Generalized Linear (GLM) and Additive Modeling (GAM) with R

An Online Course
Presented by Geoffrey S. Hubona

GLM/GAM with R: Day 1 Agenda: Linear Modeling

Linear Modeling with R:

- What are Linear Models?
- Response and Predictor Variables
- Residual Sums of Squares
- Properties of β :
    - Expected value
    - Variance
- Linear Modeling Example with R: ***How Old is the Universe ?***
    - Predictor coefficient
    - Confidence intervals
    - Testing hypotheses
- Standard (Omnibus) Linear Modeling Functions in R
- Linear Models in General: Transitioning to GLMs

GLM/GAM with R: Day 1 Agenda: GLMs

Generalized Linear Models (GLMs) with R:

- What are GLMs?
  - Error Structure
  - Linear Predictor
  - Link Function
- Count Data
  - As Proportions
  - As Frequencies
- Proportion Data
  - Binomial Errors
  - Binomial Model Example with R: **Likelihood of Heart Disease**
- Count Data as Frequencies
  - Poisson Errors
  - Poisson Model Example with R: **The Spread of a Global Epidemic**

Day 2 material below dotted line

- Categorical Data
  - Binary Response Variables
  - Contingency Table Counts or Data as Proportions: **Two Examples in R**

Linear Modeling with R

What are Linear Models ?

- Consider n observations, x_i, y_i, where y_i is an observation on random variable , Y_i , with expectation:

$$\mu_i \equiv E(Y_i) \qquad (1)$$

- Suppose that an appropriate model for the relationship between x and y is:

$$Y_i = \mu_i + \varepsilon_i \text{ where } \mu_i = x_i \beta \quad (2)$$

- Here β is an unknown parameter and the ε_i are mutually independent zero mean random variables, each with the same variance σ^2.

Response and Predictor Variables

- So the model says that Y is given by x multiplied by a constant plus a random term.
- Y is an example of a ***response variable***, while x is an example of a ***predictor variable***.
- How can β be estimated from the x_i , y_i data? One approach is to choose a value of β that makes the model fit closely to the data.
- So we must choose a measure that defines how well, or how badly, a model with a particular β fits the data.

Residual Sums of Squares

- One possible such measure is the residual sums of squares of the model:

$$s = \sum_{i=1}^{n} (y_i - \mu_i)^2 = \sum_{i=1}^{n} (y_i - x_i\beta)^2 \qquad (3)$$

- When we have chosen a good value of β, close to the true value, then the model-predicted μ_i should be relatively close to the y_i so the s should be small.
- The method of ***least squares***: β can be *estimated* by minimizing s with respect to β.
- So we have an ***estimated parameter value***, $\hat{\beta}$, which we will use as a proxy for the true parameter, β.

Desirable Properties for $\hat{\beta}$

- The estimator, $\hat{\beta}$, is a **random variable** and so we can discuss its distribution.
- To evaluate the ***reliability*** of the least squares estimate $\hat{\beta}$ it is useful to consider the ***sampling properties*** of $\hat{\beta}$.
- The ***expected value*** of $\hat{\beta}$ is:

$$E(\hat{\beta}) = E\left(\sum_{i=1}^{n} x_i Y_i / \sum_{i=1}^{n} x_i^2\right) = \sum_{i=1}^{n} x_i E(Y_i) / \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i^2 \beta / \sum_{i=1}^{n} x_i^2 = \beta$$

- So $\hat{\beta}$ is ***unbiased***, it is an estimator that 'gets it right on average', but *how good is any one particular estimate likely to be ?*

Desirable Properties for $\hat{\beta}$

- From general probability theory, if $Y_1, Y_2, \ldots, Y_n$ are *independent* random variables and $a_1, a_2, \ldots, a_n$ are real constants then: $\operatorname{var}\left(\sum_i a_i Y_i\right) = \sum_i a_i^2 \operatorname{var}(Y_i)$
- But we know: $\hat{\beta} = \sum_i a_i Y_i$ where $a_i = x_i / \sum_i x_i^2$
- And from the original model specification:

$$\operatorname{var}(\hat{\beta}) = \sum_i x_i^2 \Big/ \left(\sum_i x_i^2\right)^2 \sigma^2 = \left(\sum_i x_i^2\right)^{-1} \sigma^2$$

- So we can say that: $\hat{\sigma}^2 = \frac{1}{n-1} \sum_i \left(y_i - x_i \hat{\beta}\right)^2$
- This gives us an ***unbiased*** estimate of the variance of $\hat{\beta}$

How Old is the Universe?

