Mathematical Statistics
Stockholm University

A mixed model with repeated measures of mammographic breast density

Urban Olanders

Examensarbete 2005:10

Postal address:

Mathematical Statistics
Dept. of Mathematics
Stockholm University
SE-106 91 Stockholm
Sweden

Internet:

http://www.math.su.se/matstat

Mathematical Statistics
Stockholm University
Examensarbete **2005:10**,
http://www.math.su.se/matstat

A mixed model with repeated measures of mammographic breast density

Urban Olanders*

November 2005

Abstract

Breast density are brighter parts of a mammography x-ray film. Breast density is a risk factor for breast cancer, it increases in hormone therapy. The prognosis of breast cancer is in many cases favorable. The x-ray pictures of 28 healthy women are digitized to examine change of breast density over time. For each picture there is a histogram of the number of pixels for all gray scale values, and the proportion in the breast that represents breast density is measured. The quality of the scanned picture is low but conclusions are still possible to draw. The data is analyzed with a mixed model with repeated measures. The proportion of breast density decreases with 18% each year ($p<0.001$).

50 histograms of x-rays from University of South Florida digital mammography home page are analyzed. The histogram has a mixture of two normal distributions. The parameters of these distributions are p, μ_1, σ_1^2 , μ_2 and σ_2^2. ACR breast density rate is a subjective measure of mammographic breast density. Simple linear regression suggests that there is a negative correlation between age and ACR breast density rate and a positive correlation between ACR breast density rate and σ_2^2.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: urol5769@student.su.se. Supervisor: Mikael Andersson.

Preface

This thesis in mathematical statistics is done at Stockholm University and in collaboration with Department of Medical Epidemiology and Biostatistics at the Karolinska Institute and Department of Diagnostic Radiology, Karolinska Hospital, Solna. I would like to thank all who have been involved in this project:

Mikael Andersson, Division of Mathematical Statistics, Stockholm University.
Gunilla Svane & Edward Azavedo, Department of Diagnostic Radiology, Karolinska Hospital, Solna.
Jan-Eric Litton & Per Hall, Department of Medical Epidemiology and Biostatistics at the Karolinska Institute, Solna.
Ulf Olsson, Department of Biometry and Engineering, Swedish University of Agricultural Sciences, Uppsala.

Contents

1 Theory of mixed models

1.1 Introduction

A statistical model is a way to describe how different background variables affect an outcome variable. The model can be used to test hypotheses, make estimations or predict responses. There are parametric and non-parametric models. In a parametric model the outcome variable is assumed to belong to a certain class of distributions whereas this assumption is not necessary for a non parametric model. The non-parametric models have often lower power, i.e. low probability to reject a hypothesis when it is not true. The parametric models can be divided into linear and non-linear models. The non linear models may be either intrinsically non-linear, non-linear in the parameters or in the variables. Among the linear models are the general linear models, where the response variable has a normal distribution, and the generalized linear model, where the response variable may have other distributions than the normal distribution. There are also Bayesian models that take into account that the parameters in the distribution from which the response variable belong are stochastic variables.

1.2 The usual linear model

Let $\mathbf{y}$= (y_1, ..., y_i, ...,y_n)' be an $n \times 1$ vector of observations from some study. In matrix terms a general linear model is written

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

For observation i the model is

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_{p-1} x_{i(p-1)} + e_i$$

x_{i1}, x_{i2}, …, $x_{i,p-1}$ are independent variables. $\mathbf{X}$ is a design matrix and $\boldsymbol{\beta}$ is a vector of the parameters.The components of the $\mathbf{e}$ vector: $e_1, e_2, \ldots, e_n$ are the residuals of the model and they are normally distributed $N(0, \sigma^2)$. The least squares estimates of the parameters in the $\boldsymbol{\beta}$ vector are obtained when the sum of squared residuals is minimal. The sum of squares is

$$\mathbf{e}'\mathbf{e} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

Using the least square method the estimators of the parameters are

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

The variance of a parameter estimator $\hat{\beta}_j$ is estimated as

$$\widehat{Var}(\hat{\beta}_j) = \sum_i w_{ij}^2 \hat{\sigma}^2$$

where w_{ij} are known weights and

$$\hat{\sigma}^2 = \frac{\sum_i \hat{e}_i^2}{n-p}$$

The test statistic to verify that the parameter β_j is zero is

$$T_j = \frac{\hat{\beta}_j}{\sqrt{\widehat{Var}(\hat{\beta}_j)}}$$

Under the null hypothesis that β_j=0 the test statistic has a t-distribution with $n - p$ degrees of freedom. For further details, see McCulloch, Generalized, Linear and Mixed Models.

1.3 The mixed model

A linear mixed model contains both fixed and random factors. For the fixed factors the levels of the effects are chosen beforehand, and these levels are the levels that are of interest to the researcher. Examples of fixed effects are time, different treatments or group characteristics that one wants to compare for instance like in this thesis the difference in breast density between the left and right breast.

If the levels of the factor can be considered a random sample from a population of values which is assumed to follow a certain distribution then it is a random effect. Random effects govern the variance-covariance structure of the outcome variable vector $\mathbf{y}$. For example if we randomly select n individuals from a set of patients and measure some value then the individual would be a random factor; we are not especially interested in these n individuals, we want to draw conclusions about the whole set of patients. The number of parameters in the model reduces in the mixed model because in the fixed model we must have one parameter for each individual, but this is not the case in mixed models.

A special case of mixed models is random effects models where there are only random effects. Contrary to the common linear model, the mixed linear model contains more than one variance parameter. In the fixed models, there is only one error term.

An example of mixed models is when there are three repeated measures made on n subjects. The response vector for subject i will be

$$\mathbf{y}_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ y_{i3} \end{bmatrix}$$

In matrix form this is written

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{u}_i + \mathbf{e}_i$$

where $\mathbf{y}_i$ is a 3×1 vector of observations, $\mathbf{X}_i$ is a 3×4 design matrix for fixed factors, $\boldsymbol{\beta}$ is a 4×1 vector of unknown fixed parameters. $\mathbf{Z}_i$ is a 3×1 design matrix for random factors, $\mathbf{u}_i$ is a vector of random factors of dimension 1×1, $\mathbf{e}_i$ is a 3×1 error vector. For one individual the model will be

$$\mathbf{y}_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ y_{i3} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} (u_i) + \begin{bmatrix} e_{i1} \\ e_{i2} \\ e_{i3} \end{bmatrix}$$

Bringing together the $\mathbf{y}_i$ vectors gives

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_n \end{bmatrix}$$

The model for all n individuals is

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_n \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_n \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \mathbf{Z}_1 & 0 & \dots & 0 \\ 0 & \mathbf{Z}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{Z}_n \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_n \end{bmatrix} + \begin{bmatrix} \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_n \end{bmatrix}$$

The vectors $\mathbf{u}_i$ and $\mathbf{e}_i$ are the random components of the model. Denote the covariance matrix of $\mathbf{u}_i$ with $\mathbf{G}_i$ and the covariance matrix of $\mathbf{e}_i$ with $\mathbf{R}_i$. Then

$$Var(\mathbf{y}_i) = \mathbf{\Sigma}_i = Var(\mathbf{Z}_i\mathbf{u}_i) + Var(\mathbf{e}_i) = \mathbf{Z}_i\mathbf{G}_i\mathbf{Z}_i' + \mathbf{R}_i$$

Since the covariance between individuals is zero, the covariance matrix of $\mathbf{y}$ has the structure

$$Cov(\mathbf{y}) = \begin{bmatrix} \mathbf{\Sigma}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{\Sigma}_n \end{bmatrix}$$

It holds that

$$E\begin{bmatrix} \mathbf{u} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}$$

and

$$Var\begin{bmatrix} \mathbf{u} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{bmatrix}$$

For further details, see Applied Mixed Models in Medicine, Helen Brown.

1.3.1 Estimation of parameters

In the model fitting process, estimating fixed effects, random effects and variance parameters is done. To get estimations of fixed parameters and variance parameters, likelihood functions are used. Given the data, the likelihood of a model is the probability of the data under that model. A model can have an infinite set of different parameter values, one of which has the highest likelihood. In a model where there are n independent observations the likelihood function is the product of the n density functions of the distributions from which the observations come. However, since in a mixed model measures are made on the same subject, the observations are dependent, so the likelihood function should be based on the multivariate density function

$$L = \frac{1}{(2\pi)^{(1/2)n}|\mathbf{V}|^{1/2}} exp[-\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})]$$

where n is the number of observations and $\mathbf{V} = \mathbf{ZGZ'} + \mathbf{R}$.

Maximum likelihood

Since the maximum of the logarithm (log) of the likelihood function is attained in the same point as the maximum of the likelihood function and since the log likelihood function is simpler to work with, it is used when estimating the fixed effects. The log likelihood function is given by

$$LL = k - \frac{1}{2}[log|\mathbf{V}| + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})]$$

where k is a constant.To get estimates of the fixed parameters and the variance of the parameters in the model the log likelihood is maximized by differentiating the log likelihood function with respect to $\boldsymbol{\beta}$ and setting the resulting expression to zero.

$$\mathbf{X'V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{0}$$

Rearrangements gives

$$\hat{\boldsymbol{\beta}} = (\mathbf{X'V}^{-1}\mathbf{X})^{-1}\mathbf{X'V}^{-1}\mathbf{y}$$

The variance of $\boldsymbol{\beta}$ is

$$var(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}var(\mathbf{y})\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$$
$$= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{V}\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$$

Residual maximum likelihood (**REML**)

ML estimators of variances are biased downward. This problem will be especially significant when there are a lot of parameters compared to the number of observations. An example of this is when we have a sample of n observations from a stochastic variable, then, if $\hat{\mu}$ is a sample mean the ML-estimation of the variance would be $\Sigma_i(x_i - \hat{\mu})^2/n$ instead of the unbiased estimator $\Sigma_i(x_i - \hat{\mu})^2/(n-1)$.

