The Likelihood Ratio Test

The likelihood ratio test is a general purpose test designed evaluate nested statistical models in a way that is strictly analogous to the F-test for reduced models (RM) and full models (FM) commonly employed with linear models (see *Biostatistics* Worksheet 402). In both, failure to reject the null hypothesis results in model simplification. The likelihood ratio test works not only with linear models, shown here, but may be applied to a very wide array of problems involving Geralized Linear Models (GLM), where Maximum Likelihood (ML) or Restricted Maximum Likelihood (REML) methods are utilized to estimate model parameters. The latter methods/models include, among others, Logistic Regression (see *GLM* 020), Poisson Regression (see *GLM* 040), and Linear Mixed Models (see *LMM* 060) described in Worksheets on the *Biologist's Analytic Toolkit Website* under *Statisistical Models*. For direct comparison of results, the data set analyzed here is the same as for the general F test (*Biostatisics* Worksheet 402). As can be seen, F test and likelihood ratio tests give similar but not exactly the same results. Helpful discussion of this approach appears in Kuter et al. (KNNL) *Applied Linear Statistical Models* 5th Edition, and numerous statistics websites.

Example in R:

```
#LOG LIKELIHOOD AND LIKELIHOOD RATIO TEST
setwd("c:/DATA/Models/")
K=read.table("KNNLCh9SurgicalUnit.txt")
K
attach(K)
```

```
> K
    X1 X2  X3   X4 X5 X6     Y
1  6.7 62  81 2.59 50  0 6.544
2  5.1 59  66 1.70 39  0 5.999
3  7.4 57  83 2.16 55  0 6.565
4  6.5 73  41 2.01 48  0 5.854
5  7.8 65 115 4.30 45  0 7.759
6  5.8 38  72 1.42 65  1 5.852
7  5.7 46  63 1.91 49  1 6.250
8  3.7 68  81 2.57 69  1 6.619
9  6.0 67  93 2.50 58  0 6.962
10 3.7 76  94 2.40 48  0 6.875
11 6.3 84  83 4.13 37  0 6.613
12 6.7 51  43 1.86 57  0 5.549
13 5.8 96 114 3.95 63  1 7.361
...
44 6.5 56  77 2.85 41  0 6.288
45 3.4 77  93 1.48 69  0 6.178
46 6.5 40  84 3.00 54  1 6.416
47 4.5 73 106 3.05 47  1 6.867
48 4.8 86 101 4.10 35  1 7.170
49 5.1 67  77 2.86 66  1 6.365
50 3.9 82 103 4.55 50  0 6.983
51 6.6 77  46 1.95 50  0 6.005
52 6.4 85  40 1.21 58  0 6.361
53 6.4 59  85 2.33 63  0 6.310
54 8.8 78  72 3.20 56  0 6.478
```

Fitting Full and Reduced linear models:

Full Model:

```
#FITTING THE FULL LINEAR MODEL
FM=lm(Y~X1+X2+X3+X4+X5+factor(X6))
FMg=glm(Y~X1+X2+X3+X4+X5+factor(X6))
anova(FM)
anova(FMg)
```

Note: R's function glm() is also employed here since this function produces a data class for which the general wrapper anova() assumes anova.glm() which produces likelihood ratio results as default.

```
> anova(FM)
Analysis of Variance Table

Response: Y
           Df Sum Sq Mean Sq  F value    Pr(>F)
X1          1 0.7763  0.7763  12.5579 0.0009042 ***
X2          1 2.5888  2.5888  41.8803 5.187e-08 ***
X3          1 6.3341  6.3341 102.4704 2.157e-13 ***
X4          1 0.0246  0.0246   0.3976 0.5313820
X5          1 0.1265  0.1265   2.0460 0.1592180
factor(X6)  1 0.0522  0.0522   0.8448 0.3627348
Residuals  47 2.9053  0.0618
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$$\sigma_{FM} = \sqrt{0.0618} = 0.2486$$

^ "standard error" (standard deviation of the residuals) for the full model

