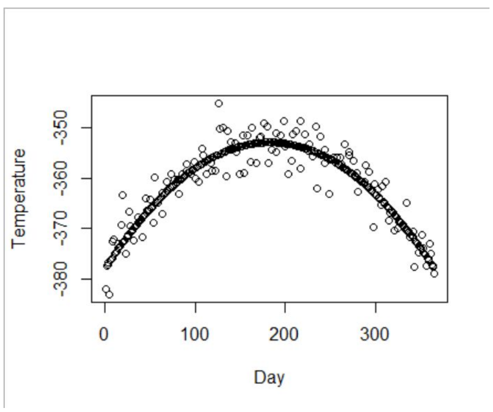
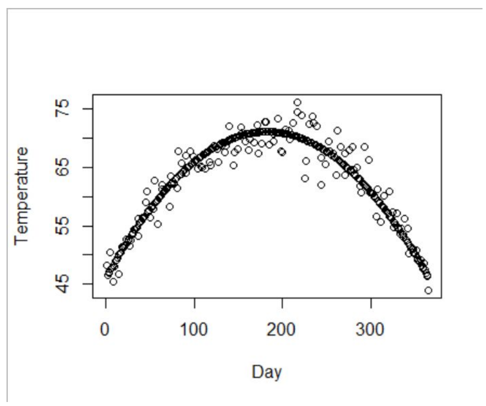
Lab 2

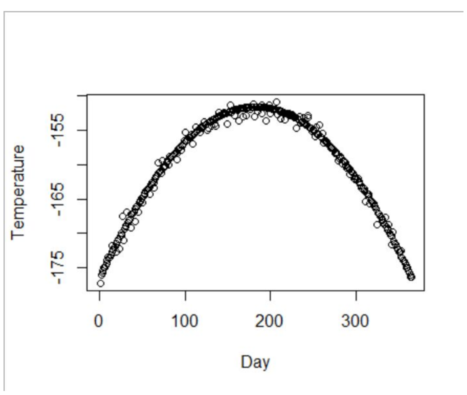
1. Linear and polynomial regression

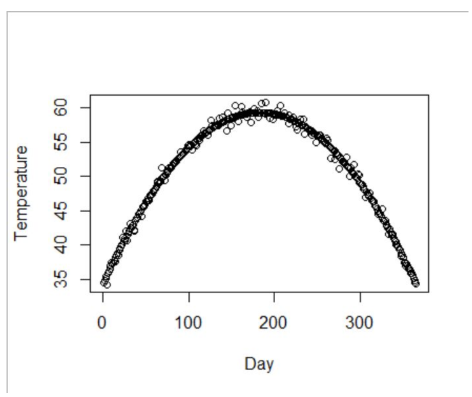
a)
Since we are measuring the temperature in sweden over the year, starting at the first of january, our prior beliefs are that the temperature at this time should be around negative ten degrees celcius. Over the year the temperature should then rise to peak around 25-30 degrees, and then fall back to the start value around negative ten degrees. This should follow a negative second degree curve.

From the given values in this exercise this was the outcome of some draws from the prior:





