[1] "volume" "girth" "height"

The girths are in centimetres but all the other data are in metres, so we convert the girths to metres at the outset:

```
girth <- girth / 100
```

Now fit the model:

```
model1 <- glm(log(volume) ~ log(girth) + log(height))
summary(model1)
```

Coefficients:
```
            Estimate  Std. Error t value  Pr(>|t|)
(Intercept)  -2.89938    0.63767  -4.547 9.56e-05 ***
log(girth)    1.98267    0.07503  26.426  < 2e-16 ***
log(height)   1.11714    0.20448   5.463  7.83e-06 ***
```

Residual deviance: 0.18555 on 28 degrees of freedom
AIC: -62.697

The estimates are reasonably close to expectation (1.11714 rather than 1.0 for log(h) and 1.98267 rather than 2.0 for log(g)).

Now we shall use `offset` to specify the theoretical response of log(v) to log(h); i.e. a slope of 1.0 rather than the estimated 1.11714:

```
model2 <- glm(log(volume) ~ log(girth) + offset(log(height)))
summary(model2)
```

Coefficients:
```
            Estimate  Std. Error t value  Pr(>|t|)
(Intercept)  -2.53419    0.01457 -174.00 < 2e-16 ***
log(girth)    2.00545    0.06287  31.88 < 2e-16 ***
```

Residual deviance: 0.18772 on 29 degrees of freedom
AIC: -64.336

Naturally the residual deviance is greater, but only by a very small amount. The AIC has gone down from -62.697 to -64.336, so the model simplification was justified.

Let us try including the theoretical slope (2.0) for log(g) in the offset as well:

```
model3 <- glm(log(volume) ~ 1 + offset(log(height) + 2*log(girth)))
summary(model3)
```

Coefficients:
```
            Estimate  Std. Error t value  Pr(>|t|)
(Intercept)  -2.53403    0.01421 -178.31 < 2e-16 ***
```

Residual deviance: 0.18777 on 30 degrees of freedom
AIC: -66.328

Again, the residual deviance is only marginally greater, and AIC is smaller, so the simplification is justified.

What about the intercept? If our theoretical model of cylindrical logs is correct then the intercept should be

```
log(1/(4*pi))
```

[1] -2.531024
This is almost exactly the same as the intercept estimated by GLM in model3, so we are justified in putting the entire model in the offset and informing GLM not to estimate an intercept from the data ($y \sim -1$):

```r
model4 <- glm(log(volume) ~ offset(log(1/(4*pi))+log(height)+2*log(girth))-1)
summary(model4)
```

No Coefficients

Residual deviance: 0.18805 on 31 degrees of freedom
AIC: -68.282

This is a rather curious model with no estimated parameters, but it has a residual deviance of just 0.18805 (compared with model1, where all three parameters were estimated from the data, which had a deviance of 0.18555). Because we were saving one degree of freedom with each step in the procedure, AIC became smaller with each step, justifying all of the model simplifications.

**Residuals**

After fitting a model to data, we should investigate how well the model describes the data. In particular, we should look to see if there are any systematic trends in the goodness of fit. For example, does the goodness of fit increase with the observation number, or is it a function of one or more of the explanatory variables? We can work with the raw residuals:

\[
\text{residuals} = \text{response variable} - \text{fitted values}
\]

With normal errors, the identity link, equal weights and the default scale factor, the raw and standardized residuals are identical. The standardized residuals are required to correct for the fact that with non-normal errors (like count or proportion data) we violate the fundamental assumption that the variance is constant (p. 389) because the residuals tend to change in size as the mean value the response variable changes.

For **Poisson** errors, the standardized residuals are

\[
\frac{y - \text{fitted values}}{\sqrt{\text{fitted values}}}
\]

For **binomial** errors they are

\[
\frac{y - \text{fitted values}}{\sqrt{\text{fitted values} \times \left[1 - \frac{\text{fitted values}}{\text{binomial denominator}}\right]}}
\]

where the binomial denominator is the size of the sample from which the $y$ successes were drawn. For **Gamma** errors they are

\[
\frac{y - \text{fitted values}}{\text{fitted values}}
\]