

Risk factor study of pododermatitis in rabbits using additive Bayesian networks

Master Thesis in Biostatistics (STA495)

by

Andrea Meier

12-724-373

supervised by

Prof. Dr. Reinhard Furrer

Dr. Sonja Hartnack, Section of Epidemiology UZH



**University of
Zurich**^{UZH}

Zurich, March 2017

Acknowledgement

I would like to thank to my thesis supervisors Prof. Dr. Reinhard Furrer, Department of Mathematics and Dr. Sonja Hartnack, Section of Epidemiology from the University of Zurich. The door was always open whenever I ran into problems or had a question about my research or writing. The nicely balanced mix between steering the project in a good direction and allowing me to do my work in an own manner made the process of accomplishing this thesis a very pleasant and enjoyable one.

I would also like to thank to Sabrina Ruchti and Dr. Sabine G. Gebhardt-Henrich from the Animal Welfare Division of the Veterinary Public Health Institute of the University of Bern for sharing their research and data with us. This cooperation gave me the possibility to have access to a very interesting and relevant research question with an elaborated and profound data basis.

This collaboration between three Institutes and two Universities had only become possible thanks to Dr. Sonja Hartnack and the plain cooperation of all the involved persons.

I would also like to acknowledge to Gilles Kratzer, Doctoral student in Epidemiology and Biostatistics at the University of Zurich, for sharing his science with me. With his knowledge and his own research about the technique of additive Bayesian networks and his changes and additions in the implementation of the R package `abn` he was an important contact person with very helpful and interesting advises. I'm very grateful for his valuable comments on this thesis.

Finally, I must express my very profound gratitude to my fellow students, my parents and my boyfriend for providing me with tireless support, great collaboration and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without you.

Thank you.

Andrea Meier

Contents

| | | |
|----------|---|-----------|
| 1 | Introduction | 1 |
| 2 | Study and Data Description | 3 |
| 3 | Methods | 11 |
| 3.1 | Additive Bayesian networks | 11 |
| 3.2 | ABN: a more statistical approach | 14 |
| 3.3 | Random forest | 16 |
| 3.4 | Software | 17 |
| 4 | Analysis and Results | 19 |
| 4.1 | Additive Bayesian networks | 19 |
| 4.2 | Random forest | 24 |
| 4.3 | Regression approach | 26 |
| 5 | Technical Remarks on the R Package abn | 29 |
| 5.1 | Peculiar posterior densities | 29 |
| 5.2 | Error with random effects | 30 |
| 5.3 | Interpretational difficulties and lack of documentation | 32 |
| 6 | Conclusion and Discussion | 33 |
| | Bibliography | 37 |
| A | Appendix | 39 |

Chapter 1

Introduction

In commercial rabbit production pododermatitis is a common and severe disease with a relevant implication for the affected animals. Pododermatitis, also called foot pad lesion, is a chronic skin disease that mainly appears on the plantar surface of the rabbits hind pad, but it can also be present on the front legs. Pododermatitis leads through several stages and normally starts with hair loss and callus formation and ends with open wounds and ulcers. Foot pad lesions are associated with pain [[Rommers and de Jong, 2009](#)], imply impaired animal welfare and are a cause for the culling of reproductive female rabbits on commercial farms. Pododermatitis is a multifactorial disease and different factors have to be considered to have an effect on the occurrence of the lesions.

The project was conducted within the framework of the testing and authorization procedure for housing systems and equipment, in particular for pens and floors for group housed rabbit. Cage systems with wire mesh floors are dominating in European rabbit housing. In Switzerland group housing of breeding rabbits in pens with slatted plastic floors and straw bedding offers an alternative and better fulfills the animals' behavioral needs. However, foot pad lesions also occur in these husbandry systems. Until today there is a lack of comprehensive data on foot pad lesions in group housed breeding rabbits in Switzerland. The identification of possible risk factors and information on severity and prevalence is important in view of developing preventive measures.

The aim of this thesis was to identify possible risk factors of pododermatitis in group housed breeding rabbits in Switzerland.

To achieve this goal the technique of additive Bayesian networks (ABN) was used. The advantage of Bayesian network analysis is that it not only identifies statistically associated variables but also distinguishes between direct and indirect dependencies between the variables. Additionally Bayesian network analysis is appropriate to complex and messy data and can handle multiple dependent variables quite well. The results of the additive Bayesian network analysis were compared to the results of an analysis with random forest and generalized linear regression. As additive Bayesian network analysis is a rather new methodology in epidemiological research this thesis aimed to establish this technique and find out potential difficulties when working with real-life data.

Chapter 2

Study and Data Description

This master thesis was done in collaboration with the Animal Welfare Division from the Veterinary Public Health Institute of the University of Bern. Data used for this thesis were collected between June and September 2016. The animal experiment authorization number was BE8/14.

The study included 17 out of 18 existing commercial rabbit farms with group housing of female rabbits from the production group “KaniSwiss” (www.kani-swiss.ch) in Switzerland. The population of female rabbits from these 17 farms was estimated as 3500 animals of which a third were assessed. Thus the data set used for this analysis comprises a total number of 1090 animals which are kept together in pens of six to ten animals. During the visits of the farms data specific to the single animal to the particulate farm were assessed resulting in a hierarchical structure of our data. Whereas variables on farm level have the same value for several animals, variables on animal level are specific for each individual and however can be assigned to a group, respectively to a farm in our case.

For each assessed animal the pododermatitis score, some information about the general condition of the animal and additional factors potentially associated with pododermatitis were examined. Figure 2.1 shows the number of assessed animals per farm. The number differs between 25 and 126.

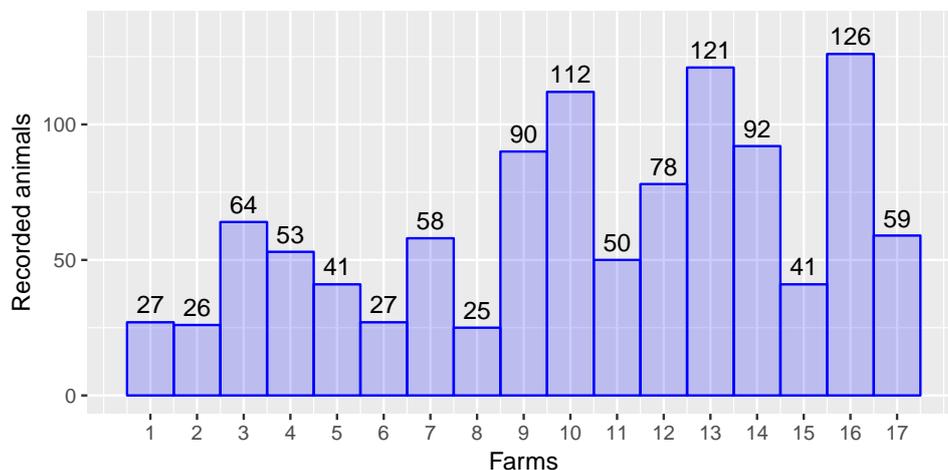


Figure 2.1: Number of recorded animals per farm.

Table 2.1: Overview about variables included into the analysis, where all variables until **Moist** were on animal level and after **Moist** on farm level. The color indicates the used distribution in the analysis, where dark blue stands for Poisson, light blue for Gaussian and grey for Binomial. For the binary variables the basis category is emphasized by the wavy line.

| Variable | Details |
|----------------|--|
| PDmeanmiddle | Mean of the pododermatitis scores from the middle spot |
| PDworst | Worst pododermatitis score of all measured spots |
| Weight | Body weight of the animal in kilograms |
| Age | Age of the animal in months |
| Claws | Condition of claws, <u>normal</u> or too long |
| Clean | Cleanliness of paws, <u>clean</u> or dirty |
| Moist | Moisture of paws, <u>dry</u> or moist |
| Holes | Average number of holes found in one pen |
| Gnaw | Average percentage of slats in one pen, where more than a third of the slat-area is gnawed |
| Wet | Average sum of normalized wet areas on litter and nest roofs in one pen in square centimeter |
| WaterSpillLitt | Water spilling on litter, <u>no</u> or yes |
| RH | Average relative humidity in barn over one day in percentage |
| Temp | Average temperature in barn over one day in degree Celsius |
| Experience | Experience of the farm manager in years |
| Mortality | Percentage of dead and culled animals in one year |
| LittChanInt | Time interval of total replacement of litter material in days |
| TotalNoRabbit | Total number of rabbits on the farm |

Per farm between two and five pens were chosen in a stratified way regarding their position in the barn and taken as representative for the farm. For each assessed pen data about the condition of the slats and the litter were collected. For the analysis these exemplary values were averaged, so resulting in one value per farm.

Together with the farmer a questionnaire was filled out to determine further animal and farm management based parameters potentially associated with pododermatitis.

Out of all collected variables, the most meaningful for pododermatitis were chosen based on existing literature and expert opinion. Table 2.1 gives an overview about the variables, which were included into the analysis. Table 2.2 and 2.3 show the descriptive statistics about all these variables.

The pododermatitis score for the hind paw was assessed with a tagged visual-analogue scale adapted to the scale by [Drescher and Schlender-Böbbis \[1996\]](#). Based on 6 stages of the disease, which are described by images and medical properties, a cross was made on a continuous scale from 1 to 6. This cross was then transformed to a score from 0 to 10, measured in centimeters whereas the whole scale was exactly 10 cm. Figure 2.2 shows the 6 stages as well as the transformation scale. The heel and the middle part of the hind paw were assessed separately, so for each animal four pododermatitis scores of the hind paws were assessed. For the analysis we either took the worst pododermatitis score of each animal (PDworst) or the mean score of the two values from the middle part

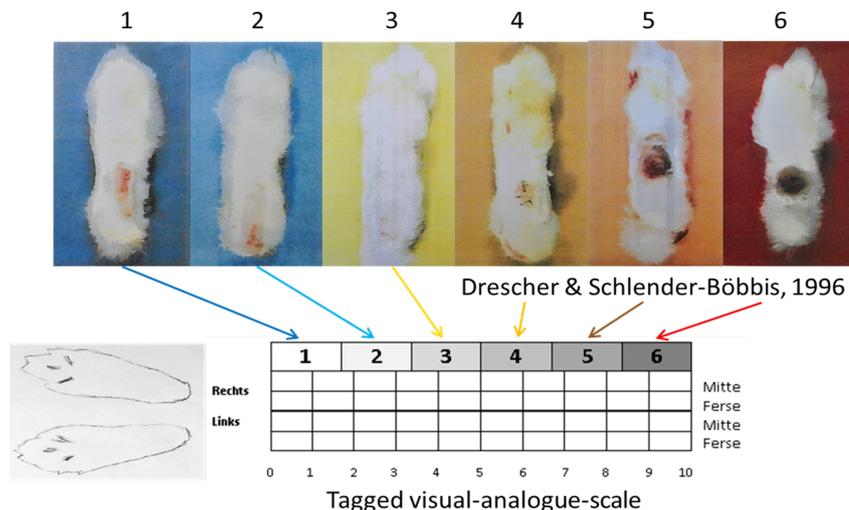
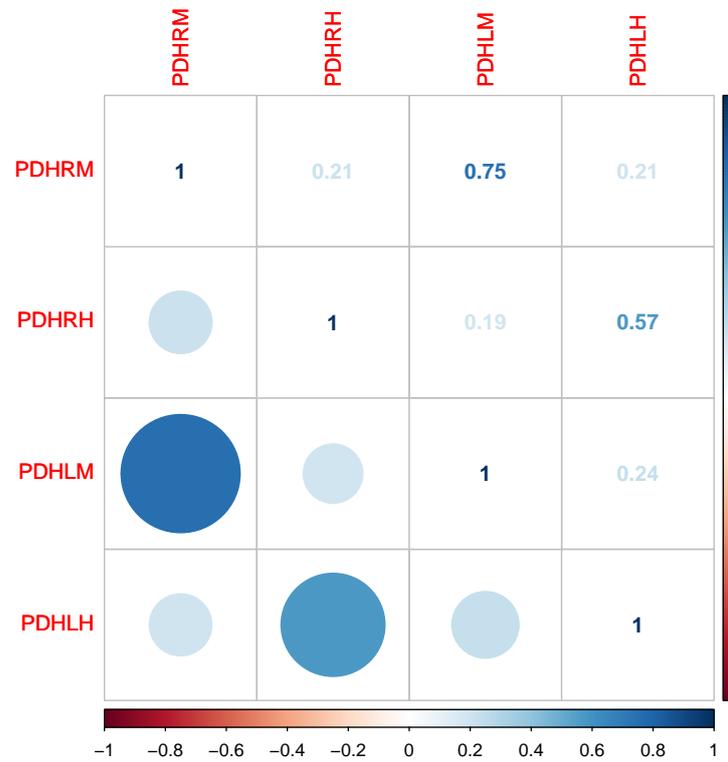


Figure 2.2: Tagged visual-analogue scale adapted from [Drescher and Schlender-Böbbitis \[1996\]](#)

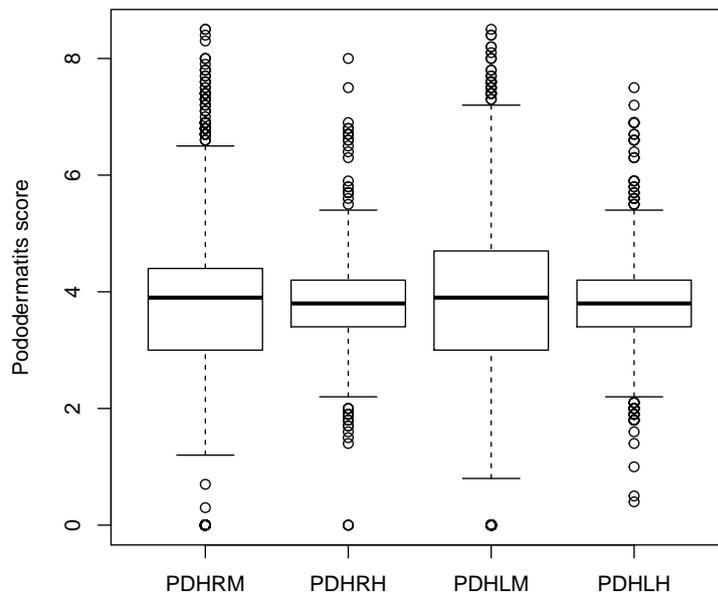
(PDmeanmiddle). We decided so because the correlation between the left and the right score from the middle was bigger than between the scores from the heel, see Figure 2.3a. Additionally we found a higher dispersion in the scores from the middle, see Figure 2.3b. This may be due to the fact that the heel almost always has some thickened skin and is hairless due to normal attrition and not specifically due to pododermatitis.

Figure 2.4 visualizes the occurrence of the mean scores of the two pododermatitis values from the middle part. A high occurrence of PDmeanmiddle values equal to zero was detected. This excess of zeros is not expected by a normal distribution. However, the amount of zeros could be traced back to a property of disease determination. If determining a disease first of all the investigator had to decide whether the disease is present or not. If the disease was present the severity and magnitude can be defined. This was done with the tagged visual-analogue scale and this led to a nicely dispersed variable. If the disease was not present, the value was just equal to zero. When first of all a binary decision has to be made this leads automatically to such an artifact. It would be interesting to work with a zero inflated model or a hurdle model, which could take into account such a superset of zeros, unfortunately ABN had not yet the possibilities to do so. Another possibility to handle such a distribution would be to simply divide the scores in two groups, a group where the disease was present and one where it was not, and then use a Binomial distribution to model it. However, such an approach would lead to a big loss of information. Therefore, the variable PDmeanmiddle was treated as a normal continuous Gaussian variable.

