Analyze the relation between HDL, gender, BMI, and age using the log-normal distribution.

Is age associated with HDL?

From a crude linear regression model, the log of the age parameter was -0.0007164, (95% CI: -0.001705, 0.0002666). The 95% confidence interval contains the null value of 0, so age was not associated with HDL in this model.

Using a multivariate linear regression model adjusting for sex and BMI, the log of the age parameter was 0.0000409, (95% CI: -0.0007952, 0.0008769). The 95% confidence interval contains the null value of 0, so age was not associated with HDL in this model either.

Is BMI associated with HDL after you adjust for sex?

Using a multivariate linear regression model adjusting for sex, the log of the BMI parameter was -0.01642, (95% CI: -0.0183913, -0.0145566). The 95% confidence interval does not contain the null value of 0, so BMI is associated with HDL in this model.

Use the model with sex and BMI to answer these questions:

1. What is the sex effect?
2. What is the BMI effect?
3. What is the predicted HDL in a woman with BMI equal to the average BMI of the sample?
4. What is the predicted HDL in a woman with BMI=25 and a man with BMI=25?

$$log(\beta_{sex}) = 0.2400$$
$$\beta_{sex} = e^{0.2400} = 1.2712$$

Comparing those of the same BMI, males have 1.2712 times the HDL levels as females.

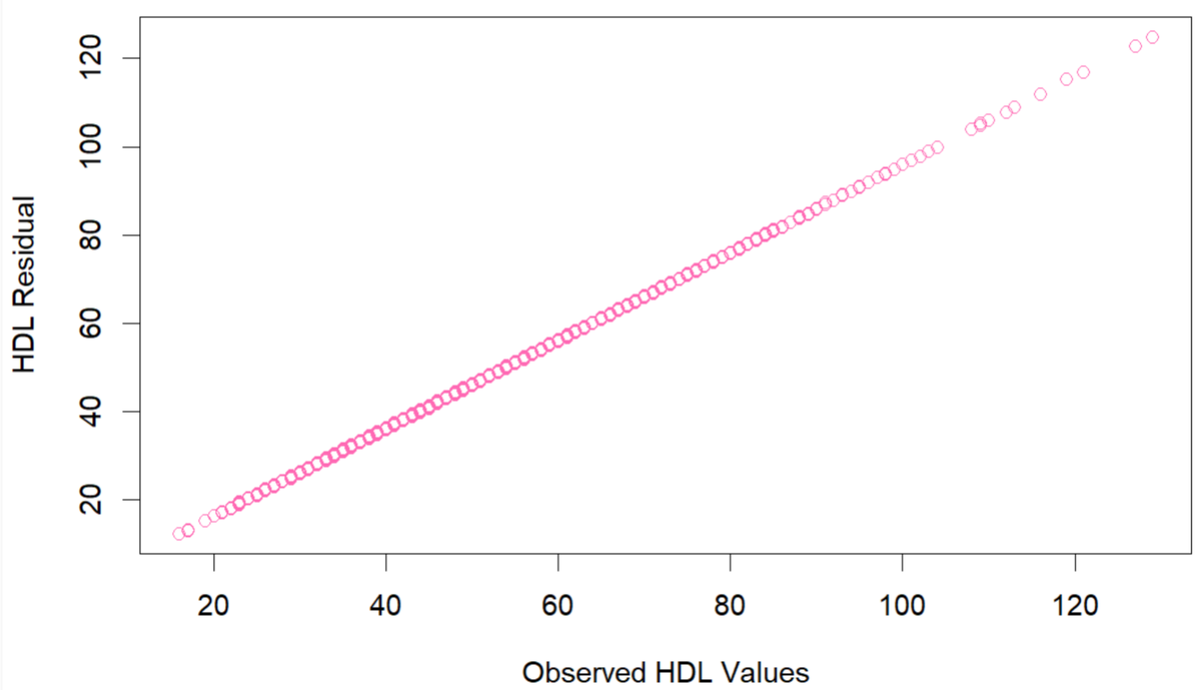
$$log(\beta_{BMI}) = 0.01642$$
$$\beta_{BMI} = e^{0.01642} = 1.0165$$

