56th Annual EAAP Meeting in Uppsala, Sweden

Session 9: **Genetics of variability G2.4** (Abstract No. 661)

A weighted regression approach for the detection of QTL effects on within-subject variability

Dörte Wittenburg[1], Volker Guiard, Norbert Reinsch

Forschungsinstitut für die Biologie landwirtschaftlicher Nutztiere
Wilhelm-Stahl-Allee 2, 18196 Dummerstorf, Germany

Abstract

Quantitative trait loci may not only affect the mean of a trait but also its variability. A special aspect is the variability between multiple measurements of genotyped animals, for example the within litter variance of birth weights. Assuming a normally distributed trait a weighted regression approach was developed, taking the transformed sample variance s^2 between repeated measurements as observation for every genotyped individual. For the daughter-design the weighted regression approach was evaluated in terms of precision of the estimation for the QTL-position, statistical power and compliance with the desired error probability under the null hypothesis.

1 Introduction

Analysis in quantitative trait loci (QTL) have been discussed by many scientists. In most cases only the mean effect of QTL was pointed out. But it may be possible that QTL affect on the variability of phenotypic values, too. For example, the advantage of increased mean of weights at birth within a litter is destroid, when its variability presumes extreme values. So we are looking for QTL which may affect the within litter variance of birth weights.

2 Theory

The considerations base on the following assumptions: The population of domestic pigs has two alleles of the QTL, say Q and q. Further we look at a fixed number N of sires in our studies. The sires are drawn by chance of the population. Every sire is mated with n unrelated dams of the population. We pick out one daughter per mating. Thus we have a constant family size of n. The daughters are mated with unrelated males of the population. The trait is multiple measured by the offsprings of every daughter. The sample variance of the weights at birth within litter is taken as characteristic for every daughter, which amounts to Nn observations.

The piglet's phenotype is normally distributed and the following model is valid.

$$Y_{ijk} = \underbrace{\mu + a_{ij} + g_{ij} + \omega}_{=:\mu_{ij}} + a_{ijk} + g_{ijk} + e_{ijk} \tag{1}$$

[1]e-mail to wittenburg@fbn-dummerstorf.de

i	index of family
j	index of daughter
k	index of piglet
μ	mean value of population
a_{ij}	maternal additive genetic effect
g_{ij}	maternal QTL effect depending on daughter's genotype
ω	random effect of litter and environment
a_{ijk}	direct polygenic effect
g_{ijk}	QTL effect depending on piglet's genotype
e_{ijk}	random deviation

μ_{ij} combinates the constant values within a litter. No dominance effect is committed. The direct polygenic effect a_{ijk} splits up into the mendelian sample distributed as $N(0, \frac{1}{2}\sigma^2_{polygene})$ and the parental breeding values b_{ij}, which are fixed for each litter. The random deviation is distributed as $N(0, \sigma^2_e)$ and the variance of the QTL effect is $\mathbb{V}(g_{ijk}) = \sigma^2_{QTL}$. The parents are not inbred. The within litter variance σ^2 depends on the daughters paternal allele. In case of Q passed on it is

$$\sigma^2 = \frac{1}{2}\sigma^2_{polygene} + \sigma^2_{QTL} + \sigma^2_e \tag{2}$$

And otherwise the within litter variance is increased by a multiplicative factor c^2.
The sample variance is calculated by the following equation. Capital letters are used for random variables.

$$S^2_{ij} = \frac{1}{n_{ij}-1}\sum_{k=1}^{n_{ij}}(Y_{ijk} - \overline{Y}_{ij.})^2 \quad \text{with} \quad \mathbb{E}(S^2_{ij}) = \sigma^2 \quad \text{and} \quad \mathbb{V}(S^2_{ij}) = \frac{2\sigma^4}{n_{ij}-1}$$

$n_{ij} = f_{ij}+1$ denotes the litter size of daughter ij. Because the variance of S^2_{ij} depends on the within litter variance σ^2, we are looking for a variance-stabilizing transformation. Such transformation is suggested by Box & Cox [1], for examples see Christensen [2, p.199]. It is with $T_{ij} = \ln S^2_{ij}$

$$\mathbb{V}(T_{ij}) = \mathbb{V}(\ln S^2_{ij}) \approx \frac{2}{f_{ij}} \quad \text{and} \quad \mathbb{E}(T_{ij}) \approx \ln \sigma^2 \tag{3}$$

And by the δ-method follows the asymptotical normal distribution

$$\mathcal{L}\left(\sqrt{f_{ij}}(\ln S^2_{ij} - \ln \sigma^2)\right) \Rightarrow N(0,2) \quad \text{with} \quad f_{ij} \to \infty$$

The variance in equation (3) still depends on the litter size n_{ij}, but we will consider this known variance by weighted examinations.
Figure (1) shows the convergence of $\sqrt{\frac{f_{ij}}{2}}\ln S^2_{ij}$ when $\sigma^2 = 1$. Even with a relative small number of degrees of freedom, the probability function of $\sqrt{\frac{f_{ij}}{2}}\ln S^2_{ij}$ approximates the standard normal distribution well, also mentioned by Lehmann [7, p.376].

In the next steps the model for the weighted regression approach is determined. The sires have the marker genotype of kind $M_{l,1}M_{l,2}$, where l denotes the marker position. The markers are reduced to two alleles at the sire denoted by $M_{l,1}$ on the paternal allele and $M_{l,2}$ on the maternal allele, respectively, for every marker position. Therefore it is not possible to determine which sire is heterozygote or homozygote a priori. After the sires are genotyped, we suppose that all daughters are full informative. That means, we only have to consider the daughters paternal allele. Each daughter inherits the haplotype from the sire, where recombination events are possible. The recombination rates are calculated by Haldane's mapping function. We consider intervals flanked

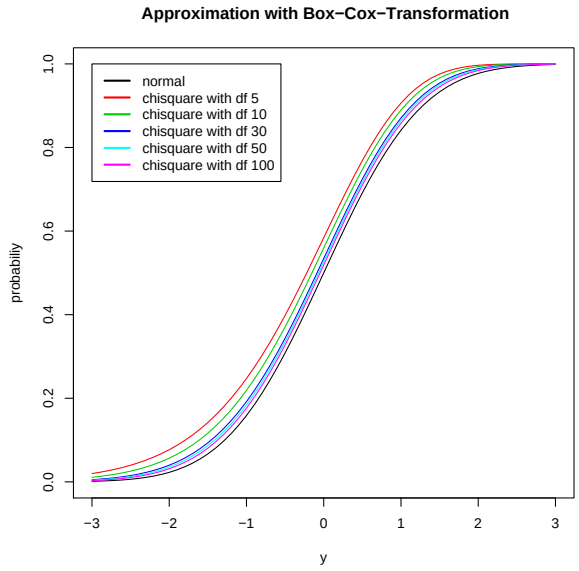

Figure 1: Approximation of the standard normal distribution

by markers of type $M_{l,r}M_{l+1,s}$ with $r, s \in \{1, 2\}$. The transmission probability p_{rs} for inheriting the QTL allele Q of the heterozygote sire with genotype Qq is determined for every desired position on the chromosome. Building the conditional variance of the repeated measurements Y_{ijk} results to the variance within litter

$$\mathbb{V}(Y_{ijk}|M_{l,r}M_{l+1,s}) = \sigma^2_{rs} \quad \text{with } r, s \in \{1, 2\} \quad \text{and} \quad \sigma^2_{rs} = \begin{cases} \sigma^2 & \text{inheriting } Q \\ (c\sigma)^2 & \text{inheriting } q \end{cases}$$