Figure 2.5 visualizes the occurrence of the worst pododermatitis scores. There was no occurrence of too many zeros but there were a lot of values around 4 and the distribution is a bit skewed to the right. Additionally there was a gap of values greater than 5 and smaller than 6.5. We had only little explanation what could cause this artifacts. Perhaps the chance of having a lesion with a higher score when having already a lesion which is above 5 is very high and accordingly the chance of having a more severe lesion if all other lesions are around or below 4 is very low. Or maybe this represents a weakness of the



(a) Pearson correlation coefficients



(b) Dispersion

Figure 2.3: Correlation and dispersion of the pododermatitis scores, where PDHRM = pododermatitis hind paw right middle, PDHRH = pododermatitis hind paw right heel, PDHLM = pododermatitis hind paw left middle and PDHLH = pododermatitis hind paw left heel.

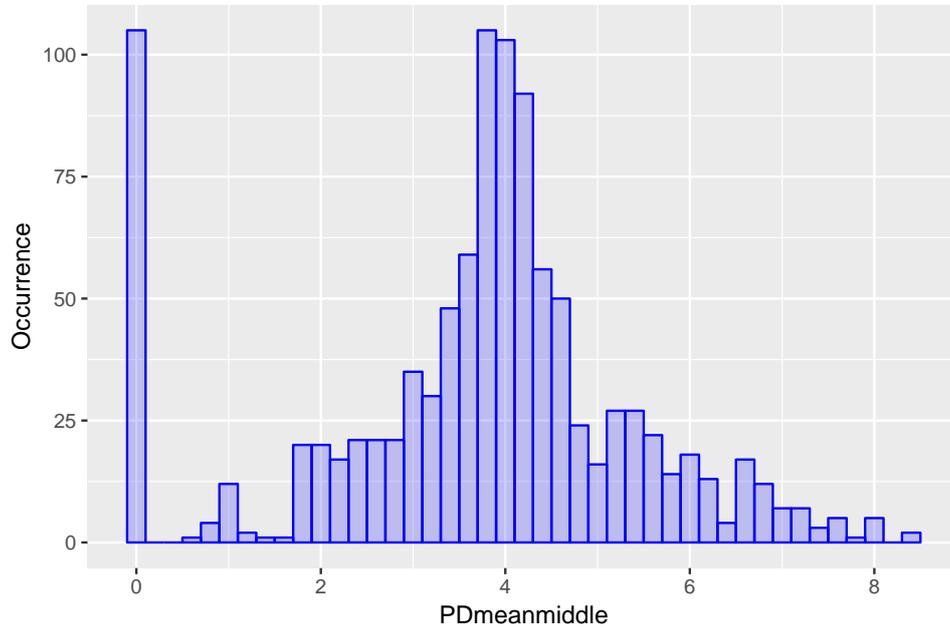


Figure 2.4: Occurrence and distribution of the mean scores of pododermatitis from the middle part of the hind paws.

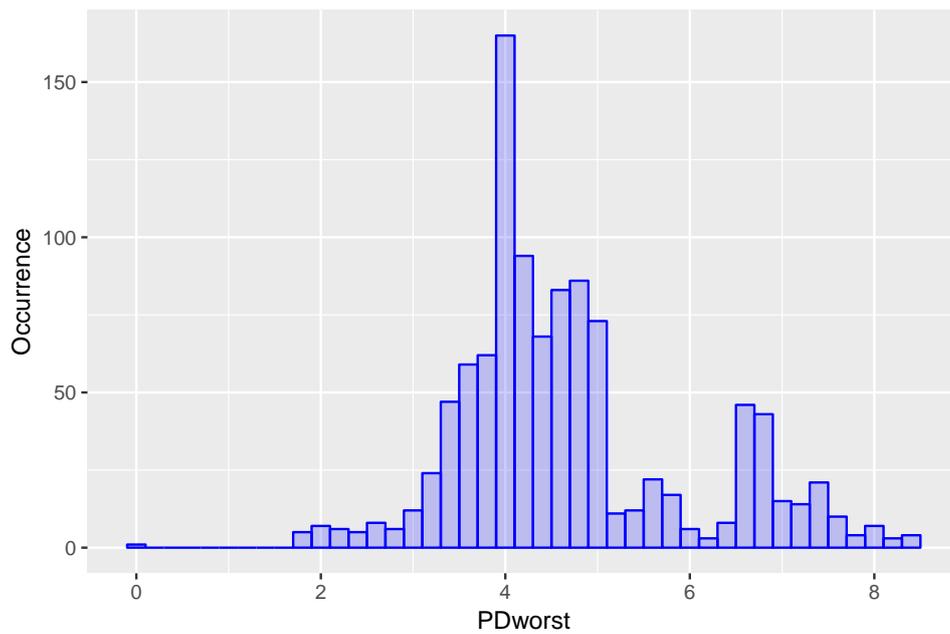


Figure 2.5: Occurrence and distribution of the worst pododermatitis scores.

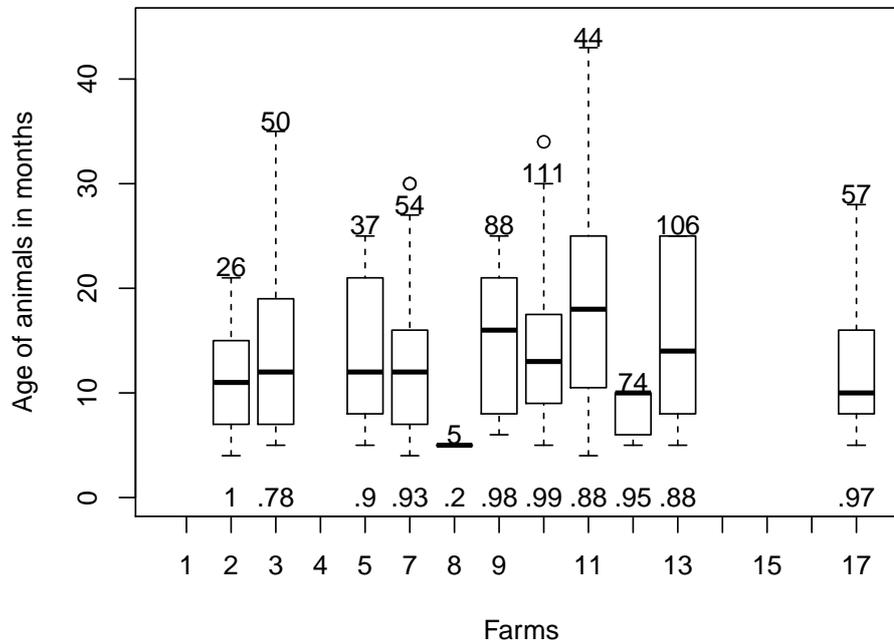


Figure 2.6: Dispersion of **Age** on the distinct farms. The values above the boxes indicate the number of animals where **Age** was recorded. The number below the boxes show the proportion of the number of animals with recorded **Age** to the total number of recorded animals per farm.

tagged visual-analogue scale where the area between 5 and 6.5 only covers a few disease stages, whereas at the area at 4 an important and common breaking point is present. As mentioned, these are only possible presumptions, the exact reason was unclear to us. Nevertheless, the variable **PDworst** was also handled as a continuous Gaussian variable in all further analysis.

The variable **Age** unfortunately included a lot of missing values, because some farmers do not record the age of their animals. However, as the age often interacts as a very important confounder regarding some diseases, we decided to split up the analysis into two parts, one with and one without the variable **Age**. When including **Age**, our data set was reduced to 661 animals. Additionally this led to a decrease in the number of included farms. The data set with **Age** included only animals coming from 11 distinct farms. Figure 2.6 illustrates the dispersion and the availability of the data about the age of the animals. In this figure we see that the age distribution of the animals is quite divers. Farm 12 has no animals older than 10 months, many farms have the upper quartile below or around 20 months and only some farms keep quite old rabbits. Fortunately we can conclude that missing respectively existing age values are completely at random and have no connection at all to the age value itself or the pododermatitis score. Farmers have animals of different age classes and a farmer normally records the age of all or none animal. Missing values from farmers which normally record the age, see percentage values below the bars in Figure 2.6, are due to undiscoverable identification cards or lost ear

tags. The only exception is farmer 8, which only knew the age of five young animals (20% of his recorded animals). Nevertheless, a missing completely at random can be assumed for the variable **Age** and hence a complete case analysis can be safely conducted.

The variable **Weight** included the body weight of each individual animal. Accordingly to Table 2.2 the average body weight of all rabbits was 4.9 kg. For each animal it was also recorded whether the claws appeared unusually long (**Claws**) and whether the paws were clean (**Clean**) and moist (**Moist**).

Moisture level of the litter is considered as one possible cause of pododermatitis [Bigler and Oester, 2003] and therefore it was noted whether waste water from the drinker was spilling on the litter (**WaterSpillLitt**) and how many and how large wet areas exist in one pen (**Wet**). A glimpse at the Table 2.2 shows the big variation of existing wet areas among the distinct farms. Another possible cause could be physical damage of the foot pads by rough or damaged parts of the plastic slats [Andrist et al., 2013]. Therefore, it was noted whether traces of gnawing marks or even holes of the slats were present (**Gnaw, Holes**). Also these variables show a rather big range of values, where many farms had no holes at all in their slats one farm had on average 4 holes in one pen. Accordingly there was a farm where on average 76% of all existing slats in one pen were gnawed in more than a third of the slat-area. The mean value of **Gnaw** is 29%, indication that out of 10 slats in one pen three slats were gnawed in more than a third of the slat-area. These were the pen specific variables which are measured in the two to five exemplary pens and then averaged to one value for all animals from a certain farm.

The relative humidity and the temperature within the barn were recorded during the farm visit and thus the day differs for each farm. Comparison of the temperature and the relative humidity between the farms has to be handled with caution. When looking at the Table 2.2 the values of **RH** vary between 47 and 84% humidity and the temperatures within the barns between 16 and 25 degree Celsius. The variables **Experience**, **Mortality**, **LittChanInt** and **TotalNoRabbit** were assessed with the help of the questionnaire, which was filled out together with the farmers. There existed farmers with very few experience, the least experienced farmer was keeping rabbits only since 6 months. On the other hand side there was a farmer with 30 years of experience. The variable **Mortality** represents the turnover of rabbits in one year. Where some farmers replaced less than half of their rabbits annually, other farmers replaced every animal even more than once a year, speaking in means of a farm. According to the litter change interval, the litter is replaced in average after six weeks, whereby this also differed between one and fifteen weeks. The total number of animals of a farm differed between 88 and 618 animals, this number included beside the reproductive female rabbits also younger female rabbits that are raised until they reach sexual maturity to replace animals that died or got culled.

Due to big differences in the accuracy and the diligence of the answers and information of one farmer compared with the others we decided to exclude this certain farm. This concerns farm number 6 in our data set and accordingly 27 animals. These observations are not taken into account in other figures and tables except Figure 2.1.

Table 2.2: Descriptive statistics of continuous and count data (**Experience**, **LittChanInt** and **TotalNoRabbit**), where **n** denotes the amount of existing values and **#NA** the amount of missing values. These two columns always add up to the total of 1063 animals. **Min** and **Max** show the minimal resp. maximal value of the variable. **q₁** and **q₃** give the upper limit of the first resp. third quantile. \tilde{x} display the median, \bar{x} the mean and **s** the standard error of the variable.

| Variable | n | #NA | Min | q ₁ | \tilde{x} | \bar{x} | q ₃ | Max | s |
|---------------|------|-----|-------|----------------|-------------|-----------|----------------|---------|--------|
| PDmeanmiddle | 1048 | 15 | 0.00 | 3.00 | 3.95 | 3.72 | 4.60 | 8.50 | 1.75 |
| PDworst | 1057 | 6 | 0.00 | 4.00 | 4.40 | 4.73 | 5.10 | 8.50 | 1.26 |
| Weight | 1062 | 1 | 2.74 | 4.55 | 4.92 | 4.91 | 5.30 | 6.69 | 0.57 |
| Age | 661 | 402 | 4.00 | 8.00 | 12.00 | 13.54 | 18.00 | 43.00 | 7.04 |
| Holes | 1063 | 0 | 0.00 | 0.00 | 0.00 | 0.64 | 0.40 | 4.00 | 1.26 |
| Gnaw | 1063 | 0 | 0.00 | 0.10 | 0.20 | 0.29 | 0.46 | 0.76 | 0.26 |
| Wet | 1063 | 0 | 41.63 | 247.39 | 762.36 | 840.13 | 876.71 | 2241.81 | 690.55 |
| RH | 1063 | 0 | 47.34 | 67.27 | 73.76 | 72.06 | 77.65 | 84.40 | 7.44 |
| Temp | 1063 | 0 | 16.28 | 19.72 | 21.28 | 21.08 | 22.11 | 24.55 | 1.88 |
| Experience | 1063 | 0 | 0.50 | 6.00 | 7.00 | 10.33 | 13.00 | 30.00 | 8.00 |
| Mortality | 1063 | 0 | 11.00 | 48.03 | 93.75 | 84.00 | 118.50 | 180.00 | 45.73 |
| LittChanInt | 1063 | 0 | 7.00 | 28.00 | 40.00 | 39.96 | 42.00 | 105.00 | 21.73 |
| TotalNoRabbit | 1063 | 0 | 88.00 | 160.00 | 285.00 | 322.85 | 460.00 | 618.00 | 166.44 |

Table 2.3: Descriptive statistics of categorical data, where **n** denotes the absolute and % the relative amount of animals in each level.

| Variable | Levels | n | % |
|----------------|----------|------|-------|
| Claws | normal | 625 | 59.1 |
| | too long | 433 | 40.9 |
| | all | 1058 | 100.0 |
| WaterSpillLitt | no | 694 | 65.3 |
| | yes | 369 | 34.7 |
| | all | 1063 | 100.0 |
| Clean | clean | 721 | 67.9 |
| | dirty | 341 | 32.1 |
| | all | 1062 | 100.0 |
| Moist | dry | 603 | 56.8 |
| | moist | 459 | 43.2 |
| | all | 1062 | 100.0 |

Chapter 3

Methods

3.1 Additive Bayesian networks

The focus of the analysis was on the work with additive Bayesian networks. Additive Bayesian networks (ABN) are a special type of Bayesian networks. A Bayesian network is a graphical model, where the graph expresses the conditional dependence structure between random variables. The random variables are represented by the nodes, the conditional dependencies by the edges. The edges are determined and drawn in a directed manner. However, when interpreting the edges and speaking about conditional dependencies one have to be aware of interpreting the direction as it can be misleading. The resulting graph is also called directed acyclic graph (DAG), which indicates that no loops are allowed in ABN. Figure 3.1 gives an example of such a simple DAG.

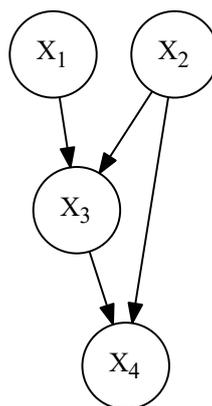


Figure 3.1: Example of a simple DAG.

The aim of Bayesian network analysis is to perform a model search on the given data to identify the optimal model, which represents the conditional dependencies in the data best. Hence a first advantage is that there is no need to define by an expert opinion how the model should look like, but the model structure is defined by the analysis itself.

Another advantage is that Bayesian network analysis is a multivariate technique without a dimension reduction. Bayesian network analysis is well suited to complex and messy data and can handle many dependent variables.

ABN is a special type of Bayesian network and involves two steps. First the best Bayesian network is searched by scoring all possible networks. Secondly the parameter estimation of the conditional dependencies are estimated with a generalized linear model (GLM). Thus, the model parameters of ABN are not just conditional probabilities but the additive components from generalized regression. The method of ABN is an analogous to a multivariate GLM. This is distinct to the usual multivariable GLM where only one response variable can be handled.