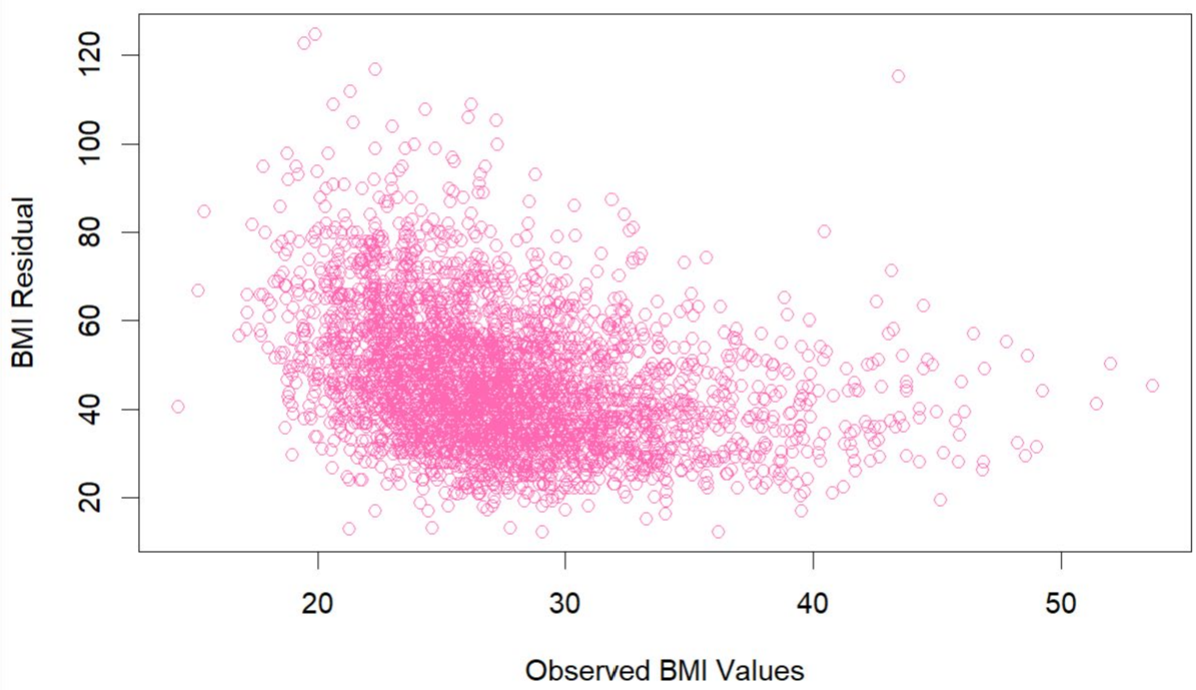
Comparing those of the same sex, for each additional unit of BMI increases HDL levels by a factor of 1.0166.

A woman with a BMI of 27.3349, which is the average BMI of the sample, will have a predicted HDL level of 42.37, (95% CI: 41.8181, 42.8849).

A woman with a BMI of 25 will have a predicted HDL level of 44.04, (95% CI: 43.4276, 44.6167). A male with a BMI of 25 will have a predicted HDL level of 56.02 (95% CI: 55.3732, 56.6469)

Produce diagnostic plots of residuals and comment on the goodness of fit of the model that relates HDL to sex and BMI.





The residuals for HDL fall almost perfectly on a positively sloped line. The larger the HDL observed, the larger the residual. The residuals for BMI are less biased but are all positive instead of being randomly distributed around the x-axis. The model that assumes a lognormal distribution of HDL does not fit the predicted values that well because the residuals follow a pattern.