- The big-bang model implies that the universe expands uniformly according to Hubble's law:

$$y = \beta x$$

- Where y is the relative velocity of any two galaxies separated by distance x, and β is "Hubble's constant".
- β^{-1} is the approximate age of the universe, but β is unknown and must be estimated from observations of x and y.

Dating the Cosmos with R

- We can use the **`lm()`** function in R to calculate the age of the universe.
- The use the Cepheid distance – velocity data for 24 galaxies stored in the data frame **`hubble`**.

```
> library(gamair) # contains 'hubble'
> data(hubble)
> hub.mod <- lm(y~x-1,data=hubble)
> summary(hub.mod)
Call:
lm(formula = y ~ x - 1, data = hubble)
Residuals:
   Min      1Q Median      3Q     Max
-736.5 -132.5  -19.0  172.2  558.0
Coefficients:
  Estimate Std. Error t value Pr(>|t|)
x   76.581      3.965   19.32 1.03e-15 ***
```

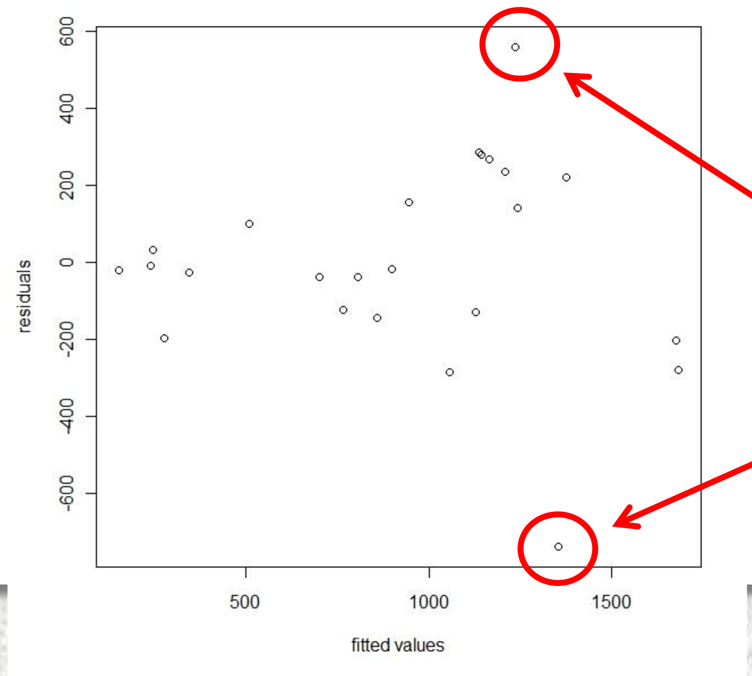
Do not include and intercept term

Beta coefficient

Must Check Assumptions

- Check if ε_i are independent and all have equal variance using plot of residuals against fitted values.

```
> plot(fitted(hub.mod),residuals(hub.mod),xlab="fitted values",ylab="residuals")
```



