HÖGSKOLAN DALARNA

School of technology and business studies

Comparison of the heritability values on pig's teat number in two methods

Supervisor: Majbritt.Felleki

Wei Jiang   v10weiji@du.se

Meina Cheng   v10meich@du.se

2011/06/06

Abstract

Teat number is an important feature in pig production since each piglet needs access to its own teat, due to a complex nursing behavior between the piglets and the sow. The number of teats can vary, but fourteen is most common. There are many factors that may effect on the teat number, some are belong to the environmental ingredient, and others may pertain to the genetic influence. The main aim of this essay is comparing the heritability values from different models to detect how the gene influences on the teat number. We apply some specific packages in R software to the real data to get the heritability value.

Keywords: Teat number; Heritability value; Linear mixed model

Content

HÖGSKOLAN DALARNA
School of technology and business studies

1 Introduction

1.1 Background

Animal breeding It is a common problem that in animal breeding process is to distinguish between effects of environment and effects of genes. As an example one can consider a very small animal and ask if it is small because of its genes or because of its environment, which could include sex, food, herd and so on. Often the environment is the same for related individuals, which makes hard to see the difference between environment effects and gene effects. For instance, the piglets from the same litter, they must share the same environment; hence the different traits among the siblings may partly come from the gene effect.

Heritability value And a common value that we used to express gene effects is the fraction of variance explained by genetics. This is called heritability value. The concept was proposed by Lush, J. L. (1949). Phenotypic variation among individuals is consisted of two independent different parts, the one is genetic factors, and the other is the environment factors. It is the proportion that the variance due to the genetic factors occupies the total phenotypic variance in a population. Heritability estimates the relative contributions of differences in genetic and non-genetic factors that we mentioned above as environmental factors to the total phenotypic variance in a population. It is widely used particularly in selective breeding and behavior genetics. An example of high heritability: a population with genotypes coding for only one hair color of black in China.

Any particular phenotype can be modelled as the sum of genetic and environmental (non-genetic) effects:

Phenotype (P) = Genotype (G) + Environment (E).

Likewise the variance in the trait $Var(P)$ is the sum of genetic effects as follows:

$$Var(P) = Var(G) + Var(E) + 2 * Cov(G, E).$$

The covariance term $Cov(G, E)$ always is assumed to 0. In this case, heritability is defined as:

HÖGSKOLAN DALARNA
School of technology and business studies

$$h^2 = \frac{Var(G)}{Var(E) + Var(G)}$$

The larger the heritability value is the stronger affection due to the genetic factors. It must be in the closed interval between 0 and 1, that is owing to at most offspring getting all the genetic characteristic from their parents，at this time there is no non-genetic effects, then $h^2 = 1$, obviously this situation does not exist in the real world, we just consider the theoretical value, or in contrast, at least posterity having no common traits in gene with their ancestors then $h^2 = 0$. [See (6) Lush J.L (1949).]

