BIOS 7345: Advanced Regression Analysis I

Andrew J. Spieker, Ph.D.

Assistant Professor of Biostatistics
Vanderbilt University Medical Center

Set 16: Hypothesis testing for GLMs

Version: 09/17/2023

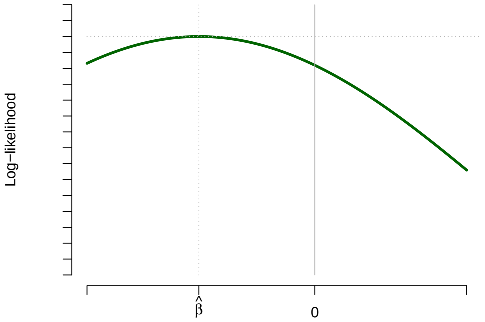
Table of Contents

Hypothesis testing

Generally:

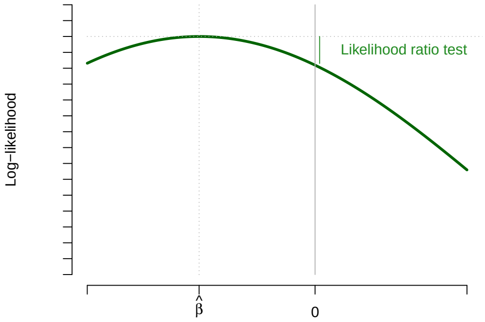
- We seek to test hypotheses $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$, where $\boldsymbol{\beta}$ denotes the regression parameters of a GLM (you may be thinking about it through the likelihood, quasi-likelihood, or estimating equations framework) and $\mathbf{C} = \mathbf{C}_{Q \times K}$ encodes Q linear hypotheses.
- The goal in this set of notes is to present and discuss different hypothesis testing methods. You'll already be familiar with many of the ideas. The three major classes of tests we will discuss are:
  - Likelihood ratio tests.
  - Score-based tests (also called Rao tests).
  - Wald-based tests.
- In correctly specified models, and under suitable regularity conditions, the three are asymptotically equivalent. This can be seen graphically.

In pictures: Log-likelihood



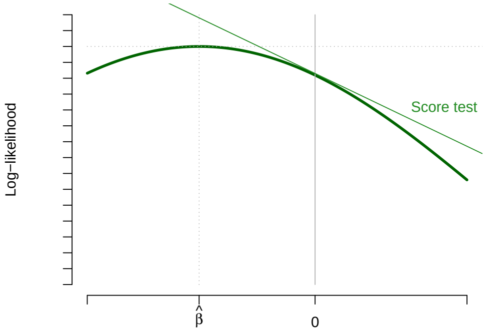
- In our settings, the log-likelihood should generally be concave in a neighborhood of $\widehat{\beta}$.

In pictures: Likelihood-ratio based tests



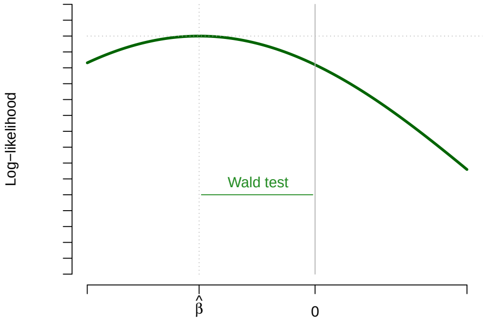
- A likelihood ratio test compares the maximum log-likelihood to that under the null.

In pictures: Score-based tests



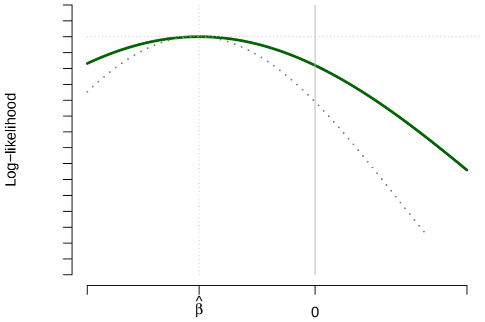
- A score-based test involves evaluating the slope of the log-likelihood under the null.

In pictures: Wald-based tests



- A Wald-based test involves evaluating how far the observed value $\widehat{\beta}$ is from the null (relative to variability).

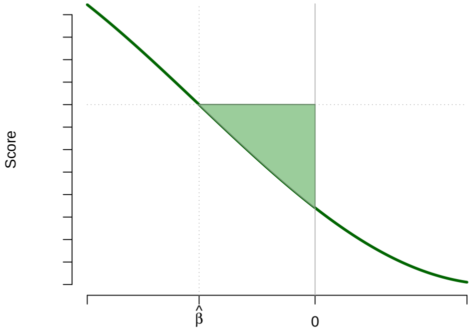
In pictures: Increasing sample size



- Greater concavity in the log-likelihood. What are the consequences for each of the tests? What are the implications if H_0 is true?