Suggests violation of constant variance

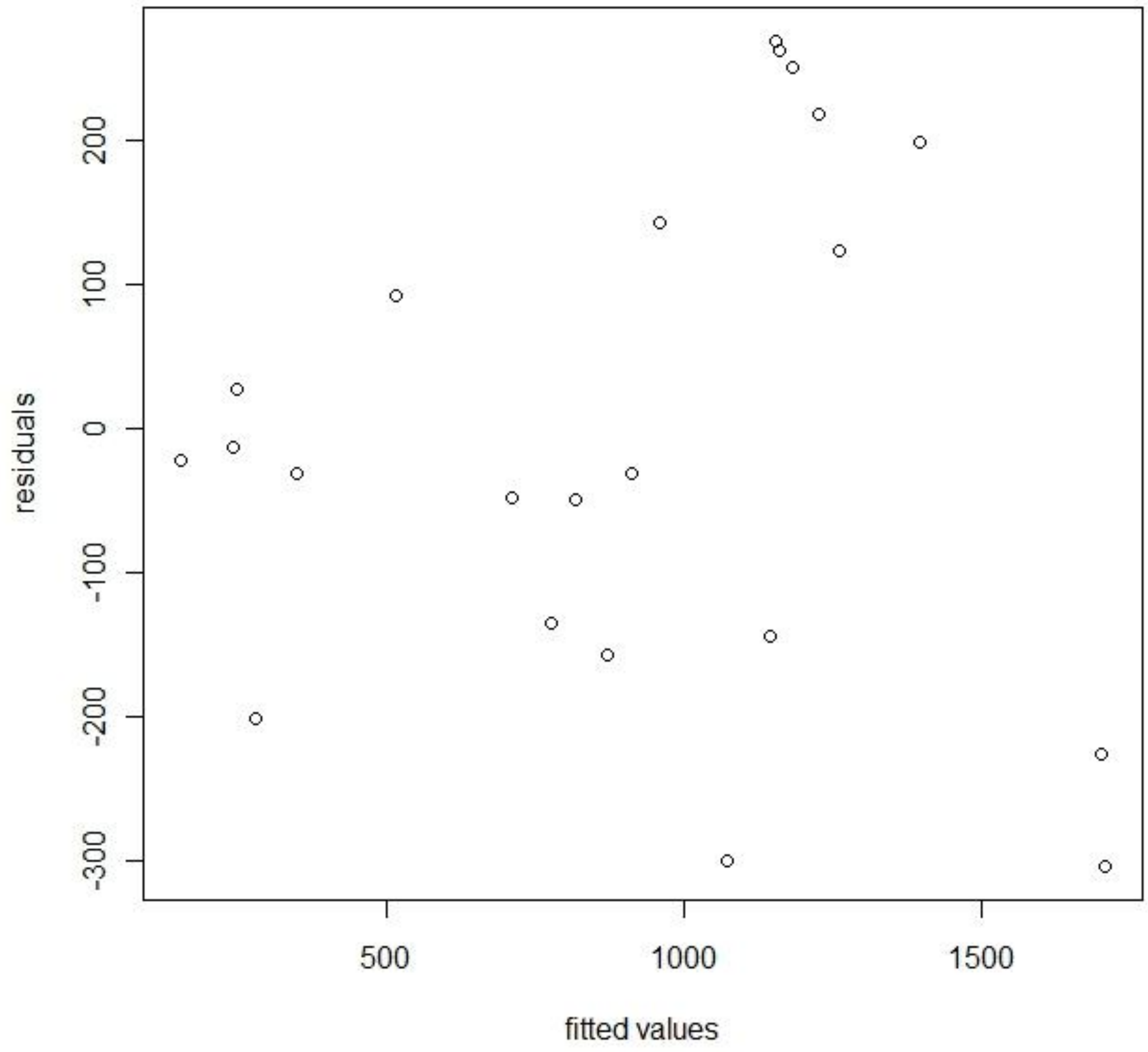
Large magnitude residuals;
Funnel shape is a problem.

Fit Same Model to New Data Set

```
> hub.mod1 <- lm(y~x-1,data=hubble[-c(3,15),])
> summary(hub.mod1)

Call: lm(formula = y ~ x - 1, data = hubble[-c(3, 15), ])

Residuals:
    Min     1Q Median     3Q    Max
-304.3 -141.9  -26.5  138.3  269.8

Coefficients:
  Estimate Std. Error t value Pr(>|t|)
x    77.67       2.97   26.15   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> plot(fitted(hub.mod1),residuals(hub.mod1),xlab="fittedvalues",ylab="residuals")
```

New Beta coefficient

Omit Observations # 3 and # 15



How Old is the Universe?

```
> hubble.const <- c(coef(hub.mod),coef(hub.mod1))/3.09e19

> age <- 1/hubble.const

> age/(60^2*24*365)

12794692825 12614854757
```

Two age estimates:
Around 13 Billion Years

The Hubble constant estimates have units of (km)s^{-1} $(Mpc)^{-1}$. A Mega-parsec Is 3.09×10^{19}km, so we need to divide β by this amount, in order to obtain Hubble's constant with units of s^{-1}. The approximate age of the universe, in seconds, is then given by the reciprocal of β.