Searching for an optimal network is referred in the literature to structure learning or structure discovery [Friedman and Koller, 2003]. Due to the super exponential search space in order to the number of nodes this searching process is not an easy task and several possible methods exists. For this thesis the established globally optimal search approach by Koivisto and Sood [2004] was used, which is implemented in the used R package `abn`. It performs an exhaustive search based on some node ordering. This approach identifies a DAG whose goodness of fit is equal to the best possible goodness of fit of any DAG. As a goodness of fit measure the marginal likelihood is used. This is the standard goodness of fit measure in Bayesian modelling [MacKay, 1992] and includes an implicit penalty for model complexity.

As ABN is a Bayesian approach, prior distributions must be defined. For this analysis a uniform structural prior was used, which means that all DAG structures were equally supported in the absence of any data. Also for the model parameters uninformative priors were used. The current implementation of the used R package `abn` only supports Gaussian, Binomial and Poisson distributed variables. This implies a recoding of multicategorical variables into Binomial variables.

After finding an optimal DAG, the quality of this model needs to be checked. To do so one have to look at each and every marginal posterior density. The marginal posterior density specifies the model parameter for each existing edge in the model. This check examines whether the data really contain sufficient information to accurately estimate all model parameters. This is approved if the density is nicely bell shaped and hence has a clear defined maximum and if the area under the curve integrates to 1, as this is a prerequisite for a density. The mode, e.g. the X-value at which the probability density function takes its maximum value, of the respective marginal posterior density specifies the parameter estimate. Based on the marginal posterior densities a credible interval can also be specified for each parameter. A 95% credible interval is defined by two real numbers l and u that fulfill

$$\int_l^u f(\theta|x)d\theta = 0.95.$$

A credible interval is similar to a confidence interval of a frequentist setting, but the interpretation is much easier. Bayesian credible intervals allow the statement that the unknown parameter lies in a certain credible interval with probability 0.95. If zero is not included in the credible interval, we know with 95% probability that our parameter is not zero and thus not meaningless. The concept of parameter significance can be based on this. If the credible interval of a parameter includes zero, a parameter is set to be non-significant. If zero is not included, a parameter can be seen as significant.

The interpretation of the model parameters is based on the regression coefficient interpretation from a GLM. Thus, comparing different model parameters is tricky as the assumption of the error distribution does not coincide. Additionally the interpretation and the scale of the model parameters differ depending if a node is set to be Gaussian, Binomial or Poisson distributed.

The parameter of an arc pointing to a Binomial node is covered by a logit transformation and an appropriate back transformation is needed to speak about the regression coefficient. This is analogous to the logistic regression. The crude coefficients are log odds and can be interpreted as log odds ratios for state 1 of the binomial variable relative to state 0, respectively basis category. When exponentiating them this results in odds ratios. For continuous or count variables, the odds ratios are in respect of a 1-unit increase.

Parameters of an arc pointing to a Poisson node need no additional transformation but are already on the parameter scale of the GLM.

A model parameter of an arc pointing to a Gaussian node represent the correlation coefficient or the regression coefficient depending whether the `centre` option in the `fitabn` function is set to `TRUE` (default) or `FALSE`. If the `centre` option is set to `TRUE` the correlation coefficient is printed, if it is set to `FALSE`, the output is the regression coefficient. Assuming a simple regression model

$$Y_i = \alpha + \beta X_i + \epsilon_i,$$

the correlation coefficient between X_i and Y_i is related to the regression coefficient β as follows:

$$\beta = \text{CORR}(Y_i, X_i) \frac{\text{SD}(Y_i)}{\text{SD}(X_i)},$$

where SD represents the standard deviation. If the option of `fitabn` is set to `centre=TRUE` observations in each Gaussian node are first standardized to mean zero and standard deviation one. This leads to a canceling of the standard deviation fraction and hence to the output of the correlation coefficient which is equal in this case to the standardized regression coefficient. All results in this thesis are calculated with the default option of `centre=TRUE`.

Additionally to the model parameters of the conditional probabilities displayed in the DAGs, a model parameter of the intercept is also estimated for each node. Furthermore, for Gaussian nodes a precision parameter exist, which is defined as the inverse of the variance. These parameters are not of big interest because they only denote a background constant.

When an optimal DAG with well-established posterior densities has been determined, the edges must be filtered for robust results to adjust for over fitting. Over fitting is an always present and critical issue in model comparison, also in ABN. To address this problem the technique of parametric bootstrapping was used. With the help of Markov chain Monte Carlo (MCMC) simulations 10'000 data sets of identical size as the original data set were generated. On each simulated data set the searching approach by [Koivisto and Sood \[2004\]](#) was performed. Based on this process the frequency and thus the robustness of an arc can be determined. Arcs present in less than 50% of the simulated DAGs were considered not to be robust and pruned out. The threshold of 50% is justified by a simple majority consensus. This process results in a single and robust graph which represents the multivariate GLM.

The method of ABN offers also the possibility to add a random effect to the model. A random effect is needed to depict a potential hierarchy in the data. It is important and interesting to include such a hierarchy in the analysis, because it allows to differentiate between variability that occurs within a certain group or across the groups. Adding a grouping variable to the model and specifying nodes which should be adjusted for this grouping variable changes the analogy of ABN to a GLM to an analogy to a GLMM (generalized linear mixed model). This means that for the defined nodes a mixed model is used. Mixed models adjust for potential within group correlation and thus support the control for unobserved heterogeneity which is correlated with independent variables, in our case for potential unrecorded differences between the farms.

Generally in Bayesian network analysis, the concept of the Markov blanket is very common and useful. The Markov blanket is a set of nodes that shield a specific target node from the rest of the network. Given the Markov blanket of A ($Mb(A)$), A is conditionally independent with other nodes B from the network [Pearl et al., 1989].

$$P(A|Mb(A), B) = P(A|Mb(A)).$$

Thus the knowledge of the nodes in the Markov blanket of a node A is the only knowledge needed to predict the behavior of the node A . In Bayesian networks, including ABNs, the Markov blanket of a node A composes of A 's parents, its children, and its children other parents.

3.2 ABN: a more statistical approach

The two steps of model learning in additive Bayesian networks are divided into structure learning and parameter learning. Structure learning comprises specifying the DAG structure, parameter learning comprises specifying the local probability distributions. Constructing an additive Bayesian network model A with a data set D , we have:

$$P(A|D) = \underbrace{P(\beta_A, S|D)}_{\text{model learning}} = \underbrace{P(\beta_A|S, D)}_{\text{parameter learning}} \cdot \underbrace{P(S|D)}_{\text{structure learning}},$$

where S denotes the structure of the DAG and β_A the model parameters.

When performing the structural learning step, the aim is to look for a network structure that represents the data set sufficiently well without being too complex. Score-based learning methods assign a score to each network structure. The score reflects how likely a structure is when comparing it to the data set at hand. Looking for the model structure with the highest score can be considered as a search problem and is the main task of score-based learning methods. To approach this search problem a score function and a search procedure are needed.

A suitable **score function** should allow a balance between the accuracy and the complexity of the structure and it should be decomposable into the sum of local scores. The marginal likelihood is a classical Bayesian approach for measuring the fitness of a network structure. The marginal likelihood is given by:

$$P(D|S) = \int_{\beta_A} P(D|S, \beta_A) \cdot \pi(\beta_A|S) d\beta_A, \quad (3.1)$$

where $\pi(\beta_A|S)$ is the prior probability distribution over the parameters, conditioned on S . The marginal likelihood can be interpreted as the probability of the data D for a given model structure S or more precise as the probability that the data D could be generated if the parameters for S were selected randomly according to the parameter prior $\pi(\beta_A|S)$.

As a result of the decomposability property of the score function, the total network score can be written as

$$P(D|S) = \prod_{j=1}^n P(D_j|S), \quad (3.2)$$

where j denotes the number of nodes in the data D .

Martha Pittavino showed in her thesis [Pittavino, 2016] that simplifications of the integral given by Equation 3.1 for standard Bayesian networks [Cooper and Herskovits, 1992, Heckerman et al., 1995] can be also obtained for an ABN model. This allows the computation of the complex integral with the help of numerical Laplace approximation technique.

To find the highest-scoring Bayesian network structure in the set of all possible network structures a **search procedure** must be defined. Beside heuristic search strategies that move around in the search space by iteratively performing small changes to the current structure, order based searches are introduced by Friedman and Koller [2003]. The order references simply to a list of the nodes, say as indexes from 1 to n . A given DAG structure is consistent with an ordering if and only if the parents of each node precede their child node in this list. This results in groups of DAG structures, groups of structures consistent with one particular ordering. This clever decomposition of the space of possible graph structures reduces the model search space from $n!2^{\binom{n}{2}}$ to only $n!$. This represents a big decrease but still it is only computationally feasible for problems smaller than 20 nodes. Two different approaches for searching across orders instead of DAGs have been proposed. The first method was proposed by Friedman and Koller [2003]. A search algorithm samples randomly across the landscape of orders and collects information about the degree of statistical support for the structural features. The results are the posterior probabilities for each arc. Later an exact algorithm which considers every order was proposed by Koivisto and Sood [2004]. Exact algorithms are designed in a way that they will find an optimal structure in a finite amount of time. In contrast, heuristics have no guarantee to find an optimal structure. Koivisto and Sood [2004] improved the algorithm by integrating a summation over orders to the algorithm. The evaluation of the sum over orders is done by a dynamic propagation algorithm. Each node receives a value from the summed up values of its parents, where each value was multiplied by a quantity that depends on the associated path. Hence, computations over different paths can be merged together. The mapping of these different path subsets is done by a so called Möbius transform. This advantage of merging overlapping parts reduces the running time even more and thus allows the exact calculations. The major advantage of this approach is that it explores all possible structures, running “only” in exponential time with respect to the number of nodes in the network. This approach is implemented in the `mostprobable` function of the R package `abn`.

After knowing the structure of a model, the aim is to estimate the parameters of the model. The model comprises of a set of conditionally independent generalized linear regressions. The model parameters are the regression coefficients based on the conditional expectation of the value of a certain node given its parents. The parameters are learned

using the principle of maximum likelihood [Held and Sabanés Bové, 2014]. These calculations also require numerical integration as used for the evaluation of the score function. Two numerical techniques are available in the `fitabn` function of the R package `abn`, standard Laplace approximation or an integrated nested Laplace approximation (INLA).

3.3 Random forest

An additional analysis was made with the technique of random forest to compare the results of the analysis with ABN.

A random forest consists of many independent tree models. In a tree model the aim is to introduce branching points in order to reduce the variance in the underlying data. By combining the results of these many independent tree models it is possible to reduce the variation of the prediction leading to a high accuracy given that the tree models are not biased.

Random forests can be used for regression or classification. Classification trees are used if the outcome is binary or discrete. Regression trees are used if the outcome is continuous. On every branching point, a split rule tries to reduce the variance regarding the outcome variable in the sample. In Figure 3.2 the blue circles denote the branching points.

Random forest uses two principles to get independent tree models. The first one is bootstrapping, which denotes the idea to use different versions of the training data for every tree. The second one includes a sampling of the variables on each split at random and then choosing the best variable out of this sample. Figure 3.2 visualizes the bootstrapping process as well as the sampling of variables on each split.

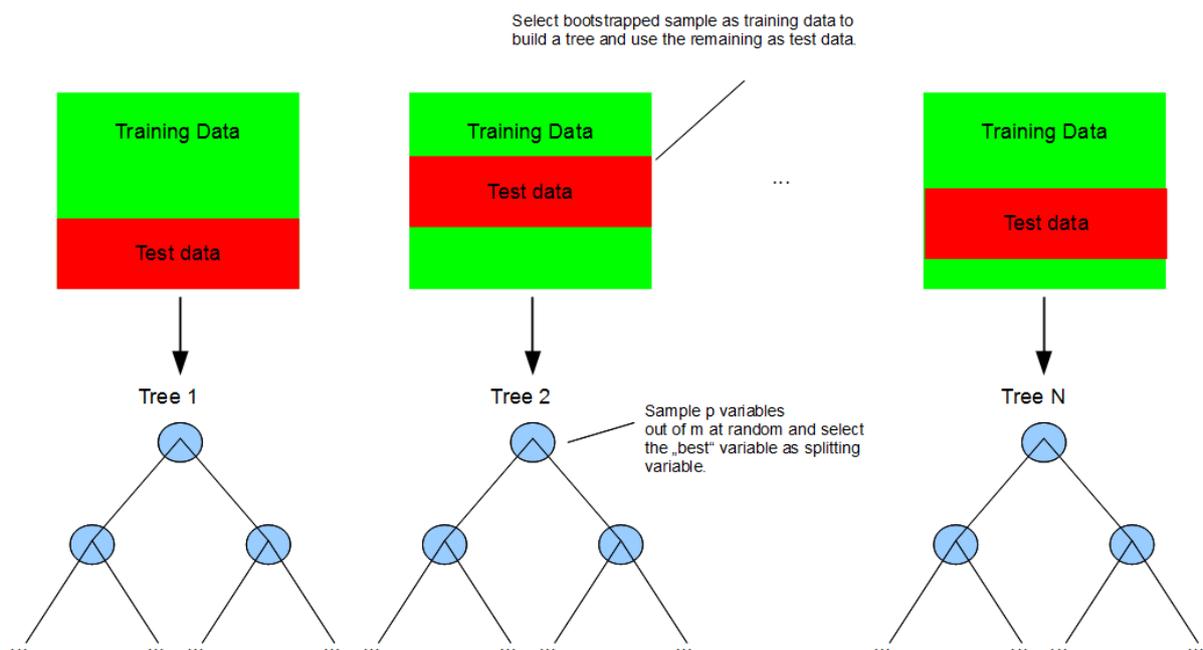


Figure 3.2: Outline of a random forest. Source: Lecture notes ‘Introduction to the Random Forest Method’, Professor B. Sick, ZHAW.

By combining the results of all the independent tree models the technique of random forest corrects for the over fitting problem, which is known for tree models. Furthermore, random forest has the advantages that it has no assumption about a certain distribution of the variables, it can handle interactions very well and it is not sensitive to a big amount of predictors [Breiman, 2001].

Due to the included bootstrapping mechanism each random forest provides an out-of-bag estimate for the explained variance and the error rate, so no additional cross validation is needed anymore. The measure of explained variance shows how well out-of-bag predictions explain the variance of training data sets.

As mentioned every split in a tree tries to minimize the variance from the sample before the split. Depending on the chosen variable at a split this reduction in variance differs. A variable importance plot indicates which variables are most important in variance reduction. For a regression tree, as it is the case in this analysis, the variable importance is defined by the average increase in the squared out-of-bag residuals when the variable is permuted. This means that one feature after the other is randomly permuted between the different observations and than the decrease of the out-of-bag classification of this manipulated observations is determined. A large decrease in classification accuracy indicates a large importance of the permuted feature.

3.4 Software

All analyses were performed in the R programming language [R Development Core Team, 2006]. The ABN part was performed with the R package `abn` [Lewis, 2016] [Kratzer et al.]. The MCMC simulations were realized with JAGS (Just Another Gibbs Sampler) [Plummer, 2003]. The random forest was performed with the function `randomForest` from the R package `randomForest` [Liaw and Wiener, 2002] [original by L. Breiman et al., 2015]. The simple GLM was performed with the function `glm` from the R package `stats` and the GLMM with the function `lmer` from the R package `lme4` [Bates et al., 2016] [Bates et al., 2015].

Chapter 4

Analysis and Results

4.1 Additive Bayesian networks

As mentioned in Chapter 2, the analysis was performed for the worst pododermatitis score as well as the mean of the pododermatitis scores from the middle part of the hind paws. Additionally, we performed an analysis with the variable **Age** and one without **Age**, because including **Age** leads to a big reduction of the data set due to the missing values. This results in four different variants of our data set.

