

# Genetische Statistik

## Präsenzübung 6: Visualisierung statistischer Konzepte

Dr. Janne Pott ([janne.pott@uni-leipzig.de](mailto:janne.pott@uni-leipzig.de))

December 07, 2021

# Fragen

Gibt es Fragen zu

- Vorlesung?
- Übung?
- Seminar?

# Plan heute

## Besprechung von RBlatt 4

- Verwandtschaft
- XY-Plots
- PCA

## Anschließend / Falls noch Zeit

- Blatt 4 - A2 (Heritabilität)
- Blatt 4 - A1 (Populationsgenetik)

# Abschnitt 1

## Verwandtschaft

# Aufgabe 1: Verwandtschaft - Hintergrund (1)

Paarweise Schätzung von Verwandtschaft:

$$\hat{k}_{i,j} = \frac{1}{M} \sum_{m=1}^M \frac{(g_{m,i} - 2 * p_{m,B})(g_{m,j} - 2 * p_{m,B})}{4 * p_{m,B} * p_{m,A}}$$

mit

- $M$  als Anzahl der betrachteten biallelischen SNPs (Allel A und B)
- $p_{m,B}$  als Allelfrequenz des SNPs  $m$  bezüglich Allel B
- $g_{m,i}$  als Genotyp des SNPs  $m$  von Person  $i$  bezüglich Allel B

# Aufgabe 1: Verwandtschaft - Hintergrund (2)

**Tabelle 1:** Verwandtschaftsmatrix mittels Schleife

0.50	0.00	0.00	0.00	0.24	0.24	0.24	0.25	0.25	0.24
0.00	0.50	0.00	0.00	0.24	0.24	0.00	0.00	0.00	0.00
0.00	0.00	0.50	0.00	0.00	0.00	0.25	0.25	0.00	0.00
0.00	0.00	0.00	0.50	0.00	0.00	0.00	0.00	0.25	0.25
0.24	0.24	0.00	0.00	0.49	0.24	0.12	0.12	0.13	0.12
0.24	0.24	0.00	0.00	0.24	0.49	0.12	0.12	0.12	0.12
0.24	0.00	0.25	0.00	0.12	0.12	0.49	0.24	0.12	0.12
0.25	0.00	0.25	0.00	0.12	0.12	0.24	0.50	0.12	0.12
0.25	0.00	0.00	0.25	0.13	0.12	0.12	0.12	0.50	0.25
0.24	0.00	0.00	0.25	0.12	0.12	0.12	0.12	0.25	0.49

# Aufgabe 1: Verwandtschaft

- a) Verwandtschaftsmatrix mittels Matrix-Operation bestimmen. Stimmt dieses Produkt mit  $K$  überein?
- b) Warum gilt:

$$\hat{k}_{i,i} \approx 0.5$$

- c) Wie viele paarweise Verwandtschaften (von Grad 1, 2,  $\dots$ , unverwandte) beobachten Sie?
- d) Welche Familienstruktur könnte die beobachteten Verwandtschaftsbeziehungen erklären?

# Aufgabe 1: Verwandtschaft - Lösung a

```
n=ncol(genotypes)
m=nrow(genotypes)
h=(genotypes-matrix(2*allelfreq,m,n))/
  sqrt(m*matrix(4*allelfreq*(1-allelfreq),m,n))
H=t(h)%*%(h)

table(round(H,4)==round(K,4))
```

```
##
```

```
## TRUE
```

```
## 100
```



# Aufgabe 1: Verwandtschaft - Lösung a & b

- $H$  und  $K$  sind identisch.
- Für die paarweise Verwandtschaft braucht man nur die obere Dreiecksmatrix.
- Auf der Diagonalen selbst sollte immer 0.5 stehen, das ist für den Kinship-Schätzer Identität oder eineiige Zwillinge.

---

$k_{i,j}$	Interpretation
0.5	Eineiige Zwillinge / Identität
0.25	erstgradige Verwandtschaft (z.B. Eltern-Kind, Geschwister)
0.125	zweitgradige Verwandtschaft (z.B. Halbgeschwister, Großeltern-Enkel, Onkel/Tante-Nichte/Neffe)

---

# Aufgabe 1: Verwandtschaft - Lösung c

Anzahl Verwandtschaften:

- n-gradig: 18 unverwandte Paare
- 2-gradig: 12 mal Großeltern-Enkel, Onkel/Tante-Nichte/Neffe oder Halbgeschwister
- 1-gradig: 15 mal Eltern-Kinder oder Geschwister

# Aufgabe 1: Verwandtschaft - Lösung c

**Tabelle 3:** Kinship Schätzer

	S1	S2	S3	S4	S5	S6	S7	S8	
S1	0.496	-0.002	0.001	-0.002	0.243	0.243	0.245	0.248	0
S2	NA	0.501	0.000	0.001	0.244	0.245	-0.002	-0.002	0
S3	NA	NA	0.500	-0.003	-0.001	-0.002	0.247	0.252	-0
S4	NA	NA	NA	0.500	0.000	0.001	-0.001	-0.003	0
S5	NA	NA	NA	NA	0.488	0.238	0.120	0.119	0
S6	NA	NA	NA	NA	NA	0.490	0.121	0.120	0
S7	NA	NA	NA	NA	NA	NA	0.492	0.244	0
S8	NA	NA	NA	NA	NA	NA	NA	0.498	0
S9	NA	NA	NA	NA	NA	NA	NA	NA	0
S10	NA	NA	NA	NA	NA	NA	NA	NA	