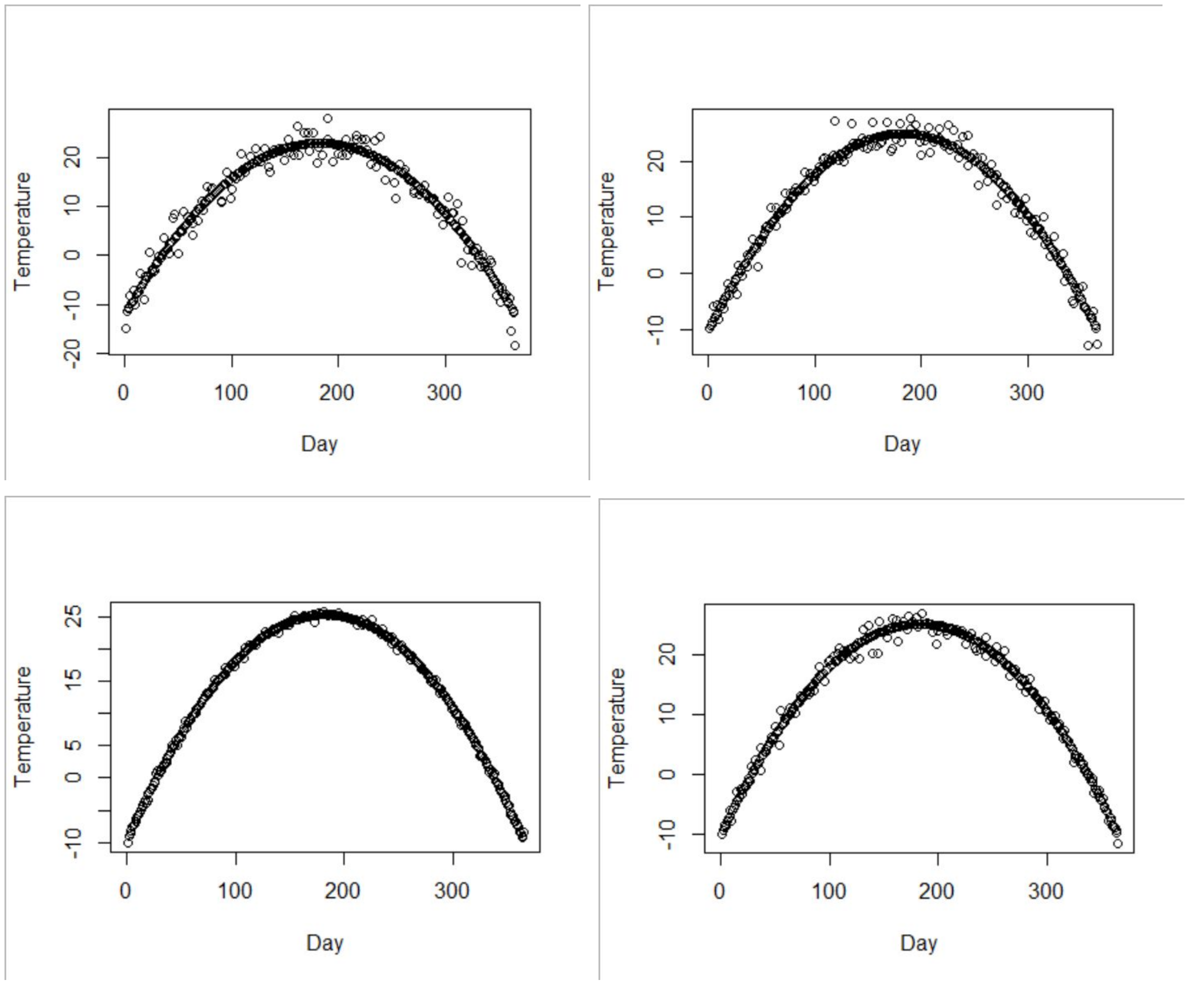




As seen in the picture the general shape curve of this prior did match our beliefs. However, the omega0 value was way too small, leading to an unreasonable high variance. This resulted in temperatures above 100 degrees and below -100.

Therefore we increased the precision (omega0) to 0.5 because we had such strong beliefs that the temperature could not vary this much. We also changed the mean for Beta to get a slightly larger range of values to include temperatures above 20 degrees.

