PLSC 503 – Spring 2020
MLE: Testing and Inference

Zoom link: https://psu.zoom.us/j/845681138

April 2, 2020

Testing: The Plan

- “The Trinity”
- An example
- Practical advice

Inference, In General

1. Pick some $\mathbf{H}_A : \boldsymbol{\Theta} = \boldsymbol{\Theta}_A$
2. Estimate $\hat{\boldsymbol{\Theta}}$
3. Determine distribution of $\hat{\boldsymbol{\Theta}}$ under $\mathbf{H}_A$
4. Use (2) and (3) $\rightarrow \hat{\mathbf{S}} \sim h(\boldsymbol{\Theta}, \hat{\boldsymbol{\Theta}})$ (*test statistic*)
5. Assess $\Pr(\hat{\mathbf{S}}|\mathbf{H}_A)$

MLEs and Inference

$$\hat{\boldsymbol{\Theta}} \overset{a}{\sim} \mathbf{N}[\Theta, \mathbf{I}(\hat{\boldsymbol{\Theta}})]$$

Means that

$$\frac{\hat{\theta}_k - \theta_k}{\sqrt{\hat{\sigma}_k^2}} \sim N(0, 1)$$

Single Coefficients: Significance Testing

- Choose θ_A
- Estimate $\hat{\theta}_k$, $\hat{\sigma}_k^2$
- Compare $z_k = \frac{\hat{\theta}_k - \theta_A}{\sqrt{\hat{\sigma}_k^2}}$ to a z-table
- (Or, just look at your output...)

Single Coefficients: Confidence Intervals

- $\alpha \in (0, 1) =$ desired level of "significance"
- $(1 - \alpha) \times 100$-percent confidence intervals for $\hat{\theta}_k$ are:

$$\hat{\theta}_k \pm \left(z_\alpha \sqrt{\hat{\sigma}_k^2} \right)$$

- (Or just look at your output...)

More General Tests: "The Trinity"

- Likelihood-Ratio ("LR")
- Wald
- Lagrangian Multiplier ("Score")

$$\mathbf{R}\boldsymbol{\Theta} = \mathbf{r}$$

$$\theta_2 = -2 \iff (0\ 1\ 0)\begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix} = -2$$

$$\mathbf{\Theta}_A : \theta_2 = 1, \theta_1 = 2\theta_3$$

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & -2 \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$r = \text{rows}(\mathbf{R}) \in [0, K]$$

$$L(\hat{\boldsymbol{\Theta}}) \geq L(\boldsymbol{\Theta}_{\mathbf{A}}), \text{ but}$$

By how much?

Odds of one thing vs. another:

$$\frac{\Pr(\text{Something})}{\Pr(\text{Something Else})}$$

LR Test

$$\frac{L(\boldsymbol{\Theta}_{\mathbf{A}})}{L(\hat{\boldsymbol{\Theta}})} \ \ (\leq 1)$$

Suggests:

$$\ln L(\boldsymbol{\Theta}_{\mathbf{A}}) - \ln L(\hat{\boldsymbol{\Theta}}) \ \ (\leq 0)$$

$$-2[\ln L(\boldsymbol{\Theta}_{\mathbf{A}}) - \ln L(\hat{\boldsymbol{\Theta}})] \ \stackrel{a}{\sim} \ \chi^2_r$$

LR Test

- Intuition: Difference in $\ln L$ under constraint(s)
- Asymptotic
- Unreliable if $r > 100$ (or so)
- Easy to compute, but
- Requires that we have $\ln L(\boldsymbol{\Theta}_{\mathbf{A}})$ *and* $\ln L(\hat{\boldsymbol{\Theta}})$

Wald Tests

Idea: If $\mathbf{\Theta_A}$, then

$$\mathbf{R\Theta} = \mathbf{r}$$

$$\mathbf{R\Theta} - \mathbf{r} = \mathbf{0}$$

Wald Tests (continued)

But...

- We have only $\hat{\boldsymbol{\Theta}}$ (from sample data)
- Possible that $\mathbf{R}\hat{\boldsymbol{\Theta}} - \mathbf{r} = \mathbf{0}$ *due to chance* (sampling variability).
- Solution: Account for *variability* in $\hat{\boldsymbol{\Theta}}$.