The conditional expected values of the transformed sample variance $\ln S^2_{ij} = T_{ij}$ are

$$\mathbb{E}(T_{ij}|M_{l,1}M_{l+1,1}) = p_{11}\mathbb{E}(T_{ij}|M_{l,1}M_{l+1,1}, Q) + (1 - p_{11})\mathbb{E}(T_{ij}|M_{l,1}M_{l+1,1}, q)$$

$$\approx p_{11} \ln \sigma^2 + (1 - p_{11}) \ln(c\sigma)^2$$

$$= \ln(c\sigma)^2 - p_{11} \ln c^2$$

$$\mathbb{E}(T_{ij}|M_{l,1}M_{l+1,2}) \approx \ln(c\sigma)^2 - p_{12} \ln c^2$$

$$\mathbb{E}(T_{ij}|M_{l,2}M_{l+1,1}) \approx \ln(c\sigma)^2 - p_{21} \ln c^2$$

$$\mathbb{E}(T_{ij}|M_{l,2}M_{l+1,2}) \approx \ln(c\sigma)^2 - p_{22} \ln c^2$$

Under the assumption the sire is heterozygote with the genotype qQ we point out the sign change in the second term of the expected values.
Therefore we can write with $t_{ij} \in \{p_{11}, p_{12}, p_{21}, p_{22}\}, i = 1, \ldots, N$ and $j = 1 \ldots, n$ for the investigated position on the chromosome

$$T_{ij} = \ln S^2_{ij} = m_i + b_i t_{ij} + \varepsilon_{ij}$$

Or in matrix way of writing with $T = (T_{11}, T_{12}, \ldots, T_{Nn})^T$

$$T = X\beta + \varepsilon \qquad \text{with } \varepsilon \text{ is normally distributed.}$$

It is m_i the mean value of family i, b_i the slope of the regression line and ε the random error of the model.

$$X = \begin{pmatrix} 1 & 0 & \cdots & 0 & t_{11} & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 & t_{12} & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & 0 & \cdots & 0 & t_{1n} & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & t_{21} & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 1 & \cdots & 0 & 0 & t_{2n} & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 1 & 0 & 0 & \ldots & t_{N1} \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & t_{Nn} \end{pmatrix}, \qquad \beta = \begin{pmatrix} m_1 \\ m_2 \\ \vdots \\ m_N \\ b_1 \\ \vdots \\ b_N \end{pmatrix} \qquad \text{and } \varepsilon = \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \vdots \\ \varepsilon_{1n} \\ \varepsilon_{21} \\ \vdots \\ \varepsilon_{Nn} \end{pmatrix}$$

Because of our assumption we know the slope has to be

$$b_i = \begin{cases} \ln c^2 & \text{genotype of sire } i \text{ is } qQ \\ 0 & \text{genotype of sire } i \text{ is } qq \\ -\ln c^2 & \text{genotype of sire } i \text{ is } Qq \\ 0 & \text{genotype of sire } i \text{ is } QQ \end{cases} \tag{4}$$

In recollaction of (3) we define a matrix of weights W. On the presumption that the daughters traits are independent, the covariance matrix is $W = \mathbb{V}(T)$. Because W is symmetric and positive definite we find the partition $W = PP^T$.

$$W = \begin{pmatrix} \frac{2}{f_{11}} & 0 & \cdots & 0 \\ 0 & \frac{2}{f_{12}} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{2}{f_{Nn}} \end{pmatrix}, \qquad P = \begin{pmatrix} \sqrt{\frac{2}{f_{11}}} & 0 & \cdots & 0 \\ 0 & \sqrt{\frac{2}{f_{12}}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\frac{2}{f_{Nn}}} \end{pmatrix}$$

The simple regression model has changed into the weighted regression model

$$\underbrace{P^{-1}T}_{=:Z} = \underbrace{P^{-1}X}_{=:Q}\beta + P^{-1}\varepsilon \qquad \text{and therefore}$$

$$Z = Q\beta + \xi \qquad \text{with } \xi \text{ is distributed as } N(0, I_{Nn}) \tag{5}$$

The parameter β is solved with the method of weighted least squares. Thus

$$\widehat{\beta} = (Q^TQ)^{-1}Q^TZ = (X^TW^{-1}X)^{-1}X^TW^{-1}T$$

Now we are going to test the hypothesis $H_0 : b_1 = \cdots = b_N = 0$, that means no difference between the marker genotyped daughters is obvious. In our consideration the failure of H_0 is equal to the existence of QTL which influences the traits.
Under the assumption of the approximately normally distributed vector Z we build with support of Seber [9, p.97] the test statistic

$$F = \frac{Nn - 2N}{N} \frac{Z^T(Q(Q^TQ)^{-1}Q^T - Q_1(Q_1^TQ_1)^{-1}Q_1^T)Z}{Z^T(I_{Nn} - Q(Q^TQ)^{-1}Q^T)Z} \tag{6}$$

$Q_1 = P^{-1}X_1$ with the reduced matrix X_1 consisting of the N first columns of X.
Under pointwise investigations and H_0 is valid, F is distributed as a F-distribution with N and $Nn - 2N$ dregrees of freedom. If H_0 fails on the presumed QTL-position, F is non-central F-distributed with the non-centrality parameter $\lambda = \beta^T K(K^T(Q^TQ)^{-1}K)^{-1}K^T\beta$. K denotes the $2N \times N$ matrix which is constructed by a N dimensional matrix of zeros and the identity matrix of range N. H_0 is rejected, if $F > f_{1-\alpha,N,Nn-2N}$, where $f_{1-\alpha,N,Nn-2N}$ is the $(1-\alpha)$-quantile of the central F-distribution with $N, Nn-2N$ degrees of freedom. α denotes the significance level, we use $\alpha = 0.05$. The pointwise p-value p_{point} is calculated by $1 - F_{N,Nn-2N}(F)$.

