Stat 414 – Day 4
Likelihood Ratio Tests (LRT), Restricted Maximum Likelihood Estimation (REML)

Last Time: Maximum Likelihood Estimation
- Alternative method for estimating slope coefficients
- Maximizes the (log) likelihood of the observed data

$$L(\mu\,,\ \sigma\,; y_i) = 1/(\sigma\sqrt{2\pi})^n\, e^{-\Sigma(y_i-\mu)^2/(2\sigma^2)}\,.$$

$$L(\beta_0,\beta_1,\sigma;y_i,x_i) = -\frac{n}{2}\ln 2\pi - \frac{n}{2}ln\,\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i-(\beta_0+\beta_1X_i))^2$$

$-\frac{1}{2}\sigma^{-2}$ SSE

$-\frac{1}{2}\,-2\sigma^{-3}$ SSE

- Regression coefficient estimates same as OLS
- Can compare models by comparing the (log) likelihood values
  - $L = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\hat{\sigma}^2 - \frac{1}{2}(\frac{SSE}{\hat{\sigma}^2})$
  - AIC = -2 x log-likelihood + 2*p* = *deviance* + 2*p*
  - May be reported differently by different software packages
  - Advice: Compute variety of measures, identify models favored by multiple criteria

Example 1: Recall the example predicting airfares from distances.

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 214.99440   69.81197    3.08   0.0116 *
distance      0.14246    0.03384    4.21   0.0018 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 83.46 on 10 degrees of freedom
Multiple R-squared:  0.6393,    Adjusted R-squared:  0.6032
F-statistic: 17.72 on 1 and 10 DF,  p-value: 0.0018
```

```
Analysis of Variance Table

Response: airfare
          Df Sum Sq Mean Sq F value Pr(>F)
distance   1 123463  123463  17.723 0.0018
Residuals 10  69662    6966
```

(a) What is $\hat{\sigma}_{OLS}$? What are the degrees of freedom for $\hat{\sigma}_{OLS}$? What is the value of SSE? What is the likelihood based on these model estimates? What is the "deviance"?

$\hat{\sigma} = 83.46$ df = 12 − 1 slope − 1 intercept = 10 $\sqrt{SSE/(n-p-1)}$ ↑ # of slopes

SSE = 69662

deviance is given by R as SSError $L_1 = -69.0$

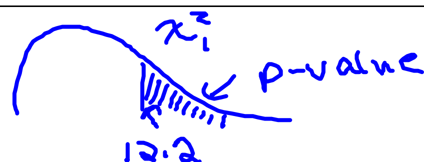
(b) What is the likelihood of this model compared to the null model? How do we compare these? $L_0 = -75.1$

$2(75.1 - 69.0) = 12.2$ test statistic

$\sim \chi^2$ df = difference in # of parameters between 2 models

Definition: The likelihood ratio test compares nested models by using the statistic -2(L0 – L1) which asymptotically follows a chi-square distribution with df = difference in number of parameters in the two model

In R: 1-pchisq(2*(75.144 – 69.023), 1)



$= .000477$

reject $H_0: \beta_{dist} = 0$ ~~as~~ in favor of $H_a: \beta_{dist} \neq 0$

F-tests vs. Likelihood ratio tests for nested models
○ Full and Reduced models (H_0: $\beta_j = \ldots = \beta_{j+k} = 0$)
○ F = [(drop in SSE)/k] / MSE (full) with df k and $n - p - 1$
○ $X^2 = 2\ln(L_{full}/L_{reduced})$ with df = k

Example 2: Recall our election data from Day 1. Fit two gls models, one with the interaction terms and one without. Compare the two models with a likelihood ratio test, what do you conclude?

$H_0: \beta_{SxR} = \beta_{IP} = \beta_{RP} = 0$, after adjusting for....

$\chi^2 = 3.579$ df = 3 p-value = .311

H_a: at least $\beta \neq 0$

Example 1 cont.

(c) But what about $\hat{\sigma}_{MLE}$? $\hat{\sigma}_{MLE} = 76.19$

$\frac{\partial}{\partial\sigma} L \Rightarrow -\frac{n}{\sigma} + \frac{SSE}{\sigma^3} = 0 \Rightarrow -n\sigma^2 + SSE = 0$

$\hat{\sigma}^2_{MLE} = \frac{SSE}{n}$ biased downward

$\hat{\sigma}^2_{OLS} = \frac{SSE}{n-p-1}$ unbiased

Maximum Likelihood Estimators for the variance are biased because they do not take into account the estimation of the "nuisance" parameters (β_0, β_1).

An alternative proposed in the 1930s is *restricted* (or *residual*) maximum likelihood, REML.

- Maximizes a different likelihood function (special matrix multiplication, some rows constrained) that doesn't depend on β's (residuals, removes the fixed effects from the model)
- With simple regression models, the slope coefficient estimates are the same
- The parameter estimate of the variances differ
- $E(\hat{\sigma}_{REML}^2) = \sigma^2$

Definition: *Restricted Maximum Likelihood Estimators (REML)* are MLEs but reflect the number of parameters estimated and are unbiased. REMLs have much better properties especially for estimating variances.

(d) Use restricted maximum likelihood estimation for model 2.

- What is the residual standard error? 83.46
- How do the slope and intercept estimates change? didn't
- How does the achieved log likelihood value change? from 69.02 to 67.48
- How does the AIC value change? 144.0 to 140.97

(e) What happens if you try to compare this to the null model?

can't use anova to compare model1 ML to model2 REML

can't compare to REML models that only differ by fixed effects

To compare models with different fixed effects, you must use maximum likelihood estimation. To compare models with different random effects, you can use either ML or REML. REML provides better standard error estimates.

Example 3: Recall the squid data from Day 2/3.

(a) Model 1 assumed Var(ε_i) = σ^2. Run model1 using gls.

- Is it using ML or REML? REML (default, mentioned at top of output)
- How many parameters are being estimated? df = 25 = 1 + DML slope + 11 month + 11 DML * month + σ

(b) Model 2 assumed DML was a variance covariate: Var(ε_i) = σ^2 /DML$_i$. Run the model. How many parameters are being estimated?

still 25

(c) Model 3 assumed the variance covariate was the month, Var(ε_{ij}) = σ^2_j where j was the month number. This actually estimates one month's variance and then all the other month variances are multiples of that. Run model 3, how many parameters are being estimated?

df = 36, add 11 $\hat{\sigma}_j$ for the 12 months

(d) What do you learn from the anova comparing the 3 models?

model 2: $\hat{\sigma}$
model 3: $\hat{\sigma}$ & 11 multipliers

model 1 & model 2 not nested, no LRT

model 2 & model 3 $\chi^2 = 28.46$ w/ df = 11

$H_0: \gamma_1 = \cdots \gamma_{11} = 0$ p-value = .0027

(e) Run model 4 which allows the power on DML to change differently for each month. Var(ε_{ij}) = σ^2 /DML^{k_j}. How many parameters? Can we compare model3 to model 4?

12 powers vs. 11 multipliers

$\chi^2 = 208.9$ w/ 1 df