# Aufgabe 1: Verwandtschaft - Lösung c

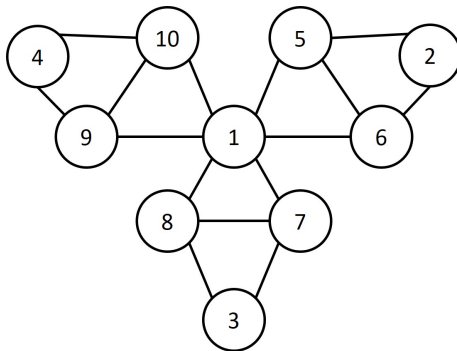
**Tabelle 4:** Verwandtschaftsgrade

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
sample1	NA	0	0	0	1	1	1	1	1	1
sample2	NA	NA	0	0	1	1	0	0	0	0
sample3	NA	NA	NA	0	0	0	1	1	0	0
sample4	NA	NA	NA	NA	0	0	0	0	1	1
sample5	NA	NA	NA	NA	NA	1	2	2	2	2
sample6	NA	NA	NA	NA	NA	NA	2	2	2	2
sample7	NA	NA	NA	NA	NA	NA	NA	1	2	2
sample8	NA	NA	NA	NA	NA	NA	NA	NA	2	2
sample9	NA	NA	NA	NA	NA	NA	NA	NA	NA	1
sample10	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

# Aufgabe 1: Verwandtschaft - Lösung d

Interpretation 1: Ein Vater (1) hat mit drei verschiedenen Müttern (2, 3, 4) je zwei Kindern (5 - 10).

Interpretation 2: Eine Mutter (1) hat mit drei verschiedenen Vätern (2, 3, 4) je zwei Kindern (5 - 10).



**Abbildung 1:** Graphische Darstellung der Verwandtschaftsbeziehungen

# Aufgabe 1: Verwandtschaft - Zusammenfassung

- Welche Grundannahme ist durch Verwandtschaft verletzt?
- Wie kann das gelöst werden?

## Abschnitt 2

### XY-Plot

## Aufgabe 2: XY-Plot - Hintergrund (1)

In genetischen Studien gibt es zwei Quellen für das Geschlecht:

- Datenbankgeschlecht: wie im Fragebogen angegeben, insbesondere auch divers
- Genetisches Geschlecht: im Genotyp-Calling bestimmt (Intensität der SNPs auf Chr. X & Y)

Mit dem XY-Plots kann man Probenvertauschungen und genetische Ausreißer entdecken. Grundannahmen:

- Intensität von X-SNPs in Frauen doppelt so stark wie in Männern
- Intensität von Y-SNPs in Frauen nur Hintergrundrauschen
- Heterozygotenrate in Frauen etwa 25%, in Männern 0%



## Aufgabe 2: XY-Plot

- a) Gesamtintensitäten pro Sample für X und Y bestimmen
- b) Plots:
  - i. X-Intensität – Y-Intensität
  - ii. X-Intensität – X-Heterozygosität
  - iii. Y-Intensität – X-Heterozygosität

## Aufgabe 2: XY-Plot - Lösung a)

```
# Mittelwert pro SNP und Sample
```

```
all<-seq(from=1,to=dim(intent)[1],by=2)
```

```
data.a<-intent[all,]
```

```
data.b<-intent[all+1,]
```

```
dataInt<-(data.a+data.b)/2
```

```
# mittlere Intensitäten pro Chromosom
```

```
dataIntX<-dataInt[,1:200]
```

```
dataIntY<-dataInt[,201:300]
```

```
IntX<-rowMeans(dataIntX)
```

```
IntY<-rowMeans(dataIntY)
```

```
# Normierung der Intensitäten nach dem 75%-Quantil
```

```
IntX2<-IntX/boxplot(IntX,plot=F)$stats[4]
```

```
IntY2<-IntY/boxplot(IntY,plot=F)$stats[4]
```

```
myDat<-data.frame(samples,IntX,IntY,IntX2,IntY2,heteroRate)
```

## Aufgabe 2: XY-Plot - Lösung a)

	sampleID	sex_datenbank	sex_computed	IntX
1:intA	1	male	male	779.373
2:intA	2	female	female	1164.601
3:intA	3	male	male	780.787

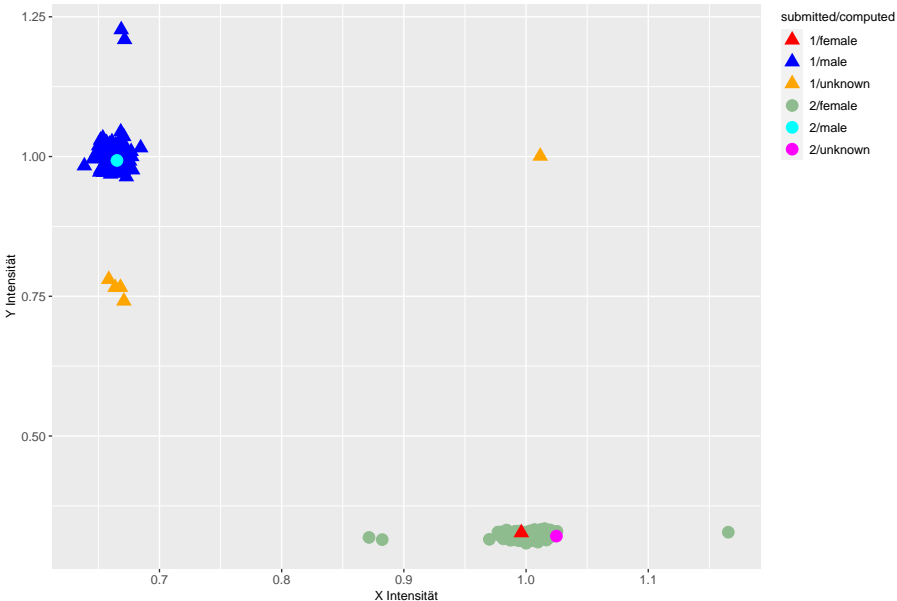
	IntY	IntX2	IntY2	heteroRate
1:intA	973.7237	0.6709741	0.9872946	0.00
2:intA	316.1990	1.0026228	0.3206059	0.22
3:intA	1003.1133	0.6721914	1.0170938	0.00