For the chromosomewise detection of a single QTL we have to consider dependencies between the marker intervals. Therefore we use the procedure of a permutation test described in detail by Good [4]. With this technique we receive an approximated distribution for the F-values over the chromosome under H_0. We follow the suggestion of Good [4, p.39] applying the blocking method. Therefore we split up the data into blocks in this way, that each family generates one block. For every block we resample the traits 100 times. The presumed phenotype - genotype connection is canceled. For every resampled dataset we construct the F-value mentioned in (6) on every chromosomal position (1 cM). The maximal F-value of this permutation is kept in mind. We repeat this for every resampling. Therefore the approximated distribution under H_0 consists of the maximal F-values of the resampled datasets. We calculate the maximal F-value of the original dataset and compare it with the treshold value. The treshold is the (1-α)-quantile of the maximal F-values of the resampled datasets. Churchill & Doerge [3] denoted this treshold by the experimentwise critical value. The null hypothesis is rejected if $F >$ treshold value. If H_0 is rejected, we presume the QTL on the position with the greatest F-value of the original dataset. The chromosomewise p-value $p_{chromosome}$ is determined by the relation of the number of maximal F-values of the resampled datasets, which are greater than the maximal F-value of the original dataset. We consider only one chromosome, therefore we apply the Bonferroni correction to calculate the genomewise p-value $p_{genome} = 1 - (1 - p_{chromosome})^{nc}$, where nc is the number of chromosomes.

The estimation of the parameter c follows from (4), therefore we conclude

$$\widehat{c}_i = \sqrt{\exp \widehat{b}_i} \qquad \text{with} \quad i = 1, \ldots, N \tag{7}$$

The estimator $\widehat{b}_i$ is asymptotically normally distributed

$$\mathcal{L}\left(\sqrt{n}(\widehat{b}_i - b_i)\right) \Rightarrow N(0, \widetilde{\sigma}_i{}^2) \quad \text{with} \quad n \to \infty \quad \text{and} \quad \mathbb{E}(\widehat{b}_i) = b_i$$

$$\widetilde{\sigma}_i{}^2 = \left[\left(X^T W^{-1} X\right)^{-1}\right]_{i,i} = \frac{\sum_{j=1}^{n} \frac{1}{w_{ij}}}{\sum_{j=1}^{n} \frac{t_{ij}^2}{w_{ij}} \sum_{j=1}^{n} \frac{1}{w_{ij}} - \left(\sum_{j=1}^{n} \frac{t_{ij}}{w_{ij}}\right)^2} \quad \text{with} \quad w_{ij} = \frac{2}{f_{ij}}$$

Thus $\widehat{c}_i^2$ is approximately log-normally distributed with the expected value

$$\mathbb{E}(\widehat{c}_i^2) \approx \begin{cases} c^2 \exp \frac{\widetilde{\sigma}_i{}^2}{2} & \text{genotype of sire } i \text{ is } qQ \\ c^{-2} \exp \frac{\widetilde{\sigma}_i{}^2}{2} & \text{genotype of sire } i \text{ is } Qq \\ \exp \frac{\widetilde{\sigma}_i{}^2}{2} & \text{sire } i \text{ is homozygote} \end{cases} \tag{8}$$

3 Simulation

When genotyping the individuals of the population we find markers in intervals of 10 cM on a chromosome of length 100 cM. So we have 11 markers at our disposal. Under the null hypothesis H_0 no QTL is segregated in the population. In the simulation we placed a single QTL at position 25 cM (between the third and fourth marker). We simulated the obove described procedure with

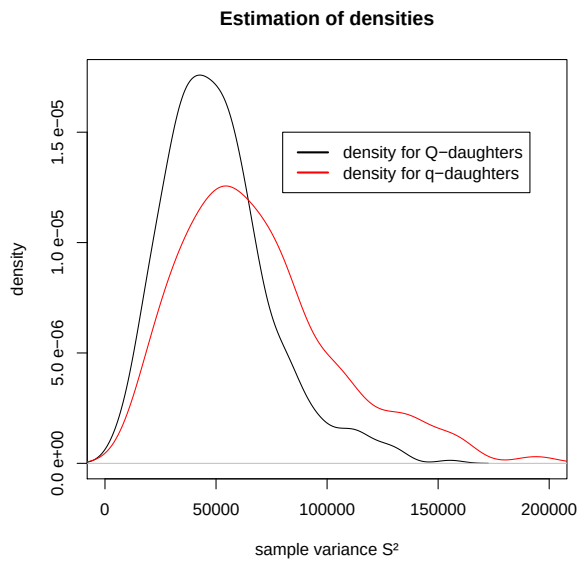

Figure 2: Variance of birth weights within litter ($c = 1.17$)

$N = 4$ sires and $n = 200$ daughters per family. The litter size is poisson distributed with mean value of 10. The piglets refering to (1) are simulated with following values. The birth weights varies from 700 gram up to 2300 gram, the population mean μ is assessed to be 1500 gram. In case of inheriting Q the phenotypic standard deviation $\sigma_{phenotype}$ is 250 gram at a rough estimate. The variance of the maternal QTL effect is assumed to be 3 % of the phenotypic variance. The piglets QTL effect takes the same value. The maternal additive genetic variance is 10 % of $\sigma^2_{phenotype}$. The direct polygenic variance is 15 % of $\sigma^2_{phenotype}$, the residual variance σ^2_e takes 64 % of $\sigma^2_{phenotype}$ and finally the variance of litter effect and environment is 5 % of $\sigma^2_{phenotype}$.
To simplify the simulation if q passed on, we modify the increased within litter variance. In this case it is $\mathbb{V}(Y_{ijk}) = \frac{1}{2}\sigma^2_{polygene} + \sigma^2_{QTL} + (c_*\sigma_e)^2$ with the multiplicative factor of random deviation $c_* = 1.0(0.1)1.4$ and thus

$$c^2 = \frac{\frac{1}{2}\sigma^2_{polygene} + \sigma^2_{QTL} + (c_*\sigma_e)^2}{\frac{1}{2}\sigma^2_{polygene} + \sigma^2_{QTL} + \sigma^2_e}$$

The simulations were repeated 100 times. The gene frequency p_Q is presumed to be $\frac{1}{2}$. Using the kernel-density-estimation of the statistic program R figure (2) shows the difference in distribution in a simulation of four families. Even we only used 100 resamples in the permutation test procedure, the results are satisfying as seen in figure (3) and (4). For example when $c = 1.17$ the simulated QTL-position was detected 10 times accurately in case of H_0 has been rejected. As figure (4) shows, the detections closely surround the correct position, they mostly differ about five positions from the right one. Precise detections increase to 21 % when $c = 1.35$.

In figure (5) it is obvious that the estimator $\widehat{c}_i$ parts. Because of the presumed gene frequency the estimated values fluctuate around 1 for homozygote sires approximately with half the density. Otherwise for heterozygote sires with genotype Qq or qQ the values of $\widehat{c}_i$ surround c and $\frac{1}{c}$, respectively, each case with quarter the density. The bias results from equation (8).