To get unbiased estimators the residual maximum likelihood REML has been developed. In REML estimation, the likelihood function is based on the residual terms $\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}$. This is done by a linear transformation of $\mathbf{y}$ to $\mathbf{w} = \mathbf{a}'\mathbf{y}$ where $\mathbf{a}' = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}'$. Then

$$E(\mathbf{a}'\mathbf{y}) = (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-\mathbf{1}}\mathbf{X}')\mathbf{X}\boldsymbol{\beta} = \mathbf{X}\boldsymbol{\beta} - \mathbf{X}\boldsymbol{\beta} = \mathbf{0}$$

In balanced mixed models the REML equations have a unique solution that is the minimum variance unbiased estimator

$$\hat{\sigma}^2 = \frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{n-p}$$

In general, however, the estimation of variations, i.e. the parameters in $\mathbf{V}$ can not be done with one explicit equation since the derivates of the log likelihood with respect to the variance parameters are non-linear, instead iterative methods like Newton-Raphson are used.

Newton- Raphson Iteration

The residual log likelihood function (RLL) is

$$RLL = -\frac{1}{2}[log|\mathbf{V}| + log|\mathbf{X}'\mathbf{V^{-1}X}| + (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})]$$

The function can be maximized with the Newton-Raphson algorithm. Let $f(\boldsymbol{\theta})$ be be the RLL function, where $\boldsymbol{\theta}$ are the parameters in $\mathbf{V}$ that shall be estimated. The RLL function will have its maximum when

$$\frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{0}$$

Setting a Taylor expansion of $\partial f(\boldsymbol{\theta})/\partial \boldsymbol{\theta}$ equal to zero gives

$$\frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = f'(\boldsymbol{\theta}) \approx f'(\boldsymbol{\theta}_0) + \frac{\partial^2 f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) = \mathbf{0}$$

Rearrangements gives

$$\boldsymbol{\theta} = \boldsymbol{\theta}_0 - \left[\frac{\partial^2 f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\right]^{-1} f'(\boldsymbol{\theta}_0)$$

So the Newton-Raphson algorithm that gives the variance parameters is

$$\boldsymbol{\theta}^{(m+1)} = \boldsymbol{\theta}^{(m)} - \left[\frac{\partial^2 f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\right]^{-1}\Bigg|_{\theta=\theta^{(m)}} f'(\boldsymbol{\theta}^{(m)})$$

Two disadvantages of the Newton-Raphson method are that it does not always give convergence, and the variance estimates can be negative. An alternative to this method is the Fischer scoring algorithm where the Hessian matrix is replaced by its expectation, the so called information matrix. The algorithm will converge even with poor starting values with the Fischer scoring. It is also possible to do the estimation of the parameters with generalized expectation maximization (GEM) algorithm, GEM is feasible when there is a large number of covariance parameters.

1.3.2 Negative variance components

The methods of estimating variance components can lead to negative values, which is not a reasonable result, since all variances are non-negative. When there are negative variance values, the real values are generally small or zero. The risk of obtaining negative values of variance increases if the number of random effect categories and the number of observations per category are small. Negative variance components are handled either by removing the corresponding random effect from the model, or to set the variance component to zero.

1.3.3 Bias in fixed and random effects standard error

When there is a large degree of imbalance in the data there will be downward bias in the standard errors. An adjustment for the bias is the "empirical" variance estimator

$$\widehat{var}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\widehat{var}(\mathbf{y})\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$$

where $\widehat{var}(\mathbf{y}) = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'$. The Bayesian method is another way to deal with the bias of standard errors.

1.3.4 Significance testing

Contrasts are used for local tests of the significance of fixed and random effects, for fixed effects the contrast is $\mathbf{C} = \mathbf{L}'\boldsymbol{\beta} = \mathbf{0}$ and for random effects $\mathbf{C} = \mathbf{L}'\mathbf{u} = \mathbf{0}$. The terms in the $\mathbf{L}$ matrix correspond to the intercept and to different treatments. For example, when comparing three treatments A, B and C, if one wants to examine the difference between treatment A and B the contrast will be

$$\mathbf{L}'\hat{\boldsymbol{\beta}} = (0 \quad 1 \quad -1 \quad 0)\hat{\boldsymbol{\beta}} = \hat{\beta}_A - \hat{\beta}_B$$

With a multiple contrast for instance

$$\mathbf{L}'\hat{\boldsymbol{\beta}} = \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix} \hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_A - \hat{\beta}_B \\ \hat{\beta}_A - \hat{\beta}_C \end{pmatrix}$$

$$\mathbf{L}'\hat{\boldsymbol{\beta}} = \begin{pmatrix} 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix} \hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_A - \hat{\beta}_B \\ \hat{\beta}_A - \hat{\beta}_C \end{pmatrix}$$

the overall equality of treatments can be tested.

The Wald statistic

To test that a multiple contrast is zero, the Wald statistic is used, which is given by

$$\begin{aligned} W &= (\mathbf{L}'\hat{\boldsymbol{\beta}})'(var(\mathbf{L}\hat{\boldsymbol{\beta}}))^{-1}(\mathbf{L}'\hat{\boldsymbol{\beta}}) \\ &= (\mathbf{L}'\hat{\boldsymbol{\beta}})'(\mathbf{L}'var(\hat{\boldsymbol{\beta}})\mathbf{L})^{-1}(\mathbf{L}\hat{\boldsymbol{\beta}}) \end{aligned}$$

for fixed effects, for random effects $\hat{\mathbf{u}}$ is used in place of $\hat{\boldsymbol{\beta}}$. W can also be written

$$W = (\mathbf{L}'\hat{\boldsymbol{\beta}})'(\mathbf{L}'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{L})^{-1}(\mathbf{L}'\hat{\boldsymbol{\beta}})$$

since $var(\hat{\boldsymbol{\beta}}) = (\mathbf{X}\mathbf{V}^{-1}\mathbf{X})^{-1}$. When the sample size is large and data are balanced W has approximately a χ^2_p distribution (p is the rank of $\mathbf{L}$) otherwise the Wald F statistic is used, this is calculated by

$$F_{DF1,DF2} = W/DF1$$

where $DF1$ is the number of linearly independent rows in $\mathbf{L}$, and $DF2$ is the denominator degrees of freedom (DF) that corresponds to the DF of the contrast variance $\mathbf{L}'var(\hat{\boldsymbol{\beta}})\mathbf{L}$.

1.3.5 Lost data

As long as data are missing at random, it is possible to make inference with the mixed model, the model is quite robust. A fixed effects model on the other hand is less flexible concerning missing data, since information on subjects with one or more values missing is completely lost. The reasons for withdrawal should be carefully examined since it is very common that data are missing for non-random reasons, this makes the interpretation of results tricky. Simple summaries of the frequency of missing data can be helpful. (See more about missing data in the chapter 5.1.) That data are missing at random means that there are no systematic causes to the loss of data. An example a systematic cause of missing data is when data that are extremely high (or low) are missing more often then average values.

2 Theory of repeated measures data

2.1 Introduction

Repeated measurements data can be obtained either by measuring some variable in a subject on multiple occasions or under multiple conditions. Measurements at different time points on the same subject are also called longitudinal data. The response variable might be either univariate or multivariate. The experimental units might be individuals or, like in this thesis, each of the breasts in a subject. Whether or not the time intervals are equidistant or not will affect the study design.

2.2 Advantages and disadvantages of repeated measurements design

In statistical model designs with repeated measurements information concerning pattern of changes in individuals can be gained. It can be measured if some data increases, is constant or decreases over time. When one wants to study changes over time, it is more effective to do repeated measurements on the same individual than to observe different individuals at each time point. When comparing different treatments, measurements on an individual can be made both under control (placebo) and treatment condition, this sharpens the estimation of relevant parameters. Practical issues can make it easier to collect data at several time points from the same individual rather than from different individuals. The more time points when measurements are made, the more missing data there may be for reasons that cannot be controlled. It has to be considered that when measurements are made on the same subject, these data are not independent. The efforts to develop models with repeated measurements have principally been directed to models with a response variable that is normally distributed.

2.3 Fixed effects models

2.3.1 Univariate methods

The multivariate response with a lot of measured data on each subject can always be transformed to a univariate response, for instance by estimating the least square regression slope of the measurements of each individual, by estimating the area under the curve (AUC) or by taking the last minus the first measured value. The advantage of this is that the issue of correlation between different time points of the same individual disappears. These methods are referred to as the summary statistic approach. An important condition is that the selected summary measure really describe the subjects data in an accurate way. If there are missing measurements this can sometimes be dealt with by extrapolation. If the data are normally distributed ANOVA and the F statistic gives the estimation of the parameters and their confidence interval, otherwise non-parametric tests like the Kruskal-Wallis test is used.

2.3.2 Multivariate methods

In the multivariate analysis of variance (MANOVA) the model is

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

now the response variable matrix $\mathbf{Y}$ has n rows and p columns where n and p are the numbers of individuals and measurement points respectively, in contrast to $\mathbf{Y}$ in univariate models where $\mathbf{Y}$ is a $n \times 1$ vector. $\mathbf{X}$ is a $n \times q$ matrix and $\boldsymbol{\beta}$ is a $q \times p$ matrix that contains the parameters where q is the number of parameters. There are several alternative test statics in

MANOVA for the hypothesis that all levels of a certain factor are equal, for example Hotellings T^2 in the case of two groups, or in the general case Wilks' lambda, Roy's largest root criterion or Pillai's trace. For details about multivariate methods, see Statistical Methods for the Analysis of Repeated Measurements, Charles Davis.