Our parameters was set to the following, due to the reasoning above: means: (-10,140,-140). Omega = 0.5*I
The simulation result are shown below:



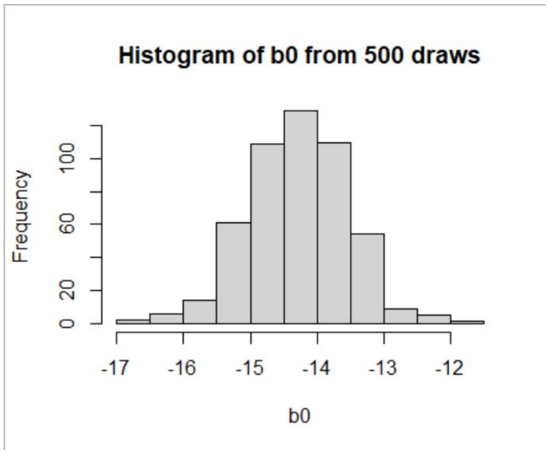
PICTURE

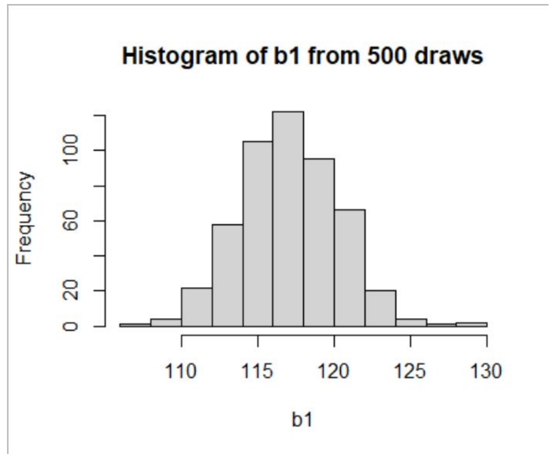
b)

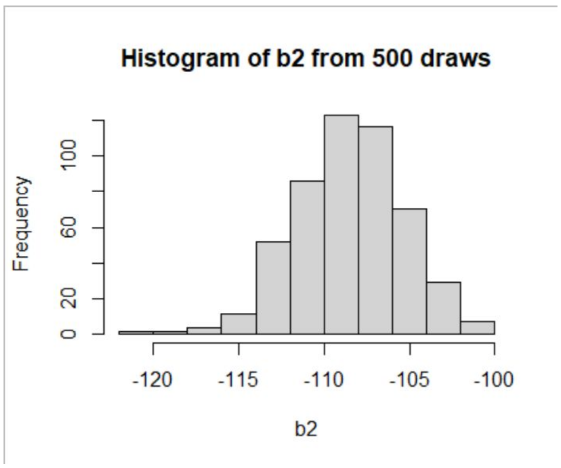
We draw samples from the marginal posterior distributions

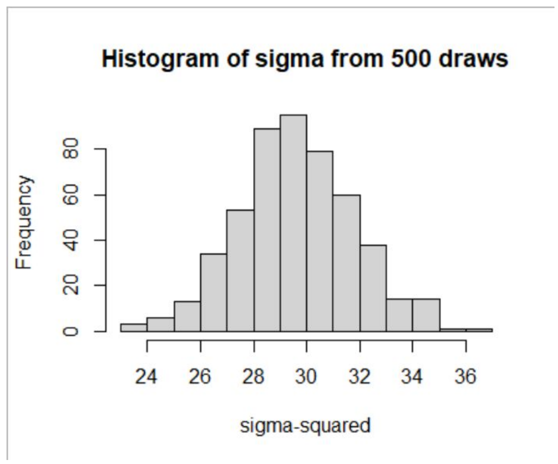
$$\beta|\sigma^2, \mathbf{y} \sim N\left[\mu_n, \sigma^2\Omega_n^{-1}\right]$$
$$\sigma^2|\mathbf{y} \sim Inv - \chi^2\left(\nu_n, \sigma_n^2\right)$$

The resulting histograms for the parameters, estimating the probability density functions, are shown below.