By repeating the simulations with an increasing ratio c the average of the p-values chromosome-wise decreases down to 0.12 %. Figure (6) displays the average of p-values over 100 repetitions depending on the ratio of standard deviation c.

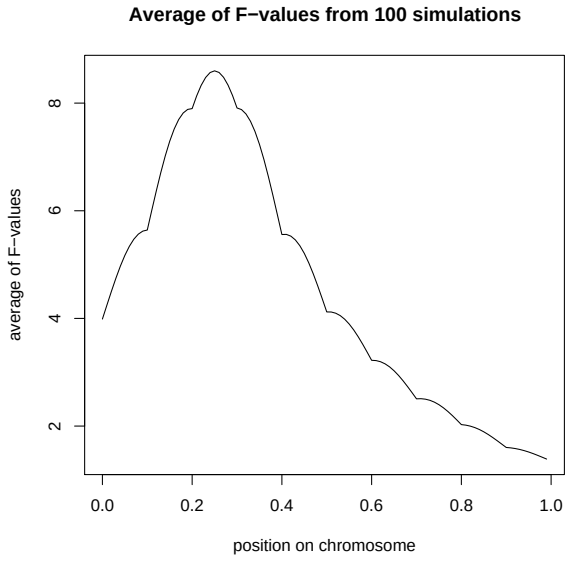

Figure 3: Average of F-values (pig, $c = 1.17$)

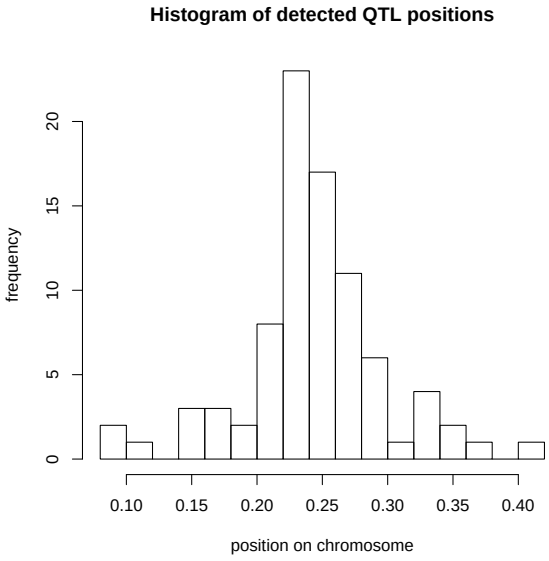

Figure 4: Detected QTL-positions (pig, $c = 1.17$)

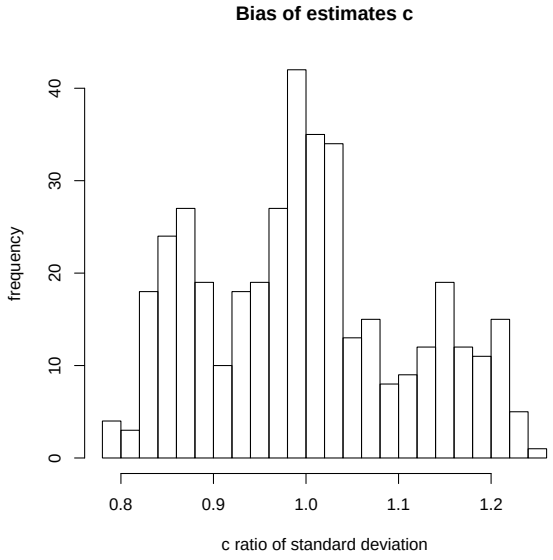

Figure 5: Bias of estimator $\widehat{c}_i$ (pig, $c = 1.17$)

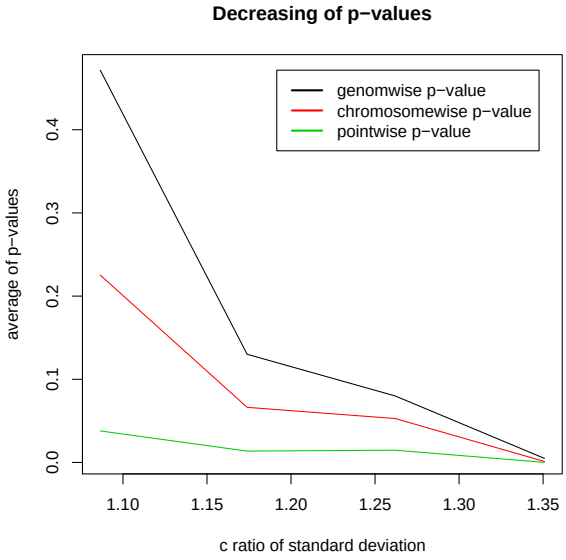

Figure 6: Decreasing of p-values

4 Power

In the chromosomewise considerations we describe the empirical power $\widehat{\pi}$ by the proportion of rejections as supported by Good [4, p.163]. In the simulation with $c = 1.17$ it is $\widehat{\pi} = 85$ %. This value doesn't include that 8 times only homozygote sire were drawn by chance. Therefore the actual power is much higher. In the repetitions with increasing ratio c up to 1.35 the empirical power is 99 % at maximum.

Now we look on detail on the power in the case of pointwise investigations on the presumed QTL-position. To calculate the power $\pi(\lambda) = \mathbb{P}(F > f_{1-\alpha,N,Nn-2N}|\lambda) = 1-F_{N,Nn-2N,\lambda}(f_{1-\alpha,N,Nn-2N})$ we have to determine the non-centrality parameter λ, which is a priori unknown. We remind $\lambda = \beta^T \underbrace{K(K^T(Q^TQ)^{-1}K)^{-1}K^T}_{=:M}\beta$. The mean values m_i containing in β could be ingnored, we set $m_1 = \ldots = m_N = 0$. Thus β has the form $(0,\ldots,0,b_1,\ldots,b_N)^T$ with $b_i \in \{0, \ln c^2\}$, $i = 1,\ldots,N$. The order and the sign of b_i are neglected because of the design of λ. For example under premise there are two of four sires heterozygte, it is $\beta = (0,0,0,0,0,0,\ln c^2, \ln c^2)^T$. Therefore we have to determine the power $\pi(\lambda)$ depending on β, that means depending on the number of heterozygote sires in our observations.
The probability $\mathbb{P}$(sire is heterozygote)$= 2p_Q(1-p_Q) = \frac{1}{2} = p$ depends on the presumed gene frequency p_Q. We use the binomial distribution to calculate the probability $\overline{p}_k$ that k of N sires are heterozygote. We denote $\beta = \beta_k$ in the case k of N sires are heterozygote.

