

# Twin analysis

Klaus Holst & Thomas Scheike

February 18, 2020

## Mets package

This document provides a brief tutorial to analyzing twin data using the **mets** package:

```
1 library("mets")
2 options(warn=-1)
```

The development version may be installed from *github*, i.e., with the *devtools* package:

```
1 devtools::install_github("kkholst/lava")
2 devtools::install_github("kkholst/mets")
```

## Twin analysis, continuous traits

In the following we examine the heritability of Body Mass Index<sup>1 2</sup>, based on data on self-reported BMI-values from a random sample of 11,411 same-sex twins. First, we will load data

```
1 data("twinbmi")
2 head(twinbmi)
```

	tvparnr	bmi	age	gender	zyg	id	num
1	1	26.33289	57.51212	male	DZ	1	1
2	1	25.46939	57.51212	male	DZ	1	2
3	2	28.65014	56.62696	male	MZ	2	1
5	3	28.40909	57.73097	male	DZ	3	1
7	4	27.25089	53.68683	male	DZ	4	1
8	4	28.07504	53.68683	male	DZ	4	2

The data is on *long* format with one subject per row. we transpose the data allowing us to do pairwise analyses

```
1 twinwide <- fast.reshape(twinbmi, id="tvparnr",varying=c("
  bmi"))
2 head(twinwide)
```

	tvparnr	bmi1	age	gender	zyg	id	num	bmi2
1	1	26.33289	57.51212	male	DZ	1	1	25.46939
3	2	28.65014	56.62696	male	MZ	2	1	NA
5	3	28.40909	57.73097	male	DZ	3	1	NA
7	4	27.25089	53.68683	male	DZ	4	1	28.07504
9	5	27.77778	52.55838	male	DZ	5	1	NA
11	6	28.04282	52.52231	male	DZ	6	1	22.30936

Next we plot the association within each zygosity group

<sup>1</sup> M. Korkeila, J. Kaprio, A. Rissanen, and M. Koskenvuo. Effects of gender and age on the heritability of body mass index. *Int J Obes*, 15(10):647–654, Oct 1991

<sup>2</sup> J. v. Hjelmborg, C. Fagnani, K. Silventoinen, M. McGue, M. Korkeila, K. Christensen, A. Rissanen, and J. Kaprio. Genetic influences on growth traits of BMI: a longitudinal study of adult twins. *Obesity (Silver Spring)*, 16(4):847–852, Apr 2008

**tvparnr** twin id

**bmi** Body Mass Index (kg/m<sup>2</sup>)

**age** Age (years)

**gender** Gender factor (male,female)

**zyg** zygosity (MZ,DZ)

---

```

1 library("cowplot")
2
3 scatterdens <- function(x) {
4   sp <- ggplot(x,
5     aes_string(colnames(x)[1], colnames(x)[2])) +
6   theme_minimal() +
7   geom_point(alpha=0.3) + geom_density_2d()
8   xdens <- ggplot(x, aes_string(colnames(x)[1], fill=1)) +
9   theme_minimal() +
10  geom_density(alpha=.5)+
11  theme(axis.text.x = element_blank(),
12    legend.position = "none") + labs(x=NULL)
13  ydens <- ggplot(x, aes_string(colnames(x)[2], fill=1)) +
14  theme_minimal() +
15  geom_density(alpha=.5) +
16  theme(axis.text.y = element_blank(),
17    axis.text.x = element_text(angle=90, vjust=0),
18    legend.position = "none") +
19  labs(x=NULL) +
20  coord_flip()
21  g <- plot_grid(xdens, NULL, sp, ydens,
22    ncol=2, nrow=2,
23    rel_widths=c(4, 1.4), rel_heights=c(1.4, 4))
24  return(g)
25 }

```

---

We here show the log-transformed data which is slightly more symmetric and more appropriate for the twin analysis (see Figure 1 and 2)

---

```

1 mz <- log(subset(twinwide, zyg=="MZ"), c("bmi1", "bmi2"))
2 scatterdens(mz)

```

---

```

1 dz <- log(subset(twinwide, zyg=="DZ"), c("bmi1", "bmi2"))
2 scatterdens(dz)

```

---

The plots and raw association measures shows considerable stronger dependence in the MZ twins, thus indicating genetic influence of the trait

---

```

1 cor.test(mz[,1], mz[,2], method="spearman")

```

---

Spearman's rank correlation rho

```

data:  mz[, 1] and mz[, 2]
S = 165460000, p-value < 2.2e-16
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.6956209

```

---

```

1 cor.test(dz[,1], dz[,2], method="spearman")

```

---

Spearman's rank correlation rho

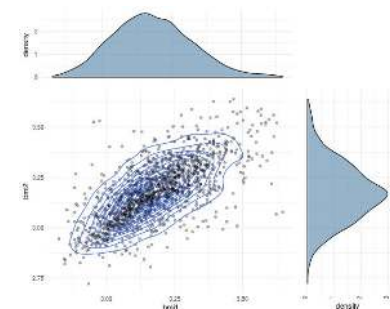


Figure 1: Scatter plot of logarithmic BMI measurements in MZ twins.