## Aufgabe 2: XY-Plot - Lösung c)

```
myPlot1 <- ggplot() +  
  geom_point(data=myDat,aes(x=IntX2,y=IntY2,color=sexLabel,  
                             shape=sexLabel),size=4) +  
  xlab("X Intensität") + ylab("Y Intensität") +  
  ggtitle("XY Plot mit 300 Samples") +  
  scale_colour_manual(name="submitted/computed",  
                      values=c("red","blue","orange",  
                               "darkseagreen","cyan","magenta")) +  
  scale_shape_manual(name="submitted/computed",  
                     values=c(17,17,17,19,19,19)) +  
  theme(legend.justification=c(1,1),  
        legend.text=element_text(size=10),  
        legend.title=element_text(size=10)) +  
  theme(axis.text=element_text(size=10),  
        axis.title=element_text(size=10),  
        plot.title=element_text(size=15))
```

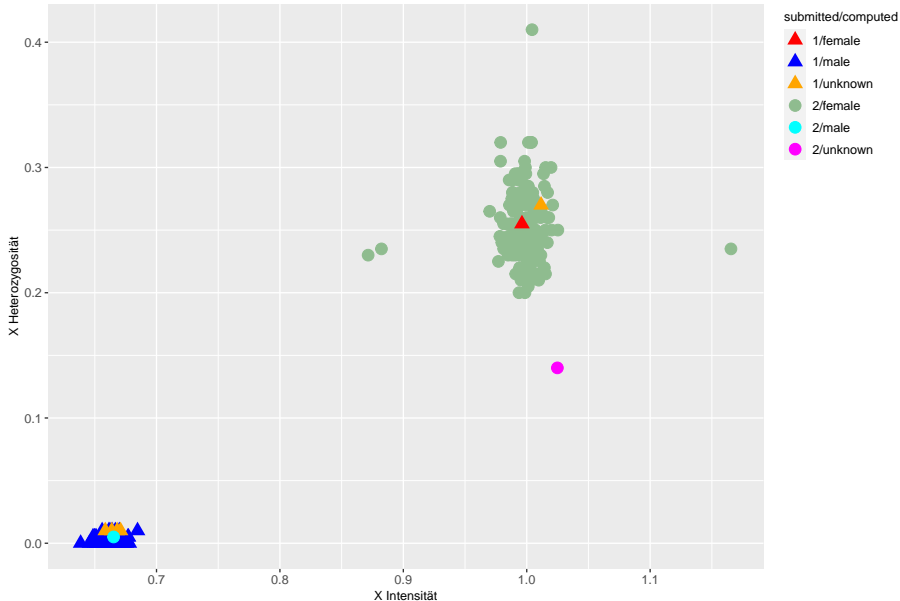
## Aufgabe 2: XY-Intensity Plot

### XY Plot mit 300 Samples



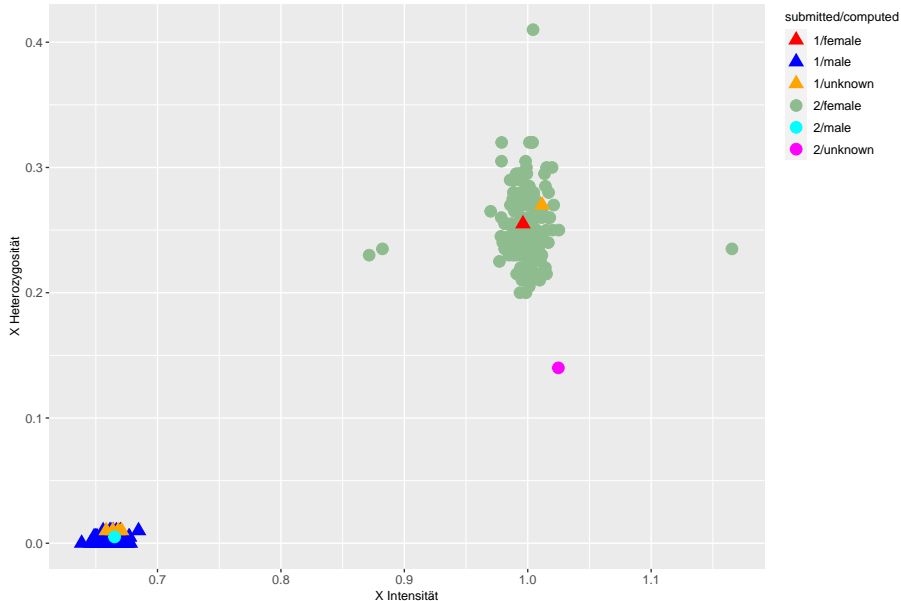
# Aufgabe 2: X-Intensity-Heterozygosity Plot