1.2 Francis Galton[1]'s (1886) Plot

Francis Galton did some researches about the genetic influence on the height trait from parents to their offspring. And he produced a line about the relationship between parents' height and children's height. When Galton did the parents-offspring regression, he found the slope of the line (0.67) approximates the heritability of the trait. And we can see from the plot, the middle line with mark M is a boundary, if mid-parents height above this boundary, then their children will lower than them and vice versa. And if we assume that the children also have their own offspring then their offspring's height will obey this rule have a lower slope than their previous generation, then generation after generation the height line will approximate the middle line.
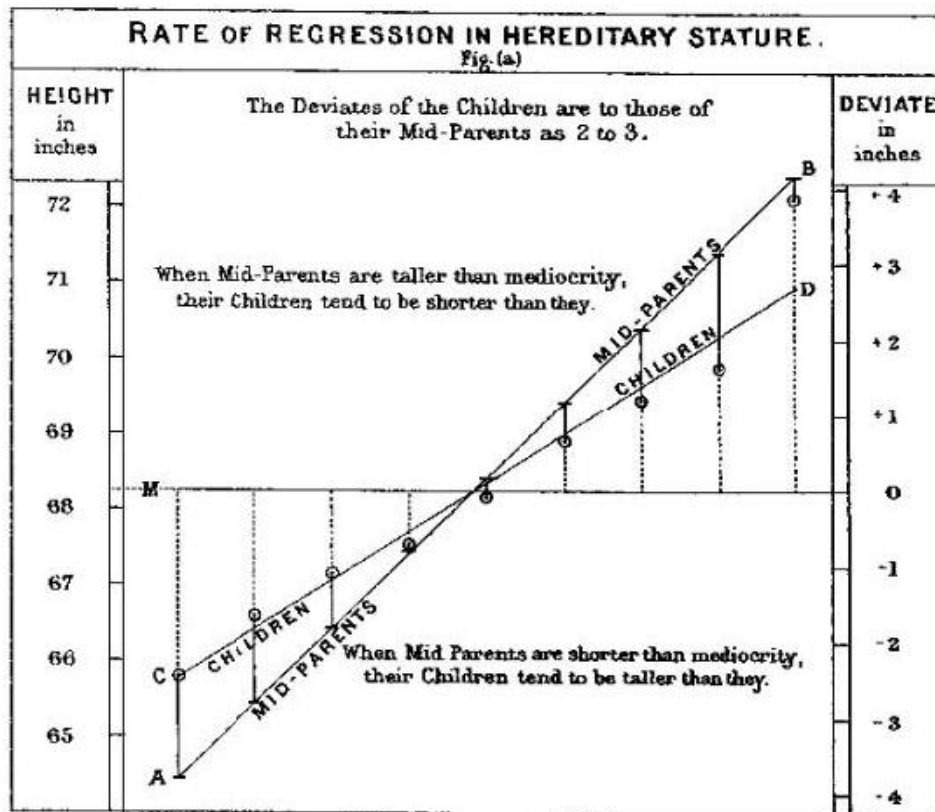
HÖGSKOLAN DALARNA
School of technology and business studies

Fig.1 Galton's Plot

1.3 Aim and interest

Genetic and environmental factors are important in human evolution and development process. Distinguishing between environmental factors and genes influence on the characteristics of offspring, we may know how to get species with better characteristics or lager litter size, and so on. It is impossible to totally separate genetic and environmental effects through a statistic model. In our essay we want to evaluate the heritability value (using the software R) from two different points of view, one is using a graphical model (do the similar plot as Galton), the other is doing regression with an sire model with the breeding information. The aim is to compare these two values, and try to get some conclusion from the similar or different aspects.

2 Materials and Methods

2.1 Data

Our data was provided by Norvdiec Genetics and data was done unidentifiable by our

supervisor M.Felleki[2] because of discrepancies, grouping of small and big litter sizes and of high parities was done, as well as removing wrong observations that teat numbers below 10 and above 19. And then we have data fill consisted of 18461 animals with their identities, litter identity, birthday, sex, herd, litter size, parity and teat number. The 18461 observations were selected from 2774 litters with 728 different birthdays. And there are 9880 males and 8581 females. The factor litter size is from 1 to 14, parity is from 1 to 4, and teat number is from 10 to 19. And we have another data frame which included individuals' identity number, father's identity and mother's identity.

2.2 Linear mixed model

Our Data was analyzed using the method of linear regression and linear mixed model. A linear mixed model (LMM) is a particular type of mixed model and also an extension to the linear model. In the linear mixed model, the main part linear predictor contains not only the usual fixed effects but also the random effects. These random effects are usually assumed to have a normal distribution. The fixed effect in statistics, represent the factors which are fixed for all individuals. And the random effects reflect the group information, the observations which are from the same group must have some specific traits, in our essay we used the genetic effects to be the random effect term, so we can see clearly that how genes affect on the pig's teat number. [See (2) Lundeheim, N., Chalkias, H., Rydhmer, L (2010).]

2.3 Animal model

An animal model is the application of linear mixed model in animal breeding process. The model is:

$$y = X\beta + Za + e \qquad (2.3.1)$$

Where, X is known, fixed, $n \times p$ matrix with rank $= r \leq \text{minimum}(n, p)$.