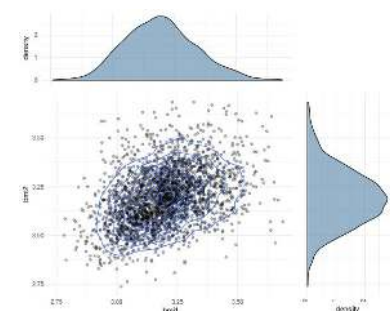


Figure 2: Scatter plot of logarithmic BMI measurements in DZ twins.

```
data: dz[, 1] and dz[, 2]
S = 2162500000, p-value < 2.2e-16
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.4012686
```

Next we examine the marginal distribution (GEE model with working independence)

```
1 l0 <- lm(bmi ~ gender + I(age-40), data=twinbmi)
2 estimate(l0, id=twinbmi$tvparnr)
```

---

	Estimate	Std.Err	2.5%	97.5%	P-value
(Intercept)	23.3687	0.054534	23.2618	23.4756	0.000e+00
gendermale	1.4077	0.073216	1.2642	1.5512	2.230e-82
I(age - 40)	0.1177	0.004787	0.1083	0.1271	1.499e-133

---

```
1 library("splines")
2 l1 <- lm(bmi ~ gender*ns(age,3), data=twinbmi)
3 marg1 <- estimate(l1, id=twinbmi$tvparnr)
```

---

```
1 dm <- Expand(twinbmi,
2   bmi=0,
3   gender=c("male"),
4   age=seq(33,61,length.out=50))
5 df <- Expand(twinbmi,
6   bmi=0,
7   gender=c("female"),
8   age=seq(33,61,length.out=50))
9
10 plot(marg1, function(p) model.matrix(l1,data=dm)%*%p,
11   data=dm["age"], ylab="BMI", xlab="Age",
12   ylim=c(22,26.5))
13 plot(marg1, function(p) model.matrix(l1,data=df)%*%p,
14   data=df["age"], col="red", add=TRUE)
15 legend("bottomright", c("Male","Female"),
16   col=c("black","red"), lty=1, bty="n")
```

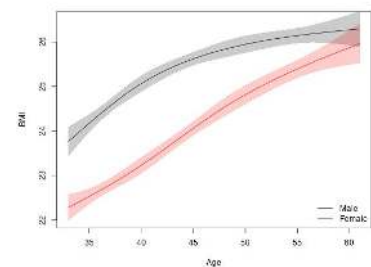


Figure 3: ...

### Polygenic model

Decompose outcome into

$$Y_i = A_i + D_i + C + E_i, \quad i = 1, 2$$

*A* Additive genetic effects of alleles

*D* Dominant genetic effects of alleles

*C* Shared environmental effects

*E* Unique environmental genetic effects

Dissimilarity of MZ twins arises from unshared environmental effects only!  $\text{Cor}(E_1, E_2) = 0$  and

$$\text{Cor}(A_1^{MZ}, A_2^{MZ}) = 1, \quad \text{Cor}(D_1^{MZ}, D_2^{MZ}) = 1,$$

$$\text{Cor}(A_1^{DZ}, A_2^{DZ}) = 0.5, \quad \text{Cor}(D_1^{DZ}, D_2^{DZ}) = 0.25,$$

$$Y_i = A_i + C_i + D_i + E_i$$

$$A_i \sim \mathcal{N}(0, \sigma_A^2), C_i \sim \mathcal{N}(0, \sigma_C^2), D_i \sim \mathcal{N}(0, \sigma_D^2), E_i \sim \mathcal{N}(0, \sigma_E^2)$$

$$\mathbf{Cov}(Y_1, Y_2) =$$

$$\begin{pmatrix} \sigma_A^2 & 2\Phi\sigma_A^2 \\ 2\Phi\sigma_A^2 & \sigma_A^2 \end{pmatrix} + \begin{pmatrix} \sigma_C^2 & \sigma_C^2 \\ \sigma_C^2 & \sigma_C^2 \end{pmatrix} + \begin{pmatrix} \sigma_D^2 & \Delta_7\sigma_D^2 \\ \Delta_7\sigma_D^2 & \sigma_D^2 \end{pmatrix} + \begin{pmatrix} \sigma_E^2 & 0 \\ 0 & \sigma_E^2 \end{pmatrix}$$