```
> anova(FMg)
Analysis of Deviance Table

Model: gaussian, link: identity

Response: Y

Terms added sequentially (first to last)


           Df Deviance Resid. Df Resid. Dev
NULL                          53    12.8077
X1          1   0.7763        52    12.0315
X2          1   2.5888        51     9.4427
X3          1   6.3341        50     3.1085
X4          1   0.0246        49     3.0840
X5          1   0.1265        48     2.9575
factor(X6)  1   0.0522        47     2.9053
```

Reduced Model:

FITTING A REDUCED LINEAR MODEL
RM=lm(Y~X1+X2+X3+X5)
RMg=glm(Y~X1+X2+X3+X5)
anova(RM)
anova(RMg)

> anova(RM)

```
Analysis of Variance Table

Response: Y
          Df Sum Sq Mean Sq  F value    Pr(>F)
X1         1 0.7763  0.7763  12.8495  0.000776 ***
X2         1 2.5888  2.5888  42.8528 3.349e-08 ***
X3         1 6.3341  6.3341 104.8499 9.118e-14 ***
X5         1 0.1484  0.1484   2.4561  0.123503
Residuals 49 2.9602  0.0604
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$\sigma_{RM} = \sqrt{0.0604} = 0.2458$

^ **"standard error" (standard deviation of the residuals) for the reduced model**

> anova(RMg)

```
Analysis of Deviance Table

Model: gaussian, link: identity

Response: Y

Terms added sequentially (first to last)


     Df Deviance Resid. Df Resid. Dev
NULL                    53    12.8077
X1    1   0.7763        52    12.0315
X2    1   2.5888        51     9.4427
X3    1   6.3341        50     3.1085
X5    1   0.1484        49     2.9602
```

Estimating Standard Error using Maximum Likelihood:

> #CALCULATING MAXIMUM LIKELIHOOD STANDARD DEVIATION
> n=length(K[,1])
> n #NUMBER OF CASES IN DATASET K
```
[1] 54
```
$n := 54$

> k=length(K)
> k #NUMBER OF VARIABLES IN FM
```
[1] 7
```
$k := 7$

> r=2 #DIFFERENCE IN NUMBER OF VARIABLES FM VS RM
> r
```
[1] 2
```
difference in number of parameters between models FM & RM > $r := 2$

> #EXTRACTING STANDARD ERRORS:
> #FOR FM:
> FMsigma = summary(FM)$sigma
> FMsigma
```
[1] 0.2486247
```
$\sigma_{FM} := 0.2486247$

> #FOR RM:
> RMsigma = summary(RM)$sigma
> RMsigma
```
[1] 0.2457874
```
$\sigma_{RM} := 0.2457874$

> #MAXIMUM LIKELIHOOD STANDARD DEVIATIONS
> #FOR FM:
> FMsigma.ML = FMsigma*sqrt((n-k)/n)
> FMsigma.ML
```
[1] 0.2319511
```
$ML\sigma_{FM} := \sqrt{\frac{n-k}{n}} \cdot \sigma_{FM}$ $\qquad ML\sigma_{FM} = 0.231951$

> #FOR RM:
> RMsigma.ML = RMsigma*sqrt((n-k+r)/n)
> RMsigma.ML
```
[1] 0.234132
```
$ML\sigma_{RM} := \sqrt{\frac{n-k+r}{n}} \cdot \sigma_{RM}$ $\qquad ML\sigma_{RM} = 0.234132$

Calculating Log Likelihoods for Each Model:

```
> # LOG LIKELIHOOD OF MODELS
> # FM:
> sum(log(dnorm(x = Y, mean = predict(FM), sd = FMsigma.ML)))
[1] 2.28368
> logLik(FM)
'log Lik.' 2.28368 (df=8)
> # RM:
> sum(log(dnorm(x = Y, mean = predict(RM), sd = RMsigma.ML)))
[1] 1.778311
> logLik(RM)
'log Lik.' 1.778311 (df=6)
```

ln likelihood value for FM > (logLik(FM))

ln likelihood value for RM > (logLik(RM))

^ ln ="natural logs" in base e

Note: log likelihoods for each model are calculated here using maximum likelihood estimates of standard error for *each model separately*. This contrasts with the use of standard error using only the FM in the test below.