$$\overline{p}_k = \binom{N}{k} p^k (1-p)^{N-k} = \binom{N}{k} \frac{1}{2^N}$$

Now we are able to calculate the power function $\pi(\lambda,\beta)$ depending on β as suggested by Lehmann [7, p.151].

$$\begin{aligned}
&\mathbb{E}_\beta\left(\pi(\lambda|\beta)\right) = \overline{p}_0\pi(\lambda|\beta_0) + \ldots + \overline{p}_N\pi(\lambda|\beta_N)\\
&= \overline{p}_0\pi(\beta_0^T M\beta_0) + \ldots + \overline{p}_N\pi(\beta_N^T M\beta_N)\\
&= \overline{p}_0(1 - F_{N,Nn-2N,\beta_0^T M\beta_0}(f_{1-\alpha,N,Nn-2N})) + \ldots + \overline{p}_N(1 - F_{N,Nn-2N,\beta_N^T M\beta_N}(f_{1-\alpha,N,Nn-2N}))
\end{aligned}$$

To adapt this determination of $\pi(\lambda,\beta)$ to the chromosomewise investigations, one could replace $f_{1-\alpha,N,Nn-2N}$ by the chromosomewise treshold value calculated by the permutation test. After some calculations with use of the Frobenius-formula to invert a sparse matrix we receive the $2N \times 2N$ matrix M. It is $w_{ij} = \frac{2}{f_{ij}}$ with $i = 1,\ldots,N, j = 1,\ldots,n$.

$$M = \begin{pmatrix}
0 & \cdots & 0 & & 0 & \cdots & 0\\
\vdots & \ddots & \vdots & & \vdots & \ddots & \vdots\\
0 & \cdots & 0 & & 0 & \cdots & 0\\
0 & \cdots & 0 & \sum_{j=1}^n \frac{t_{1j}^2}{w_{1j}} - \frac{\left(\sum_{j=1}^n \frac{t_{1j}}{w_{1j}}\right)^2}{\sum_{j=1}^n \frac{1}{w_{1j}}} & \cdots & 0\\
\vdots & \ddots & \vdots & & \vdots & \ddots & \vdots\\
0 & \cdots & 0 & & 0 & \cdots & \sum_{j=1}^n \frac{t_{Nj}^2}{w_{Nj}} - \frac{\left(\sum_{j=1}^n \frac{t_{Nj}}{w_{Nj}}\right)^2}{\sum_{j=1}^n \frac{1}{w_{Nj}}}
\end{pmatrix}$$

To organize an experiment, we may predict the power on pointwise investigations. To simplify the calculations we use a single marker next to the presumed QTL-position. Using flanking markers the power values are calculated analogously, but not shown here. Thus we construct the entries of M assuming they are drawn by chance. We replace the entries by their expected values. We have to calculate the transmission probabilities on the presumed QTL-position

$$p_r = \mathbb{P}(Q|M_{l,r}) \qquad \text{with } r \in \{1,2\}, l \in \{1,\ldots,10\}$$

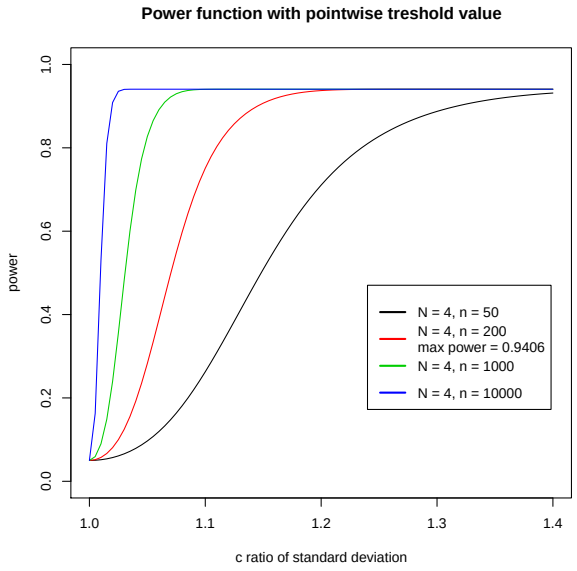

Figure 7: Power depending on number of daughters

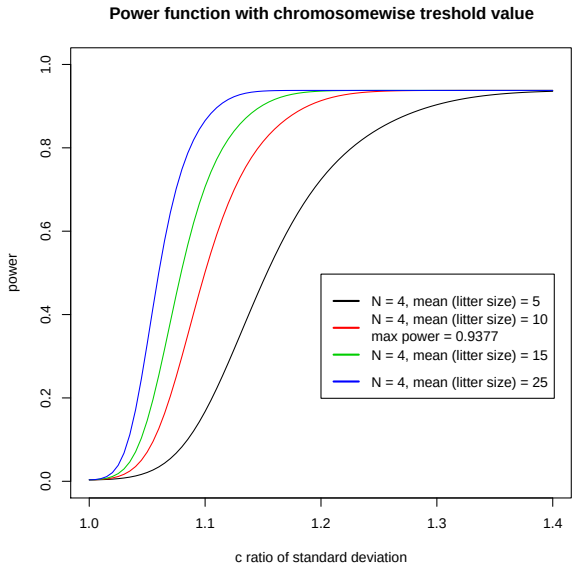

Figure 8: Power depending on litter size, $n = 200$

We apply the worst case, the QTL is located in the middle of two markers. We may determine $p_1 = 1 - \theta$ and $p_2 = \theta$, where θ is the recombination rate between $M_{l,1}$ and QTL calculated with Haldane's mapping function. The probability of inheriting the paternal allele of the sire is $\frac{1}{2}$. X_{ij} may be a random variable which is realized by $t_{ij} \in \{p_1, p_2\}$, thus X_{ij} has a two-point-distribution

$$\mathbb{P}(X_{ij} = p_1) = \frac{1}{2} \qquad \text{and} \qquad \mathbb{P}(X_{ij} = p_2) = \frac{1}{2}$$

We fix the litter size at the mean value $n_{ij} = f + 1$. It follows with $w_{ij} = \frac{2}{f}, i = 1, \ldots, N, j = 1, \ldots, n$

$$\mathbb{E}\left(X_{ij}\right) = \mathbb{P}(X_{ij} = p_1)p_1 + \mathbb{P}(X_{ij} = p_2)p_2 = \frac{1}{2}(1-\theta) + \frac{1}{2}\theta = \frac{1}{2}$$

$$\mathbb{E}\left(X_{ij}^2\right) = \frac{1}{2}p_1^2 + \frac{1}{2}p_2^2 = \frac{1}{2}(1 - 2\theta + \theta^2) + \frac{1}{2}\theta^2 = \frac{1}{2} - \theta + \theta^2$$

$$\mathbb{E}\left(\sum_{j=1}^{n} X_{ij}\right)^2 = n\mathbb{V}(X_{ij}) + (n\mathbb{E}X_{ij})^2 = n\left(\frac{1}{2} - \theta + \theta^2 + (n-1)\frac{1}{4}\right)$$

$$\mathbb{E}\left(\sum_{j=1}^{n} \frac{X_{ij}^2}{w_{ij}}\right) = n\frac{f}{2}\mathbb{E}\left(X_{ij}^2\right) = n\frac{f}{2}\left(\frac{1}{2} - \theta + \theta^2\right)$$

$$\mathbb{E}\left(\frac{\left(\sum_{j=1}^{n} \frac{X_{ij}}{w_{ij}}\right)^2}{\sum_{j=1}^{n} \frac{1}{w_{ij}}}\right) = \frac{f}{2}\left(\frac{1}{2} - \theta + \theta^2 + (n-1)\frac{1}{4}\right)$$