For all four variants first of all the optimal limit of allowed parents must be determined. To allow generally as many parents as variables are available in the data set is not really effective because the searching space increases super exponentially with every additional node. Instead, one starts with a low number of allowed parents and increases this number until the graph is saturated and allowing more edges does not increase the fit, e.g. the marginal likelihood, anymore. Figure 4.1 shows the result of such a searching process, where the saturated maximum was reached at 11 allowed parents. For the two data sets without **Age** the optimal limit of allowed parents is 11, for the two data sets with **Age** it is 12.

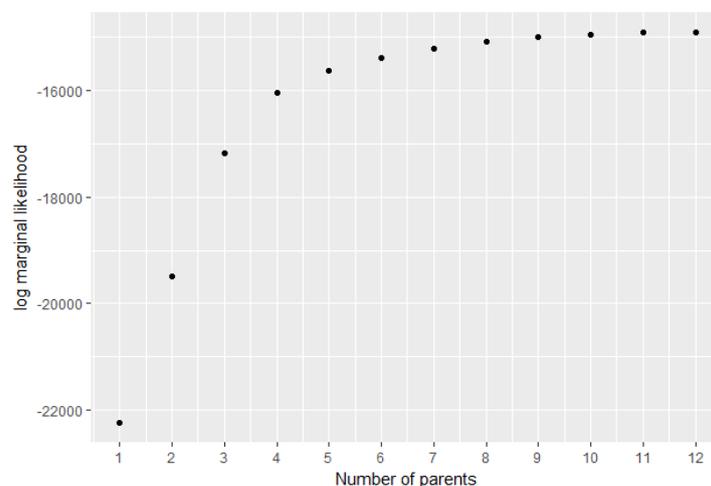


Figure 4.1: Comparison of goodness of fit with increasing number of parents.

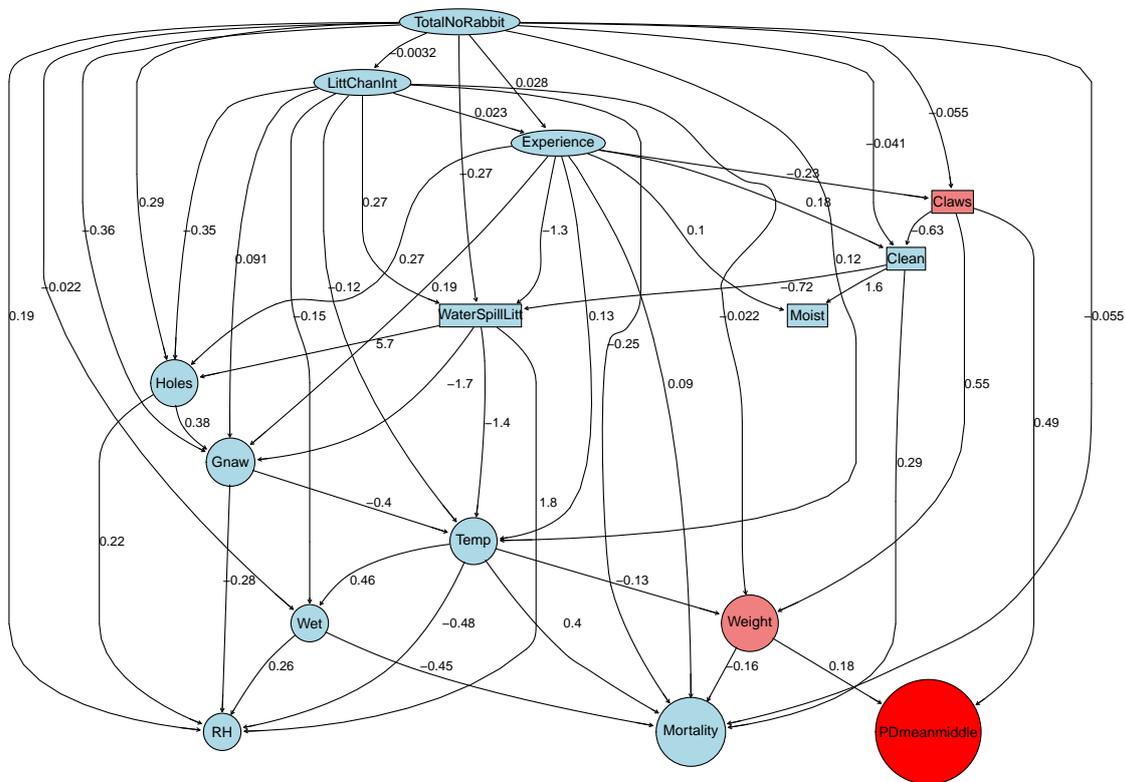


Figure 4.2: Final DAG of the analysis with the mean score of pododermatitis at the middle part, excluding age, $n = 1042$. The Markov blanket of `PDmeanmiddle` is emphasized in red. The shape of the nodes visualizes the chosen distribution, where a circle stands for the Gaussian distribution, a square for the Binomial and an ellipse for the Poisson distribution.

When reaching the maximum of the marginal likelihood we have the optimal model meaning this model represents the structure of the given data best. Nevertheless, we have to check the quality of this model by looking at each marginal posterior density. As described in Section 5.1 a bell shaped density with a clear defined maximum is not always the case and the options to deal with such irregularities are very limited. The only helpful solution is to exclude a variable if the marginal posterior densities look irregular. After having some problems with the first selection of variables, we were able to find a selection of meaningful variables, which is presented in this thesis, with well defined posterior density. The only additional irregularity we had to take care of was the upcoming problem with the variable `WaterSpillLitt` when including `Age` into the model. Section 5.1 approaches this problem further. In all further analysis with ABN of the data set with `Age` is excluding the variable `WaterSpillLitt`.

If all posterior densities are well shaped and integrate to one, we can definitely speak about an optimal model. The only remaining issue is now the issue of over fitting. So based on the optimal model we performed a bootstrap simulation and created 10'000 new data sets with the same number of observations and searched again for the optimal models of these generated data sets. At the end we discovered the support of all arcs and removed

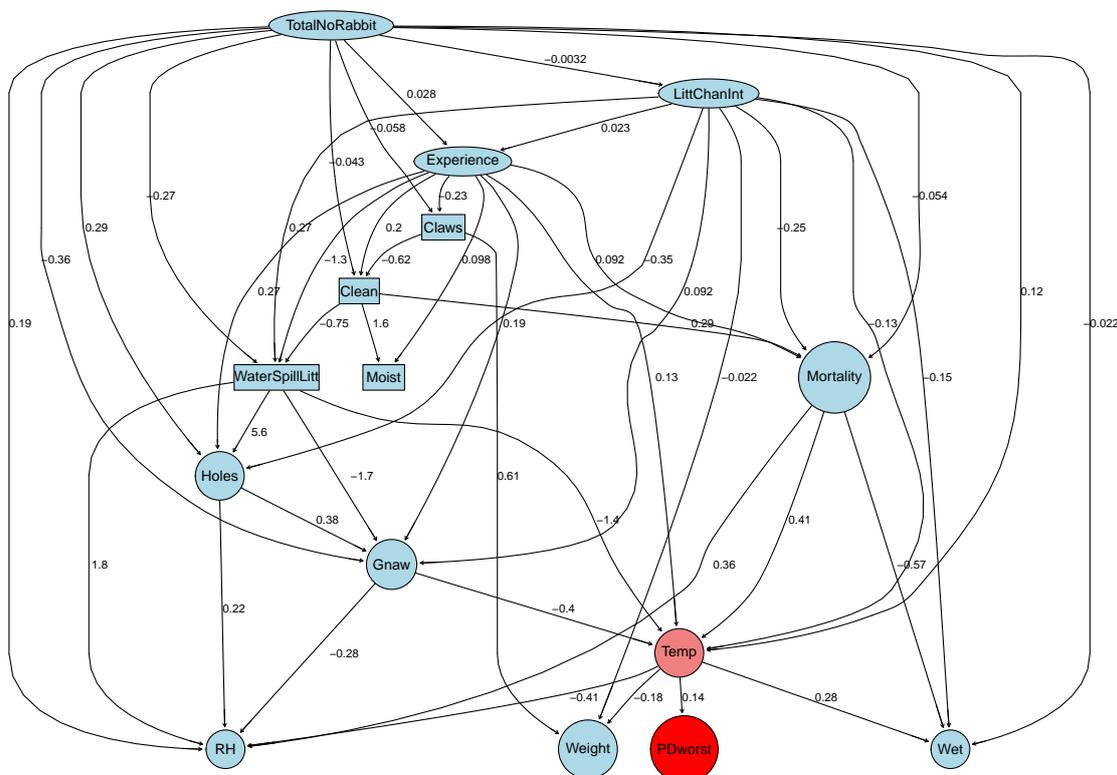


Figure 4.3: Final DAG of the analysis with the worst pododermatitis score, excluding age, $n = 1051$. The Markov blanket of **PDworst** is emphasized in red. The shape of the nodes visualizes the chosen distribution, where a circle stands for the Gaussian distribution, a square for the Binomial and an ellipse for the Poisson distribution.

these arcs which were only supported in less than 5'000 cases.

Figures 4.2 to 4.5 show the final and robust versions of the DAGs for all four variants of the data set. The shape of the nodes visualizes the chosen distribution, where a circle stands for the Gaussian distribution, a square for the Binomial and an ellipse for the Poisson distribution. In the graphs the Markov blanket of our target variables, the pododermatitis score, is colored in red. The Markov blanket of **PDmeanmiddle** or **PDworst** composes of their parents, their children and their children other parents. The knowledge of the nodes in the Markov blanket of **PDmeanmiddle** or **PDworst** is the only knowledge needed to predict their behavior, because they are conditionally independent from the rest of the network.

Figure 4.2 is the DAG of the analysis with the mean score of pododermatitis at the middle part of the hind paws, excluding the variable **Age** and thus including 1042 animals. This DAG contains 49 arcs. Two arcs were deleted after the bootstrapping process due to missing support in more than 5'000 bootstrapping samples. The deleted arcs were from **Claws** to **LittChanInt** and from **WaterSpillLitt** to **Claws**. In this figure the Markov blanket of the pododermatitis score is quite simple, as it only has two parents, **Weight** and **Claws**.

Figure 4.3 is the DAG of the analysis with the worst pododermatitis score, again excluding the variable **Age** and therefore including 1051 animals. This DAG includes 47

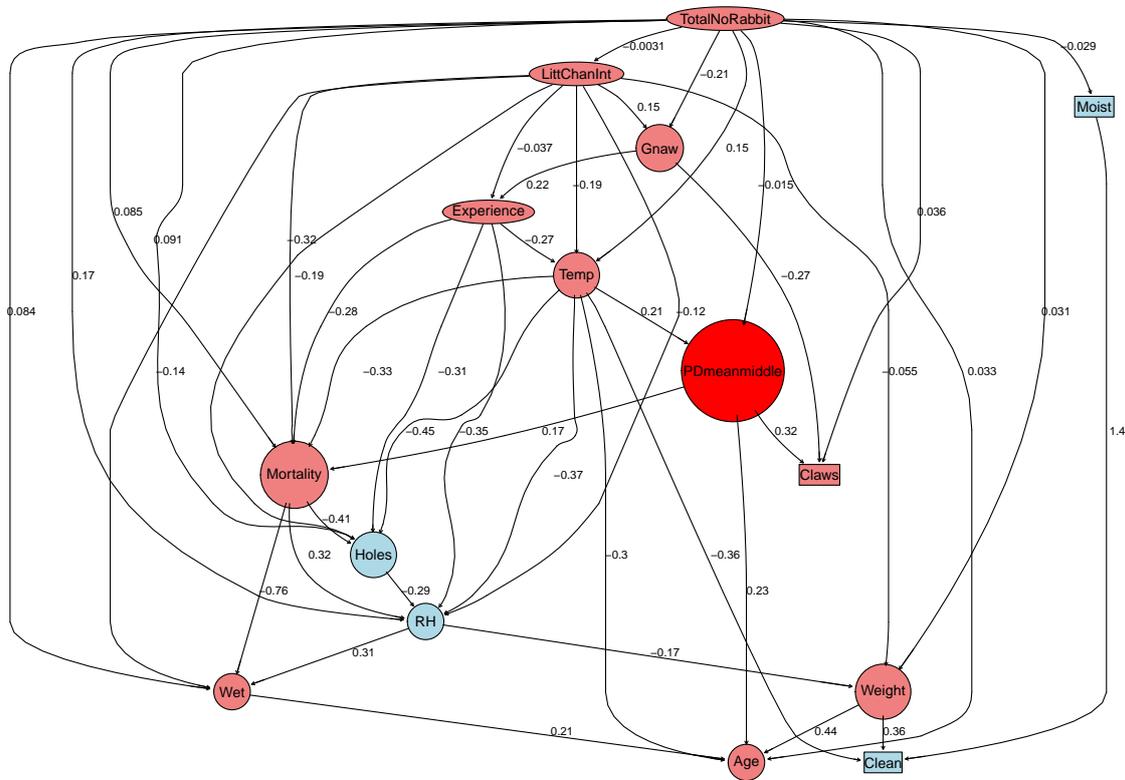


Figure 4.4: Final DAG of the analysis with the mean score of pododermatitis at the middle part, including age, excluding WaterSpillLitt, $n = 649$. The Markov blanket of PDmeanmiddle is emphasized in red. The shape of the nodes visualizes the chosen distribution, where a circle stands for the Gaussian distribution, a square for the Binomial and an ellipse for the Poisson distribution.

arcs. Also two arcs were deleted after the bootstrapping process. In this DAG it concerned again the arc from Claws to LittChanInt and additionally the arc from Wet to RH. The Markov blanket in this figure is even more simple, as the pododermatitis score is only connected to Temp.

Figure 4.4 and Figure 4.5 show the DAGs of the analyses with the mean score of pododermatitis at the middle part respectively with the worst pododermatitis score when including the variable Age. This results in almost a bisection of the data set, namely to 649 respectively 652 individuals. This implies also the exclusion of the variable WaterSpillLitt. In both cases one arc has to be trimmed after the bootstrapping process and this affected in both cases the arc from Claws to Age. The final DAG with PDmeanmiddle has 45 arcs. The final DAG with PDworst has 43 arcs. The Markov blanket in the analysis with PDmeanmiddle looks more complicated. The variable of interest is much more connected to the rest of the network and also has some children that were not present in the analysis without Age. For the analysis with PDworst there are again no children of PDworst but three parents Temp, LittChanInt and TotalNoRabbit.