β is fixed, $p \times 1$ vector generally known.

Z is known, fixed, $n \times q$ matrix with elements of 0 or 1.

a is a random, $q \times 1$ vector. $u \sim N(0, G)$

e is a random, $n \times 1$ vector. $e \sim N(0, R)$

HÖGSKOLAN DALARNA
School of technology and business studies

The variance-covariance matrix of a is G, a $q \times q$ symmetric matrix that is usually non-singular. It is also called a relationship matrix or animal matrix. Hereafter we shall use the notation of $Var(a)$ to present the variance-covariance matrix of the random effects.

$Var(e) = R$ is an $n \times n$ symmetric matrix usually non-singular, and we also assumed that $Cov(u, e^{'}) = 0$.

And we can calculate the statistical values of y:

$$E(y) = X\beta$$

$$Var(y) = ZGZ^{'} + R$$

From these two values we can easily do the variance components analysis in the results part. [See (4) Henderson, C.R. (1984).]

So that the heritability value in animal model is that:

$$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2}$$

2.4 Sire model

The application of a sire model implies that only sires are being evaluated using progeny records. Most early applications of best linear unbiased prediction of the breeding values were based on a sire model. The main advantage with a sire model is that the number of equations is reduced compared with an animal model since only sires are evaluates. However, with a sire model, the genetic merit of the mate (dam of progeny) is not accounted for. It is assumed that all mates are of similar genetic merit and this can result in bias in the predicted breeding values if there is preferential mating. But in our essay the main point is to fit the model and get the heritability value, not the prediction.

The sire model in matrix notation is:

HÖGSKOLAN DALARNA
School of technology and business studies

$$y = Xb + Zs + e \qquad (2.4.1)$$

All terms in this formula are defined for (1) and s is the vector of random sire effects, Z now relates records to sires and:

$$\text{var}(s) = A\sigma_s^2$$

$$\text{var}(y) = ZAZ'\sigma_s^2 + R$$

Where A is the numerator relationship matrix for sires, $\sigma_s^2 = 0.25\sigma_a^2$ and $R = I\sigma_e^2$. So that the heritability value in sire model is that: [See (5) Mrode , R. A., Thompson, R. (2005).]

$$h^2 = \frac{4\sigma_s^2}{4\sigma_s^2 + \sigma_e^2} \qquad (2.4.2)$$

2.5 Relationship matrix

Estimating genetic value from family data: One of the largest and most successful applications of the mixed models is in the animal breeding, where the main objective is to identify animals of high genetic merit. The genetic merit of animal is modeled as a random parameter, and the identification process involves estimation of this parameter. Some analytical complexity is introduced by the different degrees of relationship between the animals. A typical application of animal breeding, for example to identify dairy pigs that have high genetic merits, involves thousands of animals with hundreds of thousands of records. We will show the process as follows, let y represents the response factor, in our essay, y stands for the teat number of the observations. And we can get the relationship matrix in theoretical way like that:

1. First, the relationship matrix is a symmetric matrix.

2. For the diagonal elements:

$$r_{ii} = 1 + \frac{1}{2}r_{sire,dam}$$

Where 'sire' and 'dam' are those of the i'th animal. The second term on the right-hand side takes account of inbreeding，if there is no inbreeding then $r_{ii} = 1$.

1. For off-diagonal elements r_{ij}, where j is of an older generation compared with i and

• Both parents of i are known

$$r_{ij} = \frac{1}{2}\left(r_{j,sire} + r_{j,dam}\right)$$

• Only one parent of i is known

$$r_{ij} = \frac{1}{2}\left(r_{j,sire}\right)$$

or

$$r_{ij} = \frac{1}{2}\left(r_{j,dam}\right)$$

• Both parents of i are unknown

$$r_{ij} = 0$$

And then we can get the relationship matrix easily. Furthermore, we can estimate the $\hat{\beta}$ and $\hat{a}$ in theoretical based on the $E(Y|b) = X\beta + Zb$. [See (3) Pawitan, Yudi (2001).]