Wald Tests (continued)

Test:

$$\mathbf{W} = (\mathbf{R}\hat{\Theta} - \mathbf{r})' \left[\mathbf{R}\,\mathsf{Var}(\hat{\Theta})\,\mathbf{R}'\right]^{-1} (\mathbf{R}\hat{\Theta} - \mathbf{r})$$

$$\mathbf{W} \overset{a}{\sim} \chi^2_r$$

Two-Handed Wald Tests

(+) Easy, fast

(+) No need for $\ln L(\boldsymbol{\Theta}_{\mathbf{A}})$

(-) Uses $\text{Var}(\hat{\boldsymbol{\Theta}})$, not $\text{Var}(\boldsymbol{\Theta}_{\mathbf{A}})$ (potentially poor coverage)

(-) Can yield nonsensical results

Lagrange Multiplier (LM) Tests

Idea: If $\boldsymbol{\Theta}_{\mathbf{A}}$, then

$$\left.\frac{\partial \ln L}{\partial \theta}\right|_{\boldsymbol{\Theta}_{\mathbf{A}}} \approx \mathbf{0}$$

Consider a new problem:

$$\max_{\boldsymbol{\Theta}} \left[L(\boldsymbol{\Theta}) - \boldsymbol{\lambda}(\boldsymbol{\Theta} - \boldsymbol{\Theta}_{\mathbf{A}}) \right]$$

LM Tests

Yields:

$$\tilde{\boldsymbol{\Theta}} = \boldsymbol{\Theta}_{\mathbf{A}}$$

$$\tilde{\boldsymbol{\lambda}} = \mathbf{g}(\tilde{\boldsymbol{\Theta}})$$

Suggests

$$LM = \mathbf{g}(\tilde{\boldsymbol{\Theta}})' \, \mathbf{I}(\tilde{\boldsymbol{\Theta}})^{-1} \mathbf{g}(\tilde{\boldsymbol{\Theta}})$$

LM Tests

$$LM \overset{a}{\sim} \chi^2_r$$

Note: No need for $\hat{\boldsymbol{\Theta}}$!

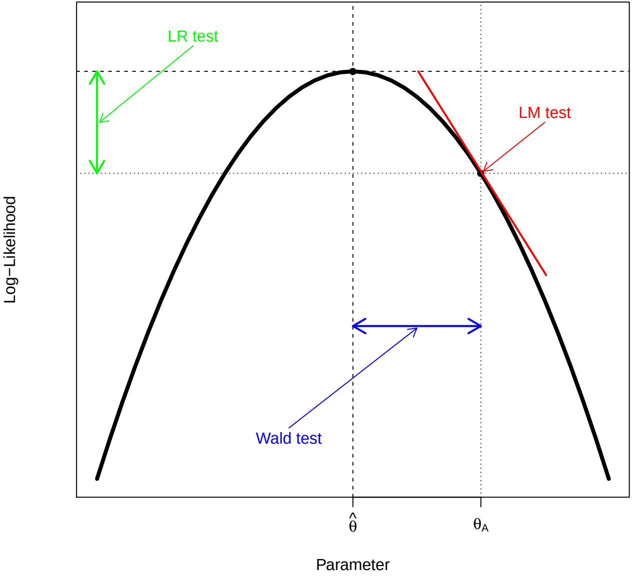
Tests, Conceptually (C. Franklin remix)

- The LR asks, "Did the likelihood change much under the null hypotheses versus the alternative?"
- The Wald test asks, "Are the estimated parameters very far away from what they would be under the null hypothesis?"
- The LM test asks, "If I had a less restrictive likelihood function, would its derivative be close to zero here at the restricted ML estimate?"