In the final DAGs every arc is labelled with a value. These values correspond to the mode of the respective marginal posterior density and therefore represent the model pa-

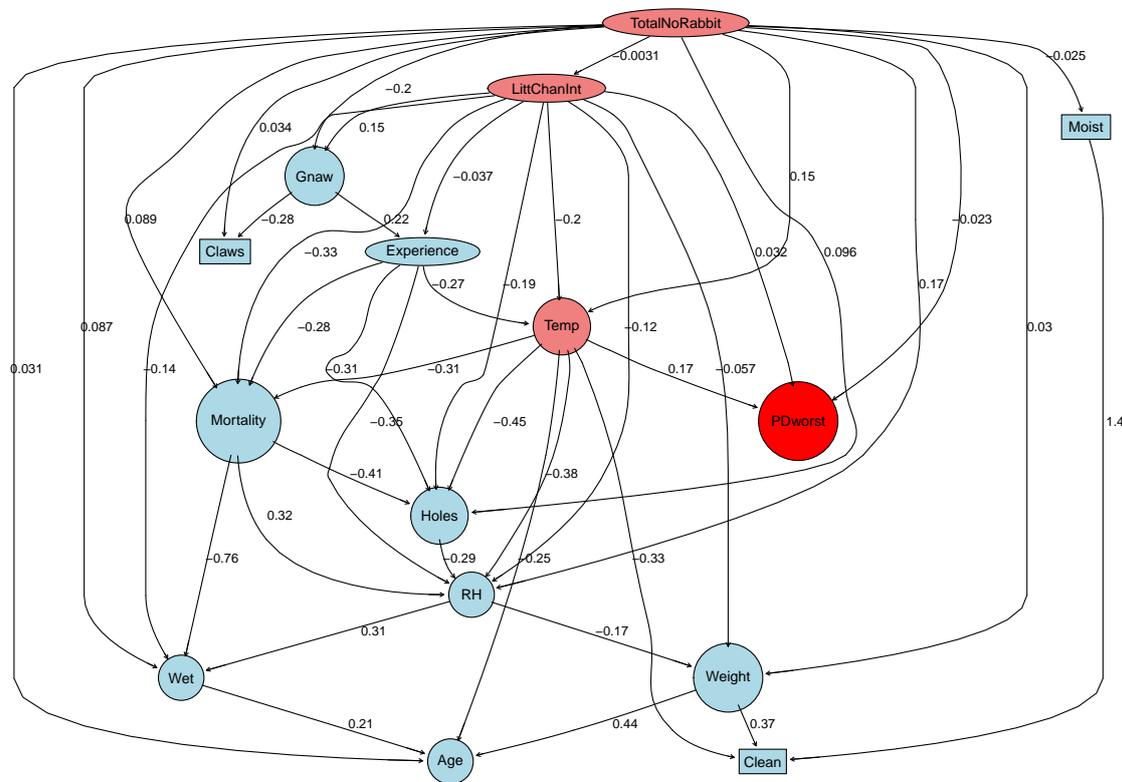


Figure 4.5: Final DAG of the analysis with the worst pododermatitis score, including age, excluding WaterSpillLitt, $n = 652$. The Markov blanket of PDworst is emphasized in red. The shape of the nodes visualizes the chosen distribution, where a circle stands for the Gaussian distribution, a square for the Binomial and an ellipse for the Poisson distribution.

parameter estimates. Depending from the distribution of the node an arc is pointing to the values should be interpreted differently, see Section 3.1. Focusing on the values on the arcs pointing to PDmeanmiddle from Figure 4.2, as PDmeanmiddle is set be a Gaussian node they can be interpreted as standardized regression coefficients or correlation coefficients between Weight and PDmeanmiddle respectively Claws and PDmeanmiddle. Both values are positive and denote thus a positive correlation. Heavier body weight is associated with higher pododermatitis scores and too long claws are also associated with higher pododermatitis scores. Whereas the correlation between Claws and PDmeanmiddle is almost three times stronger than the correlation between Weight and PDmeanmiddle. Figure 4.4 shows a more complex situation. Again focusing on the node PDmeanmiddle two arcs are ending and three arcs are leaving. Two of the three leaving arcs are also pointing to Gaussian nodes, one arc points to Claws, e.g. a Binomial node. Consequently, all four values from arcs to a Gaussian node can be compared, they all denote a standardized regression coefficient, respectively the correlation. The correlation from TotalNoRabbit to PDmeanmiddle is negative. Therefore, we associate heavier pododermatitis scores with smaller farms, although we see that this correlation is really small and minor. The correlations of Temp, Mortality and Age do not vary greatly in size and can be considered as equally strong.

The parameter of `PDmeanmiddle` to `Claws` has to be transformed before comparing it to the others. Taking the inverse logit of 0.32 this results in 0.58. This value denotes now also a regression coefficient and can be compared to the others. The coefficient between `Claws` and `PDmeanmiddle` is stronger than the other coefficients related to `PDmeanmiddle` and about the same size as the coefficient from `PDmeanmiddle` to `Claws` in Figure 4.2.

None of the credible levels of the parameter estimates include zero, see Table A.1 to A.4 in the appendix for the exact values. This means that all parameters are significant. This is not guaranteed to happen, as this criteria was not part of the model selection process. All parameters in the model were only justified by the use of maximum likelihood and bootstrapping. The estimates of the intercepts and the precision parameters are not included into these tables in the appendix, as they only denote some background constants.

Another idea to analyse the results of ABN after the bootstrapping is to look at the exact number of retrieval for each arc in the 10'000 bootstrap samples. The number of retrieved arcs evaluates the robustness and strength of the arcs independently from the parameter estimation. In the next release of the R package `abn` there will be an additional option which allows to visualize this information.

As mentioned in Section 2 the data contain a hierarchical structure with two levels, the level of the individual animals and the level of the farms. Specifying this data structure in the function of `abn` unfortunately leads to meaningless or no results. A detailed documentation of the errors, crashes and odd findings resulting by using ABN with a nesting factor is written in Section 5.2. As the R package `abn` is still in development, these findings are nevertheless useful and helpful to implement a more robust and better usable version of the R package `abn`.

4.2 Random forest

All four variants of the data set were performed in an analysis with random forest. The aim of such an analysis is to find a tree model which is able to predict the outcome variable, in our case the pododermatitis score, as good as possible with the help of the existing explanatory variables. The pododermatitis score is handled as a continuous outcome, so the random forest was performed for regression. Additionally to the known variables, we added randomly distributed variables, one Gaussian (`randomG`), one Poisson (`randomP`) and one Binomial (`randomB`) random variable. These random variable help to compare the results and judge them in a relative manner. The parameters for the random variables were selected accordingly to the means of the parameters of the real variables. The success probability for the Binomial random variable was determined by the mean of the four success probabilities from the real binary variables, see Table 2.3, resulting in a value of 0.623. The rate parameter (λ) for the Poisson random variables was determined by the mean of the three means from the real count variables, see Table 2.2, resulting in a value of 124. The standard deviation for the Gaussian random variable was determined by the squared mean of the standard errors from the real continuous variables, see Table 2.2, resulting in a value of 5741.699. The mean of the random Gaussian variable is set to be

zero. The formula for the random forest was formulated as follows:

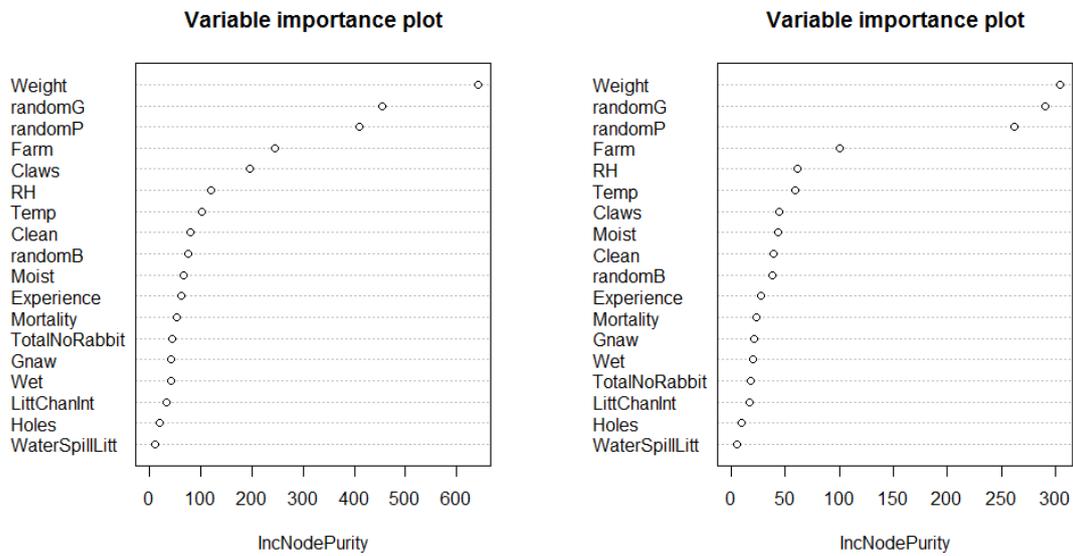
$$\begin{aligned} \text{PD} \sim & (\text{Age}) + \text{Weight} + \text{Claws} + \text{Clean} + \text{Moist} + \text{Wet} + \text{Gnaw} + \text{Holes} + \\ & \text{RH} + \text{Temp} + \text{Mortality} + \text{Experience} + \text{WaterSpillLitt} + \text{LittChanInt} + \\ & \text{TotalNoRabbit} + \text{Farm} + \text{randomG} + \text{randomP} + \text{randomB}. \end{aligned} \quad (4.1)$$

The most interesting results from this analysis with random forest is the percentage of explained variance and the variable importance plot. In Table 4.1 the mean of the squared residuals and the percentage of explained variance is displayed for all four random forests. The mean of the squared residuals denote the deviance of the regressed pododermatitis score of a certain group to the actual pododermatitis scores which end up in this particular group. This measure describes the inaccuracy of a model, where higher values are associated with less accurate models. Whereas, higher values of the percentage of explained variance are associated with a better model that is able to explain more variance of the outcome variable. It is important to note that these numbers change slightly for every new run of a random forest. So for the interpretation a bouncing of these numbers around the written values should be taken into account. The means of the squared residuals are more or less equal for all four models and do not represent a very good accuracy of our models. The percentage of explained variance differ between the models. Generally, we can say that the models with `PDmeanmiddle` perform better than the models with `PDworst` and that the models with `Age` perform better than the models without `Age`. The best model is thus the model with `PDmeanmiddle` and `Age` with almost 40% variance explained. Nevertheless this result is not very convincing as it is still less than 50% which coincide with explaining half of the variance. This indicates that important variables which could explain the variance of the pododermatitis score are missing in this model.

Judging the variable importance plots, see Figure 4.6, `Weight` and `Age` seem to be the most important variables. However, the supplementary random variable `randomG` is almost always nearly as important as `Weight` and/or `Age`. Additionally the random variable `randomP` is always in a higher position than the other count variables (`Experience`, `LittChanInt`, `TotalNoRabbit`). `randomB` is at a rather low position in the importance plots. Some real binary variables, like `Claws` and `Moist` are in higher positions, but when looking at the corresponding x-values, the differences are negligible and all of these lower variables have to be considered as not important. These results indicate again that the model is not really above the threshold of predicting our target variable better than just by random. The only exception would be the variables `Weight` and `Age`.

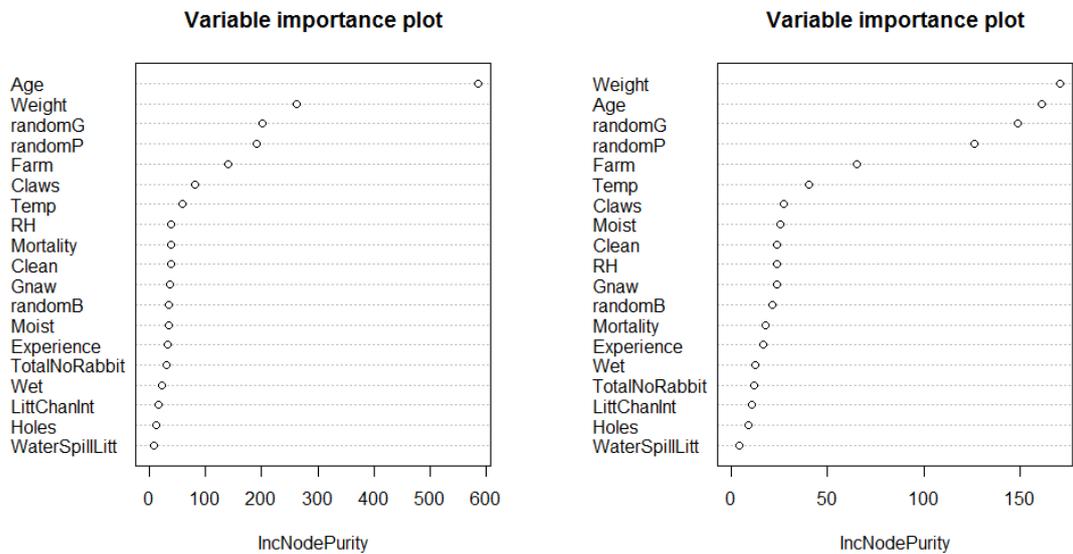
Table 4.1: Mean of squared residuals and percentage of explained variance for the random forests, where Forest 1 denotes the analysis with `PDmeanmiddle` and without `Age`, Forest 2 the analysis with `PDworst` and without `Age`, Forest 3 the analysis with `PDmeanmiddle` and `Age` and Forest 4 the analysis with `PDworst` and `Age`.

| | Forest 1 | Forest 2 | Forest 3 | Forest 4 |
|---------------------------|----------|----------|----------|----------|
| Mean of squared residuals | 2.6 | 1.5 | 2.0 | 1.5 |
| % variance explained | 14.2 | 3.1 | 37.4 | 8.7 |



(a) Analysis PDmeanmiddle without Age, $n = 1042$

(b) Analysis PDworst without Age, $n = 1051$



(c) Analysis PDmeanmiddle with Age, $n = 649$

(d) Analysis PDworst with Age, $n = 652$

Figure 4.6: Variable importance plots of the four random forests, where IncNodePurity denotes the total increase in node purities from splitting on the variable.

4.3 Regression approach

When running a simple GLM with the same Formula 4.1 on our four data sets, except the random variables, the variables indicated in Table 4.2 were denoted as significant on a

Table 4.2: Significant variables on a 95% significance level and the decrease of the deviance (null deviance – residual deviance) when performing a GLM, where Model 1 denotes the analysis with `PDmeanmiddle` and without `Age`, Model 2 the analysis with `PDworst` and without `Age`, Model 3 the analysis with `PDmeanmiddle` and `Age` and Model 4 the analysis with `PDworst` and `Age`.

| | Coefficients | Estimate | Std. Error | P-value | Decrease of deviance |
|---------|----------------|----------|------------|---------|----------------------|
| Model 1 | Weight | 0.464 | 0.096 | <0.001 | 348.4 |
| | Claws | 0.845 | 0.113 | <0.001 | |
| | WaterSpillLitt | 0.420 | 0.176 | 0.017 | |
| | Temp | 0.104 | 0.042 | 0.014 | |
| Model 2 | Claws | 0.254 | 0.084 | 0.003 | 69.5 |
| | Temp | 0.122 | 0.032 | <0.001 | |
| Model 3 | Age | 0.070 | 0.012 | <0.001 | 428.6 |
| | Weight | 0.332 | 0.123 | 0.007 | |
| | Claws | 0.607 | 0.150 | <0.001 | |
| | Temp | 0.323 | 0.072 | <0.001 | |
| | Mortality | 0.008 | 0.003 | 0.026 | |
| Model 4 | Claws | 0.308 | 0.116 | 0.008 | 108.81 |
| | RH | 0.029 | 0.011 | 0.010 | |
| | Temp | 0.336 | 0.056 | <0.001 | |
| | Mortality | 0.008 | 0.003 | 0.002 | |

95% significance level. `Claws` and `Temp` are in every model significant. `Weight` seems only to be important for the models with `PDmeanmiddle`. A look at the decrease of deviance, which is calculated by subtracting the residual deviance from the null deviance, confirms the observation that `PDmeanmiddle` is better explainable than `PDworst` by our models. Again, the models including `Age` perform better than the ones without `Age`.

Changing the GLM setting to a GLMM setting and including a random effect for all variables measured on animal level, the results look a bit different. The formula used for the GLMM setting looks as follows:

$$\begin{aligned}
 \text{PD} \sim & (\text{Age}) + \text{Weight} + \text{Claws} + \text{Clean} + \text{Moist} + \text{Wet} + \text{Gnaw} + \text{Holes} + \\
 & \text{RH} + \text{Temp} + \text{Mortality} + \text{Experience} + \text{WaterSpillLitt} + \text{LittChanInt} + \\
 & \text{TotalNoRabbit} + (1 + (\text{Age}) + \text{Weight} + \text{Claws} + \text{Clean} + \text{Moist} | \text{Farm}) .
 \end{aligned} \tag{4.2}$$

According to this formula a random effect for the intercept, for `Age`, if `Age` is included into the analysis, `Weight`, `Claws`, `Clean` and `Moist` is allowed. A random effect allows to differentiate between variability that occurs within a certain group or across the groups. This adjusts for potential within group correlation.