2.6 The first stage

The number of teats is fitted by the linear model:

$$y = \mu + x\beta + \varepsilon$$

and in the beginning x includes the effects of six independent factors which are herd, sex, birthday, litter size, litter identity and parity and we can find them from the data

which we have got, so we do the following model to check the result.

So we try to fit the first linear model with all six factors in R:

model1 <- lm(teat ~ as.factor(birthday) + as.factor(litter_id) + as.factor(sex) + as.factor(litter_size) + as.factor(parity) + as.factor(herd), data = data_teats)

We got the results from R with a lot of missing value in the coefficients of birthday term and litter identity term, and it took so long time to operate this model. Then we have to do the variable selecting to these factors due to some of these factors may with so many levels that we cannot get the ideal result. So if we want to get the reasonable interpretation, we have to select some factors from them.

Then we decided to delete the factor birthday of the individuals and their litter identity, though we had the way to deal with them to become a fewer level form, we also have reason to believe that these two factors seem not to be the important effects, they are inessential.

Thereafter we selected the remaining four factors as the independent variables in the linear model for getting a better result in R:

model2 <- lm(teat ~ as.factor(sex) + as.factor(litter_size) + as.factor(parity) + as.factor(herd), data = data_teats);

This is the first step in our model, next we want to start the second step, we have to select the offspring whom parents' identities exist in the second data frame and these parents' identities also could be found in the identity of the first data frame, if we could pick out this kind offspring and parents' identities, we could get their residuals from the linear model easily. Following, we start to calculate the value called corrected teat numbers which equals to the intercept plus the residual term.

Firstly the corrected teat numbers of offspring's:

$$n = y_i - x_i\beta = \mu + e_i \qquad (2.6.1)$$

While the same, we calculate the parents' corrected teat numbers from the same model

with the function:

$$m = \frac{y_j - x_j\beta + y_k - x_k\beta}{2} = \frac{\mu + e_j + \mu + e_k}{2} = \mu + \frac{e_j + e_k}{2} \qquad (2.6.2)$$

Which j, k represent the dam and sire of individual i.

After finishing these steps, we will plot the pairs (m, n), and added the fitted the line of the linear model, y axis stands for the corrected offspring teat numbers and x axis stands for the average corrected parents' teat numbers.

Based on the pairs $\left(\dfrac{y_j - x_j\beta + y_k - x_k\beta}{2}, y_i - x_i\beta\right)$, we plotted the relationship figure between offspring and their parents.

After the first stage, we should start the second stage, in our next step, we use the animal model to analyses the teat numbers.

2.7 The second stage

In our essay, we use the sire animal model likes this:

$$y = x\beta + zs + e$$

Based on what we talked above, the number of teats y is fitted by this animal model; the vector β contains the fixed effects of four independent factors which are herd, sex, litter size, parity. The vector s is a genetic random effect which corresponded to the father identities in the pedigree; we choose the father identities to be the random effect, because father identities can tell us how much genetic merit the offspring shared from their parents. We group the observations with the father identities so that differences among groups must caused by the genetic factors. [See (1) Felleki, M., Chalkias, H. (2010).]

And try to evaluate the effect of genetic ingredient without thinking about the environmental factor (litter identity).

Afterwards we must do the regression with the breeding information that means trying to fit an animal model with father identity as random effect so it includes the relationship matrix and genetic information.

In this step we use a new package in R, which is "pedigreemm". While the relationship matrix is pretty large so that it could not be stored in R (because the ped data is very large). Thus we have to deal with the pedigree data with the method of "last father", that is a method help us delete some individuals whose identity numbers are larger than the max father identity in sire column. After that, we could use this package in R to fit the animal model.

We fitted the model in R:

model3<-pedigreemm(data_teats$teat~as.factor(data_teats$sex)+as.factor(data_teat s$litter_size)+as.factor(data_teats$parity)+as.factor(data_teats$herd) + (1 | SireID), pedigree = list(SireID = ped))

Next we choose the litter identity as the second random effect. So in the next model, there has two random effect, one is father identity, the other one is litter identity. So the model is as follows:

$$y = x\beta + wp + zs + e$$

In this model, we add the other random effect, the vector p is a permanent environmental random effect which corresponded to the litter identities, we choose the litter identities to be the random effect firstly that due to the individuals with the same litter identities meaning that they also having the same ancestor and environment, both father and mother, they should have some same genes, approximate half. We group the observations with the litter identities so that differences among groups must caused by the environmental factors. [See (1) Felleki, M., Chalkias, H. (2010).]