2.4 Mixed models

The linear mixed model for repeated measurements is

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{u}_i + \mathbf{e}_i$$

$i = 1, 2, \ldots, n$. (see section 1.3). The $\mathbf{y}_i$ vectors has normal distribution with $E(\ \mathbf{y}_i)=\ \mathbf{X}_i\boldsymbol{\beta}$ and $Var(\ \mathbf{y}_i)=\mathbf{V}_i$ where $\mathbf{V}_i\ =\ \mathbf{Z}_i\mathbf{G}\mathbf{Z}_i' + \mathbf{R}_i$. The diagonal-matrix $\mathbf{R}$ looks like

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_1 & 0 & \ldots & 0. \\ 0 & \mathbf{R}_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \mathbf{R}_n \end{bmatrix}$$

Mixed models are very suitable for modeling repeated measurements because all individuals do not have to be measured at the same time points, missing data can be coped with as long as the data are missing at random and they allow several different patterns of the correlation over time e.g. we can determine whether there is a constant correlation between all the time points, or whether the pattern of correlation varies with time.

2.4.1 Covariance structures

The $\mathbf{R}_i$ matrix mentioned above is a $n \times n$ matrix where n and i are the number of measurement time points and experimental units, respectively. The $\mathbf{R}_i$ matrix can either be allowed to be free (**general** or **unstructured**) covariance structure or have a more restricted structure. General/Unstructured covariance structure looks as follows:

$$\mathbf{R}_i = \begin{bmatrix} \sigma_1^2 & \theta_{12} & \theta_{13} & \theta_{14} \\ \theta_{12} & \sigma_2^2 & \theta_{23} & \theta_{24} \\ \theta_{13} & \theta_{23} & \sigma_3^2 & \theta_{34} \\ \theta_{14} & \theta_{24} & \theta_{34} & \sigma_4^2 \end{bmatrix}$$

A reason for not choosing the general covariance pattern is that when there are many measurement time points on each subject, there will be a lot of parameters and this can cause the iteration algorithm of estimation not to converge. In the **first-order autoregressive** model are all the variances the same, and the co-variances decrease exponentially with time. It seems right that the correlation between a few time points is larger than between many time points. The model is especially suitable when time points are evenly spaced but it can also be used in trials when many measurements in short intervals are done just in the beginning of a trial and with increasingly separated intervals later on. First-order autoregressive covariance structure looks like:

$$\mathbf{R}_i = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

In the **compound symmetry** covariance model all covariances are equal. The compound symmetry and the first order autoregressive models contain only two parameters. Compound symmetry covariance structure looks as follows:

$$\mathbf{R_i} = \begin{bmatrix} \sigma^2 & \theta & \theta & \theta \\ \theta & \sigma^2 & \theta & \theta \\ \theta & \theta & \sigma^2 & \theta \\ \theta & \theta & \theta & \sigma^2 \end{bmatrix}$$

The **Toeplitz** model has a separate covariance for each level of separation. This means that the covariance between e.g. time point one and two will be the same as between time point two and three, and that the covariance between time point one and three will be the same as between two and four. Toeplitz covariance structure looks like:

$$\mathbf{R_i} = \begin{bmatrix} \sigma^2 & \theta_1 & \theta_2 & \theta_3 \\ \theta_1 & \sigma^2 & \theta_1 & \theta_2 \\ \theta_2 & \theta_1 & \sigma^2 & \theta_1 \\ \theta_3 & \theta_2 & \theta_1 & \sigma^2 \end{bmatrix}$$

In the **spatial power** covariance structure, correlation between two observations is proportional to the distance in time between them. If variability in a measurement differs between time points this can be handled with the **heterogeneous** covariance structure, where there are different parameters for the variance of each time point. The example shows heterogeneous first-order autoregressive but the heterogeneous structure can also be applied to the models above. Heterogeneous first-order autoregressive covariance structure looks as follows:

$$\mathbf{R_i} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & \rho^2\sigma_1\sigma_3 & \rho^3\sigma_1\sigma_4 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 & \rho\sigma_2\sigma_3 & \rho^2\sigma_2\sigma_4 \\ \rho^2\sigma_1\sigma_3 & \rho\sigma_2\sigma_3 & \sigma_3^2 & \rho\sigma_3\sigma_4 \\ \rho^3\sigma_1\sigma_4 & \rho^2\sigma_2\sigma_4 & \rho\sigma_3\sigma_4 & \sigma_4^2 \end{bmatrix}$$

2.5 Choice of covariance structure

The likelihood statistic is a measure of model fit, the higher likelihood the better is the fit of the model to the given data. But the likelihood increases when more covariance parameters are added, so the likelihood can only be used directly when comparing two models with the same number of parameters. When comparing models with different covariance structures the **likelihood ratio test** can be used if the models are nested within each other. Nesting of covariance parameters means that if the covariance parameters in a more complex model are restricted to a covariance pattern with a simpler structure, then these models are nested. It holds that approximately

$$2(\log(L_1) - \log(L_2)) \sim \chi^2_{DF}$$

where L_1 and L_2 are the likelihoods of the more complex and the simpler model respectively, and DF is the difference in numbers of covariance parameters. If the test statistic is significantly large on say 5% level then the more complex structure should be chosen. If the covariance parameters in the models are not nested, a comparison of each model with a simpler model which is nested within both models can be done. The model with the best improvement wins.

An alternative when the covariance parameters are not nested is the **Akaike information criterion** (AIC) which is given by

$$AIC = -2\log(L) + 2q$$

where q is the number of parameters in the model. When more parameters in the model are chosen q increases, and so does L, so AIC takes into account both that the likelihood should be large and that the model should be simple, i.e. that there are few parameters. The choice of covariance structure can give additional knowledge to the topic studied, it gives estimates of covariances between different time points.

2.6 Graphical presentation for repeated measures

A first approach to a statistical analysis is the graphical display of the repeated measures. Visualizing data makes it possible to see trends and differences between groups. All data shall be shown rather then data summaries. The graphical presentation shall contain both cross-sectional and longitudinal data, and outliers and unusual observations shall be easy to observe. This is done by using individual profile plot where the development over time is shown for each individual, the mean profile plot that shows the mean values and standard error of the mean for different groups in different occasions, and box-plots that give a more detailed information of the spread of data.

3 A longitudinal study of breast density

3.1 Introduction

Mammographic breast density is a feature of the mammography x-ray that is seen as brighter parts of the x-ray. The reason for these brighter parts is that different types of breast tissues are different radio translucent. Fibroglandular tissue is less radio translucent then fat tissue. In a breast with high proportion of fibroglandular tissue the x-ray-beams do not pass through the breast so easily, this will cause a brightness on the x-ray. Mammographic breast density can be measured as the percentage of the mammography x-ray that appears bright. Almost all women have breast density, ranging from less then 10% to more then 90%. High breast density is a risk factor for developing breast cancer. The purpose of this work is to study changes mammographic breast density over time, in a group who has not developed breast cancer, to see if breast density increases or decreases with age. For further information about breast density see Lundström et al. Am. J. Obstet Gynecol 2002;186:717-772

3.2 Method

Mammography pictures from 28 subjects that had done mammography screening 1990-1995 at Radiumhemmet at Karolinska hospital were decoded, the name and personal identity number were removed from the film. Each subject had pictures from two or three mammography screening occasions. The examination were normally done with an interval of two years. In each screening occasion four pictures were taken, two projections on each breast. The birth year of the subjects were between 36-42. (fig. 1). There were a total of 135 pictures. With a HP scanjet 3970 scanner, the images were digitalized. The digitalized picture is built up by squares each called a pixel. The measure of the size of a pixel is d.p.i, dot per inch , the resolution of the pictures were 100 d.p.i.The digitized x-ray consists of 256 gray-levels, and each pixel has a certain gray level value ranging from 1/256 to 256/256 where 1/256 is the darkest gray level black and 256/256 the brightest white. A digitalized x-ray has a matrix that contains all gray level values of the pixels. A certain position in the matrix corresponds to the pixel gray level in the same position in the picture. (E.g. If there is a low value (7/256) in the first row and first column in the matrix, the first pixel in the upper left corner of the picture is dark.) With an option in the computer program that places a polygon on the picture we select only pixels within the breast tissue i.e. the background in the picture and the breast muscle is excluded. A histogram is made over the gray level values. We see how many pixels there are in the picture that have gray level value 1/256 or 2/256 (black) as the first bar to the left in

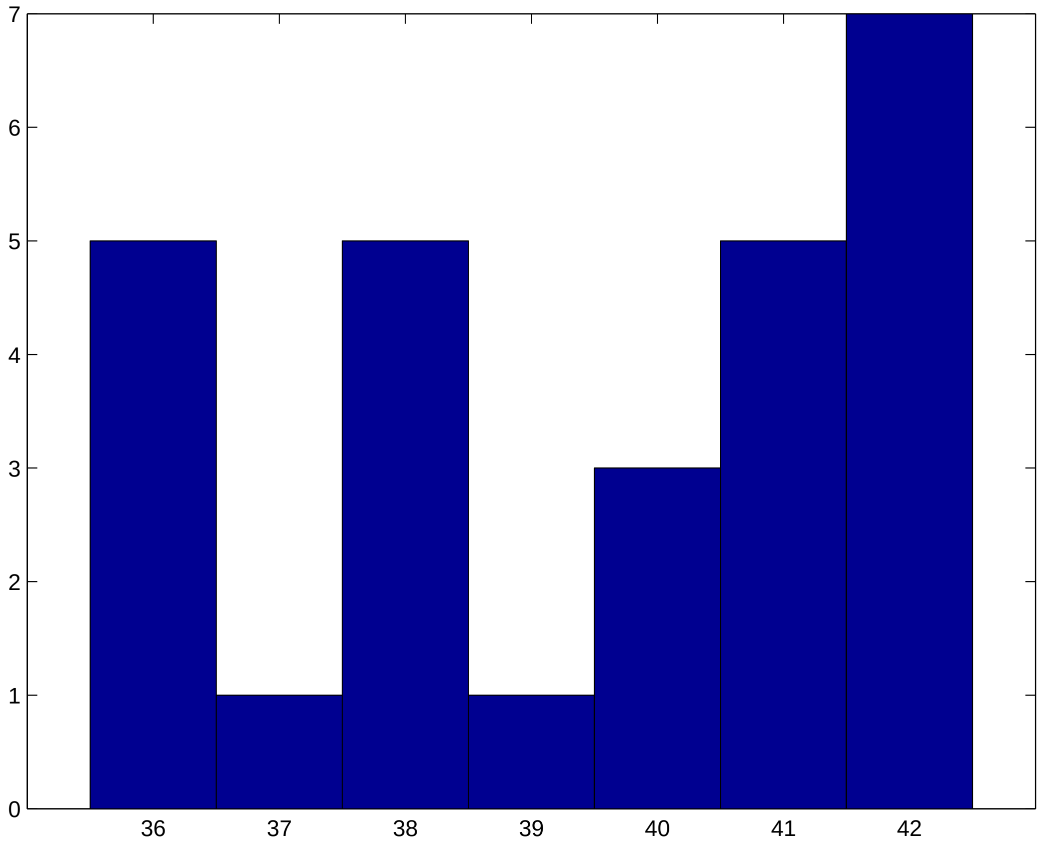

Figure 1: Age distribution of the women, birth year on the x-axis, number on the y-axis.