Tests, Conceptually (h.t.: Buse 1982)

- LR test $\approx$ manic mountaineer
- Wald test $\approx$ tired mountaineer
- LM test $\approx$ lazy mountaineer

Tests, Conceptually (A Picture)



Tests, Practically

- All are asymptotically identical...
- Require different estimates, but similar information
- Generally, LR $>$ Wald $>$ LM

Software: R

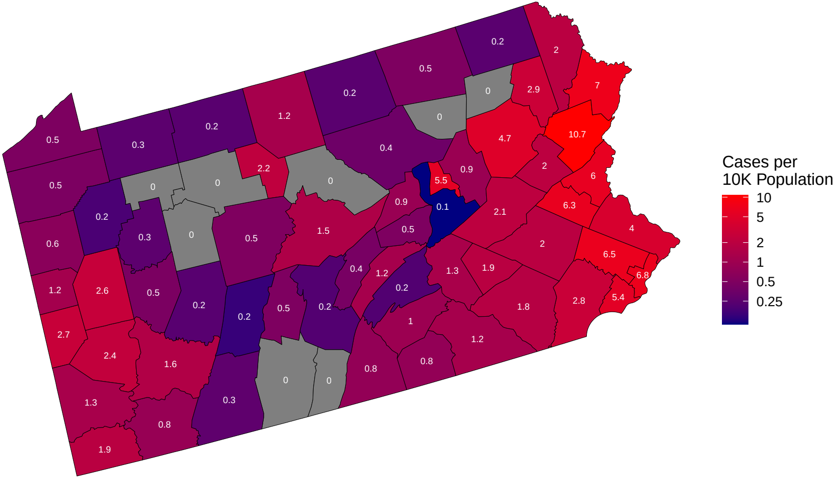
- Wald tests: `waldtest` (in `lmtest`), `wald.test` (in `aod`), etc.
- LR tests: `lrtest` (in `lmtest`), `RLRsim`, many others
- "by-hand" straightforward...

Example: COVID-19 in Pennsylvania

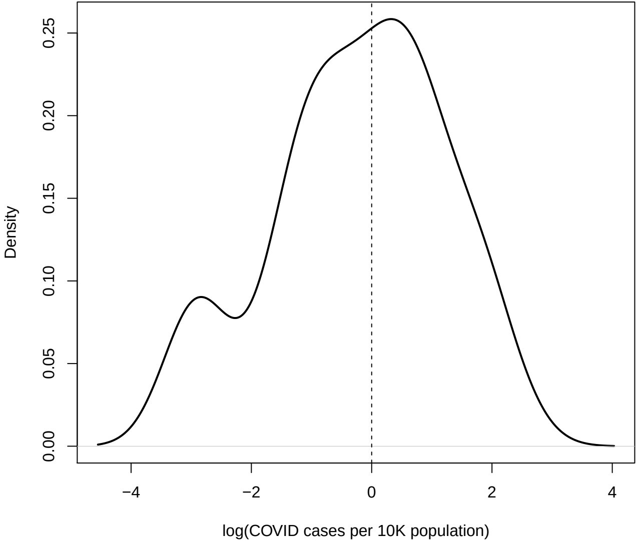
- COVID-19 cases, 67 counties, as of $3/30/2020$
- Source: `https://github.com/nytimes/covid-19-data`
- (Badly) Skewed $\rightarrow$ logged
- We're guessing $\sim N(\mu, \sigma^2)$...

PA COVID−19 Cases (per 10,000 population) by County, through March 30, 2020

Source: https://github.com/nytimes/covid−19−data



PA COVID-19 Cases per 10K Population, 3/30/2020 (logged)



```
> library(RCurl)
> library(maxLik)
> library(aod)
> library(lmtest)

# Get COVID data:

> temp<-getURL("https://raw.githubusercontent.com/PrisonRodeo/
   PLSC503-2020-git/master/Data/COVID-PA.csv")
> COVID<-read.csv(text=temp, header=TRUE)

# log-lik function:

> COVIDll <- function(param) {
+  mu <- param[1]
+  sigma <- param[2]
+  ll <- -0.5*log(sigma^2) - (0.5*((x-mu)^2/sigma^2))
+  ll
}

> x<-log(COVID$CasesPer10K+0.055)
```

```
> hats <- maxLik(COVIDll, start=c(0,1))
> summary(hats)
--------------------------------------------
Maximum Likelihood estimation
Newton-Raphson maximisation, 5 iterations
Return code 2: successive function values
  within tolerance limit
Log-Likelihood: -56.4
2  free parameters
Estimates:
     Estimate Std. error t value Pr(> t)
[1,]   -0.217      0.172   -1.26    0.21
[2,]    1.407      0.122   11.58  <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
--------------------------------------------
```