We fitted the model in R:

model4 <- pedigreemm(data_teats$teat ~ as.factor(data_teats$sex) + as.factor(data_teats$litter_size) + as.factor(data_teats$parity) + as.factor(data_teats$herd) + (1| data_teats$litter_id) + (1 | SireID), pedigree

HÖGSKOLAN DALARNA
School of technology and business studies

= list(SireID = ped))

3 Results and Model checking

3.1 The first stage

Based on the linear regression model (model2), we had got the part of coefficients in the table below:

Table 1: the part coefficients of the linear regression model 2

Residuals:				
Min	1Q	Median	3Q	Max
-4.7369	-0.5579	-0.3512	0.5326	4.6013
Coefficients:				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	14.517339	0.113129	128.325	< 2e-16 ***
as.factor(sex)2	-0.047464	0.013840	128.325	0.000606 ***
……				
as.factor(litter_size)2	-0.055303	0.049559	-1.116	0.264481
……				
as.factor(parity)2	0.116172	0.015793	7.356	1.97e-13 ***
……				
as.factor(herd)2	-0.020723	0.110762	-0.187	0.851589
……				

At last, we fitted a model with all these four factors and teat number. After modeling, we find the intercept we cared best is 14.517339. It seems reasonable according to the common teat number as fourteen.

According to the formula (2.6.1) and (2.6.2), we got the plot of the relationship figure between offspring and their parents as follow:
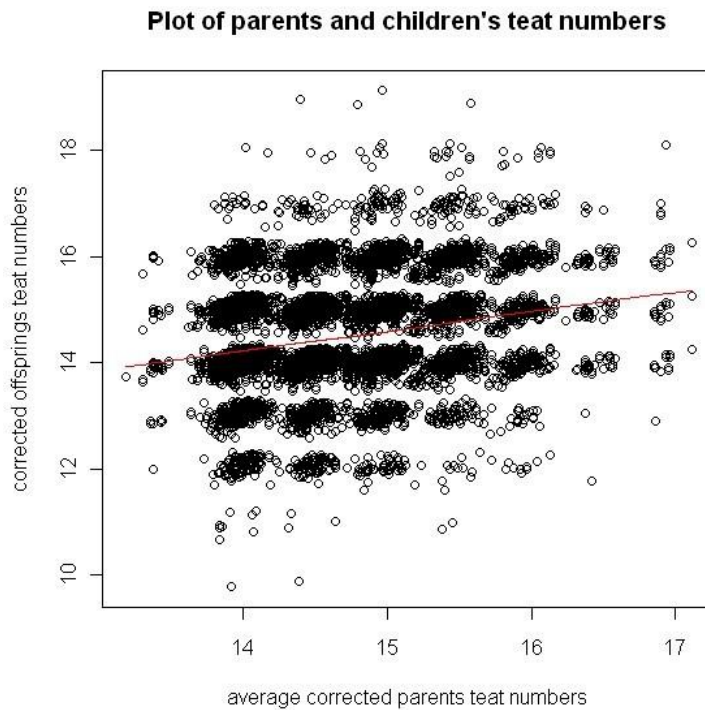
Figure 2: the relationship between offspring and parents' corrected teat number

From Figure.2 we mentioned above, we can find the slope of the line with the code lm(y~x) in R:

The summary (lm(y~x)) as follows:

Table 2: summary the coefficients of the linear model lm(y~x)

coefficients

	Estimate	Std. Error	t value	Pr(>|t|)
(intercept)	9.14159	0.15474	59.08	<2e-16 ***
x	0.36563	0.01053	34.71	<2e-16 ***

The slope of this line is 0.36563 which means the heritability:

$$h^2 = 0.36563$$

So we can argue that the genetic effect is about 36.563% in the whole effects.