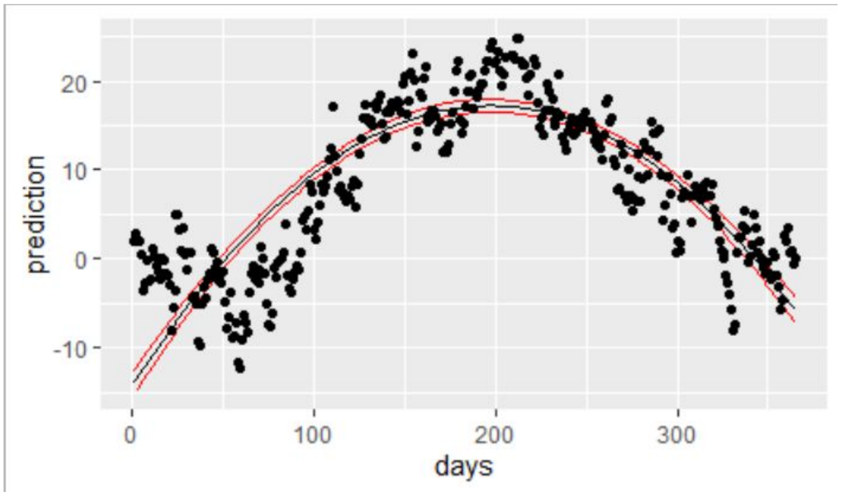




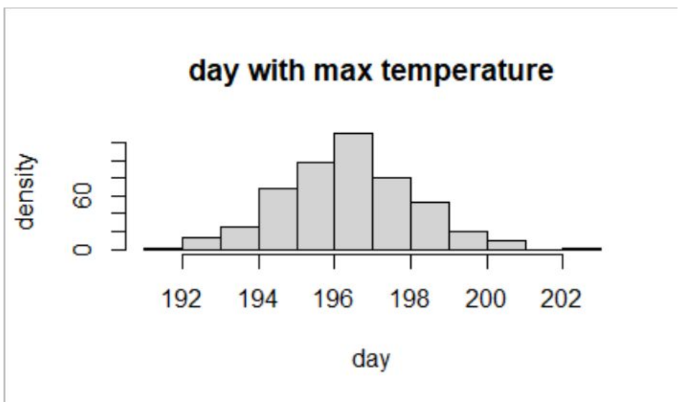
By saving all predicted y values for each draw of the 500 simulations of the beta posterior, we can estimate the ETI for each x value from 500 predictions of this x value. Therefor we have calculated the ETI for each f(time). We overlay this as red lines in the plot.

The posterior median is found by taking the median for each beta from our 500 simulations and using these to predict the y values from x.

The confidence interval curves do not contain most of the data points but neither should they. This is because we are not computing the confidence intervals för the data points but rather for each of the trend lines from every simulation. So if we were to look at the computed trend lines for every simulation, most of them would fit within our ETI.



c)



Median: 197
The posterior distribution of x tilder is calculated by saving the x-value corresponding to the max y value for each simulation of **B**. The distribution is displayed in the histogram above.

d)
When choosing a prior to meet the criterias in d) our solution is to still use a Normal distributed joint prior for B and sigma squared on the form:

$$\beta|\sigma^2 \sim N\left(\mu_0, \sigma^2\Omega_0^{-1}\right)$$

We keep the same parameters for the first three betas from before, to preserve the previous prior beliefs of the curve shape and only adding the higher degree polynomials to the model.

We set the mean vector for the higher order beta values to 0. We also increase the precision for each higher order beta value. When the variance decreases, the higher the likelihood that the beta value is close to 0. This can be done by specifying omega as a unit matrix with its diagonal units increasing in size. For instance, linear, or exponential, depending on which rate you wish to punish the betas corresponding to higher degrees. We chose for simplicity's sake to increase the prevision linearly.