Adding a Distributional Assumption

- To find confidence intervals for β or to test hypotheses, we need to make an additional distributional assumption. We have assumed that:

$\varepsilon_i \sim N\left(0, \sigma^2\right)$ for all i, which implies $Y_i \sim N\left(x_i\beta, \sigma^2\right)$

- Further, we know that $\hat{\beta}$ is the weighted sum of Y_i which we assume to be a random normal variable.
- So the estimator $\hat{\beta}$ must also be a random normal variable. So:

$$\hat{\beta} \sim N\left(\beta, \left(\sum x_i^2\right)^{-1} \sigma^2\right)$$

Testing Hypotheses About β

- We want to test that the age of the universe is only 6,000 years. This implies that $\beta = 163\times10^6$.
- We empirically examine the probability, or ***p-value***, that we would have observed our value for $\hat{\beta}$ if the true value was actually $\beta = 163\times10^6$.
- R code to evaluate the **p-value** for H_0: ***the Hubble constant is 163000000***.

Distribution function for student's t-value

```
> cs.hubble <- 163000000
> t.stat<-(coef(hub.mod1)-
+ cs.hubble)/summary(hub.mod1)$coefficients[2]
> pt(t.stat,df=21)*2 # multiply by 2 because want |T|
             x
3.906388e-150 # This is the t-stat for H0
```

Confidence Intervals

- What ***range of values*** for β would be consistent with the proposition that the universe is only 6,000 years old?
- The **R** function **`qt()`** can be used to find these ranges:

```
qt(c(0.25, 0.975),df=21) # returns the range of the middle 95%

-------------------------------------------------

> sigb <- summary(hub.mod1)$coefficients[2]
> h.ci<-coef(hub.mod1)+qt(c(0.025,0.975),df=21)*sigb
> h.ci
[1] 71.49588 83.84995
> h.ci<-h.ci*60^2*24*365.25/3.09e19 # convert to 1/years
> sort(1/h.ci) 71.49588 83.84995
[1] 11677548698 13695361072
```

95% CI for β

95% CI for age of universe

Some Standard Linear Modeling Functions

Function in R	Description of Function in R
lm	Estimates a linear model by least squares. Returns a fitted model of class **`lm`** containing parameter estimates plus other auxiliary results for use by other functions.
plot	Produces model checking plots from a fitted model object.
summary	Produces summary information about a fitted model, including parameter estimates, associated standard errors, p-values, r^2 etc.
anova	Used for model comparison based on F-ratio testing.
AIC	Extract Akaike's information criterion for a model fit.
residuals	Extract an array of model residuals from a fitted model.
fitted	Extract an array of fitted values from a fitted model object.
predict	Obtain predicted values from a fitted model, either for new values of the predictor variables, or for the original values. Standard errors of the predictions can also be returned.

Linear Models in General

- We can generalize the simple linear model by allowing the response variable to depend on multiple predictor variables (plus an additive constant).
- The extra predictors can be transformations of the original predictors. Here are some examples:

(1) $$\mu_i = \beta_0 + x_i\beta_1, Y_i = \mu_i + x_i\varepsilon_i,$$

(2) $$\mu_i = \beta_0 + x_i\beta_1 + x_i^2\beta_2 + x_i^3\beta_3,$$

(3) $$\mu_i = \beta_0 + x_i\beta_1 + z_i\beta_2 + \log(x_i z_i)\beta_3,$$

Linear Models in General

(1) $\mu_i = \beta_0 + x_i\beta_1, Y_i = \mu_i + x_i\varepsilon_i,$

(2) $\mu_i = \beta_0 + x_i\beta_1 + x_i^2\beta_2 + x_i^3\beta_3,$

(3) $\mu_i = \beta_0 + x_i\beta_1 + z_i\beta_2 + \log(x_i z_i)\beta_3,$

- Each of these is a linear model because the ε_j terms and the model parameters, β_j , enter the model in a linear way.
- But the predictor variables can enter the model non-linearly.
- Like the simple model, the parameters of these models can be estimated by finding the β_j values which make the models best fit the observed data in the sense of minimizing $\sum_i (y_i - \mu_i)^2$.