≡ Mean-Only Linear Model

```
> COVIDLM<-lm(x~1)
> summary(COVIDLM)

Call:
lm(formula = x ~ 1)

Residuals:
   Min     1Q Median     3Q    Max
-2.684 -0.972  0.163  0.969  2.595

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   -0.217      0.173   -1.25     0.22

Residual standard error: 1.42 on 66 degrees of freedom
```

Moving parts...

```
> hats$estimate
[1] -0.217  1.407
```

```
> hats$gradient
[1] 0.00000000422 0.00000223621
```

```
> hats$hessian
      [,1]  [,2]
[1,] -33.8   0.0
[2,]   0.0 -67.7
```

More moving parts...

```
> -(solve(hats$hessian))
       [,1]   [,2]
[1,] 0.0296 0.0000
[2,] 0.0000 0.0148
```

```
> sqrt(-(solve(hats$hessian)))
      [,1]  [,2]
[1,] 0.172 0.000
[2,] 0.000 0.122
```

Wald test...

```
> wald.test(Sigma=vcov(hats),b=coef(hats),Terms=1:2,verbose=TRUE)
Wald test:
----------

Coefficients:
[1] -0.22  1.41

Var-cov matrix of the coefficients:
     [,1]  [,2]
[1,] 0.030 0.000
[2,] 0.000 0.015

Test-design matrix:
   [,1] [,2]
L1    1    0
L2    0    1

Positions of tested coefficients in the vector of coefficients: 1, 2

H0:  -0.217 = 0; 1.407 = 0

Chi-squared test:
X2 = 135.6, df = 2, P(> X2) = 0.0
```

More Wald tests

```
> wald.test(Sigma=vcov(hats),b=coef(hats),Terms=1:2,H0=c(0,2))
Wald test:
----------

Chi-squared test:
X2 = 25.4, df = 2, P(> X2) = 0.0000031
```

```
> wald.test(Sigma=vcov(hats),b=coef(hats),Terms=1:2,H0=c(-0.2,1.5))
Wald test:
----------

Chi-squared test:
X2 = 0.59, df = 2, P(> X2) = 0.74
```

A Nonsensical Wald Test: $\mu = -0.2, \sigma = -0.1$

```
> wald.test(Sigma=vcov(hats),b=coef(hats),
  Terms=1:2,H0=c(-0.2,-0.1))
Wald test:
----------

Chi-squared test:
X2 = 153.7, df = 2, P(> X2) = 0.0
```

Restricted model: fix $\mu = 0$:

```
> COVIDllAlt <- function(param) {
+     sigma <- param[1]
+   ll <- -0.5*log(sigma^2) - (0.5*((x-0)^2/sigma^2))
+   ll
+ }
```

```
> hatsF <- maxLik(COVIDll, start=c(0,1))
> hatsR <- maxLik(COVIDllAlt, start=c(1))
```

LR tests

```
> hatsF$maximum
[1] -56.4

> hatsR$maximum
[1] -57.2
```

```
> -2*(hatsR$maximum-hatsF$maximum)
[1] 1.57

> pchisq(-2*(hatsR$maximum-hatsF$maximum),df=1,lower.tail=FALSE)
[1] 0.21
```

LR tests (continued)

```
> library(lmtest) # install as necessary

> lrtest(hatsF,hatsR)
Likelihood ratio test

Model 1: hatsF
Model 2: hatsR
  #Df LogLik Df Chisq Pr(>Chisq)
1   2  -56.4
2   1  -57.2 -1  1.57       0.21


> # Compare to Wald:
>
> wald.test(Sigma=vcov(hats),b=coef(hats),Terms=1:1,H0=0)
Wald test:
----------

Chi-squared test:
X2 = 1.6, df = 1, P(> X2) = 0.21
```