My0 = [-10, 140, -140, 0, 0, 0, 0, 0]
Omega0 =

0.5	0	0	0	0	0	0	0
0	0.5	0	0	0	0	0	0
0	0	0.5	0	0	0	0	0
0	0	0	0.7	0	0	0	0
0	0	0	0	0.9	0	0	0
0	0	0	0	0	1.1	0	0
0	0	0	0	0	0	1.3	0
0	0	0	0	0	0	0	1.5

2. Posterior approximation for classification with logistic regression

a)

Multivariate distribution for B

$$\beta|\mathbf{y}, \mathbf{X} \sim N\left(\tilde{\beta}, J_{\mathbf{y}}^{-1}(\tilde{\beta})\right),$$

Table of posterior 8-dim parameter vector, Beta mode

b0	b1	b2	b3	b4	b5	b6	b7
0.62672884	-0.01979113	0.18021897	0.16756670	-0.14459669	-0.08206561	-1.35913317	-0.02468351

Table of J^{-1}(theta)

	V1	V2	V3	V4	V5	V6	V7	V8
1	2.266022568	3.338861e-03	-6.545121e-02	-1.179140e-02	0.0457807243	-3.029345e-02	-0.1887483542	-0.0980239285
2	0.003338861	2.528045e-04	-5.610225e-04	-3.125413e-05	0.0001414915	-3.588562e-05	0.0005066847	-0.0001444223
3	-0.065451206	-5.610225e-04	6.218199e-03	-3.558209e-04	0.0018962893	-3.240448e-06	-0.0061345645	0.0017527317
4	-0.011791404	-3.125413e-05	-3.558209e-04	4.351716e-03	-0.0142490853	-1.340888e-04	-0.0014689508	0.0005437105
5	0.045780724	1.414915e-04	1.896289e-03	-1.424909e-02	0.0555786706	-3.299398e-04	0.0032082535	0.0005120144
6	-0.030293450	-3.588562e-05	-3.240448e-06	-1.340888e-04	-0.0003299398	7.184611e-04	0.0051841611	0.0010952903
7	-0.188748354	5.066847e-04	-6.134564e-03	-1.468951e-03	0.0032082535	5.184161e-03	0.1512621814	0.0067688739
8	-0.098023929	-1.444223e-04	1.752732e-03	5.437105e-04	0.0005120144	1.095290e-03	0.0067688739	0.0199722657

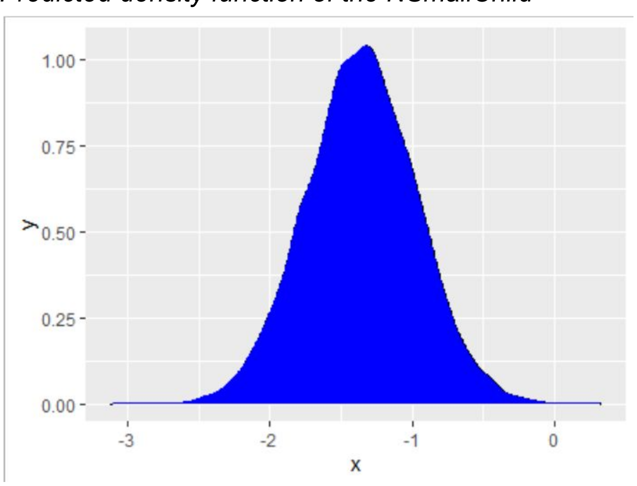
As seen from the table of the posterior 8-dim parameter vector, Beta mode, the absolute value of the seventh coefficient, corresponding to the NSmallChild column, is by far the largest of all Betas. Furthermore, when looking at the predicted density function of the NSmallChild parameter the upper 97,5 quantile has the absolute value of 0.6, only slightly below the second to largest absolute value in the posterior Beta parameter vector (~0.62). This indicates that it has a large influence on whether the mother works or not. This is also reasonable in logical terms since a mother of a young child should be more likely to be on maternity leave (aka mammaledig).