3.2 The second stage

Based on the sire model (model3), we use the package (pedigreemm) to fit the linear mixed animal model and achieve the result.

We had got the part of coefficients in the table below:

Table 3: the part coefficients of the sire model 3

AIC	BIC	logLik	deviance	REMLdev
48097	48355	-24016	47873	48031

Random effects:

Groups	Name	Variance	Std Dev
SireID	(intercept)	0.10200	0.31937
Residual		0.76752	0.87608

Number of obs:18461, groups:SireID,271

Fixed effects:

	Estimate	Std. Error	T value
(intercept)	14.502607	0.050268	287.87
as.factor(data_teats$sex)2	-0.052189	0.013345	3.91
……			
as.factor(data_teats$litter_size)2	-0.118685	0.047993	-2.47
……			
as.factor(data_teats$parity)2	0.053803	0.015990	3.36
……			
as.factor(data_teats$herd)2	0.064243	0.110834	0.58
……			

And then we can calculate the heritability value of the sire model from formula (2.4.2):

$$h^2 = \frac{4 * Var(s)}{4 * Var(s) + Var(e)} = \frac{4 * 0.10200}{4 * 0.10200 + 0.76752} = 0.3471$$

So the genetic effect is about 34.71% in the whole effects to the offspring teat number.

Following, when we add the litter identity as the second random effect, then we get the model4. We use the package (pedigreemm) to fit the linear mixed animal model again and achieve the new result.

Based on the sire model with another random effect litter id (model4) we have got, we could get the part of coefficients in the table below:

Table 4: the part coefficients of the sire model 4 with environmental effect

AIC	BIC	logLik	deviance	REMLdev
47266	47532	-23599	47059	47198

Random effects:

Groups	Name	Variance	Std Dev
data_teats$litter_id	(intercept)	0.105382	0.32463
SireID	(intercept)	0.088094	0.29681
Residual		0.668967	0.81790

Number of obs:18461, groups: data_teats$litter_id, 2774; SireID, 271

Fixed effects:

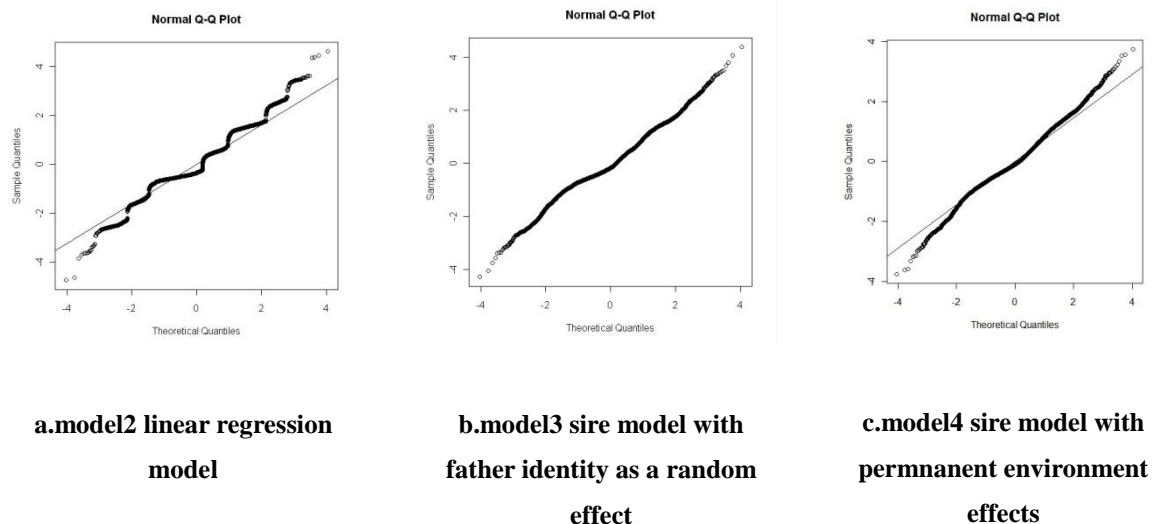|	Estimate	Std. Error	T value
(intercept)	14.500233	0.132959	109.06
as.factor(data_teats$sex)2	-0.057673	0.012906	-4.47
……			
as.factor(data_teats$litter_size)2	-0.093175	0.061502	-1.51
……			
as.factor(data_teats$parity)2	0.059926	0.022564	2.66
……			
as.factor(data_teats$herd)2	0.062319	0.122387	0.51
……			