XX Plot mit 300 Samples



# Aufgabe 2: Y-Intensity-Heterozygosity Plot

XX Plot mit 300 Samples



## Aufgabe 2: XY-Plots - Lösung b)

Man kann folgende Ausreißer erkennen:

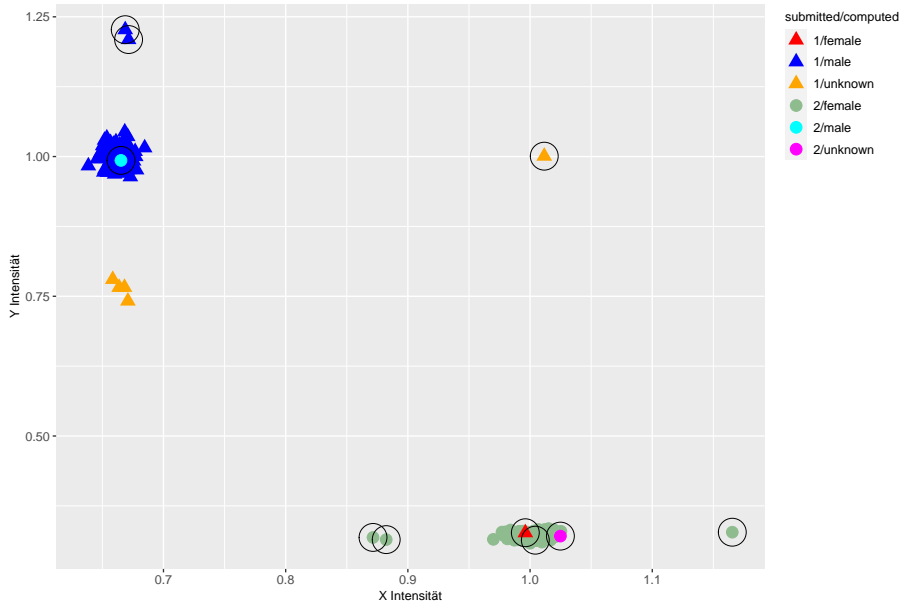
- ① Frauen mit zu hoher oder zu niedriger X-Intensität (Mono-X oder Triple-X Frauen)
- ② Männer mit zu hoher Y-Intensität (Doppel-Y Männer)
- ③ Männer mit zu hoher X-Intensität (Doppel-X Männer)
- ④ Frauen mit zu hoher oder zu niedriger X-Heterozygotie
- ⑤ Samples mit Sex-Mismatches zwischen Datenbank und Berechnung

1)-4) Samples sollten für gonosomale Analysen gefiltert werden (autosomal ok). 5) Sex-Mismatches müssen immer gefiltert werden, auch für autosomale Analysen!