Therefore the random entries of the second part of the diagonal M for $i = 1, \ldots, N$ are

$$\mathbb{E}(M_i) = n\frac{f}{2}\left(\frac{1}{2} - \theta + \theta^2\right) - \frac{f}{2}\left(\frac{1}{2} - \theta + \theta^2 + (n-1)\frac{1}{4}\right) = (n-1)\frac{f}{2}\left(\frac{1}{4} - \theta + \theta^2\right)$$

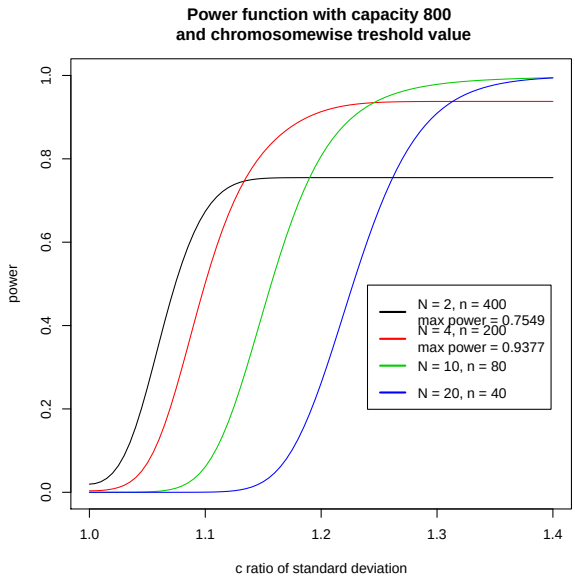

Figure 9: Power depending on given capacity

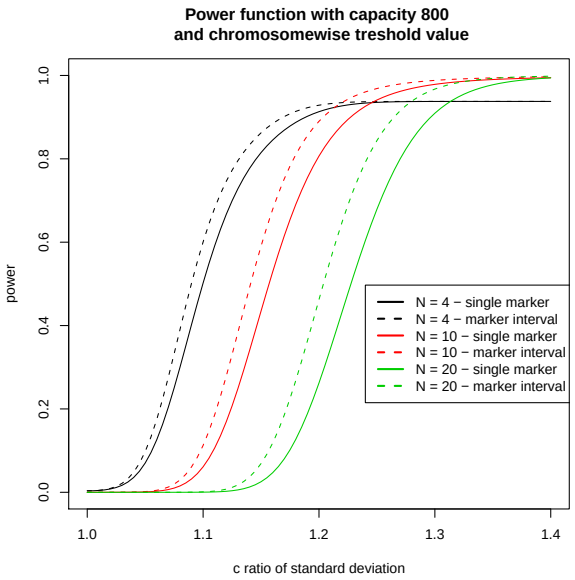

Figure 10: Power depending on marker analysis

In figure (9) one can see, with usage of only two families the power never exceeds 75.49 %. From this point of view it is useful to apply the test procedures on more than four families to obtain an acceptable power. The used chromosomewise treshold value is determined as the average of the chromosomewise tresholds from 100 reruns. The chromosomewise treshold is larger than the $(1-\alpha)$-quantile of the central F-distribution. Therefore the power is essentially less than 5 % under H_0 using this treshold. Analysis between two linked markers improves the results. Difference in power values could be recognized, see figure (10). Considering two linked markers the power depending on the ratio c increases faster than under investigations of a single marker.

In the simulations under the alternative hypothesis with $c = 1.09$ up to $c = 1.35$ the power increases to 94.06 %. Minimal differences are obvious when using either the actual degrees of freedom per litter or the mean value. The deviation within the range of the simulation is less than 0.9 % and decreases to zero.

5 Second model for application

We consider N sires with n daughters of one population of dairy cattle. Normally distributed traits are assumed in repeated measurements of wither heights, where the daughters are presumed to be full-grown. Only a small number of measurements is taken, so the daughters don't get used to this procedure and no trend of measurements is committed. We are looking for QTL, which may influence the behaviour. For every daughter the following model is valid.

$$Y_{ijk} = \underbrace{\mu + a_{ij} + g_{ij} + \omega_{ij}}_{=:\mu_{ij}} + e_{ijk} \tag{9}$$

i index of family
j index of daughter
k index of measurement
μ mean value of population
a_{ij} parental breeding values and mendelian sample

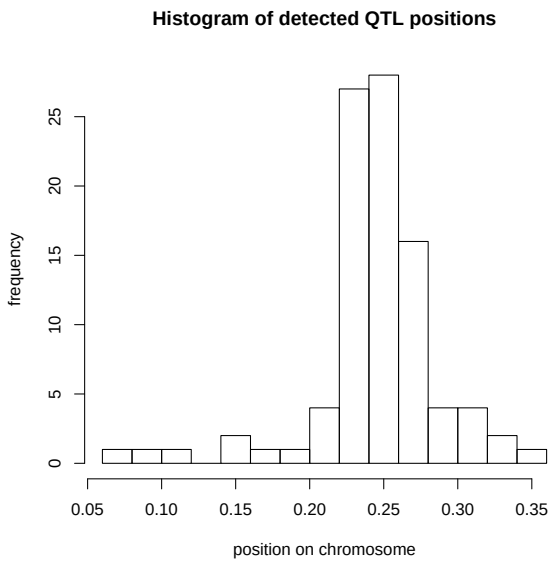

Figure 11: Detected QTL-positions (cow, $c = 1.2$)

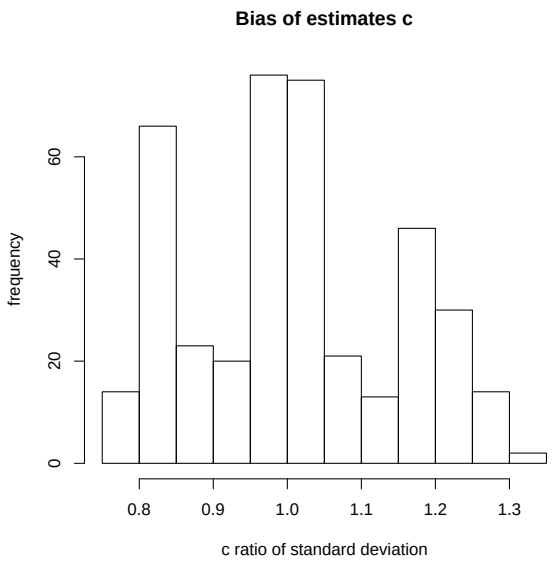

Figure 12: Bias of estimator $\widehat{c}_i$ (cow, $c = 1.2$)

g_{ij}	QTL effect depending on daughter's genotype
ω_{ij}	random effect of environment
e_{ijk}	random deviation