the histogram, how many pixels there are with value 2/256 or 3/256 as the second bar on the left side of the histogram and so on to the bar most to the right in the histogram, that shows how many pixels there are that have gray level value 255/256 or 256/256 (white). Totally there are 128 bars in the histogram. In most pictures, brighter parts of the picture are centered in the region close to the nipple of the breast, especially when a large portion of the breast has high density, this feature is obvious.

When only the dense parts in the picture are selected, the histogram shows unimodal distribution with large mean value (bright), and when only the non-dense part is selected the histogram shows a unimodal distribution with small mean value (dark). The histogram of the picture that includes both the dense parts and the non-dense parts has a bimodal distribution, which can be regarded as a mixture of two distributions, with means and variance that coincide with the two separate distributions for dense and non dense areas respectively, as expected.

3.3 Missing values, selection

The data consisted of 14 subjects that had x-ray pictures from two examination occasions, and 14 subjects that had pictures from three occasions. From each occasion the medio lateral oblique MLO projection (from the side) of the left and right breast were selected, in a mammographic examination a craniocaudal CC (from upside) projection is also done but these projections were not included. The feature of the breast density of the two projections is similar, so the results of the craniocaudal projection should be similar also to the results of this investigation. The time interval between examinations was approximately two years. In a few cases the time interval was one or three years. The x-ray films were selected from a population of healthy subjects, who had undergone mammography screening in a regular screening

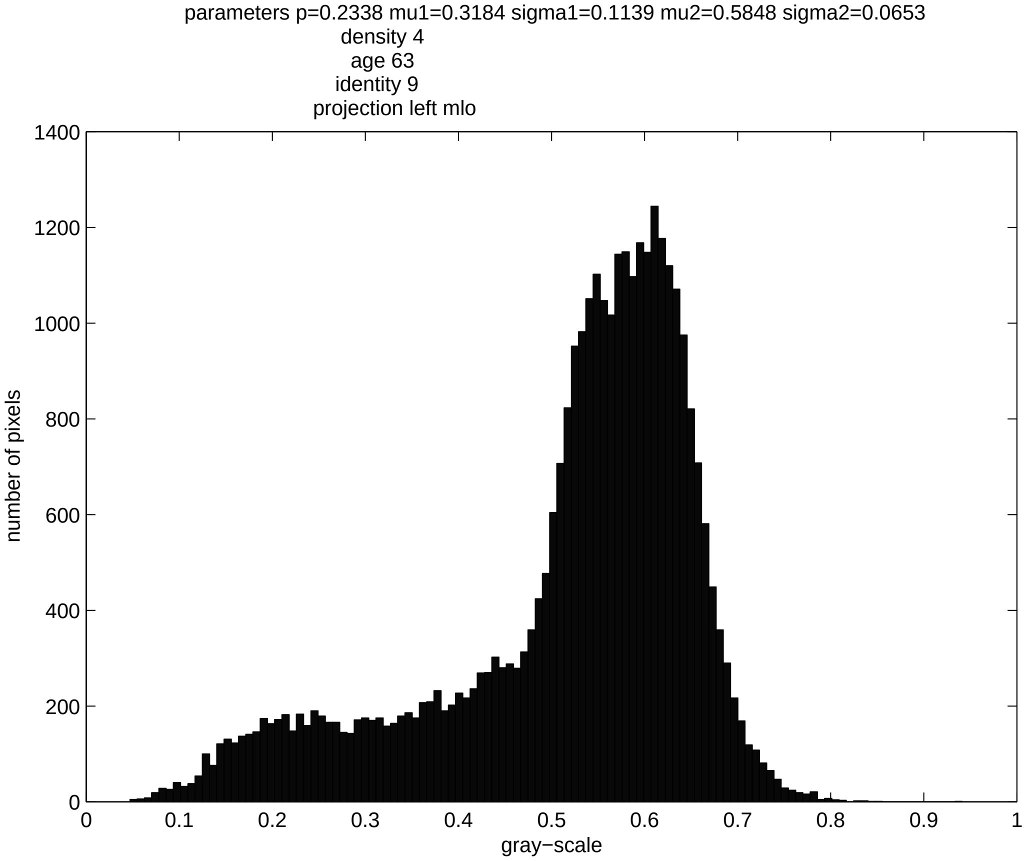

Figure 2: Example of histogram of a x-ray where the scanning has been optimal. The picture is from the University of South Florida digital mammography home page. The data are fitted to two mixed normal distributions. The distribution to the left has μ_1=0.318 and σ_1=0.114 and the distribution to the right has μ_2=0.585 and σ_2=0.085. p=0.2338 is the proportion of numbers of pixels in the left mode of the distribution compared to all the pixels in the whole histogram. A high pixel value means that the pixel is bright, i.e. that there is high breast density. Many pixels with high gray scale value, i.e. that p is low or μ_2 is high should indicate high density in the breast.

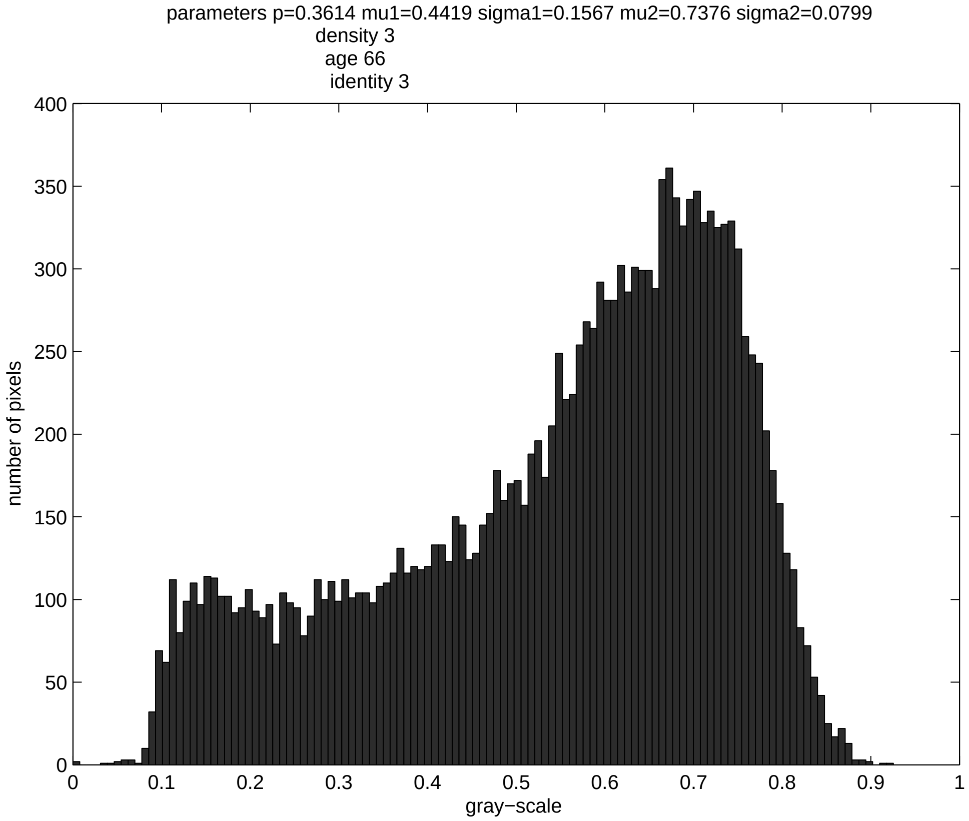

Figure 3: Example of histogram of a x-ray from the where the scanning has been optimal. The picture is from the University of South Florida digital mammography home page.

program, and who had not later been stricken with breast cancer.The selection was done so that there were more subjects with high density in the sample of 28 then in the population from which the sample was drawn from, i.e. some pictures with low breast density was sorted out. This for reasons that it probably should be easier to investigate breast density in subjects with high breast density, and that breast density is a risk factor for breast cancer. There is no special reason, why there are different numbers of occasions of examination, two or three within the subjects. Since not all the subjects have three examinations there will be missing values in the repeated measurements design for those subjects with two examinations. It can be assumed that those with three examinations do not differ considering breast density, from those with two, though this can be questioned, so the values that are missing are assumed missing at random. An example how there can be a difference between the two groups is that those with two examinations have participated in fewer examinations, and this for reasons that is associated with specific pattern of breast density change over time. It can also be that that those with two examinations had also a third examination, but that this was not chosen. One subject's pictures were not used, because the feature of the histograms differed a lot from all other subjects' pictures. The reason for this was probably that these pictures (which was taken about ten years earlier than all other pictures) were produced with other techniques or other film type. One picture was lost in the scanning procedure.