Forced expiratory volume (FEV, in liter/sec) were measured in 654 pediatric patients, and age (in years, between 3 and 19 years), height (in inches), sex (1=female, 2=male) and smoking status (1=yes, 0=no) of patients were also recorded. The following model was implemented in JAGS:

```
```{r}
M0 <- "model{
  for (i in 1:N){
        fev[i] ~ dlnorm(mu[i], tau)
        mu[i] <- b[1] +
                 b[2]*(age[i]-mean(age[])) +
                 b[3]*(height[i]-mean(height[])) +
                 b[4]*smoke[i] +
                 b[5]*(sex[i]-1)
                 b[6]*smoke[i]*(sex[i]-1)
        }

#Prior Distribution
for (j in 1:6){b[j] ~ dnorm(0, 0.01)}
tau ~ dgamma(1,1)

#Parameters
parameter[1] <- exp(b[4])
parameter[2] <- exp(b[4] + b[6])
parameter[3] <- exp(b[4] + b[3]*(60-mean(height[])) + b[4])
parameter[4] <- exp(b[4] + b[3]*(60-mean(height[])) + b[4] + b[5] + b[6])
}
"
```
```

What does the parameter[1]=exp(b[4]) represent?

Parameter1 represents the OR of a female smoker's FEV compared to a female nonsmoker's FEV, assuming they're the same age and height.

What does the parameter[2]=exp(b[4]+b[6]) represent?

Parameter2 represents the OR of a male smoker's FEV compared to a male nonsmoker's FEV, assuming they're the same age and height.

The model was fitted using Gibbs Sampling, and the summary of the Bayesian estimates is:

```
##
## Iterations = 1001:10000
## Thinning interval = 1
## Number of chains = 1
## Sample size per chain = 9000
##
## 1. Empirical mean and standard deviation for each variable,
##    plus standard error of the mean:
##
##                   Mean       SD  Naive SE Time-series SE
## b[1]         0.816652 0.053197 5.607e-04      8.524e-03
## b[2]         0.023380 0.003507 3.697e-05      1.150e-04
## b[3]         0.042772 0.001780 1.877e-05      5.249e-05
## b[4]         0.044111 0.027901 2.941e-04      4.333e-03
## b[5]         0.026659 0.083881 8.842e-04      1.517e-02
## b[6]         0.001249 0.043468 4.582e-04      8.173e-03
## parameter[1] 1.045505 0.029157 3.073e-04      4.520e-03
## parameter[2] 1.046952 0.033909 3.574e-04      3.902e-03
## parameter[3] 2.252834 0.058573 6.174e-04      9.170e-03
## parameter[4] 2.316936 0.072311 7.622e-04      7.530e-03
##
## 2. Quantiles for each variable:
##
##                   2.5%      25%        50%     75%   97.5%
## b[1]          0.717886  0.77895  0.8165231 0.85477 0.91495
## b[2]          0.016588  0.02101  0.0233379 0.02574 0.03040
## b[3]          0.039280  0.04158  0.0427529 0.04395 0.04637
## b[4]         -0.008492  0.02450  0.0442286 0.06362 0.09646
## b[5]         -0.133133 -0.03144  0.0304162 0.08862 0.17682
## b[6]         -0.076720 -0.03074 -0.0008728 0.03124 0.08505
## parameter[1]  0.991544  1.02480  1.0452213 1.06569 1.10126
## parameter[2]  0.982963  1.02314  1.0456863 1.07011 1.11582
## parameter[3]  2.144132  2.21054  2.2524737 2.29465 2.36528
## parameter[4]  2.176714  2.26592  2.3162371 2.36679 2.46101
```

How many samples were generated in the MCMC?

9000 samples were generated by Markov chain Monte Carlo simulations.

Test the null hypothesis that the interaction between sex and smoking is 0, versus the alternative hypothesis that it is different from 0.