The components mentioned above in μ_{ij} are fixed for every observation per daughter. The variance of repeated measurements depends on the daughter's paternal allele. If Q passed on, then $\mathcal{L}(e_{ijk}) = N(0, \sigma_e^2)$, otherwise $\mathcal{L}(e_{ijk}) = N(0, (c\sigma_e)^2)$ with a multiplicative factor c. Using these considerations under the condition of flanking markers $M_{l,r}M_{l+1,s}$ with $r, s \in \{1, 2\}$ the variance within individual is

$$\mathbb{V}(Y_{ijk}|M_{l,r}M_{l+1,s}) = \sigma^2_{e,rs} \quad \text{with} \quad r, s \in \{1, 2\} \quad \text{and} \quad \sigma^2_{e,rs} = \begin{cases} \sigma_e^2 & \text{inheriting } Q \\ (c\sigma_e)^2 & \text{inheriting } q \end{cases}$$

Expecting a heterozygote sire with genotype Qq it is with $\ln S^2_{ij} = T_{ij}$

$$\mathbb{E}(T_{ij}|M_{l,r}M_{l+1,s}) \approx \ln(c\sigma_e)^2 - t_{ij} \ln c^2$$

Therefore we apply the weighted regression model (5) with a fixed number of measurements per daughter. The degrees of freedom per individual are $f_{ij} = f = 9, i = 1, \ldots, N, j = 1, \ldots, n$. Equation (7) is used to estimate the parameter c. The result of simulation is shown in figure (11) for $c = 1.2$. The correct QTL-position has been detected 13 times properly. This figure displays, that the detections disperse around the correct position very closely. The positions deviate about two or three positions from the right one. The bias of $\widehat{c}_i$ is shown in figure (12).
The empirical power $\widehat{\pi} = 93$ %, but in 6 cases of non-significance exclusive homozygote sires were drawn by chance. The power on pointwise investigations enormously depends on the number of measurements, analogous to figure (8). Therefore we repeated the simulation with $c = 1.3$ and three times measuring the wither height expecting a loss of empirical power. In fact $\widehat{\pi} = 66$ % and the actual QTL-position has been detected 6 times accurately. Results are shown in figure (13) and (14).

The F-test used for pointwise investigations behaves robust against non-normality, see Lehmann [7, p.401]. The permutation test is a non-parametric approximation of the distribution under H_0.

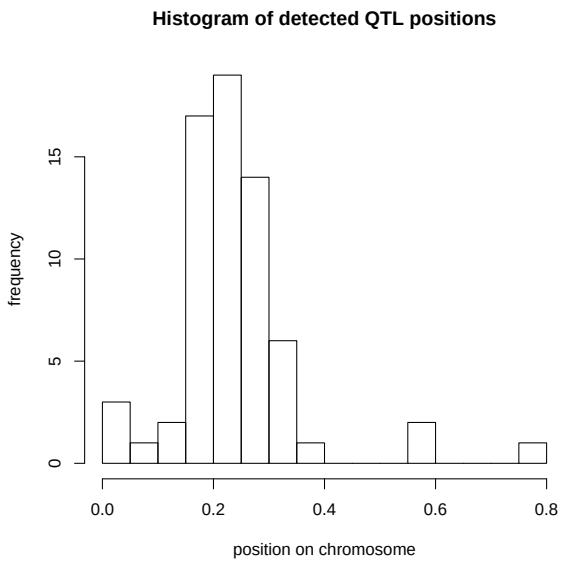

Figure 13: Detected QTL-positions (cow, $c = 1.3$)

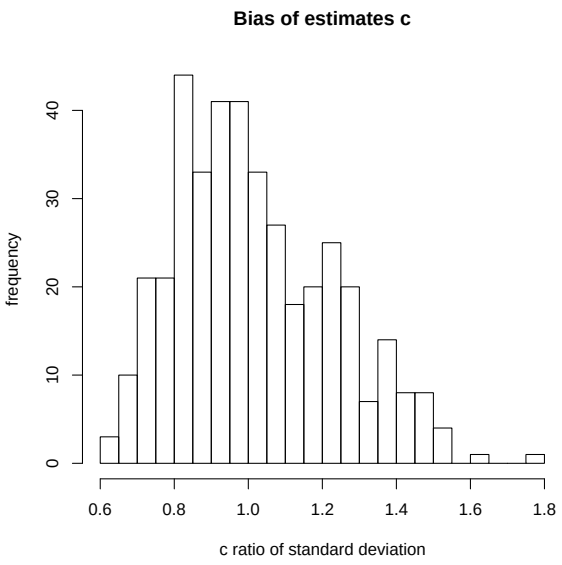

Figure 14: Bias of estimator $\widehat{c}_i$ (cow, $c = 1.3$)

Thus we repeated the simulation of the model (9) using the non-normally distributed random deviation e_{ijk}. We applied the Gamma-distribution $\Gamma(\alpha, \beta)$ with $\alpha = \sigma_e^2$ and $\beta = 1$. Under this circumstance the test procedure behaves robust as seen in figure (15) and (16). Using c=1.2 the empirical power is $\widehat{\pi} = 94$ %, but 6 simulations create homozygote sires only.

6 Discussion

Refering to Good [4, p.26] the permutation test is a powerful and unbiased test to check the hypothesis H_0: $\beta = \beta_0$. Therefore we have constructed at least an asymptotic α-test when the number of permutations increases to all possibilities of permutations $(Nn)!$.
Churchill & Doerge [3] suggested to use at least 1,000 resamples of the datasets. As the empirical power of our simulations shows, applying the permutation test with only 100 resamples of the original dataset is quite successful.

We mention that Haley & Knott [5] suggest to use one additional degree of freedom for the presumed QTL-position in the divisor of equation (6). It is questionable, if those adjustment is necessary. Under H_0 the additional parameter for the QTL-position is not defined. Lander & Botstein [6] recommend the application of an Ornstein-Uhlenbeck diffusion process to determine the distribution under the null hypothesis.
In our investigations using the procedure of the permutation test the number of degrees of freedom only serve as a scaling factor and could be neglected.

