Miniproject: Birthweight of babies

Søren Højsgaard Created: November 7, 2011

1 Birthweight of babies

TOPICS: Linear normal model

The `births` data in the `Epi` package contains recordings of birtweights of babies in a London hospital together with recordings of several possible explanatory variables.

The purpose of this exercise is to establish a model for understanding how the birthweight is related to the explanatory variables.

```
library(Epi)
data(births)
head(births)
```

```
  id bweight lowbw gestwks preterm matage hyp sex
1  1    2974     0   38.52       0     34   0   2
2  2    3270     0      NA      NA     30   0   1
3  3    2620     0   38.15       0     35   0   2
4  4    3751     0   39.80       0     31   0   1
5  5    3200     0   38.89       0     33   1   1
6  6    3673     0   40.97       0     33   0   2
```

```
dim(births)
```

```
[1] 500   8
```

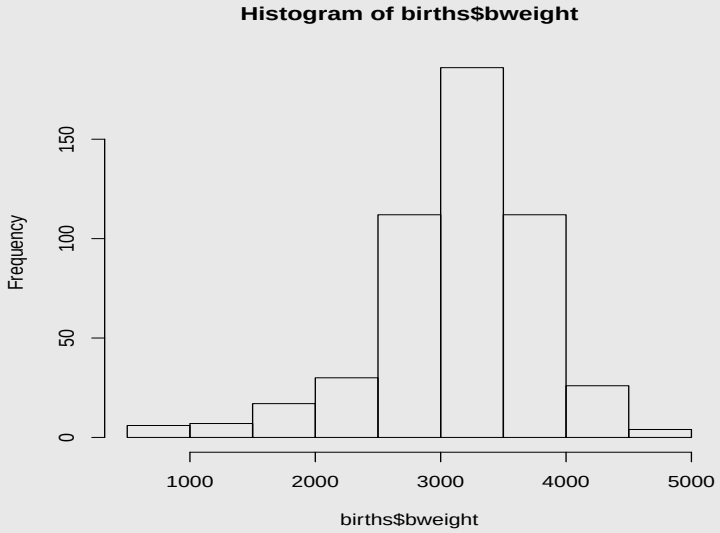
```
sapply(births, class)
```

```
       id   bweight     lowbw   gestwks   preterm    matage       hyp       sex
"numeric" "numeric" "numeric" "numeric" "numeric" "numeric" "numeric" "numeric"
```

1. (a) Plot the birth weights of babies and observe the distribution. Comments?

Solution:

```
hist(births$bweight)
```



(b) See how birth weight of babies are related to gestation period, and maternal age by scatter plot. Differentiate metween males and females in the plot. (Hint: `xyplot()` could be your friend.) Comments?

(c) Make a scatterplot of all pair of variables. (Hint: `pairs()` could be your friend). Comments?

2. (a) Assume that number of boys follows a $bin(N = 500, \theta)$–distribution.

(b) Estimate θ from observed data. Estimate the variance of $\hat{\theta}$ from data.

(c) What is the probability of observing a male child?

(d) What is the probability of observing at least 26 babies of the same sex in 44 births? (Use above estimated value of θ.)

3. Fit a linear normal model M_1 for $bweight \sim gestwks + matage$ to the `births` data.

(a) Write down in precise mathematical terms what your model is.

(b) Write down in your own words what the assumptions behind a linear normal model are. Investigate the residuals of M_1. Comment on whether you find that these assumptions are likely to be satisfied for these data.

(c) Write down in your own words what a least squares fit of a model to data is.

(d) From model M_1 report the estimated parameters and their standard deviations.

(e) In model M_1, calculate the values of fitted values (use function `predict()`) and plot the fitted line over the observations.

(f) Test the effect of each term in the model M_1 using the function `drop1()`.

(g) Predict the birth weights at weeks $10, 15, 20, 25, 30, 35, 40, 45, 50$ for women of age 20, 30 and 40. Include confidence intervals in your prediction. Present the results as a table and as illustrative graphics. Comment on these predictions.

4. Fit an interaction model M_2 for $bweight \sim gestwks + matage + gestwks * matage$ to the `births` data.
5. Write down M_2 in precise mathematical terms.
6. Investigate whether M_2 fits well to data.
7. Test whether the additive model is more appropriate or the interactive model.
8. Fit a model M_3 that differentiate the two slope for sexes.
9. Write down M_3 in precise mathematical terms.
10. Investigate whether M_3 fits well to data.
11. Investigate whether it is necessary to have two slopes for each sex by comparing the models M_1 and M_3.