ETI interval values for Beta(7), NSmallChild

95% ETI

[-2.12, -0.60]

Predicted density function of the NSmallChild



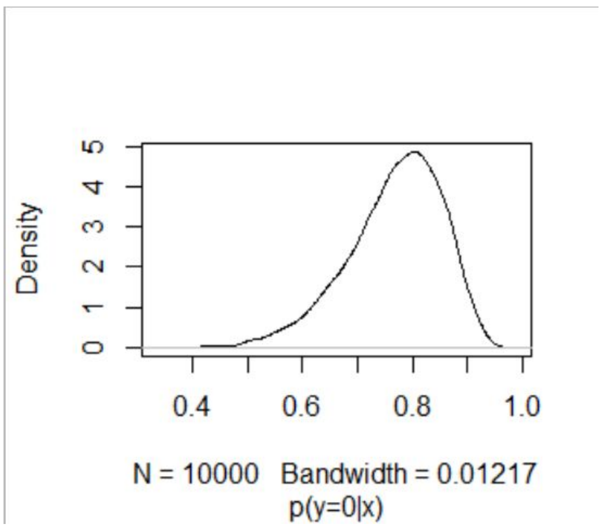
b)

We draw n samples of beta from the approximated normal posterior distribution given above. The Work data given in the exercise resulted in the sample vector **x** = (1, 10, 8, 10, 1, 40, 1, 1). These n beta draws and the sample vector is plugged in to the probability function below, giving us n probabilities simulating the predicted likelihood of the woman not working, p(y=0|**x**). From this simulation, it is reasonable to make the prediction that this woman is not working. There is only 0.5% likelihood of her working given this distribution.

Probability function

$$Pr(y = 1|x) \; = \; \frac{1}{1 + exp\,(x^{T}B)}$$

*Simulated predicted likelihood of p(y=0|**x**)*

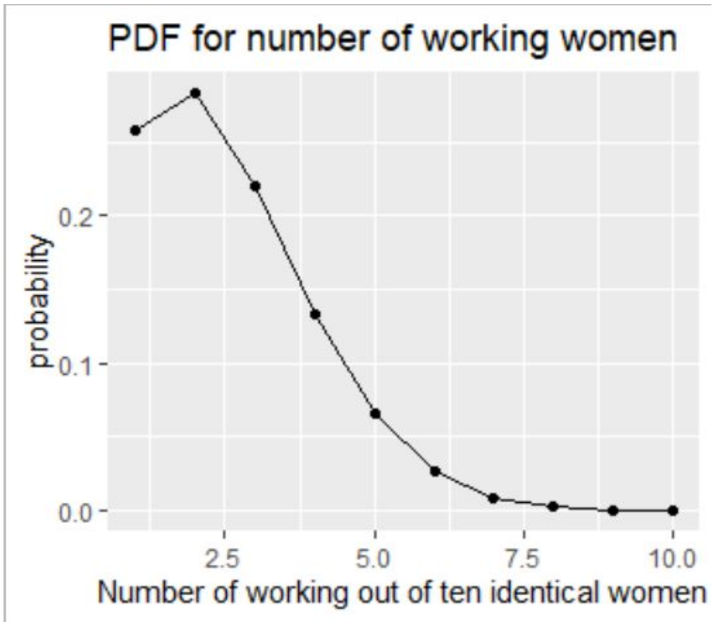


2 c)

For each sample of beta we calculate the probability p(y=1) by $\frac{exp\,(x^{T}B)}{1 + exp\,(x^{T}B)}$. This probability is used with the binomial distribution to calculate the probability of j women working out of 10, where j is (1:10). ( ) We do this by using the binomial pdf for 10 draws and probability p(y=1).

$$\binom{10}{j} p^{j} * (1-p)^{10-j}$$

When all probabilities for each p and each j are calculated, all probabilities for j women working are summed and divided by n. We then normalize over all j probabilities. The resulting PDFfor j women working out of 10 is shown below.



PDF for number of working women
probability
0.0
0.1
0.2
2.5
5.0
7.5
10.0
Number of working out of ten identical women