The parameter c^2 declares the ratio of within litter variance when inheriting the allele q from the heterozygote sire to the within litter variance when inheriting the allele Q. If the parameter c is significantly different from 1, it is ambigious whether the within litter variance is affected by the random deviation or by the enlarged polygenic variance or by an increased QTL variance, see equation (2). When a non-significant result is observed, a constant within litter variance could also be generated by e.g. an increased random deviation and decreased polygenic variance. The parameter c is therefore just a net-effect for any changes of the components of the within litter variance. In the investigations of the diary cattle the within individual variance is well-defined by

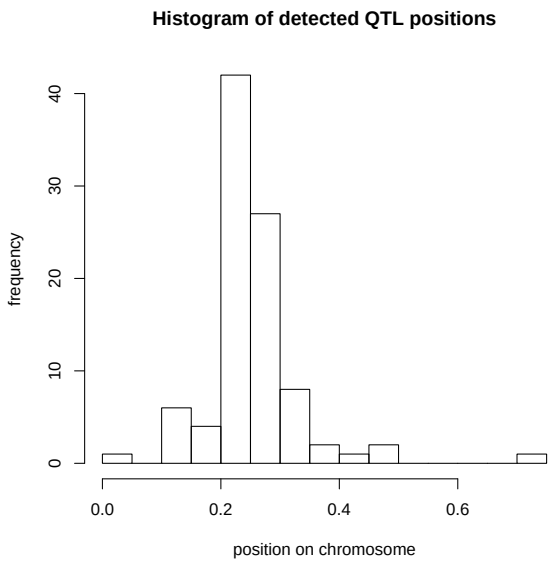

Figure 15: Detected QTL-positions (gamma)

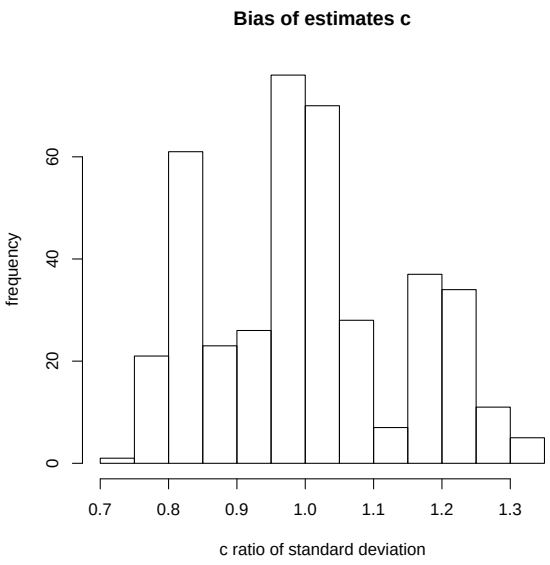

Figure 16: Bias of estimator $\widehat{c}_i$ (gamma, $c = 1.2$)

the existence of only one component.
Weller [10] is engaged in this topic with main emphasis on selective genotyping. He pointed out the economically importance of variance effects. For example, harvesting of fruits is efficient, if all individuals are ripe at the same time.

The applied model of the daughter-design is expandable to the grand-daughter-design with some modifications. Assuming N grandsires, each of them is mated to n unrelated grandams of the population. We select one son per mating. These sons are mated to m unrelated dams. One daughter per mating is chosen to investigate the within litter variance. For each grand-daughter we have to calculate the sample variance S^2_{ijk} of birth weights, $i = 1, \ldots, N, j = 1, \ldots, n, k = 1, \ldots, m$. Therefore we may assign to every sire the pooled variance of sample variances of his daughters $S^2_{ij} = \frac{1}{\sum_{k=1}^{m} f_{ijk}} \sum_{k=1}^{m} f_{ijk} S^2_{ijk}$. f_{ijk} denotes the degrees of freedom per litter. Thus we take as observation for every sire the pooled variance and apply the techniques as mentioned above with substitution of f_{ij} by $\sum_{k=1}^{m} f_{ijk}$ in the covariance matrix W.
The F-2 or backcross design may be treated as special cases of this consideration.

In our examinations we presumed the gene frequency $p_Q = \frac{1}{2}$. If we apply a decreased gene-frequency e.g. $p_Q = 0.3$, the prediction of the pointwise power is declined, see figure (17). In our pointwise investigations using $N = 4$ sires the power is expected to be 88.72 % at maximum.

In further studies we still have to calculate the confidence interval to assess the quality of the estimator $\widehat{\beta}$ and therefore also $\widehat{c}$.

References

[1] Box and Cox, *An Analysis of Transformations*, Journal of the Royal Statistical Society (B) 26: 211-252, 1964

[2] Christensen, *Analysis of Variance, Design and Regression*, Chapman and Hall, 1998

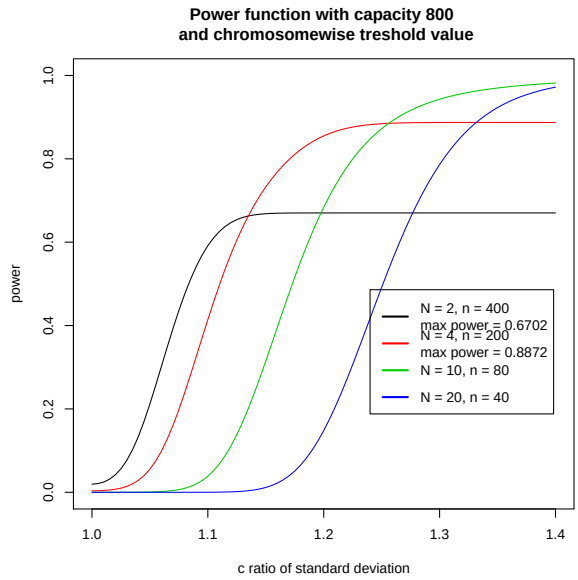

Figure 17: Power depending on given capacity, $p_Q = 0.3$

[3] Churchill and Doerge, *Empirical Treshold Values for Quantitative Trait Mapping*, Genetics Society of America 138: 963-971, 1994

[4] Good, *Permutation Tests*, Springer Series in Statistics, 1993

[5] Haley and Knott, *A simple regression method for mapping quantitative trait loci in lines crosses using flanking markers*, Heredity 69: 315-324, 1992

[6] Lander and Botstein, *Mapping Mendelian Factors Underlying Quantitative Traits Using RFLP Linkage Maps*, Genetics Society of America 121: 185-199, 1989

[7] Lehmann, *Testing Statistical Hypothesis*, Springer Texts in Statistics, 1997

[8] Searle, Casella and McCulloch, *Variance Components*, Wiley Series in Probability and Mathematical Statistics, 1992

[9] Seber, *Linear Regression Analysis*, Wiley Series in Probability and Mathematical Statistics, 1977

[10] Weller, *Power of different sampling strategies to detect quantitative trait loci variance effects*, Theoretical and Applied Genetics 83: 582-588, 1992