## [32] "Comments, bugs: goncalo@umich.edu"
## [33] ""
## [34] "The following models will be evaluated..."
```

```
## [35] "  NULL MODEL"
## [36] "      Means = Mu + SEX + AGE"
## [37] " Variances = Ve + Vg"
## [38] ""
## [39] "  FULL MODEL"
## [40] "      Means = Mu + SEX + AGE"
## [41] " Variances = Ve + Vg + Va"
## [42] ""
## [43] "Testing trait:                                Y"
## [44] "==========================================="
## [45] ""
## [46] "Testing marker:                             C10"
## [47] "-------------------------------------------"
## [48] ""
## [49] " Allele   df(0) -LnLk(0)   df(V) -LnLk(V)   ChiSq          p"
## [50] "   All   22423 30196.57   22422 30196.57    0.00          (22428 probands)"
## [51] ""
## [52] ""
## [53] "Run completed on Thu Jul 23 16:01:54 2015"
## [54] "1 tests carried out"
## [55] ""
## [56] "The most significant result refers to:"
## [57] "    Trait: Y"
## [58] "   Marker: C10"
## [59] "    ChiSq: 0.000"
## [60] "  p-value: 1.0000"
## [61] ""
## [62] ""
```

```r
# print off variance info
cat(readChar("regress.tbl", 1e5))
```

```
## Trait: Y              Marker: C10
## ================================================================
## Total Probands: 22428
## Family #1 Phenotypes - scores :   0.470  -1.325
##
## NULL HYPOTHESIS
## ---------------
##
## Family #1 var-covar matrix terms [2]...[[Ve]][[Vg]]
## Family #1 regression matrix...
##        [linear] =
##        [2 x 3]     Mu     SEX     AGE
##            1.3   1.000   1.000  16.000
##            1.4   1.000   1.000  16.000
##
## Some useful information...
##             df : 22423
##   log(likelihood) : 30196.57
##        variances :   0.080   0.894
##            means :   0.079   0.019   -0.002
##
## FULL HYPOTHESIS
```

```
## ---------------
##
## Family #1 var-covar matrix terms [3]...[[Ve]][[Vg]][[Va]]
## Family #1 regression matrix...
##         [linear] =
##           [2 x 3]       Mu      SEX      AGE
##               1.3    1.000    1.000   16.000
##               1.4    1.000    1.000   16.000
##
## Some useful information...
##                 df : 22422
##   log(likelihood) : 30196.57
##         variances :    0.080    0.894    0.000
##             means :    0.078    0.019   -0.002
##
##
```

Appendix

`getwd()`

```
## [1] "C:/Users/vonholle/Documents/grad_school/misc/training/sisg-2015/module23/sisg-2015-module23/pra
```