3.4 The parameters

A mixture of two normal distribution has five parameters, p, μ_1, σ_1^2 , μ_2 and σ_2^2 .The parameter p is the proportion of the distribution closest to zero compared to both distributions, i.e. if p is small the first distribution is small compared to the second, μ_1, μ_2 , σ_1^2 and σ_2^2 are the mean values and variance for the distribution closest to zero and one, respectively. It is reasonable to think that in a picture with low breast density the p value is high reflecting that most gray level values are low (dark) and thus belongs to the first distribution, the one nearest zero with mean value μ_1 and variance σ_1^2 . When there is little breast density μ_2 and the difference between μ_1 and μ_2 should be low. These changes can also be expected when breast density changes from one mammography occasion to another when one subject is examined, it is reasonable that changes in density should be reflected in changes in the parameters of the mixture distribution. The change of the parameters over time may be almost the same in both breasts or they may be different. The risk of developing breast cancer increases with age and breast density is a risk factor for breast cancer so considering this, breast density should increase with age. It could also be that breast density with age decreases more in those who don't develop cancer compared with those who do. On the other hand the glandular tissue stimulating sex hormones decreases after menopause causing the ratio fat tissue/fibroglandular tissue to increase leading to less mammographic density.

Since the gray level values are larger then zero and smaller then one the data can also be modeled as a mixture of two beta-distributions. The fit of such a model is probably better then the normal distribution. (Fig. 2, 3)

3.5 Non optimal scanning

When comparing histogram from the scanned pictures in this study with pictures scanned under optimal conditions from the University of South Florida digital mammography home page with for instance a Lumisys 200 laser densitometer scanner it is apparent that the scanning in this study has severe deficiencies that distorts the histogram, the appearance of two distinct mixed distributions is lost in the scanning, some values has inexplicable many numbers of pixels (the peaks in the histogram) due to a shade on the scanned pictures. Something in the scanning process interprets the x-ray as if it is a very common gray level value, which it in reality is not, there was no shade on the x-ray films. When doing ML-estimation of the mixed distribution this single extreme value has a great impact so the ML-estimation in the first distribution will have a mean about 0.1 and with a variance close to zero, representing the extreme value, the second distribution representing the rest of the histogram, this is of course not the correct two distributions, the parameters of these two distributions do not give any information about the breast density. Doing a ML-estimation with three distributions and putting a restriction on the ML-estimation that $\sigma_1^2 > 0.001$ and $\mu_3 > 0.4$ only results in that σ_1^2 becomes 0.001 or that μ_3 becomes 0.4 which is not either a reasonable result. Scanning with 300 and 600 d.p.i. does not improve the quality of the scanned pictures.

3.5.1 The parameter in this study

Since the lack of quality in the scanned pictures makes it impossible to estimate the suggested parameters p, μ_1, μ_2, σ_1^2 and σ_2^2, the proportion parameter, defined as the number of pixels with gray scale values > 0.4/total number of pixels is the outcome variable in this study. 0.4 is measured as the limit for high breast density. Since the quality of the scanned pictures is so imperfect the values of the outcome variable is probably not the same in pictures scanned under optimal conditions, the proportion is probably higher in the latter pictures. (Fig. 4). The data are assumed to be log-normal, the proportion decreases with time, (Fig. 5, 6).

3.5.2 An alternative parameter

The absolute number of pixels > 0.4 is a compatible parameter to the proportion parameter. A comparison is made on six subjects, five measurements are

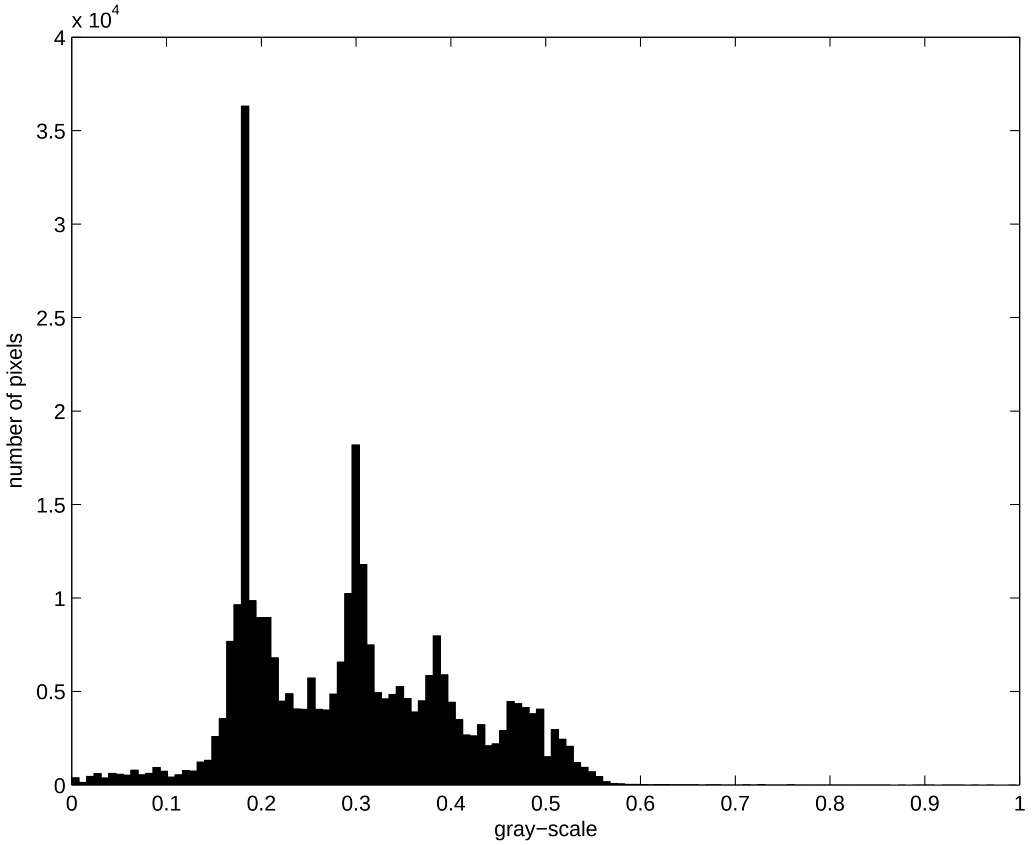

Figure 4: Example of histogram from this investigation. The feature of a mixed distribution is lost due to non-optimal scanning. A shade more or less randomly scattered over the scanned picture surface gives rise to the spike about 0.19. Gray scale value higher then 0.4 is chosen as a limit value for high breast density. The outcome variable is the number of pixels with gray scale value higher then 0.4 divided with the number of all pixels in the picture, i.e. the proportion of the picture that has high breast density. Since the resolution is higher, there are more pixels in each bar than in the pictures from the University of South Florida digital mammography home page.

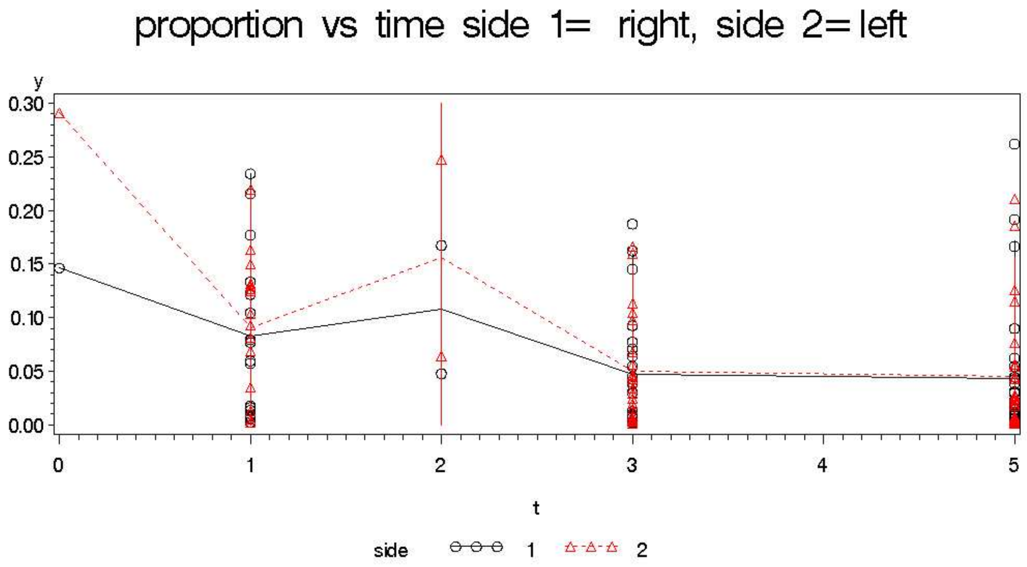

Figure 5: Proportion vs. time. There is a decrease of the proportion with time.

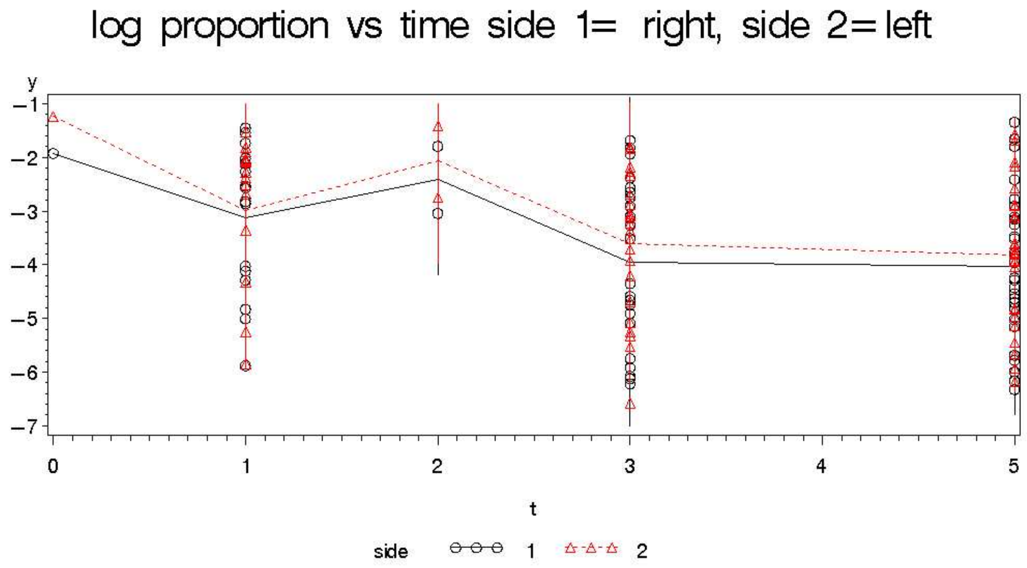

Figure 6: log. proportion vs. time.