The significant variables of the mixed Model 1 are `Weight`, `Claws`, `Experience`, `LittChanInt` and `TotalNoRabbit`. For Model 2 with `PDworst` as the outcome variable the significance level of 95% is reached by `Weight`, `Claws` and `RH`. Model 3 has `Age`, `Weight`, `Claws`, `Holes` and `RH` as significant variables, whereas Model 4 only has `Temp`. Additionally to the significant variables and its estimate, the variance decomposition of the GLMM is important. This indicates how much variance can be explained by group differences

Table 4.3: Variance decomposition of the random effects from the mixed models, where Model 1 denotes the analysis with PDmeanmiddle and without Age, Model 2 the analysis with PDworst and without Age, Model 3 the analysis with PDmeanmiddle and Age and Model 4 the analysis with PDworst and Age.

| Groups | Variables | Model 1 | Model 2 | Model 3 | Model 4 |
|----------|-----------|---------|---------|---------|---------|
| Farm | Intercept | 0.283 | 0.160 | 0.752 | 0.433 |
| | Age | - | - | 0.377 | 0.229 |
| | Weight | 0.020 | 0.013 | 0.009 | 0.009 |
| | Claws | 0.025 | 0.025 | 0.057 | 0.112 |
| | Clean | 0.033 | 0.027 | 0.085 | 0.055 |
| | Moist | 0.033 | 0.037 | 0.080 | 0.046 |
| Residual | | 0.816 | 0.900 | 0.661 | 0.812 |

and how much variance remains across the groups respectively farms. Table 4.3 gives an overview about this variance decomposition. We can see that the residual variance remains the biggest part of the total variance, meaning that this is the variance across the groups, so not explained by differences between the farms. In the second place is the intercept of the random effect, this number indicates the variability which can be explained by differences between the farms. The variance explained by allowing a random slope for Age should not be neglected because this could be an indication of the different distributions of this variable on the distinct farms, see Figure 2.6. An interaction between the appearance of pododermatitis and these different age distributions is easy to imagine.

Chapter 5

Technical Remarks on the R Package abn

5.1 Peculiar posterior densities

After doing the first analysis with our most interesting variables we detected some strange marginal posterior densities. Strange means in this context, that these densities have no bell shape and consequently no well defined maximum. Additionally these densities are very wide, meaning the values of the x-axis cover a big area and the y-values are accordingly low. This indicates as well, that there is no explicit and precise maximum. One can find some examples in Figure 5.1. Figure 5.1d shows an example where the density has even negative y-values. This is not valid for a density function and makes all following calculations impossible.

Estimating marginal posterior densities for individual parameters, can run into trouble as this presupposes that the data contains sufficient information to accurately estimate the density for every individual parameter in the model. This is a strong requirement and more demanding than just being able to estimate an overall network structure and its goodness of fit metric. So generally this feature indicates that our data does not contain enough information to estimate these marginal posterior densities. Trying to find some more detailed explanations we found several aspects. First of all these peculiarities exclusively concern Binomial variables. Mostly the problems arise from complete or partial data separation. The risk of having such a data separation increases when having a lot of Binomial variables in the model and when having very unbalanced Binomial variables. Another aspect is the constellation of parents and children that a certain (Binomial) node has. The best example for this manner is the node `WaterSpillLitt` in the analyses with and without `Age`. When comparing the settings of the analyses without and with `Age` for the worst and the mean of the middle pododermatitis scores, the only difference is that `Age` is included in the analysis and that based on this a lot of observations had to be excluded due to missing values in `Age`. The resulting DAG differs and the marginal posterior densities from `WaterSpillLitt` are messy when including `Age`. In both cases (worst and mean of the middle part pododermatitis score) `WaterSpillLitt` is having more parents in the analysis with `Age` (6 parents) than without `Age` (4 parents). And in both cases some parents are becoming children and some children are becoming parents when including `Age` (e.g. `Clean` and `Experience` or `Temp`). This example from the analysis with the mean of

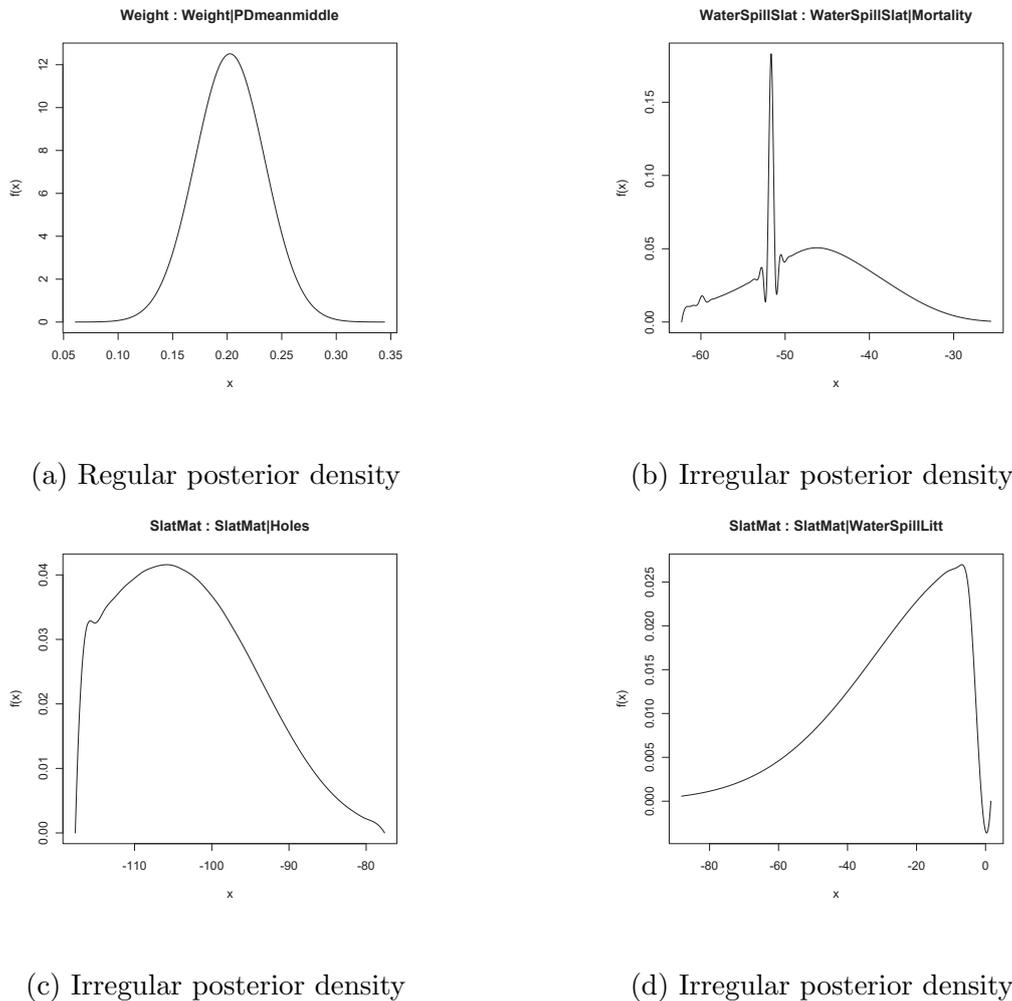


Figure 5.1: Peculiar marginal posterior densities compared to a bell shaped and well defined marginal posterior density.

the pododermatitis scores from the middle part is visualized in the appendix. Figure A.1 shows the DAG from the analysis without `Age` and Figure A.2 with `Age`. These figures also visualize that the Markov blanket of `WaterSpillLitt` increases when adding the node `Age`. Where all posterior densities of `WaterSpillLitt` in the analysis without `Age` looked like regular densities (see Example 5.1a), Figure A.3 shows how the densities look like after including `Age`. What exactly causes the destruction of all marginal posterior densities of `WaterSpillLitt` is still unclear, but should be closely related to these facts.

However, the only encountered solution to such irregular densities is to exclude the concerning nodes. The bootstrapping process is not possible or completely misleading on the basis of such densities.

5.2 Error with random effects

When trying to run the ABN analysis with a random effect for `Farm`, we run into several problems.

A first problem occurred when trying to run a simple model allowing a random effect for `Farm` on the variable `Weight` on a computer with Windows as operating system. Shortly after starting the `buildscorecache` function the following error message is displayed and the calculations get aborted.

```
Error in try(res.c <- .Call("fit_single_node", data.df, as.integer(child), :  
oops - got an NAN! in g_rv_g_inner_gaus-----  
Error: object 'error.modes' not found
```

Luckily when running the same simple model on a computer with operating system Linux or Apple OS X there are no error messages and the calculations result in a meaningful outcome.

Based on the same data as used for the analysis without a random effect and presented in Section 4.1 we specified the model for the analysis with ABN and a random effect similar as in Equation 4.2. For both data sets without `Age`, e.g. once with `PDmeanmiddle` and once with `PDworst`, it was possible to start the calculations with the function `buildscorecache`. There are five variables (`Weight`, `Claws`, `Clean`, `Moist` and `PDmeanmiddle` or `PDworst`) for which a mixed model should be used to adjust for within group correlation. During the calculations you get a feedback on which node the function is running. Unfortunately the running time of this function took almost two weeks only for one node. This means that the running time of these two model is estimated to more than two months. This makes the selection process for the optimal number of allowed parents as well as the bootstrapping process almost unfeasible. Unhappily it definitely burst the possibilities during this thesis. Additionally to the super long running time the function returned constantly the two error messages below. Unfortunately we could not really figure out the reasons and the meaning of these warnings and errors and neither a solution.

```
ERROR: ABNORMAL_TERMINATION_IN_LNSRCH at node 2
```

```
Zero finding warning: internal--- epsilon status = iteration is not \\  
making progress towards solution
```

For both data sets with `Age`, again once with `PDmeanmiddle` and once with `PDworst`, the problems were even more dire. Shortly after starting the `buildscorecache` function, R crashes and displays the following message

```
Abortion of R session: Fatal error.
```

A reboot of the R session is the only left solution. We have some obvious differences, like having here six variables for which a mixed model should be used, allowing 12 parents instead of 11 and having only half as much observations as in the case without `Age`. But as defining only five variables for which a mixed model should be used, allowing only 11 parents and duplicating the data set, did not change anything on the fact that the fatal error occurs shortly after starting the function, we failed to find a proper explanation. The only left possible explanation is that the reason can be located to a special data constellation which occurs in the data set with `Age` but not without `Age`.

5.3 Interpretational difficulties and lack of documentation

When reading through literature and documentations of ABN [Kratzer et al.] [Pittavino, 2016] [Lewis and Ward, 2013], a lot of interesting textes are available and I got a big knowledge about using and understanding the technique of ABN. Nevertheless, I had to realize that there remains a gap between the mathematical and statistical understanding of the method and the interpretational understanding of some results, more precisely the interpretation of the model parameters. It would be very nice if these model parameters could be directly transferred to more common techniques used in statistical analyses. However, therefore a detailed documentation about these model parameters would be necessary.

As reflected also in Section 3.2, the method of ABN is described clearly and additional information can be looked up in several referenced literature until the step of parameter learning and its results. I was not successful in finding a detailed answer what the outcome, e.g. the fitted values from the `fitabn` function denote and how they are derived. In the documentations and publications about ABN this part is mostly missing or very reduced. When interpreting the outputted values the authors simply write about odds ratios [Kratzer et al.] [Pittavino et al., 2016] [Lewis and McCormick, 2012]. Mostly they only work with Binomial nodes and hence there is no need to distinguish if a Binomial, Gaussian or Poisson distribution was assumed for the parent node or the child node. Only in Lewis and Ward [2013] they write about marginal mean posterior effects when having an arc to a Gaussian node. Unfortunately I could not find any justification, which would explain exactly how these estimates are calculated, if additional transformations to the outputted results from the function `fitabn` were applied or not and why they denote exactly odds ratios and mean effects.

On the other hand side, Gilles Kratzer was able to show that the interpretation presented in Section 3.1 is needed to get the right context between the parameters estimated from ABN and established coefficients, e.g. the regression coefficients. His results which are unpublished at this time are based on simulated data, where the true parameters are well known. Although citing a yet unpublished ABN tutorial written by Gilles Kratzer et al. with the appropriate results I decided to apply this interpretation of the parameter estimates, because I got insights into a reasonable, traceable and argumentative pathway, which explains the proposed interpretation.

However, ABN is mainly about network structure discovery and probably should also primarily be used for this task, at least at the moment until further knowledge about the exact and detailed parameter interpretation is available.

Chapter 6

Conclusion and Discussion

In summary, similar outcomes across all used methods and some more specific outcomes which not clearly coincide with the other methods were found. Focusing first on the similar outcomes, the worst pododermatitis score of an animal is harder to explain than the mean score of the pododermatitis from the middle part of the hind paws. This conclusion is supported by the results from the random forest as well as from the GLM and the GLMM, where Model 2 always performs worse than Model 1 and Model 4 worse than Model 3, where the performance is linked in this case to the percentage of explained variance for the random forest and to the decrease in deviance for the regression models. This fact can be explained by the understanding that one extreme value can be rather due to special and specific circumstances, than a mean value which is more likely to be dependent on the general conditions an animal lives in. Assessing housing conditions like moisture of litter for a few pens per farm might have contributed to this. Special and animal-specific circumstances are moreover often very difficult to record.

We also see, that including **Age** into our model increases its power, although the data set is then very reduced. This conclusion is also supported by the results from the random forest, the GLM and the GLMM, where Model 3 always perform better than Model 1 and Model 4 better than Model 2. This fact supports the general understanding that with increasing age the risk of having pododermatitis increases and age should consequently always be considered as confounder to, e.g., body weight which normally increases with age. For this reason Model 3 with **PDmeanmiddle** and including **Age** is preferred.

Age is more important for **PDmeanmiddle** than **PDworst**. This result is support by the edges of the DAGs, the variable importance plot of the random forest and the significant variables of the GLM respectively GLMM, where for Model 4 **Age** is never significant, important or connected to **PDworst**. Indeed this is the case for Model 3, **PDmeanmiddle** is connected to **Age** in the DAG, **Age** is the most important variable in the random forest and it is significant for the GLM and the GLMM. This suggests that the worst pododermatitis score is more likely to be affected by special happenings, whereas **PDmeanmiddle** represents more the general condition of the animal and its disease.

The body weight of the animals also appears as very important variable albeit confounded with age. Likewise there is a stronger connection to **PDmeanmiddle** than to **PDworst**. This positive association can be explained quite easily. Commercially breeding female rabbit are actually always pregnant and/or lactating. Depending if the fertilization of the female rabbits is done artificially or naturally with a male rabbit, the females are

fertilized one respectively ten days after they gave birth. This leads to the fact that these female rabbits are most of the time beyond the normal body weight.

To find out a bit more specifically which variables are additionally important ones for the pododermatitis score, there is quite a big diversity between all the results. However, the models are not straight forward comparable. Where ABN has the liberty to define the model structure by itself, the models for the analysis with the random forest, the GLM and the GLMM are predefined. Nevertheless, we can detect another pattern when looking at the effect sizes of **Claws**. **Claws** seems also to be directly connected to **PDmeanmiddle** but indirectly or generally said less strong to **PDworst**. This conclusion results from looking at the effect sizes of the GLM and at the existing versus non-existing edges in the DAGs. This connection between the pododermatitis score and **Claws** is additionally a good example to remind that all these analysis are only about association but not about causation. This means that although there is an association between longer claws and higher pododermatitis scores and we have a certain effect size for it, we do not know if pododermatitis influences the length of the claws or vice versa. Both directions would have a biologically reasonable justification. Painful pododermatitis lesions could lead to impaired movement of the animals and thus to longer claws or longer claws could lead to an altered weight distribution which might favors pododermatitis.