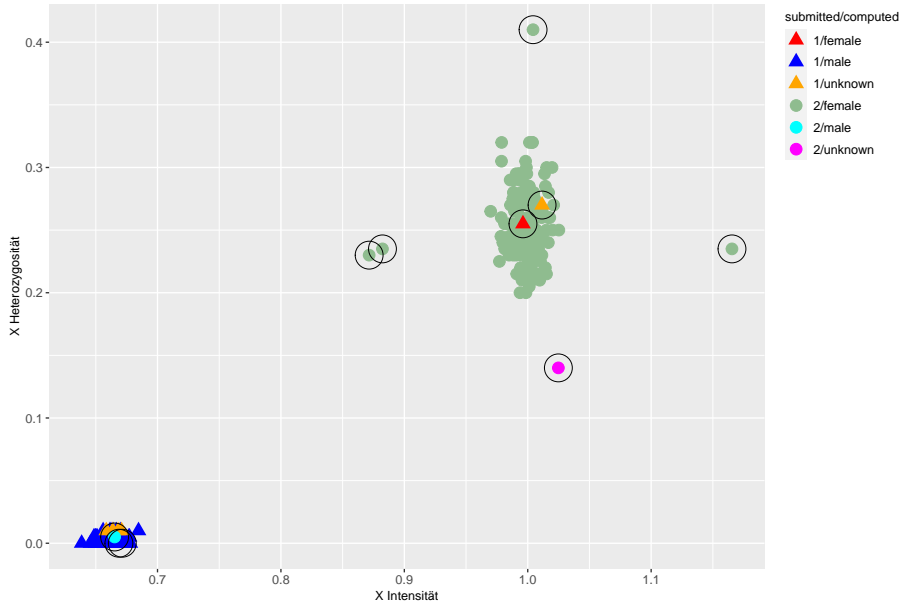
# Aufgabe 2: XY-Intensity Plot

XY Plot mit 300 Samples



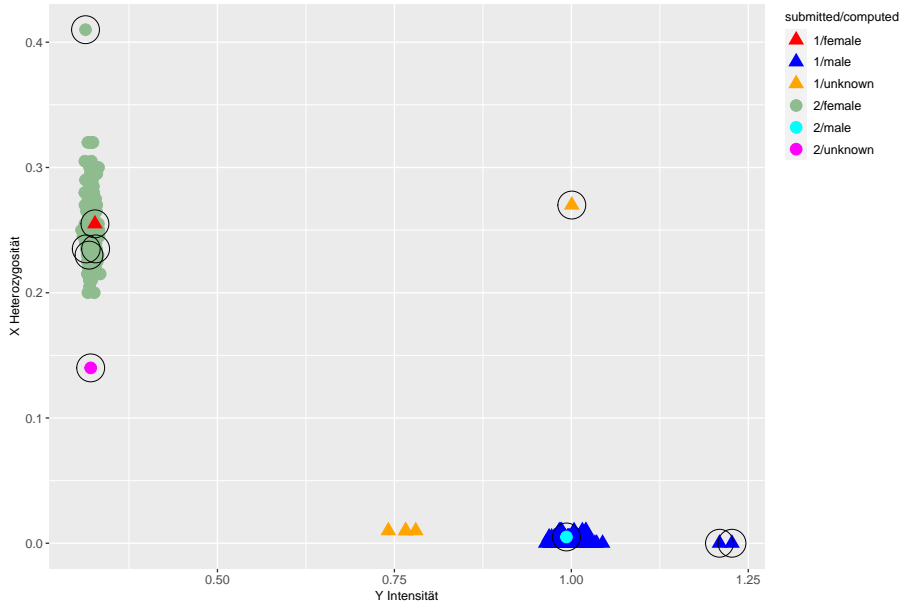
# Aufgabe 2: X-Intensity-Heterozygosity Plot

XX Plot mit 300 Samples



# Aufgabe 2: Y-Intensity-Heterozygosity Plot

YX Plot mit 300 Samples



## Aufgabe 2: XY-Plot - Zusammenfassung

- Welches Grundproblem wird bei einem XY-Plot betrachtet, und warum kann das nur unzureichend gelöst werden?
- Warum werden drei Parameter hier betrachten / warum reichen zwei nicht aus?

## Abschnitt 3

### PCA

# PCA 1 - Datenvorbereitung - SNPs filtern

**Hinweis:** Es sollten am Ende 206,233 SNPs sein!

```
myTab<-read.table("../Exercises_R/data2/mySnps.txt")
rslist<-fread("../Exercises_R/data2/1KG_PCA.bim",
              sep="\t",stringsAsFactors=F)
table(is.element(myTab$V1,rslist$V2))
```

```
##
##  FALSE    TRUE
##  18225 206233
```

```
filt<-is.element(myTab$V1,rslist$V2)
dummy<-as.character(myTab$V1[filt])
write.table(dummy,file="PCA/mySnps_filtered.txt",
            quote=F,row.names=F,col.names=F)
```

## PCA 2 - Datenvorbereitung - Samples filtern

```
fam.data<-read.table("../Exercises_R/data2/1KG_PCA.fam",  
                      stringsAsFactors=F,sep=" ")  
ethno<-substr(fam.data$V2,1,3)  
v.ethno<-c("AFR","ASN","EUR")  
n.ethno<-min(table(ethno)[v.ethno])  
samp.auswahl<-rep(F,length(ethno))  
set.seed(2)  
for(i in v.ethno){  
  samp.auswahl[ethno==i] <- 1:sum(ethno==i) %in%  
    sample(sum(ethno==i),n.ethno)  
}  
table(ethno[samp.auswahl])
```

```
##  
## AFR ASN EUR  
## 246 246 246
```

## PCA 2 - Datenvorbereitung - Samples filtern

**Hinweis:** Es sollten am Ende 3\*246 Individuen sein!

```
fam.data.restr<-fam.data[samp.auswahl,]  
  
write.table(fam.data.restr,file="PCA/mySamples.txt",  
            quote=F,row.names=F,col.names=F)
```



## PCA 3 - Datenvorbereitung - SNPs prunen

**Hinweis:** Es sollten am Ende 117,351 SNPs sein.

```
call1<-paste(plink_call,  
              "--bfile ../Exercises_R/data2/1KG_PCA",  
              "--extract PCA/mySnps_filtered.txt",  
              "--keep PCA/mySamples.txt",  
              "--indep-pairwise 50 5 0.2",  
              "--out PCA/pruning_filter",  
              sep=" ")  
system(call1)
```

## PCA 4 - Datenvorbereitung - Datensatz erstellen

```
call2<-paste(plink_call,  
             "--bfile ../Exercises_R/data2/1KG_PCA",  
             "--extract PCA/pruning_filter.prune.in",  
             "--keep PCA/mySamples.txt",  
             "--make-bed",  
             "--out PCA/pruned_data",  
             sep=" ")  
system(call2)
```

## PCA 5 - Eigentliche PCA berechnen

```
call3<-paste(plink_call,  
             "--bfile PCA/pruned_data",  
             "--pca",  
             "--out PCA/pca_out",  
             sep=" ")  
system(call3)
```

## PCA 6 - PCA auswerten

```
pca2values<-read.table("PCA/pca_out.eigenval")$V1  
pca2vector<-read.table("PCA/pca_out.eigenvec",  
                        stringsAsFactors=F,sep="\t")
```

```
(pca2values[1])/sum(pca2values)
```

```
## [1] 0.50733
```

```
(pca2values[1]+pca2values[2])/sum(pca2values)
```

```
## [1] 0.8610275
```