$\overline{x}$= number of pixels > 0.4	std/$\overline{x}$	$\overline{y}$=fraction > 0.4	std/$\overline{y}$
32780	0.000694	0.2009	0.0311
25880	0.0015	0.2285	0.0074
8100	0.0232	0.0587	0.0294
5900	0.0215	0.0474	0.0138
2120	0.1217	0.0189	0.0952
1976	0.082	0.0724	0.058

Table 1: The data indicates that when there is little breast density, i.e small proportion or little numbers of pixels with high values, the standard deviation/mean value (variation coefficient) is lower for the proportion measurement then for the absolute number measurement while the opposite is true for much breast density.

made on each. The differences in the variation coefficients are partly due to that when there are few pixels>0.4 the impact of artefacts becomes greater, and that the placing of the polygon is less crucial when there is high breast density. (table 1)

3.6 The choice of model

For a mixed model, we choose the covariance structure of the **V** matrix and the fixed effects must be decided. The possible fixed effects that can affect the outcome variable proportion is time, side, age and interactions. One model for both the right and the left breast is chosen and not one model for each side.Two models for each of the breast were also considered and the results confirm the results in the model chosen here. Data is transformed by taking the logarithm of the proportion, that data is log-normal is shown by examining normal plots and histograms of the transformed data. Other transformations like the square root is also possible to do.

We want to decide the model that best fits the data. We look at models that includes all fixed effects and interactions and where time is modeled as a class variable. (model 1-6). The model with the smallest AIC is model 5. From model 5 we remove the effects with the highest p-values until only effects with significant (on a 5% level) effects remain. In model 5E time is highly significant. Backward elemination is also done in models where time is modelled as a continuous and not as a class variable (model 8A-9F). 9D is the model with the lowest AIC. Both side and time are significant effects. We chose model 9D even though 5D has lower AIC, because in 9D there is a parameter that describes how much breast density decreases every year, the interpretation is more direct. (table 2)

3.7 The covariance structure

The **G** matrix models the random effect of the individuals and the covariance between the two breasts in each individual. **G** has dimensions 27×27 with σ_G^2 in the diagonal and zeros off-diagonal. **Z** is a 324×27 matrix.

R models the variance-covariance between different time points in an individual. The subject that is repeated in the model is individual*side. There are measurements on six time points but each individual has two or three time points measured so the $\mathbf{R}_i$ matrix dimensions are is 4×4 and 6×6. In, for example, a compound symmetry or ar(1) covariance structure for an individual with two measurements $\mathbf{R}_i$ is

compound symmetry

$$\mathbf{R}_i = \begin{bmatrix} \sigma^2+\sigma_1 & \sigma_1 & 0 & 0 \\ \sigma_1 & \sigma^2+\sigma_1 & 0 & 0 \\ 0 & 0 & \sigma^2+\sigma_1 & \sigma_1 \\ 0 & 0 & \sigma_1 & \sigma_1^2+\sigma_1 \end{bmatrix}$$

where σ_1 is the covariance within the same breast for different time points.

Model	Class	log(y)=	# par.	Cov. type	-2LL	AIC
1	i s t	s a t s*a s*t a*t	41	un	380.4^{a}	462.4
2	i s t	s a t s*a s*t a*t	22	cs	389.3	433.3
3	i s t	s a t s*a s*t a*t	22	ar(1)	393.8^{a}	437.8
4	i s t a	s a t s*a s*t a*t	76	un	$-^{b}$	-
5	i s t a	s a t s*a s*t a*t	57	cs	314.8	428.8
6	i s t a	s a t s*a s*t a*t	57	ar(1)	$-^{c}$	-
7	i s	s a t s*a s*t a*t	29	un	$-^{c}$	-
8	i s	s a t s*a s*t a*t	10	cs	416.4	436.4
9	i s	s a t s*a s*t a*t	10	ar(1)	407.0	427.0
2A	i s t	s a t s*a a*t	17	cs	392.8	426.8
2B	i s t	s a t a*t	16	cs	389.9	421.9
2C	i s t	s a t	11	cs	388.9	410.9
2D	i s t	s t	10	cs	386.4	406.4
2E	i s t	t	9	cs	386.9	404.9
5A	i s t a	s a t t*a s*a	52	cs	317.7	421.7
5B	i s t a	s a t t*a	46	cs	327.6	419.6
5C	i s t a	s a t	16	cs	372.2	404.2
5D	i s t a	s t	10	cs	386.4	406.4
5E	i s t a	t	9	cs	386.9	404.9
8A	i s	s a t a*s a*t	9	cs	413.4	431.4
8B	i s	s a t a*s	8	cs	407.7	423.7
8C	i s	s a t	7	cs	404.8	418.8
8D	i s	s t	6	cs	402.2	414.2
8E	i s	t	5	cs	402.8	412.8
8F	i s	t*t t	6	cs	410.4	420.4
9A	i s	s a t t*s a*t	9	ar(1)	404.0	422.0
9B	i s	s a t a*t	8	ar(1)	401.2	417.2
9C	i s	s a t	7	ar(1)	395.7	409.7
9D	i s	s t	6	ar(1)	393.3	405.3
9E	i s	t	5	ar(1)	396.0	406.0
9F	i s	s t t*t	7	ar(1)	394.7	408.7

Table 2: i=individual, s=side, t=time, a=age, a=Hessian matrix not positive definite, b=iteration stopped because of infinite likelihood, c=iteration did not converge. The model with the best (smallest) Akaike information criteria is 5C, but the model 9D is chosen because the interpretation of this model where age and time are continuous variables is simpler.

The correlation coefficient between left and right breast in an individual is $\sigma_G^2/(\sigma_G^2 + \sigma^2 + \sigma_1)$ and the correlation coefficient between time points in an individual is $(\sigma_G^2 + \sigma_1)/(\sigma_G^2 + \sigma^2 + \sigma_1)$.

autoregressive (1)

$$\mathbf{R}_i = \sigma^2 \begin{bmatrix} 1 & \rho^d & 0 & 0 \\ \rho^d & 1 & 0 & 0 \\ 0 & 0 & 1 & \rho^d \\ 0 & 0 & \rho^d & 1 \end{bmatrix}$$

where d is the number of years between the time points measured. The correlation coefficient between left and right breast in an individual is $\sigma_G^2/(\sigma_G^2 + \sigma^2)$ and the correlation coefficient between time points in an individual is $(\sigma_G^2 + \sigma^2\rho^d)/(\sigma_G^2 + \sigma^2)$.

The covariance structure that is chosen is autoregressive (1). The $\mathbf{R}_i$ matrix for a person with two measurements that are done with a time distance of four years is

$$\widehat{\mathbf{R}}_i = 0.8875 \begin{bmatrix} 1 & (-0.697)^4 & 0 & 0 \\ (-0.697)^4 & 1 & 0 & 0 \\ 0 & 0 & 1 & (-0.697)^4 \\ 0 & 0 & (-0.697)^4 & 1 \end{bmatrix}$$

The estimate for the individual effect (the $\mathbf{G}$ matrix) is $\hat{\sigma}_G^2 = 1.0107$. We could also choose a covariance structure with a parameter (θ) for the covariance between the two breasts at the same time point. This covariance pattern looks like

$$\mathbf{R}_i = \sigma^2 \begin{bmatrix} 1 & \rho^d & \theta & 0 \\ \rho^d & 1 & 0 & \theta \\ \theta & 0 & 1 & \rho^d \\ 0 & \theta & \rho^d & 1 \end{bmatrix}$$

The correlation coefficient between the right and the left breast is

$$\widehat{Corr}(y_{is_1t}, y_{is_0t}) = \frac{\widehat{Cov}(y_{is_1t}, y_{is_0t})}{\sqrt{\widehat{Var}(y_{is_1t})\widehat{Var}(y_{is_0t})}} = \frac{1.0107}{1.0107 + 0.8875} = 0.532$$

where i is individual, s is the left or the right breast and t is the time point. The correlation coefficient between time points in one individual is (with a time distance of four years)

$$\widehat{Corr}(y_{ist_0}, y_{ist_3}) = \frac{\widehat{Cov}(y_{ist_0}, y_{ist_3})}{\sqrt{\widehat{Var}(y_{ist_0})\widehat{Var}(y_{ist_3})}} = \frac{1.0107 + 0.8875(-0.697)^4}{1.0107 + 0.8875} = 0.6428$$

With a model where θ is estimated a comparison of these correlation coefficients is adequate. Then the correlation coefficient between right and left breast is

$$\widehat{Corr}(y_{is_1t}, y_{is_0t}) = \frac{\widehat{Cov}(y_{is_1t}, y_{is_0t})}{\sqrt{\widehat{Var}(y_{is_1t})\widehat{Var}(y_{is_0t})}} = \frac{1.0107 + 0.8875\hat{\theta}}{1.0107 + 0.8875}$$

3.8 Model checking

In the normal plot and the predicted values vs. residuals plot we see that the assumptions that the residuals are normally distributed and have constant variance are satisfied. That the data is above the line in the right part of the figure indicates that the right tail of the distribution of the residuals is thinner than the normal distribution, (Fig.7).