Likelihood Ratio Test:

Assumptions:

- Standard Linear Regression depends on specifying in advance which variable is to be considered 'dependent' and which 'independent'. This decision matters as changing roles for Y & X usually produces a different result.

- $Y_1, Y_2, Y_3, \ldots, Y_n$ (dependent variable) is a random sample.

Note: Although a Normal distribution is assumed here for Y in a linear model, in other instances of the likelihood ratio test, this assumption doesn't apply.

- $X_1, X_2, X_3, \ldots, X_n$ (independent variable) with each value of X_i matched to Y_i

Within this setup, two models for the relationship between X and Y variables are explicitly compared:

Full Model:

$$Y_i = \beta_0 + \Sigma\beta_j X_i + \varepsilon_i$$

Reduced Model:

$$Y_i = \beta_0 + \Sigma\beta_k X_i + \varepsilon_i$$

where: Y_i and $[X_1,X_2, \ldots X_i]$ are matched dependent and independent variables, and

β_0 is the y intercept of the regression line (translation)

β_j are slope coefficients for the *full set* of independent variables $X_1,X_2, \ldots X_j$

β_k are slope coefficients for a smaller *set* of independent variables *within* X_i

ε_i is the error factor in prediction of Y_i and a random variable $\sim N(0,\sigma^2)$.

Hypotheses:

H_0: coefficients in j but NOT INCLUDED in k = 0.

Note: this is always the more parsimonious (i.e., smaller) model

H_1: at least some of these coeficients not 0

Degrees of Freedom:

n = 54 < **n = number of matched observations in dataset**

k = 7 < **k = number of variables in FM**

r = 2 < **r = differences in number of variables between FM & RM**

Sum of Squares and Standard Error for FM:

```
> anova(FM)
Analysis of Variance Table

Response: Y
           Df Sum Sq Mean Sq  F value    Pr(>F)
X1          1 0.7763  0.7763  12.5579 0.0009042
X2          1 2.5888  2.5888  41.8803 5.187e-08
X3          1 6.3341  6.3341 102.4704 2.157e-13
X4          1 0.0246  0.0246   0.3976 0.5313820
X5          1 0.1265  0.1265   2.0460 0.1592180
factor(X6)  1 0.0522  0.0522   0.8448 0.3627348
Residuals  47 2.9053  0.0618
```

```
> anova(RM)
Analysis of Variance Table

Response: Y
          Df Sum Sq Mean Sq  F value    Pr(>F)
X1         1 0.7763  0.7763  12.8495  0.000776 ***
X2         1 2.5888  2.5888  42.8528 3.349e-08 ***
X3         1 6.3341  6.3341 104.8499 9.118e-14 ***
X5         1 0.1484  0.1484   2.4561  0.123503
Residuals 49 2.9602  0.0604
```

```
> #LIKELIHOOD RATIO TEST:
> #SUM OF SQUARES ERROR FOR MODELS:
> SSE.FM = sum((Y-predict(FM))^2) #SSE for FM
> SSE.FM
[1] 2.90527
> SSE.RM = sum((Y-predict(RM))^2) #SSE for RM
> SSE.RM
[1] 2.960161
```

```
> #STANDARD ERROR  FOR FM:
> s=summary(FM)$sigma
> s
[1] 0.2486247
> s=sqrt(summary(FMg)$dispersion)
> s
[1] 0.2486247
```