From formula (2.4.2) we can get the heritability value of the sire model:

$$h^2 = \frac{4 * Var(s)}{4 * Var(s) + Var(e) + Var(p)} = \frac{4 * 0.088094}{4 * 0.088094 + 0.668967 + 0.105382} = 0.3127$$

So the genetic effect is about 31.27% in the whole effects to the offspring teat number.

This model contains the genetic effect with the environmental effect, because in this model, the litter identity as a permanent environmental random effect.

3.3 QQ-plot of studentized residuals

| a.model2 linear regression model | b.model3 sire model with father identity as a random effect | c.model4 sire model with permnanent environment effects |

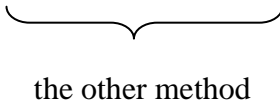QQ-plots for the three models are found in Figure 3.

Figure 3. QQ-plots of studentized residuals from the model 2 to model 4. All three models are not fitting the data very well and the residuals from all three models re not independent identically normal distributed.

4 Conclusion and discussion

The heritability values we got were as table 5 shown:

HÖGSKOLAN DALARNA
School of technology and business studies

Table 5: Comparison heritability values of three types models in two methods

Methods	Linear model (model2)	Sire model	
		Sire model (model3)	Sire model with permanent environment effects (model4)
	One method	the other method	
Heritability (h^2)	0.36563	0.3471	0.3127

Comparing the different heritability values we have got from the three different models, we can see that all the three values look like the same, which means the two methods which are linear regression as Galton and a kind of animal model (sire model) perform the same results. And the heritability of sire model with litter identity as random effect is smaller than the sire model only with father identity as random effect. We think the reason is that in model4 we pay more attention on the environment effect, it will reduce the result we got of the influence of the genetic effect.

We got the conclusion that the genetic effect is about 35% in all the effects which contain genetic effect and environmental effect.

And there must be some problems in our essay. First one, the animal model which is using the individual identity as the random effect is not possible to fit in R because of the memory problems. Second one, the QQ-plots of studentized residual for all three models are not independent identically normal distributed, data is not fitted well by the linear mixed model. That means the assumption of independent identically normal distributed is not suitable for our data. But the normal distribution is commonly used in animal breeding, and a concern about this assumption is above the scope of this essay. A better fit has been proved by structuring the residual variance.

5 Reference

(1) Felleki, M., Chalkias, H. (2010). A Double Hierarchical Generalized Linear Model for teat number in Pigs.

(2) Lundeheim, N., Chalkias, H., Rydhmer, L (2010). Teat number, fattening performance and litter traits in pigs, -estimates of genetic correlations.

(3) Pawitan, Yudi (2001). In all likelihood: an applied approach. Lund: Studentlitteratur

(4) Henderson, C.R. (1984) Applications of Linear Models in Animal Breeding, University of Guelph, Guelph, ON, Third Edition edited by L.R. Schaeffer

(5) Mrode, R.A., Thompson (2005)   Linear models for the prediction of animal breeding values

(6) Lush, J. L. (1949), Heritability of Quantitative Characters in Farm Animals. Hereditas, 35: 356–375.