We could not approve any connection between the condition of the slats (**Holes** and **Gnaw**) and the moisture of the litter (**Wet**) to the severity of pododermatitis. These variables were never directly connected to the pododermatitis score and mostly not even a member of its Markov blanket. The only exception is in the final DAG of the analysis with **PDmeanmiddle** and **Age** where **Wet** and **Gnaw** are parents of shared children from **PDmeanmiddle** and thus in the Markov blanket of **PDmeanmiddle**. The analyses with the random forest as well as with GLM or GLMM gave no hint of an important association between these three variables and the pododermatitis score.

In case that there is a connection between **TotalNoRabbit** and the pododermatitis score, this effect is very week and it is negative. This means that the more rabbits a farmer has, the less pododermatitis they have. Also the link between **Mortality** and pododermatitis has mostly a very week effect size. So these and more detailed and specific observations seem to be a bit far-fetched and would need further investigations to conclude profound and established results.

When looking a bit closer to the several results, unfortunately we realize that some or many important variables which affect pododermatitis may be missing in our data set. Looking at the DAGs the pododermatitis score is almost always at the border of our network. This is not desired, the goal would be to have the target variable in the middle of the network and accordingly exclude all variables which are at the border of the network and do not influence the target variable. The only exception here is again Model 3, where **PDmeanmiddle** is quite integrated in the network and only some variables are not in its Markov blanket, hence not influencing the target variable. The results of the random forests firmly support this conclusion. The percentage of explained variance of our target variable is always very low, indicating that we are not successful at all in explaining and predicting the pododermatitis score. Even the value of our best model, Model 3, is below 50%. Also the high position of the randomly distributed variables in the variable importance plots confirm this conclusion. Additionally the results from the GLM and the GLMM do not provide a great fit of our model to the data.

Coming back to the fact that our data contains a hierarchical structure and looking at the results of the GLMM, it would be interesting to run an ABN with a grouping variable for **Farm**. The GLMM shows that quite a substantial part of the variance could be explained by not recorded differences between the farms. Also the variable importance plot of the random forest show the variable **Farm** at third or fourth place, although always behind a rather big gap. But this indicates as well that just the fact from which farm an animal comes from, helped to explain the pododermatitis score.

Generally said this study gave rise to really interesting facts about pododermatitis. We showed and explained the advantages and disadvantages of using the worst pododermatitis score versus using a mean value. We could also establish profound results about the role of the age and the body weight of the animals for the development of pododermatitis. Additionally this analysis shows some other interesting traces and hints what could be relevant for further studies in the area of pododermatitis. Currently the Animal Welfare Division from the University of Bern is performing a longitudinal study, where some animals, which are also included into this study, are observed and recorded monthly over one year. This study will certainly give rise to some other interesting facts and results.

Additionally the work of this thesis serves as an example how additive Bayesian networks can be used and what advantages and disadvantages exists when working with ABN. This thesis helped to detect some current weaknesses and problems of the R package `abn` as well as of the method of ABN in general. As the R package as well as the method in general are still in research and development and the findings from this thesis were directly delivered to the responsible persons, they help to have soon a more practical, robust and comprehensible version of the R package `abn`. Hopefully all this will lead to a more common use of ABN, as the advantages of this method should not be neglected.

Bibliography

- J.M. Rommers and I.C. de Jong. Three mm wire floors with plastic mats prevent footpad injuries in rabbit does. *Tagungsband der 16. Internationale Tagung über Haltung und Krankheiten der Kaninchen, Pelztier und Heimtiere 2007*, pages 113–119, 2009. [1](#)
- B. Drescher and I. Schlender-Böbbis. Pododermatitis beim Kaninchen. *Kleintierpraxis (Deutschland)*, Band 41:99–104, 1996. [4](#), [5](#)
- L.M. Bigler and H. Oester. Ergebnisse einer Überprüfung der Tiergerechtigkeit eines Haltungssystems für Hauskaninchen-Zuchtgruppen. *Übersichtsaufnahme*, 2003. [9](#)
- C.A. Andrist, B.H.P. van den Borne, L.M. Bigler, T. Buchwalder, and B.A. Roth. Epidemiologic survey in Swiss group-housed breeding rabbits: extent of lesions and potential risk factors. *Preventive veterinary medicine*, 108(2):218–224, 2013. [9](#)
- N. Friedman and D. Koller. Being Bayesian about network structure. A Bayesian approach to structure discovery in Bayesian networks. *Machine learning*, 50(1-2):95–125, 2003. [12](#), [15](#)
- M. Koivisto and K. Sood. Exact Bayesian structure discovery in Bayesian networks. *Journal of Machine Learning Research*, 5(May):549–573, 2004. [12](#), [13](#), [15](#)
- D.J.C. MacKay. Bayesian interpolation. *Neural computation*, 4(3):415–447, 1992. [12](#)
- J. Pearl, D. Geiger, and T. Verma. Conditional independence and its representations. *Kybernetika*, 25(7):33–44, 1989. [14](#)
- M. Pittavino. *Additive Bayesian Networks for Multivariate Data: Parameter Learning, Model Fitting and Applications in Veterinary Epidemiology*. PhD thesis, University of Zurich, 2016. [15](#), [32](#)
- G.F. Cooper and E. Herskovits. A Bayesian method for the induction of probabilistic networks from data. *Machine learning*, 9(4):309–347, 1992. [15](#)
- D. Heckerman, D. Geiger, and D.M. Chickering. Learning Bayesian networks: The combination of knowledge and statistical data. *Machine learning*, 20(3):197–243, 1995. [15](#)
- L. Held and D. Sabanés Bové. Applied statistical inference. *Springer, Berlin Heidelberg*, doi, 10:978–3, 2014. [16](#)
- L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001. [17](#)

- R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2006. URL <http://www.R-project.org>. ISBN 3-900051-07-0. 17
- F.I. Lewis. *abn: Modelling Multivariate Data with Additive Bayesian Networks*, 2016. URL <https://CRAN.R-project.org/package=abn>. R package version 1.0.2. 17
- G. Kratzer, M. Pittavino, F.I. Lewis, and R. Furrer. *abn: an R package for modelling multivariate data using additive Bayesian networks*. URL <http://www.r-bayesian-networks.org/>. Version 1.0.2. 17, 32
- M. Plummer. Jags: A program for analysis of Bayesian graphical models using Gibbs sampling, 2003. URL <http://mcmc-jags.sourceforge.net/>. 17
- A. Liaw and M. Wiener. Classification and Regression by randomForest. *R News*, 2(3): 18–22, 2002. URL <http://CRAN.R-project.org/doc/Rnews/>. 17
- Fortran original by L. Breiman, A. Cutler, R port by A. Liaw, and M. Wiener. *randomForest: Breiman and Cutler's Random Forests for Classification and Regression*, 2015. URL <https://CRAN.R-project.org/package=randomForest>. R package version 4.6-12. 17
- D. Bates, M. Mächler, B. Bolker, and S. Walker. *lme4: Linear Mixed-Effects Models using Eigen and S4*, 2016. URL <https://CRAN.R-project.org/package=lme4>. R package version 1.1-12. 17
- D. Bates, M. Mächler, B. Bolker, and S. Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015. doi: 10.18637/jss.v067.i01. 17
- F.I. Lewis and M.P. Ward. Improving epidemiologic data analyses through multivariate regression modelling. *Emerging themes in epidemiology*, 10(1):4, 2013. 32
- M. Pittavino, A. Dreyfus, C. Heuer, J. Benschop, P. Wilson, J. Collins-Emerson, P. Torgerson, and R. Furrer. Comparison between Generalized Linear Modelling and Additive Bayesian Network. Identification of Factors associated with the Incidence of Antibodies against *Leptospira interrogans* sv Pomona in Meat Workers in New Zealand. *submitted to BMC Veterinary Research*, 2016. 32
- F.I. Lewis and B.JJ. McCormick. Revealing the complexity of health determinants in resource-poor settings. *American journal of epidemiology*, page kws183, 2012. 32

Appendix A

Appendix

Table A.1: Boundaries of 95% credible interval for all parameters of the analysis with PDmeanmiddle, excluding Age.

| Parameter | 2.5% | 50% | 97.5% |
|-------------------------------|--------|--------|--------|
| Weight Mortality | -0.221 | -0.161 | -0.102 |
| Weight PDmeanmiddle | 0.124 | 0.184 | 0.243 |
| Claws Weight | 0.410 | 0.554 | 0.702 |
| Claws PDmeanmiddle | 0.347 | 0.493 | 0.643 |
| Claws Clean | -0.926 | -0.631 | -0.340 |
| Holes Gnaw | 0.321 | 0.378 | 0.434 |
| Holes RH | 0.163 | 0.220 | 0.276 |
| Gnaw RH | -0.343 | -0.284 | -0.226 |
| Gnaw Temp | -0.460 | -0.400 | -0.341 |
| Wet RH | 0.203 | 0.256 | 0.309 |
| Wet Mortality | -0.500 | -0.446 | -0.393 |
| WaterSpillLitt Holes | 4.821 | 5.778 | 6.866 |
| WaterSpillLitt Gnaw | -2.124 | -1.739 | -1.387 |
| WaterSpillLitt RH | 1.495 | 1.850 | 2.227 |
| WaterSpillLitt Temp | -1.730 | -1.414 | -1.121 |
| Temp Weight | -0.184 | -0.134 | -0.083 |
| Temp Wet | 0.405 | 0.462 | 0.519 |
| Temp RH | -0.536 | -0.483 | -0.431 |
| Temp Mortality | 0.347 | 0.403 | 0.460 |
| Experience Claws | -0.272 | -0.231 | -0.191 |
| Experience Holes | 0.233 | 0.268 | 0.302 |
| Experience Gnaw | 0.165 | 0.185 | 0.205 |
| Experience WaterSpillLitt | -1.359 | -1.277 | -1.198 |
| Experience Temp | 0.107 | 0.126 | 0.145 |
| Experience Mortality | 0.068 | 0.090 | 0.112 |
| Experience Clean | 0.141 | 0.184 | 0.227 |
| Experience Moist | 0.064 | 0.105 | 0.145 |
| LittChanInt Weight | -0.032 | -0.022 | -0.012 |
| LittChanInt Holes | -0.362 | -0.347 | -0.333 |
| LittChanInt Gnaw | 0.080 | 0.091 | 0.103 |
| LittChanInt Wet | -0.158 | -0.147 | -0.136 |
| LittChanInt WaterSpillLitt | 0.239 | 0.267 | 0.295 |
| LittChanInt Temp | -0.135 | -0.125 | -0.115 |
| LittChanInt Experience | 0.022 | 0.023 | 0.025 |
| LittChanInt Mortality | -0.262 | -0.250 | -0.237 |
| TotalNoRabbit Claws | -0.063 | -0.055 | -0.048 |
| TotalNoRabbit Holes | 0.286 | 0.293 | 0.299 |
| TotalNoRabbit Gnaw | -0.366 | -0.361 | -0.355 |
| TotalNoRabbit Wet | -0.028 | -0.022 | -0.017 |
| TotalNoRabbit WaterSpillLitt | -0.280 | -0.268 | -0.255 |
| TotalNoRabbit RH | 0.181 | 0.186 | 0.191 |
| TotalNoRabbit Temp | 0.111 | 0.116 | 0.122 |
| TotalNoRabbit Experience | 0.028 | 0.028 | 0.029 |
| TotalNoRabbit Mortality | -0.061 | -0.055 | -0.049 |
| TotalNoRabbit LittChanInt | -0.004 | -0.003 | -0.003 |
| TotalNoRabbit Clean | -0.048 | -0.041 | -0.033 |
| Clean WaterSpillLitt | -1.044 | -0.726 | -0.417 |
| Clean Mortality | 0.139 | 0.287 | 0.436 |
| Clean Moist | 1.297 | 1.584 | 1.877 |

Table A.2: Boundaries of 95% credible interval for all parameters of the analysis with PDworst, excluding Age.

| Parameter | 2.5% | 50% | 97.5% |
|-------------------------------|--------|--------|--------|
| Claws Weight | 0.471 | 0.613 | 0.759 |
| Claws Clean | -0.910 | -0.622 | -0.340 |
| Holes Gnaw | 0.319 | 0.375 | 0.431 |
| Holes RH | 0.161 | 0.217 | 0.273 |
| Gnaw RH | -0.341 | -0.282 | -0.224 |
| Gnaw Temp | -0.462 | -0.403 | -0.345 |
| WaterSpillLitt Holes | 4.736 | 5.672 | 6.734 |
| WaterSpillLitt Gnaw | -2.106 | -1.726 | -1.378 |
| WaterSpillLitt RH | 1.481 | 1.834 | 2.207 |
| WaterSpillLitt Temp | -1.746 | -1.431 | -1.137 |
| Temp Weight | -0.236 | -0.182 | -0.128 |
| Temp Wet | 0.230 | 0.285 | 0.339 |
| Temp RH | -0.465 | -0.410 | -0.356 |
| Temp PDworst | 0.085 | 0.138 | 0.192 |
| Experience Claws | -0.273 | -0.232 | -0.192 |
| Experience Holes | 0.231 | 0.265 | 0.300 |
| Experience Gnaw | 0.165 | 0.185 | 0.205 |
| Experience WaterSpillLitt | -1.353 | -1.272 | -1.193 |
| Experience Temp | 0.108 | 0.128 | 0.147 |
| Experience Mortality | 0.070 | 0.092 | 0.114 |
| Experience Clean | 0.155 | 0.197 | 0.239 |
| Experience Moist | 0.058 | 0.098 | 0.138 |
| Mortality Wet | -0.618 | -0.566 | -0.514 |
| Mortality RH | 0.302 | 0.357 | 0.411 |
| Mortality Temp | 0.351 | 0.405 | 0.459 |
| LittChanInt Weight | -0.032 | -0.022 | -0.012 |
| LittChanInt Holes | -0.361 | -0.346 | -0.331 |
| LittChanInt Gnaw | 0.080 | 0.092 | 0.103 |
| LittChanInt Wet | -0.157 | -0.146 | -0.135 |
| LittChanInt WaterSpillLitt | 0.239 | 0.267 | 0.295 |
| LittChanInt Temp | -0.135 | -0.125 | -0.115 |
| LittChanInt Experience | 0.022 | 0.023 | 0.025 |
| LittChanInt Mortality | -0.262 | -0.249 | -0.237 |
| TotalNoRabbit Claws | -0.065 | -0.058 | -0.050 |
| TotalNoRabbit Holes | 0.285 | 0.292 | 0.298 |
| TotalNoRabbit Gnaw | -0.364 | -0.359 | -0.353 |
| TotalNoRabbit Wet | -0.027 | -0.022 | -0.016 |
| TotalNoRabbit WaterSpillLitt | -0.281 | -0.268 | -0.255 |
| TotalNoRabbit RH | 0.181 | 0.186 | 0.191 |
| TotalNoRabbit Temp | 0.111 | 0.117 | 0.123 |
| TotalNoRabbit Experience | 0.028 | 0.028 | 0.029 |
| TotalNoRabbit Mortality | -0.060 | -0.054 | -0.048 |
| TotalNoRabbit LittChanInt | -0.003 | -0.003 | -0.003 |
| TotalNoRabbit Clean | -0.051 | -0.043 | -0.035 |
| Clean WaterSpillLitt | -1.066 | -0.750 | -0.443 |
| Clean Mortality | 0.142 | 0.288 | 0.435 |
| Clean Moist | 1.275 | 1.559 | 1.848 |