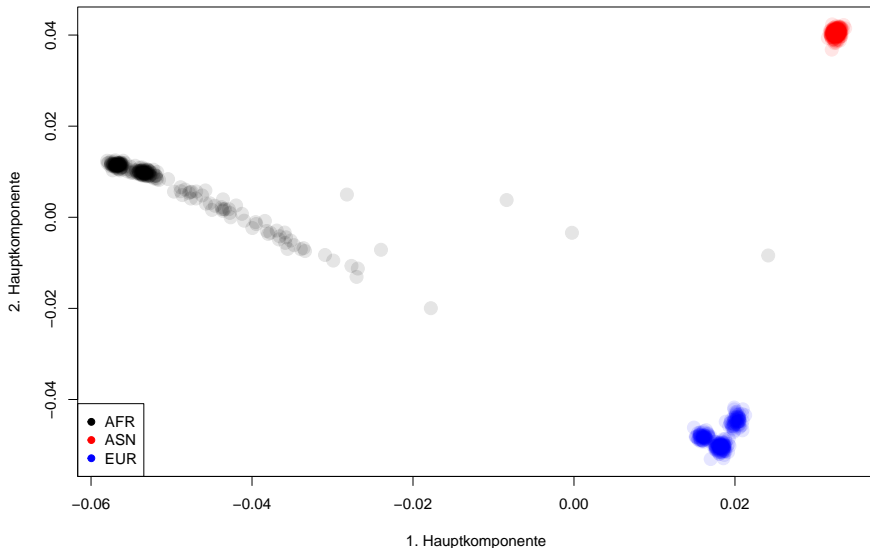
```
xmin<-min(pca2vector[,3]);xmax<-max(pca2vector[,3])  
ymin<-min(pca2vector[,4]);ymax<-max(pca2vector[,4])
```

## PCA 6 - PCA Plot der ersten 2 EVs

```
myMain1="PCA 1000Genomes (3*246 Samples, 121970 geprüfte SNPs)"
plot(0,0,col="white",xlim=c(xmin,xmax),ylim=c(ymin,ymax),
     main=myMain1,
     xlab="1. Hauptkomponente",ylab="2. Hauptkomponente")
lines(pca2vector[substr(fam.data.restr$V2,1,3)=="AFR",c(3,4)],
      col=alpha("black",0.1),type="p",pch=19,cex=1.9)
lines(pca2vector[substr(fam.data.restr$V2,1,3)=="ASN",c(3,4)],
      col=alpha("red",0.1),type="p",pch=19,cex=1.9)
lines(pca2vector[substr(fam.data.restr$V2,1,3)=="EUR",c(3,4)],
      col=alpha("blue",0.1),type="p",pch=19,cex=1.9)
legend("bottomleft",legend=v.ethno,col=c("black","red","blue"))
```

# PCA 6 - PCA Plot der ersten 2 EVs

PCA 1000Genomes (3\*246 Samples, 121970 geprüfte SNPs)

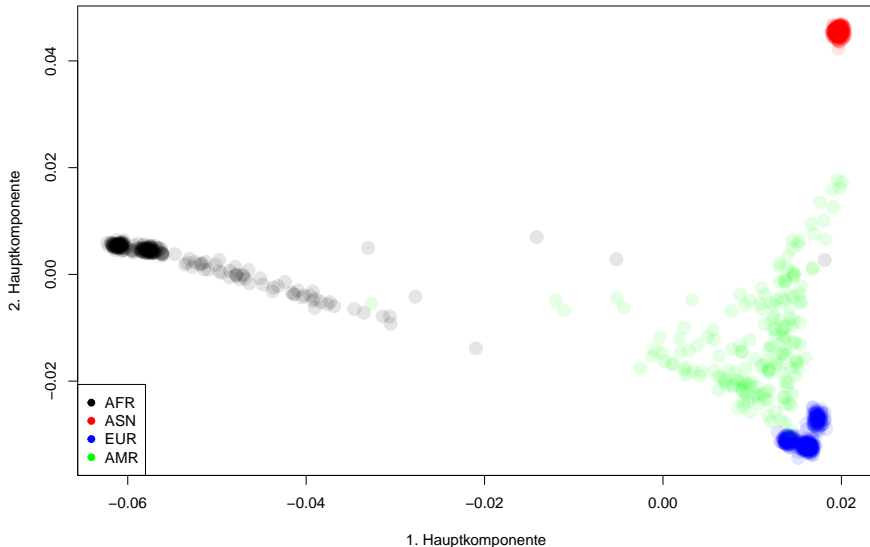


# PCA - Interpretation

- Die ersten zwei Hauptkomponenten trennen die Ethnien auf.
- Beide Vektoren erklären etwa 78% der Varianz in den Genetik-Daten.
- Wenn man das ganz für alle Samples wiederholt erklären die ersten beiden Eigenwerte 84% der genetischen Varianz.

# PCA - Alle Samples

PCA 1000Genomes (1092 Samples, 115204 geprunte SNPs)





# Aufgabe 3: PCA - Zusammenfassung

- Welche Grundannahme ist durch gemischte Populationen verletzt?
- Wie kann das gelöst werden?

## Abschnitt 4

# Heritabilität

# Heritabilität - Definition

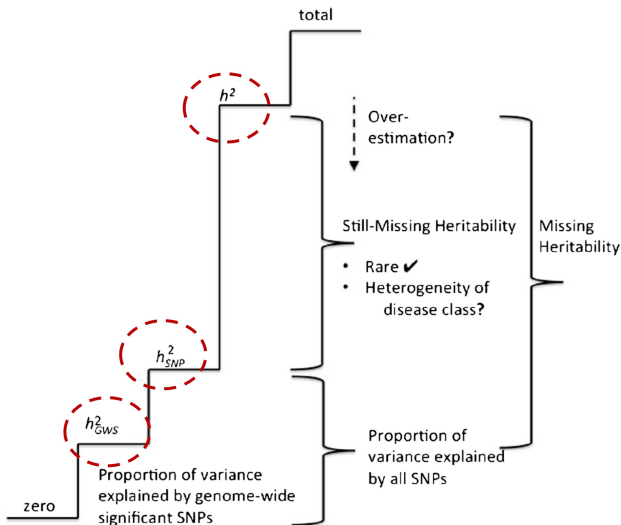
**Heritabilität:** Anteil der Varianz eines Merkmals, der durch die Genetik erklärt wird. Beantwortet in wie fern Gene den Unterschied (Varianz) einer Eigenschaft erklären, **NICHT** welche Gene die Eigenschaft beeinflussen.