Using a multivariate linear regression model, the log of the interaction term was -0.0008728, (95% CI: -0.076720, 0.08505). The 95% confidence interval contains the null value of 0, so the interaction between smoking and sex was not associated with FEV in this model.

The data was analyzed again using a model without the interaction term between sex and smoking:

````
```{r}
M1 <- "model{
  for (i in 1:N){
        fev[i] ~ dlnorm(mu[i], tau)
        mu[i] <- b[1] +
                 b[2]*(age[i]-mean(age[])) +
                 b[3]*(height[i]-mean(height[])) +
                 b[4]*smoke[i] +
                 b[5]*(sex[i]-1)
        }

#Prior Distribution
for (j in 1:6){b[j] ~ dnorm(0, 0.01)}
tau ~ dgamma(1,1)

#Parameters
parameter[1] <- exp(b[4])
parameter[2] <- exp(b[1] + b[3]*(60-mean(height[])) + b[4])
parameter[3] <- exp(b[1] + b[3]*(60-mean(height[])) + b[4] + b[5])
}
"
```
````

9000 iterations of the Gibbs sampling after 1000 burn-ins produced the following summary statistics:

```
##
## Iterations = 1001:10000
## Thinning interval = 1
## Number of chains = 1
## Sample size per chain = 9000
##
## 1. Empirical mean and standard deviation for each variable,
##    plus standard error of the mean:
##
##                   Mean       SD  Naive SE Time-series SE
## b[1]           0.80914 0.045019 4.745e-04      0.0061261
## b[2]           0.02343 0.003553 3.745e-05      0.0001098
## b[3]           0.04281 0.001788 1.884e-05      0.0000503
## b[4]           0.04805 0.023428 2.469e-04      0.0031832
## b[5]           0.02909 0.012130 1.279e-04      0.0002760
## b[6]          -0.19611 9.943543 1.048e-01      0.1048141
## parameter[1]   1.04951 0.024539 2.587e-04      0.0033280
## parameter[3]   2.24452 0.050393 5.312e-04      0.0059905
## parameter[4]   2.31083 0.054109 5.704e-04      0.0059647
##
## 2. Quantiles for each variable:
##
##                     2.5%      25%      50%     75%   97.5%
## b[1]            0.721435  0.77920  0.80777 0.83704  0.90710
## b[2]            0.016542  0.02104  0.02340 0.02578  0.03050
## b[3]            0.039279  0.04163  0.04281 0.04399  0.04633
## b[4]           -0.001818  0.03334  0.04842 0.06399  0.09321
## b[5]            0.005562  0.02091  0.02899 0.03750  0.05291
## b[6]          -19.862706 -7.03173 -0.18497 6.56585 19.20373
## parameter[1]    0.998183  1.03390  1.04962 1.06608  1.09769
## parameter[3]    2.148537  2.21022  2.24239 2.27596  2.35164
## parameter[4]    2.209987  2.27338  2.30881 2.34623  2.42432
```

What is the effect of smoking on FEV?

$$log(\beta_{smoke}) = 0.04842$$
$$\beta_{smoke} = e^{0.04842} = 1.0496$$

Comparing those of the same age, height, and sex, smokers FEV are 1.0496 times larger than the FEV of nonsmokers.

What is the effect of age on FEV?

$$log(\beta_{age}) = 0.02340$$
$$\beta_{age} = e^{0.02340} = 1.0237$$

Comparing those of the same height, smoking status, and sex, each additional year in age increases FEV by a factor of 1.0237.

What is the predicted FEV of a male, smoker, of average age, and 60 inches height?

A 60-inch smoking man of average age will have a predicted FEV of 2.30881 (95% CI: 2.209987, 2.42432).

What is the predicted FEV of a female, smoker, of average age, and 60 inches height?

A 60-inch smoking woman of average age will have a predicted FEV of 2.24239 (95% CI: 2.148537, 2.35164).