Table A.3: Boundaries of 95% credible interval for all parameters of the analysis with PDmeanmiddle, including Age.

| Parameter | 2.5% | 50% | 97.5% |
|-----------------------------|--------|--------|--------|
| Weight Age | 0.373 | 0.442 | 0.511 |
| Weight Clean | 0.212 | 0.362 | 0.513 |
| Holes RH | -0.366 | -0.291 | -0.218 |
| Gnaw Claws | -0.365 | -0.274 | -0.184 |
| Gnaw Experience | 0.204 | 0.216 | 0.228 |
| Wet Age | 0.136 | 0.211 | 0.287 |
| RH Weight | -0.245 | -0.171 | -0.098 |
| RH Wet | 0.236 | 0.309 | 0.383 |
| Temp Age | -0.369 | -0.301 | -0.235 |
| Temp Holes | -0.518 | -0.446 | -0.375 |
| Temp RH | -0.441 | -0.375 | -0.309 |
| Temp Mortality | -0.401 | -0.329 | -0.257 |
| Temp PDmeanmiddle | 0.139 | 0.205 | 0.271 |
| Temp Clean | -0.494 | -0.355 | -0.218 |
| Experience Holes | -0.348 | -0.311 | -0.274 |
| Experience RH | -0.385 | -0.355 | -0.325 |
| Experience Temp | -0.306 | -0.275 | -0.243 |
| Experience Mortality | -0.312 | -0.275 | -0.239 |
| Mortality Holes | -0.450 | -0.407 | -0.365 |
| Mortality Wet | -0.801 | -0.758 | -0.716 |
| Mortality RH | 0.276 | 0.320 | 0.364 |
| LittChanInt Weight | -0.069 | -0.055 | -0.042 |
| LittChanInt Holes | -0.205 | -0.185 | -0.165 |
| LittChanInt Gnaw | 0.126 | 0.148 | 0.170 |
| LittChanInt Wet | -0.161 | -0.141 | -0.121 |
| LittChanInt RH | -0.137 | -0.117 | -0.097 |
| LittChanInt Temp | -0.206 | -0.189 | -0.172 |
| LittChanInt Experience | -0.045 | -0.038 | -0.030 |
| LittChanInt Mortality | -0.348 | -0.324 | -0.300 |
| TotalNoRabbit Age | 0.027 | 0.033 | 0.039 |
| TotalNoRabbit Weight | 0.026 | 0.031 | 0.036 |
| TotalNoRabbit Claws | 0.025 | 0.036 | 0.046 |
| TotalNoRabbit Holes | 0.083 | 0.091 | 0.099 |
| TotalNoRabbit Gnaw | -0.212 | -0.206 | -0.199 |
| TotalNoRabbit Wet | 0.073 | 0.084 | 0.094 |
| TotalNoRabbit RH | 0.161 | 0.169 | 0.176 |
| TotalNoRabbit Temp | 0.141 | 0.148 | 0.155 |
| TotalNoRabbit Mortality | 0.073 | 0.085 | 0.096 |
| TotalNoRabbit LittChanInt | -0.003 | -0.003 | -0.003 |
| TotalNoRabbit PDmeanmiddle | -0.020 | -0.015 | -0.010 |
| TotalNoRabbit Moist | -0.039 | -0.029 | -0.020 |
| PDmeanmiddle Age | 0.149 | 0.232 | 0.314 |
| PDmeanmiddle Claws | 0.152 | 0.316 | 0.481 |
| PDmeanmiddle Mortality | 0.100 | 0.173 | 0.247 |
| Moist Clean | 1.081 | 1.435 | 1.794 |

Table A.4: Boundaries of 95% credible interval for all parameters of the analysis with PDworst, including Age.

| Parameter | 2.5% | 50% | 97.5% |
|----------------------------|--------|--------|--------|
| Weight Age | 0.373 | 0.442 | 0.511 |
| Weight Clean | 0.217 | 0.367 | 0.516 |
| Holes RH | -0.365 | -0.291 | -0.217 |
| Gnaw Claws | -0.366 | -0.276 | -0.186 |
| Gnaw Experience | 0.204 | 0.216 | 0.228 |
| Wet Age | 0.137 | 0.212 | 0.287 |
| RH Weight | -0.244 | -0.170 | -0.098 |
| RH Wet | 0.235 | 0.308 | 0.381 |
| Temp Age | -0.312 | -0.248 | -0.183 |
| Temp Holes | -0.518 | -0.445 | -0.374 |
| Temp RH | -0.443 | -0.375 | -0.309 |
| Temp Mortality | -0.382 | -0.310 | -0.238 |
| Temp PDworst | 0.107 | 0.171 | 0.235 |
| Temp Clean | -0.472 | -0.333 | -0.196 |
| Experience Holes | -0.348 | -0.310 | -0.274 |
| Experience RH | -0.384 | -0.354 | -0.324 |
| Experience Temp | -0.307 | -0.275 | -0.244 |
| Experience Mortality | -0.312 | -0.275 | -0.239 |
| Mortality Holes | -0.450 | -0.406 | -0.364 |
| Mortality Wet | -0.799 | -0.756 | -0.714 |
| Mortality RH | 0.274 | 0.318 | 0.363 |
| LittChanInt Weight | -0.071 | -0.057 | -0.044 |
| LittChanInt Holes | -0.206 | -0.186 | -0.166 |
| LittChanInt Gnaw | 0.123 | 0.145 | 0.168 |
| LittChanInt Wet | -0.159 | -0.139 | -0.119 |
| LittChanInt RH | -0.140 | -0.120 | -0.100 |
| LittChanInt Temp | -0.215 | -0.198 | -0.181 |
| LittChanInt Experience | -0.045 | -0.037 | -0.030 |
| LittChanInt Mortality | -0.352 | -0.328 | -0.304 |
| LittChanInt PDworst | 0.019 | 0.032 | 0.045 |
| TotalNoRabbit Age | 0.025 | 0.031 | 0.037 |
| TotalNoRabbit Weight | 0.025 | 0.030 | 0.035 |
| TotalNoRabbit Claws | 0.024 | 0.034 | 0.045 |
| TotalNoRabbit Holes | 0.088 | 0.096 | 0.104 |
| TotalNoRabbit Gnaw | -0.209 | -0.202 | -0.196 |
| TotalNoRabbit Wet | 0.076 | 0.087 | 0.097 |
| TotalNoRabbit RH | 0.163 | 0.171 | 0.179 |
| TotalNoRabbit Temp | 0.146 | 0.153 | 0.160 |
| TotalNoRabbit Mortality | 0.078 | 0.089 | 0.101 |
| TotalNoRabbit LittChanInt | -0.003 | -0.003 | -0.003 |
| TotalNoRabbit PDworst | -0.028 | -0.023 | -0.018 |
| TotalNoRabbit Moist | -0.035 | -0.025 | -0.016 |
| Moist Clean | 1.084 | 1.436 | 1.792 |

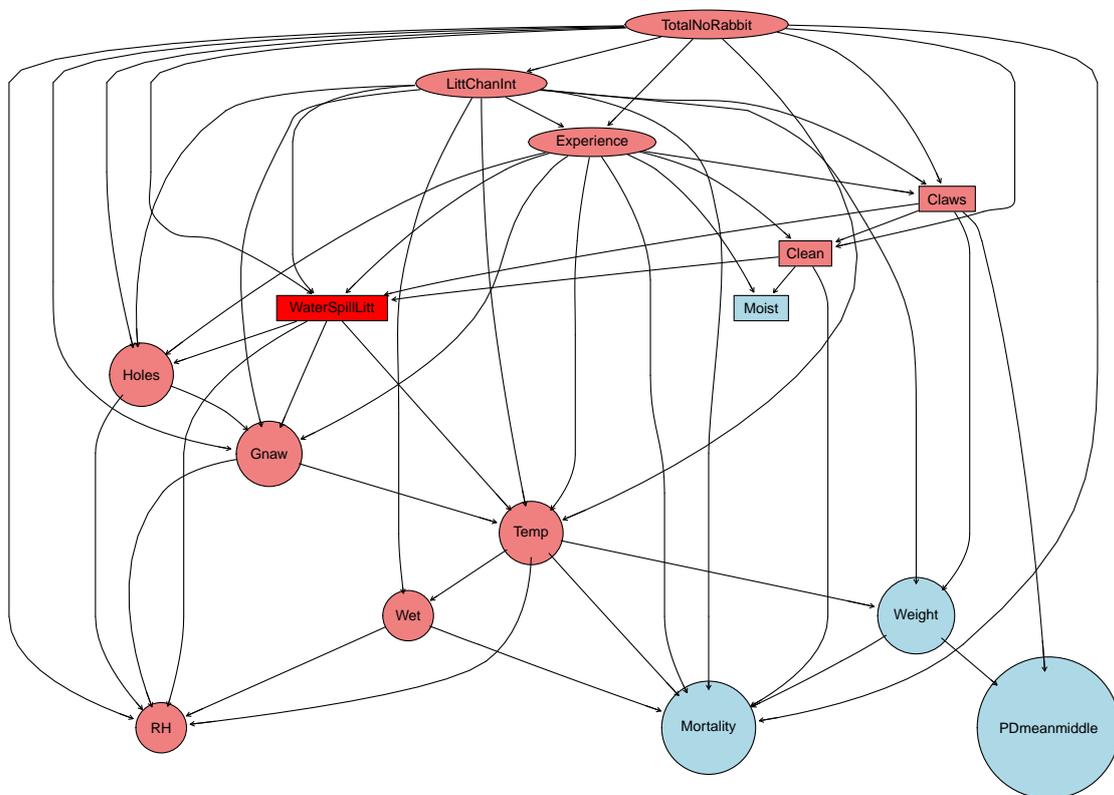


Figure A.1: Optimal DAG from the analysis without Age, $n = 1042$. The Markov blanket of the node WaterSpillLitt is indicated in red.

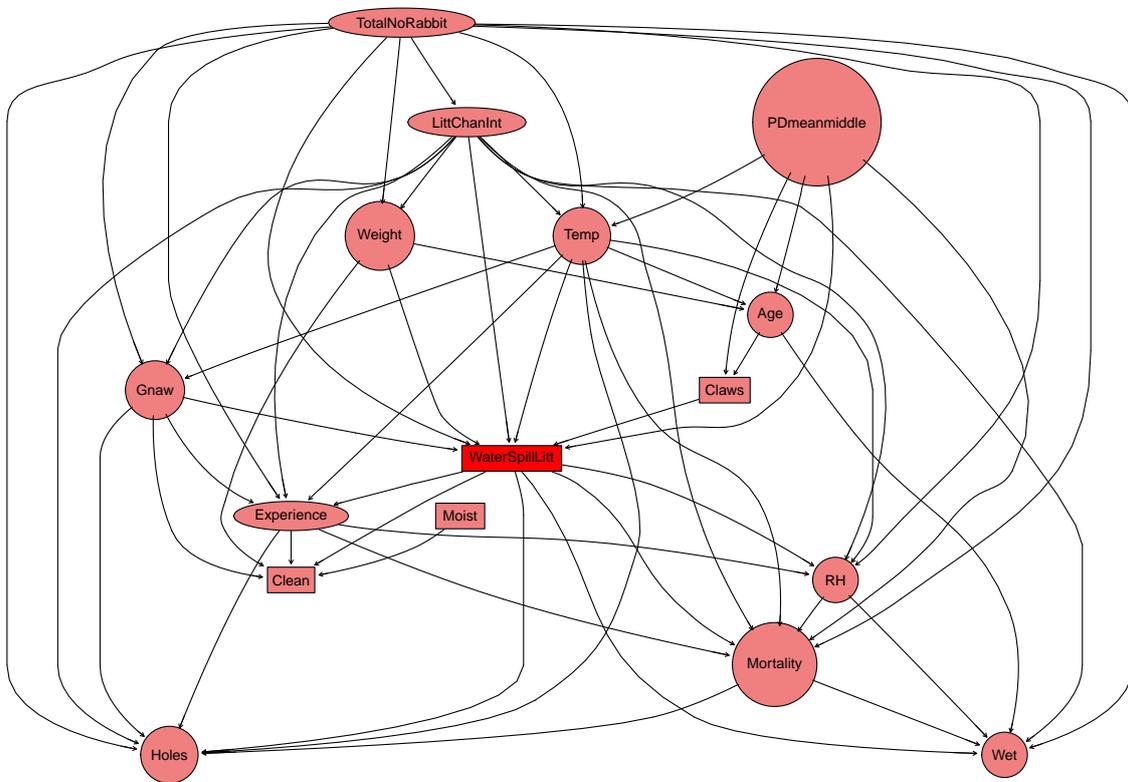


Figure A.2: Optimal DAG from the analysis with Age, $n = 649$. The Markov blanket of the node WaterSpillLitt is given in red.

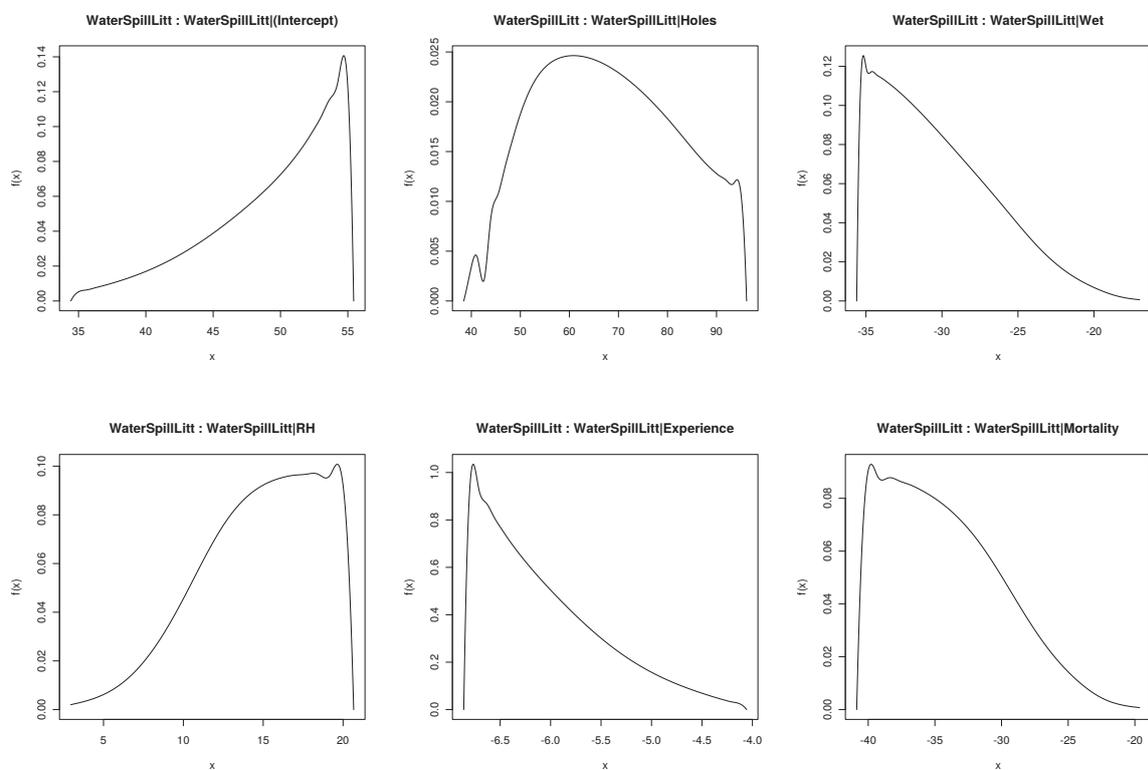


Figure A.3: Strange marginal posterior densities of `WaterSpillLitt` after inclusion of `Age` into the model.