$$h^2 = \frac{\text{Var}(\text{Genetik})}{\text{Var}(\text{Merkmal})} = \frac{\text{Var}(\text{Gen.})}{\text{Var}(\text{Gen.}) + \text{Var}(\text{Umw.}) + 2 \cdot \text{Cov}(\text{Gen.}, \text{Umw.})}$$

Einfachste Methode zur Bestimmung von  $h^2$ : Zwillingsstudie & Falconers Formel  $h^2 = 2 \cdot (r(MZ) - r(DZ))$  (Vergleich der Merkmalskonkordanz zwischen monozygoten (MZ) und dizygoten (DZ) Zwilligen).

Alternative: GCTA, LDHub

# Heritabilität - Missing heritability



**Abbildung 2:** Missing heritability

# Heritabilität - Aussage 1

- a) Falls eine Person die Veranlagung einer Krankheit hat, die eine Heritabilität von 1 besitzt, wird diese Person auch die Krankheit erleiden.

# Heritabilität - Aussage 1

- a) Falls eine Person die Veranlagung einer Krankheit hat, die eine Heritabilität von 1 besitzt, wird diese Person auch die Krankheit erleiden.

**Fast immer falsch.** Bsp. Phenylketonurie (PKU, angeborene Stoffwechselstörung, autosomal-rezessiv, >400 Mutationen im Gen Phenylalaninhydroxylase bekannt, Mutationen beeinflussen das Ausmaß der Aktivitätseinschränkung) – hat Heritabilität 1, aber bei geeigneter Diät bricht die Krankheit nicht aus.

# Heritabilität - Aussage 2

- b) Die Heritabilität Finger an jeder Hand zu haben ist 1 (oder fast 1).

## Heritabilität - Aussage 2

- Die Heritabilität Finger an jeder Hand zu haben ist 1 (oder fast 1).

**Falsch**, sie liegt nahe bei 0. Ursache ist hier fast immer Fehlbildungen aufgrund Medikamente / andere Substanzen in der Embryonalphase („Teratogens“) oder Unfälle im Erwachsenenalter



# Heritabilität - Aussage 3

- Die Begriffe „Heritabilität“ und „ererbte“ bedeuten fast das Gegenteil.

## Heritabilität - Aussage 3

- ❑ Die Begriffe „Heritabilität“ und „ererbte“ bedeuten fast das Gegenteil.

**Richtig.** Je mehr ein Merkmal ererbt wird, desto niedriger ist dessen Heritabilität.

## Heritabilität - Aussage 4

- d) In Amerika der 1950er Jahre war die Heritabilität für das Tragen von Ohrringen sehr hoch.

## Heritabilität - Aussage 4

- d) In Amerika der 1950er Jahre war die Heritabilität für das Tragen von Ohrringen sehr hoch.

**Richtig.** Fast nur Frauen haben in dieser Zeit Ohrringe getragen → stellt quasi die Heritabilität vom Geschlecht dar.

# Heritabilität - Aussage 5

- e) Die Heritabilität von eineiigen Zwillingen ist 1.

# Heritabilität - Aussage 5

e) Die Heritabilität von eineiigen Zwillingen ist 1.

**Falsch**, sie haben eine Heritabilität von 0. Jede Variation kommt durch die Umwelt zustande.

# Heritabilität - Aussage 6

- ① Je mehr sich die Umwelt für verschieden Populationen mit unterschiedlicher Heritabilität angleicht, desto höher wird die (Gesamt-)Heritabilität.

# Heritabilität - Aussage 6

- ❶ Je mehr sich die Umwelt für verschieden Populationen mit unterschiedlicher Heritabilität angleicht, desto höher wird die (Gesamt-)Heritabilität.

**Richtig.** Je ähnlicher die Umwelt wird, desto niedriger wird deren Varianz und der Anteil der Genetik steigt.



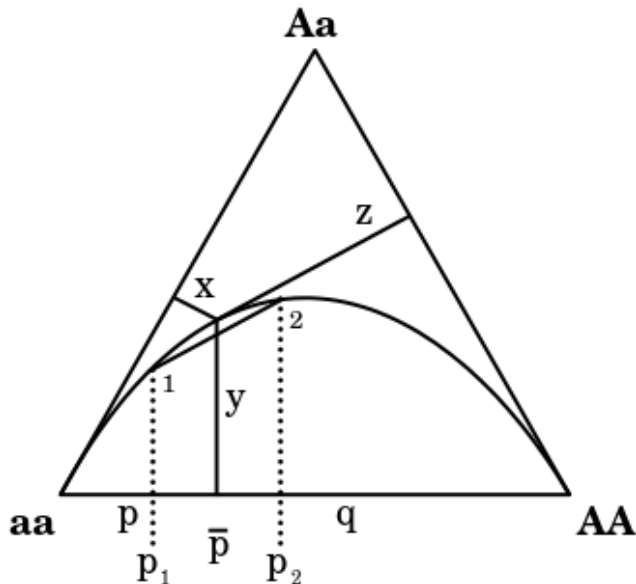
# Aufgabe A1: Heritabilität - Zusammenfassung

- Wie kann Heritabilität geschätzt werden
- Wodurch entsteht die “missing heritability”?