The residuals are observed values minus predicted values. From the plot of predicted values vs. residuals we see that in the model low values of breast density is estimated as larger than they really are and large values are estimated as lower than they really are. The variation is constant, (Fig. 8).

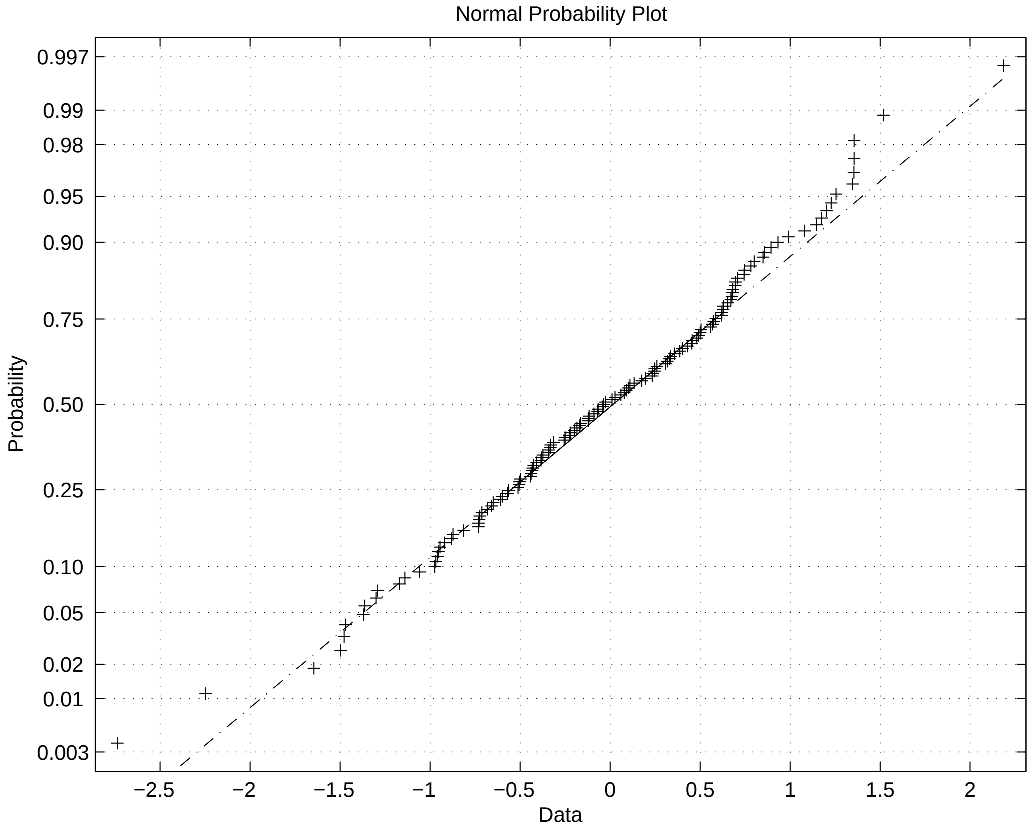

Figure 7: Normal probability plot for the model.

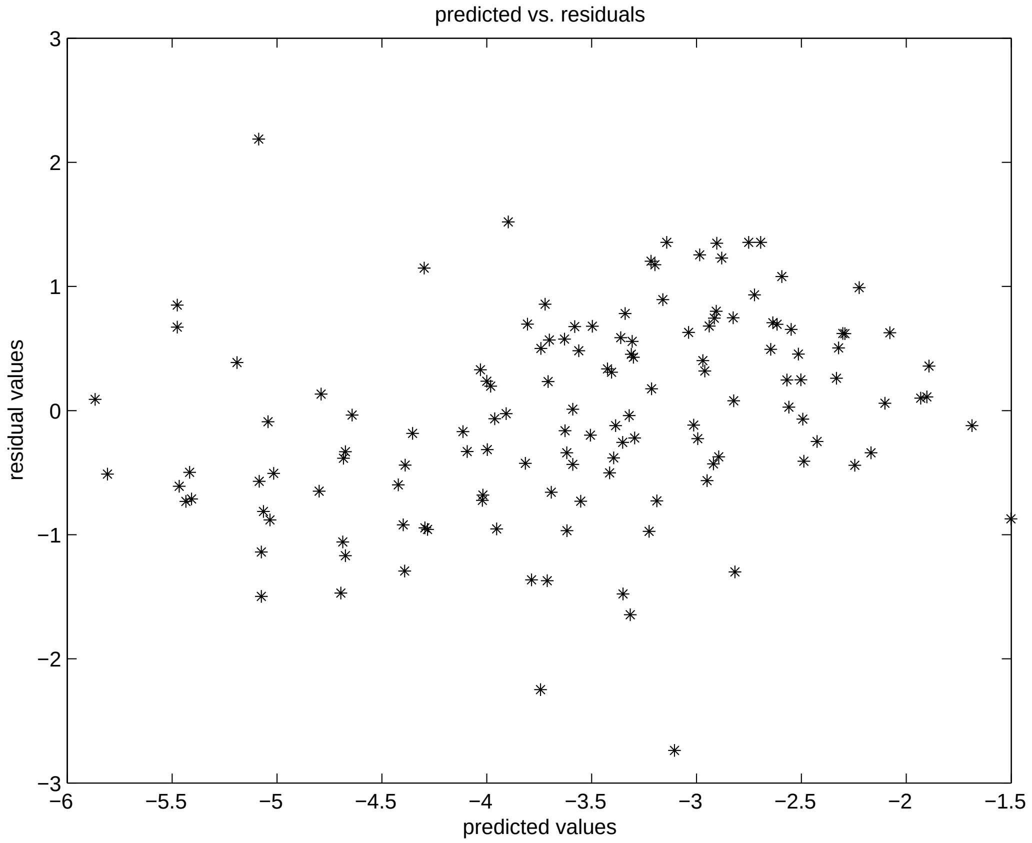

Figure 8: Predicted values vs. residuals.

Effect	side	Estimate	Error	DF	t Value	Pr> $\|t\|$
μ=intercept		-2.7082	0.2789	26	-9.71	< 0.0001
β=time		-0.2004	0.04560	106	-4.40	< 0.0001
γ=right breast	1	-0.3913	0.1879	106	-2.08	0.0397

Table 3: Estimates of fixed variables and p-values.

Effect	Alpha	Lower	Upper
μ=intercept	0.05	-3.2815	-2.1348
β=time	0.05	-0.2908	-0.1100
γ=side	0.05	-0.7639	-0.01882

Table 4: Estimates confidence intervals.

3.9 Results

The model chosen is

$$\begin{aligned} log(y_{ts}) &= \mu + \beta t + \gamma s \\ y_{ts} &= e^{\mu+\beta t+\gamma s} \\ &= e^{\mu} * e^{\beta t} * e^{\gamma s} \end{aligned}$$

where $t = 0, 1, \ldots, 5$ years and $s = 1$ for the right breast and $s = 0$ for the left breast. The mean value of the proportion for the left breast year zero is estimated as $e^{-2.2708} = 0.0667$, with a 95% confidence-interval of 0.038-0.118. The parameter for time, β, is significant, with point estimator $\hat{\beta} = -0.2004$. Each year the proportion of breast density decreases with $1 - e^{-0.2004} = 18\%$, with a 95% confidence-interval of 10%-25%. The estimates of the parameter for side effect γ is -0.391 so there is $1 - e^{-0.391} = 32\%$ less proportion in the right breast then in the left, with a 95% confidence interval of 2%-53%. (Table 3, 4)

For the left breast year zero the breast density is $0.0667 = e^{-2.7082} * e^{-0.2004*0} * e^{-0.391*0}$, or for year five in the right breast the density is $0.0166 = e^{-2.7082} * e^{-0.2004*5} * e^{-0.391*1}$

Birth year is not found significant, if the spread of birth years is larger than 36-42 differences might be found.

4 Simple linear regression on pictures from South Florida digital mammography home page

Digitalized pictures of high technical quality that are available at the University of South Florida digital home page were analyzed with simple linear regression on 50 subjects that had undergone one mammographic examination with the parameters p, μ_1, σ_1, μ_2, σ_2 and μ_2-μ_1 as dependent variable, and the given data age and the ACR breast density rating as independent variables. ACR (The American Collage of Radiology) breast density rating is a coding system for breast density that is determined by visual inspection of the x-ray picture. The scale is from one to four, where one is low and four is high breast density. The procedure to take 50 samples were done by taking the first 50 pictures on the web-site. For further details, see http://marathon.csee.usf.edu/Mammography/Database.html

4.1 Results

In the Bonferroni method the significance level is 0.05/n=0.0038 (n=13 is the number of tests made) so in none of the tests differs the slope significantly from zero. There is a suggested negative correlation between age and ACR breast density rating and a suggested positive correlation between

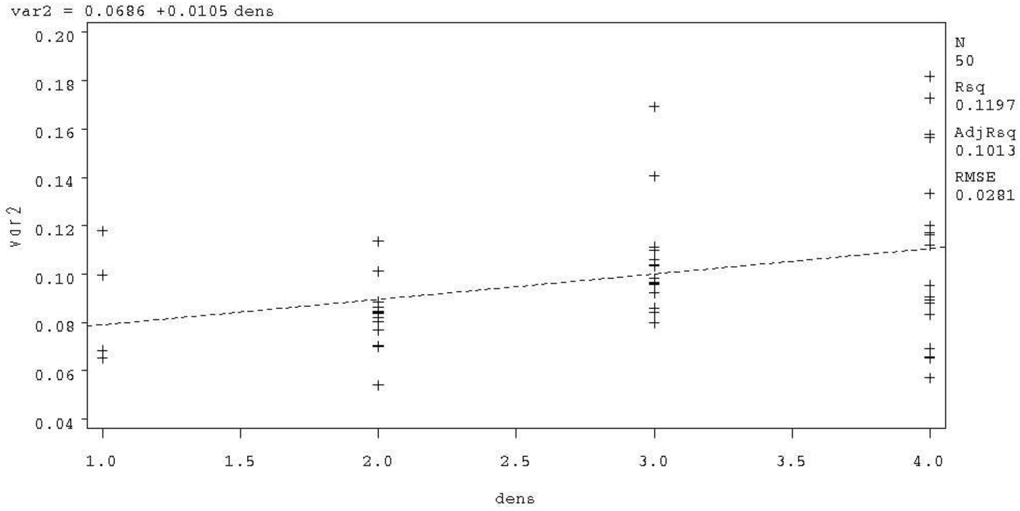

Figure 9: ACR breast density rating vs. σ_2^2

ACR breast density rating vs. σ_2, that is the variance of the right part of the bimodal distribution increases with increased breast density. A correlation between p and density, μ_2 - μ_1 and density and between density and age was expected but not found. (Figure 9, Table 5)