HÖGSKOLAN DALARNA
School of technology and business studies

6 Appendix (code in R)

```
#### read data and package#####
data_teats<-read.table("C:\\Users\\jiangwei\\Desktop\\Dlevel\\20110421_spenar_dat
a.dat",na.string=".")
names(data_teats)<-c("id","litter_id","birthday","sex","herd","litter_size","parity","t
eat")
data_ped<-read.table("C:\\Users\\jiangwei\\Desktop\\Dlevel\\20110421_spenar_ped.
dat",na.string=".",header=FALSE)
names(data_ped)<-c("id","Sire","Dam")
library(lme4)
library(pedigree)


#### linear model and plot ####
model1 <- lm(teat~as.factor(birthday) + as.factor(litter_id) + as.factor(sex) +
as.factor(litter_size) + as.factor(parity) + as.factor(herd), data = data_teats)
model2 <- lm(teat ~ as.factor(sex) + as.factor(litter_size) + as.factor(parity) +
as.factor(herd), data = data_teats);
summary(model2);
rdMat <- numeric(0);
pedMat <- numeric(0);
for (i in 1:nrow(data_teats)){
    tID <- data_teats$id[i];
    pLine <- which(data_ped$id == tID);
    SireID <- data_ped[pLine, 2];
    DamID <- data_ped[pLine, 3];
    SireL <- which(data_teats$id == SireID);
    DamL <- which(data_teats$id == DamID);
    if (length(SireL) != 0 && length(DamL) != 0) {
     rdMat    <-    rbind(rdMat,    c(model2$resid[i],    model2$resid[SireL],
model2$resid[DamL]));
     pedMat <- rbind(pedMat, c(tID, SireID, DamID));
     }
}

u <- coef(model2)[1];
m <- (rdMat[, 2] + rdMat[, 3] + u*2)/2
n <- rdMat[, 1] + u
plot(m, n, xlab = "average corrected parents teat numbers", ylab = "corrected
offsprings teat numbers", main="Plot of parents and children's teat numbers");
panel.smooth(m, n);
```

HÖGSKOLAN DALARNA
School of technology and business studies

```
qqnorm(model2$resid);
qqline(model2$resid);



##### sire model ####
k <- which(data_ped[,2] == -97)
j <- which(data_ped[,3] == -97)
is.na(data_ped[,2]) <- k
is.na(data_ped[,3]) <- j
u <- max(na.omit(data_ped[,2]))
i <- which(data_ped[,1] >= u)
newped <- data_ped[-i,]
names(newped)<-c("id","Sire","Dam")
ped <- numeric(0)
for (j in 1:nrow(newped)){
    tID <- newped$id[j]
    pLine <- which(data_ped$id == tID);
    SireID <- data_ped[pLine, 2];
    DamID <- data_ped[pLine, 3];
    ped <- rbind(ped, c(tID, SireID, DamID));
}
ped1 <- numeric(0)
Sireid <- numeric(0)
for (i in 1:length(ped[,2])){
    if (ped[i,2] %in% t[,1]){Sireid[i] = ped[i,2]}
    else {Sireid[i] = NA}
    ped1 <- rbind(ped1,c(ped[i,1], Sireid[i], ped[i,3]))
}
ped2 <- numeric(0)
Damid <- numeric(0)
for (i in 1:length(ped1[,3])){
    if (ped1[i,3] %in% ped1[,1]){Damid[i] = ped1[i,3]}
    else {Damid[i] = NA}
    ped2 <- rbind(ped2,c(ped1[i,1], ped1[i,2], Damid[i]))
}
pedMat <- numeric(0);
for (i in 1:nrow(data_teats)){
    tID <- data_teats$id[i];
    pLine <- which(data_ped$id == tID);
    SireID <- data_ped[pLine, 2];
    DamID <- data_ped[pLine, 3];
    pedMat <- rbind(pedMat, c(tID, SireID, DamID));
```

```
}
SireID <- pedMat[,2]
ped <- pedigree(sire = ped2[, 2], dam = ped2[, 3], label = ped2[, 1])
model3    <-    lmer(data_teats$teat    ~    as.factor(data_teats$sex)    +
as.factor(data_teats$litter_size)    +    as.factor(data_teats$parity)    +
as.factor(data_teats$herd)   + (1 | SireID))
model4   <-    pedigreemm(data_teats$teat    ~    as.factor(data_teats$sex)    +
as.factor(data_teats$litter_size)    +    as.factor(data_teats$parity)    +
as.factor(data_teats$herd)   + (1 | SireID),   pedigree = list(SireID = ped))
summary(model3)
summary(model4)
qqnorm(resid(model3))
qqline(resid(model3))
qqnorm(resid(model4))
qqline(resid(model4))
```