## Abschnitt 5

# Populationsgenetik

# Populationsgenetik - Hintergrund



# Populationsgenetik - Aufgabe

- a) Bestimmung von  $p_i$  und  $q_i$
- b) Berechnung von Inzuchtskoeffizient  $F_i$
- c) Warum Varianz = Heterozygotität?
- d) Bestimmung von  $H_I$ ,  $H_S$  und  $H_T$
- e) Berechnung des Fixationsindex  $F_{ST}$
- f) Interpretation

Genotyp	AA	AB	BB
Population 1	125	250	125
Population 2	50	30	20
Population 3	100	500	400

## Populationsgenetik - Lösung a)

$$p = \begin{pmatrix} (2AA_1 + AB_1)/2n_1 \\ (2AA_2 + AB_2)/2n_2 \\ (2AA_3 + AB_3)/2n_3 \end{pmatrix} = \begin{pmatrix} 500/1000 \\ 130/200 \\ 700/2000 \end{pmatrix} = \begin{pmatrix} 0.5 \\ 0.65 \\ 0.35 \end{pmatrix}, q = 1 - p = \begin{pmatrix} 0.5 \\ 0.35 \\ 0.65 \end{pmatrix}$$

$$\begin{aligned} \bar{p} &= \frac{2 \cdot (AA_1 + AA_2 + AA_3) + (AB_1 + AB_2 + AB_3)}{2 \cdot (n_1 + n_2 + n_3)} \\ &= \frac{2 \cdot 275 + 780}{2 \cdot 1600} = 0.416 \\ \bar{q} &= 0.584 \end{aligned}$$

# Populationsgenetik - Lösung b)

Beobachtete Heterozygotität:

$$p_{obs}(AB) = \begin{pmatrix} AB_1/n_1 \\ AB_2/n_2 \\ AB_3/n_3 \end{pmatrix} = \begin{pmatrix} 250/500 \\ 30/100 \\ 500/1000 \end{pmatrix} = \begin{pmatrix} 0.5 \\ 0.3 \\ 0.5 \end{pmatrix}$$

Erwartete Heterozygotität:

$$p_{exp}(AB) = 2 \cdot p \cdot q = \begin{pmatrix} 2 \cdot 0.5 \cdot 0.5 \\ 2 \cdot 0.65 \cdot 0.35 \\ 2 \cdot 0.35 \cdot 0.65 \end{pmatrix} = \begin{pmatrix} 0.5 \\ 0.455 \\ 0.455 \end{pmatrix}$$

Inzuchtskoeffizient:

$$F = \frac{p_{exp}(AB) - p_{obs}(AB)}{p_{exp}(AB)} = \begin{pmatrix} (0.5 - 0.5)/0.5 \\ (0.455 - 0.3)/0.455 \\ (0.455 - 0.5)/0.455 \end{pmatrix} = \begin{pmatrix} 0 \\ 0.34 \\ -0.10 \end{pmatrix}$$

# Populationsgenetik - Lösung c)

- Binomialverteilung  $B(k|n, p)$ :
- Allel A zählt als Erfolg, Allel B als Misserfolg.
- $n=2$ , weil pro Genotyp zweimal gezogen
- Erfolgswahrscheinlichkeit entspricht der Allelfrequenz ( $p$ ).
- Bei zwei Treffern ( $k = 2$ ) erhält man den Genotyp AA ( $P(AA) = B(2|2, p)$ )
- Die Varianz unter Binomialverteilung ist immer  $Var(X) = n \cdot p \cdot q = 2pq = p_{exp}(AB)$  im HWE.

## Populationsgenetik - Lösung d) & e)

$$H_I = \frac{p_{ops}^T \cdot n}{N_{total}} = \frac{250 + 30 + 500}{1600} = 0.4875$$

$$H_S = \frac{p_{exp}^T \cdot n}{N_{total}} = \frac{0.5 \cdot 500 + 0.455 \cdot 100 + 0.455 \cdot 1000}{1600} = 0.470$$

$$H_T = 2 \cdot \bar{p} \cdot \bar{q} = 0.486$$

$$F_{ST} = 1 - \frac{H_S}{H_T} = 0.034$$

$$F_{IT} = 1 - \frac{H_I}{H_T} = -0.0031$$



# Populationsgenetik - Lösung f)

Interpretation:

- Population 1 ist im HWE
- Population 2 hat weniger Heterozygote als erwartet → Hinweis für *inbreeding* (Inzucht; Verletzung von HWE weil keine zufällige Partnerwahl, sondern eher Verwandte)
- Population 3 hat mehr Heterozygote als erwartet → Hinweis für *outbreeding* (Auszucht; Verletzung von HWE weil keine zufällige Partnerwahl, sondern alle Verwandten ausgeschlossen)
- Subpopulationen sind für etwa 3.4% der gesamten genetischen Variation verantwortlich
- Die Gesamtpopulation zeigt keine Anzeichen für Inzucht

# Aufgabe A1: Populationsgenetik - Zusammenfassung

- Wie kann Heritabilität geschätzt werden
- Wodurch entsteht die “missing heritability”?

# Abschnitt 6

## Zusammenfassung

# Zusammenfassung

- Warum kann PCA und Verwandtschaft zu Adjustierung auf Stratifikationsbias genutzt werden?
- Welche Ausreißer in einem XY-Plot müssen gefiltert werden?
- Was ist Heritabilität?