---

```

1 dd <- na.omit(twinbmi)
2 l0 <- twinlm(bmi ~ age+gender, data=dd,
3             DZ="DZ", zyg="zyg", id="tvparnr", type="sat")

```

---

```

1 l <- twinlm(bmi ~ ns(age,1)+gender, data=twinbmi,
2            DZ="DZ", zyg="zyg", id="tvparnr", type="cor", missing=
3            TRUE)
4 summary(l)

```

---

```

-----
Group 1

```

	Estimate	Std. Error	Z value	Pr(> z )
Regressions:				
bmi.1~ns(age, 1).1	4.16937	0.16669	25.01334	<1e-12
bmi.1~gendermale.1	1.41160	0.07284	19.37839	<1e-12
Intercepts:				
bmi.1	22.53618	0.07296	308.87100	<1e-12
Additional Parameters:				
log(var)	2.44580	0.01425	171.68256	<1e-12
atanh(rhoMZ)	0.78217	0.02290	34.16186	<1e-12

```

-----
Group 2

```

	Estimate	Std. Error	Z value	Pr(> z )
Regressions:				
bmi.1~ns(age, 1).1	4.16937	0.16669	25.01334	<1e-12
bmi.1~gendermale.1	1.41160	0.07284	19.37839	<1e-12
Intercepts:				
bmi.1	22.53618	0.07296	308.87100	<1e-12
Additional Parameters:				
log(var)	2.44580	0.01425	171.68256	<1e-12
atanh(rhoDZ)	0.29924	0.01848	16.19580	<1e-12

```


```

	Estimate	2.5%	97.5%
Correlation within MZ:	0.65395	0.62751	0.67889
Correlation within DZ:	0.29061	0.25712	0.32341

```

'log Lik.' -29020.12 (df=6)
AIC: 58052.24
BIC: 58093.29

```

A formal test of genetic effects can be obtained by comparing the MZ and DZ correlation:

---

```

1 estimate(1,contr(5:6,6))

```

---

```

                                Estimate Std.Err   2.5%  97.5%   P-value
[atanh(rhoMZ)@1] - [a....    0.4829 0.04176 0.4011 0.5648 6.325e-31

```

Null Hypothesis:

```
[atanh(rhoMZ)@1] - [atanh(rhoDZ)@3] = 0
```

```

1  l <- twinlm(bmi ~ ns(age,1)+gender, data=twinbmi,
2      DZ="DZ", zyg="zyg", id="tvparnr", type="cor", missing=
      TRUE)
3  summary(l)

```

-----

Group 1

	Estimate	Std. Error	Z value	Pr(> z )
Regressions:				
bmi.1~ns(age, 1).1	4.16937	0.16669	25.01334	<1e-12
bmi.1~gendermale.1	1.41160	0.07284	19.37839	<1e-12
Intercepts:				
bmi.1	22.53618	0.07296	308.87100	<1e-12
Additional Parameters:				
log(var)	2.44580	0.01425	171.68256	<1e-12
atanh(rhoMZ)	0.78217	0.02290	34.16186	<1e-12

-----

Group 2

	Estimate	Std. Error	Z value	Pr(> z )
Regressions:				
bmi.1~ns(age, 1).1	4.16937	0.16669	25.01334	<1e-12
bmi.1~gendermale.1	1.41160	0.07284	19.37839	<1e-12
Intercepts:				
bmi.1	22.53618	0.07296	308.87100	<1e-12
Additional Parameters:				
log(var)	2.44580	0.01425	171.68256	<1e-12
atanh(rhoDZ)	0.29924	0.01848	16.19580	<1e-12

	Estimate	2.5%	97.5%
Correlation within MZ:	0.65395	0.62751	0.67889
Correlation within DZ:	0.29061	0.25712	0.32341

'log Lik.' -29020.12 (df=6)

AIC: 58052.24

BIC: 58093.29

*Twin analysis, censored outcomes*

*Twin analysis, binary traits*

*Time to event*

*References*

- [1] J. v. Hjelmborg, C. Fagnani, K. Silventoinen, M. McGue, M. Korkeila, K. Christensen, A. Rissanen, and J. Kaprio. Genetic influences on growth traits of BMI: a longitudinal study of adult twins. *Obesity (Silver Spring)*, 16(4):847–852, Apr 2008.
- [2] M. Korkeila, J. Kaprio, A. Rissanen, and M. Koskenvuo. Effects of gender and age on the heritability of body mass index. *Int J Obes*, 15(10):647–654, Oct 1991.