$s := 0.2486247$ ^ **Standard errors are the square root of MSE, see above.**

$SSE_{FM} := 2.90527$

$SSE_{RM} := 2.960161$

Relative Likelihoods: see eq 1.26 in KNNL

$$LFM := e^{\frac{-1}{2}\cdot\frac{SSE_{FM}}{s^2}} \qquad LFM = 6.2241 \times 10^{-11}$$

$$LRM := e^{\frac{-1}{2}\cdot\frac{SSE_{RM}}{s^2}} \qquad LRM = 3.9926 \times 10^{-11}$$

$$C := \frac{1}{\left(2\cdot\pi\cdot s^2\right)^{\frac{n}{2}}} \qquad C = 1.2296 \times 10^{11}$$

```
> #RELATIVE LIKELIHOODS FOR THE MODELS:
> LFM = exp(-(1/2)*(SSE.FM/s^2)) #TIMES CONSTANT C
> LFM
[1] 6.224145e-11
> LRM = exp(-(1/2)*(SSE.RM/s^2)) #TIMES CONSTANT C
> LRM
[1] 3.992573e-11
```

```
> #CONSTANT C:
> n #NUMBER OF CASES IN DATASET K
[1] 54
> C=1/((2*pi*s^2)^(n/2))  #CONSTANT IN EQ 1.26 IN KNNL
> C
[1] 122956414826
```

Likelihoods:

$$\Lambda_{FM} := C \cdot LFM \qquad \Lambda_{FM} = 7.653$$

$$\Lambda_{RM} := C \cdot LRM \qquad \Lambda_{RM} = 4.9091$$

```
> LCFM=C*LFM
> LCFM
[1] 7.652985
> LCRM=C*LRM
> LCRM
[1] 4.909125
```

Likelihood Ratio Test Statistic:

$SSE_{FM} := 2.9053$

$SSE_{RM} := 2.9602$

$$LRT := \frac{(SSE_{RM} - SSE_{FM})}{s^2}$$ $LRT = 0.8881$ **< difference here due to rounding...**

```
> #LOG LIKELIHOOD RATIO STATISTIC:
> LRT=(SSE.RM - SSE.FM)/s^2
> LRT    #LOG LIKELIHOOD RATIO STATISTIC
[1] 0.8880001
```

Critical Value of the Test:

$\alpha := 0.05$ **< Probability of Type I error must be explicitly set**

$CV := \text{qchisq}(1 - \alpha, r)$ $CV = 5.9915$ **< note degrees of freedom reflect difference between the models**

Decision Rule:

IF F > CV, THEN REJECT H_0 OTHERWISE ACCEPT H_0

$LRT = 0.8881$ $CV = 5.9915$

Probability Value:

$P := 1 - \text{pchisq}(LRT, r)$ $P = 0.6414$

```
> #PROBABILITY OF NULL HYPOTHESIS RM
> P=1-pchisq(LRT,2)
> P   #PROBABILITY
[1] 0.6414654
```

IMPORTANT NOTE:

FALURE to reject H_0 in this test means that the MORE PARSIMONIOUS model RM is PREFERRED!

Prototype in R:

```
#LIKELIHOOD RATIO TEST:
anova(RM,FM,test="LRT")
anova(RMg,FMg,test="LRT")
```

```
> #LIKELIHOOD RATIO TEST:
> anova(RM,FM,test="LRT")
Analysis of Variance Table

Model 1: Y ~ X1 + X2 + X3 + X5
Model 2: Y ~ X1 + X2 + X3 + X4 + X5 + factor(X6)
  Res.Df    RSS Df Sum of Sq Pr(>Chi)
1     49 2.9602
2     47 2.9053  2  0.054891   0.6415

> anova(RMg,FMg,test="LRT")
Analysis of Deviance Table

Model 1: Y ~ X1 + X2 + X3 + X5
Model 2: Y ~ X1 + X2 + X3 + X4 + X5 + factor(X6)
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1        49     2.9602
2        47     2.9053  2 0.054891   0.6415
```