5 Miscellaneous topics

5.1 Dropouts from mammography screening

The participation in mammography screening, as also gynecological cervical cancer screening, is high in Sweden. The reasons for not participating in the screening can be many, for instance

- Lack of time.

x-axis	y-axis	intercept	p-value	coefficient	p-value
age	p	0.3922	0.0002	-0.0019	0.2617
age	μ_1	0.2882	< 0.0001	-0.0002	0.8187
age	σ_1	0.1234	< 0.0001	-0.0003	0.4677
age	μ_2	0.716	< 0.0001	-0.0018	0.1355
age	σ_2	-0.0003	< 0.0001	-0.0003	0.4089
age	ACR density	4.6707	< 0.0001	-0.0291	0.0270
age	μ_2-μ_1	0.4280	< 0.0001	-0.0015	0.0575
ACR density	p	0.2488	< 0.0001	0.0113	0.5255
ACR density	μ_1	0.2952	< 0.0001	-0.0071	0.5164
ACR density	σ_1	0.1063	< 0.0001	-0.0013	0.8084
ACR density	μ_2	0.5822	< 0.0001	0.0098	0.4468
ACR density	σ_2	0.0686	< 0.0001	0.0105	0.0139
ACR density	μ_2-μ_1	0.2871	< 0.0001	0.0169	0.0510

Table 5: Results from simple linear regression of parameters from University of South Florida digital mammography home page.

- Economical reasons.
- Worry that the radiation causes illness or damages the lactation.
- The risk of a false positive answer that can cause unnecessary investigations and operations.
- Underestimation of the risk of developing cancer.
- Denial of the risk of developing cancer or fear of the treatment if cancer is shown.
- Pain or discomfort due to the compression of the breast.
- Concerns that the compression will affect the shape of the breasts,damage the breasts or fear that if there is cancer,it will spread due to the compression of the breast.
- Issues due to breast-implants.

- Feelings of unpleasantness and shame caused by the nakedness in the examination situation especially if there are male x-ray staff.
- Avoidance of showing bruises, scares, tattoos, anomalies or disfiguring skin affections.
- Threats from a family member if participating.
- Low knowledge about diseases.
- A general distrust to authorities, or a feeling of not being respected by the health care system.
- Earlier maltreatment.
- Religious beliefs.
- Opinions that x-ray examination is technocratic and unnatural.

Among immigrants, where the participation is lower then on average, the reason can be that there are no screening programs in the countries from which they come. Due to ethnic differences of breast cancer incidence some ethnic groups may not participate in mammography screening because they feel that the risk for them is smaller. Many immigrants come from countries where human rignts are denied, and contacts with government organizations including health care system can have been harmful, so screening programs can be sceptically viewed. Immigrants without residence permit probably don't participate in mammography screening. For further reading see Demographic predictors of mammography and PaP smear screening i US women, Am J Public Health, Calle et al.

Advertising in media and events about advantages of screening, mobile x-ray equipment for rural areas, drop-in-hours at evening and at weekends, free examinations, information in connection with other health care occasions in adequate language and information on web-sites increase participation. Information through immigrant organizations should be more effective if the proportion of women in the committees for immigrant organizations increased. Information about monthly self-examination of the breasts the day after menstruation should also be given. Breast cancer organizations can also play an important role. Another way to increase participation in mammography screening can be to offer, at the same occasion other health care, for instance screening for heart and blood vessel disease (i.e. cardiac infarction and stroke) risk factors by measuring body mass index, blood pressure and serum-lipids, or screening for other common disease groups like other cancers, mental or psycho-social illness, or diseases in the skeleton-locomotor system.

5.2 Randomization of order of measurement

When deciding where to put the region of interest polygon there is a risk that the part of the picture that is covered with the polygon varies in a systematic way between repeated measures. Of each subject all pictures were analyzed in succession, the order was randomized. The identity and time of examination of each picture were not coded but the risk of any bias due to this procedure is not very high. Ideally the pictures should have been coded. The placing of the polygon on subsequent pictures of one breast of a subject might be more correct in the later pictures since the appearance of the picture becomes clearer after having studied the picture for a while. On the other hand fatigue and getting into a rut can make the former measure of pictures better then the latter.

5.3 Model discussion

The variation of the parameters can have several causes. There can be variations caused by properties of the subjects, e.g. age, hormonal and other medical and para-medical treatments, previous diseases including diseases of the breast, day in the menstrual cycle the examination was performed or if the subject is pre- or post-menopausal, metabolism of sex hormones, changes of the size of the breast, breast surgery and implants, socio-economic status, number of deliveries, environmental factors, life style factors like diet, weight, smoking, drinking, narcotic drug use, physical exercise, and other factors.

There can also be variations caused by technical matters, e.g. x-ray film-type, time of exposure, voltage, type of x-ray machine, raster, practical matters when performing the examination and the subjects ability to cooperate, the angle of the x-ray beam in relation to the breast, the part of the breast that is examined, and the compression of the breast when the mammography picture is taken. It is uncertain if there is an automatic x-ray dose adjustment that increases the dose if the breast density increases, if it is so an increase of the density in the breast will be underestimated in the x-ray, since a higher x-ray dose makes the picture darker.

5.4 Ethical considerations

It is important that subjects that are included in medical studies are informed about the study and that they have given their consent to participate. This is especially important when the subjects health is at risk. A subject has the right to chose not to participate in a study for whatever reasons. The purpose of the study must have intentions that are ethically acceptable, that is, do good and not cause harm, and the performance in all respects of a study must be high to guarantee that meaningful conclusions can be drawn.

Ideally the analysis of the pictures shall be done in a place where not other students and staff can see the pictures. See also www.nih.gov/sigs/bioethics/

5.5 The use of the parameters in future studies

Since the parameter σ_2 has a positive correlation with breast density, σ_2 can be compared between two groups in a t-test, the group with the lowest σ_2 has the lowest mammographic breast density, for instance the t-test can be used to see which of some oestrogen drug that affects breast density least, or to see if alcohol consumption affects breast density. The change of σ_2 between two mammography examinations can be investigated to see if the change is within the normal range according to some limiting α-value. Each strata considering age, hormone therapy, etc. will have it's own distribution considering the change of σ_2. If the radiation from the mammographic examination affects breast density can be investigated.

Since about 50% of all breast cancer is located in the upper outer quadrant of the breast it could be of interest to do a mapping of the extent of breast density in the different parts of the breast to see if there is more breast density there.

One could investigate if breast density decreases slower, or increases in those who develop breast cancer than among those who don't, and if there is a difference between the groups in both breasts or in the affected breast.

The relationship between cysts, calcifications and lymph nodes density, size and numbers in the breast and breast density can be investigated. Counts data can be modelled with a generalized linear model. A generalized linear model with breast-cancer or no cancer diagnosis as outcome variable and the parameters as covariates can also be made to see if any of the parameters are predictors of breast density.

Features of biopsies or histo-pathological preparations (e.g. the extent of oedema or how many glandular cells there are per square millimeter) can be compared with the value of the mammographic breast density on the place where the biopsy is taken from with for instance simple linear regression to explain breast density.

A way to estimate the breast density in a more exact way is to add the bars in the histogram of the CC projection with the bars in the histogram of the ML projection and then decide the parameters. The average of the two projections is a better measurement then one of them.

An evaluation of a three dimensional histogram where x- and y-axis are the locations in the x-ray picture and z is the gray level value can give additional information of the location of the breast density.

The quality of the x-ray e.g. the extent of contrast can be assessed with the histogram. The less contrast there is the more the histogram looks like a single distribution.

If one wants to compare if there is differences in breast density measured as the parameter σ_2 between groups estimation of what power such a study has shall be made. The power will depend on the variation of σ_2 in each group and the sample size. It takes appr. one minute to analyze one digitalized picture, so it is possible to have a large sample size.

Investigating the change of the statistical distribution of pixel gray level values in digitalized pictures can of course be used in a lot of applications, in principle all biological and medical changes that is reflected in a change of nuances in a picture can be assessed. Some examples are the healing process of bone fracture or development of pneumonitis seen on x-ray, or changes in photos of histological preparations.

How breast cancer and it's treatment affects breast density can be investigated.

References

Helen Brown, Robin Prescott, Applied Mixed Models in Medicine, John Wiley & Sons Ltd. 1999
Calle et. al. Demographic predictors of mammography and PaP smear screening i US women Am J Public Health Charles S. Davis, Statistical Methods for the Analysis of Repeated Measurements, Springer-Verlag 2002
B.S. Everitt, D.J. Hand, Finite Mixture Distributions, Chapman & Hall, 1981
Lundström E, Christow A, Kersemackers W, Svane G, Azavedo E, Söderqvist G, Mol-Arts M, Barkfeldt J, von Schoultz B. Effects of tibolone and continuous combined hormone replacement therapy on mammographic breast density. Am J Obstet Gynecol 2002;186:717-772
Charles E. McCullouch, Shayle R. Searle, Generalized, Linear, and Mixed Models, John Wiley & Sons, Inc, 2001

http://marathon.csee.usf.edu/Mammography/Database.html
www.nih.gov/sigs/bioethics/