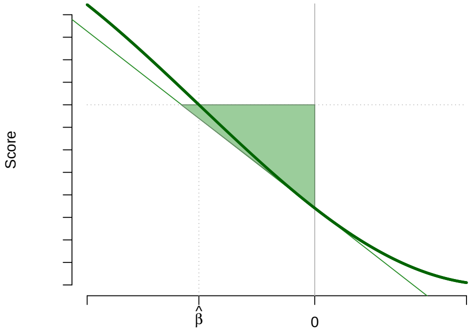
Hypothesis testing

In pictures: Likelihood ratio-based tests



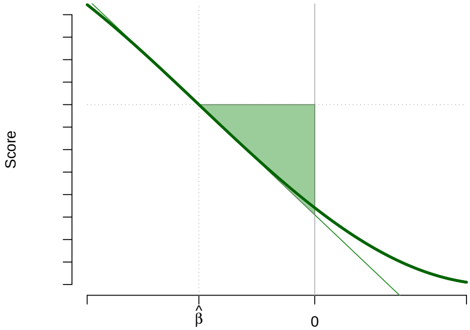
- The likelihood ratio test statistic is twice the shaded area.

In pictures: Score-based tests



- The score statistic is twice the shaded area.

In pictures: Wald-based tests



- The Wald statistic is twice the shaded area.

For GLMs: Wald statistic

- The following is a form for the Wald statistic (we've encountered it before):

$$X_W^2 = (\mathbf{C}\widehat{\boldsymbol{\beta}})^T \left(\mathbf{C}\widehat{\text{Cov}}[\widehat{\boldsymbol{\beta}}]\mathbf{C}^T\right) (\mathbf{C}\widehat{\boldsymbol{\beta}}).$$

- $\widehat{\text{Cov}}[\widehat{\boldsymbol{\beta}}]$ is any reasonable and appropriate estimator (model-based, sandwich, quasi-likelihood, bootstrap-based).
- Under H_0, we have $X_W^2 \longrightarrow_d \chi_Q^2$ (I state this without proof; we have already done a rigorous treatment of hypothesis testing for linear regression, so I don't wish to belabor the point).

Score tests

For GLMs: Score statistic

- Let $(\widehat{\boldsymbol{\beta}}^0, \widehat{\phi}^0)$ denote parameter estimates under H_0 (i.e., subject to the constraint $\mathbf{C}\boldsymbol{\beta} = \mathbf{0}$).
- The score takes the form $\mathbb{S}_N(\boldsymbol{\beta}, \phi) = \mathbf{D}^T \mathbf{V}^{-1}(\mathbf{y} - \boldsymbol{\mu})/\phi$.
- The information takes the form $\mathcal{I}_N(\boldsymbol{\beta}, \phi) = \mathbb{A}_N(\boldsymbol{\beta})/\phi$.
- The following is a form for the score statistic:

$$X_S^2 = (\mathbb{S}_N(\widehat{\boldsymbol{\beta}}^0, \widehat{\phi}^0))^T \left(\mathcal{I}_N(\widehat{\boldsymbol{\beta}}^0, \widehat{\phi}^0)\right)^{-1} (\mathbb{S}_N(\widehat{\boldsymbol{\beta}}^0, \widehat{\phi}^0)).$$

- This test is model-based, although use of the *observed* information may offer some robustness under certain circumstances.
  - This is a rabbit hole I'm electing not to go down.
- Under H_0, we have $X_S^2 \longrightarrow_d \chi_Q^2$.

For GLMs: Likelihood ratio test statistic

- Let $(\widehat{\boldsymbol{\beta}}^0, \widehat{\phi}^0)$ denote parameter estimates under H_0 (i.e., subject to the constraint $\mathbf{C}\boldsymbol{\beta} = \mathbf{0}$).
- The following is a form for the likelihood ratio test statistic:

$$X^2_{LR} = -2\left(\ell(\widehat{\boldsymbol{\beta}}^0, \widehat{\phi}^0) - \ell(\widehat{\boldsymbol{\beta}}, \widehat{\phi})\right)$$

- This test is model-based.
- Under H_0, we have $X^2_{LR} \longrightarrow_d \chi^2_Q$.

Table of Contents

Model setup: For simulation

- Exposure: $X = 0, 1, 2$, each with probability $1/3$.
- Outcome: $Y \sim \text{Bernoulli}(p = \text{expit}(\beta_0 + \beta_1 1(x = 1) + \beta_2 1(x = 2)))$.
- Let $\beta_0 = -1$, $\beta_1 = 0.4$ and $\beta_2 = 0.6$.
- Let $N = 150$.
- Consider the following hypothesis tests:
  1. $H_0 : \beta_1 = 0$ (groups 0 and 1 the same).
  2. $H_0 : \beta_2 - \beta_1 = 0$ (groups 1 and 2 the same).
  3. $H_0 : \beta_1 = \beta_2 = 0$ (all three groups the same).

Simulated example: Generate data

```
## Important function
expit <- function(x)
{
  expitx <- exp(x)/(1 + exp(x))
  return(expitx)
}

## Set seed and generate data
set.seed(7345)
n <- 150
beta <- matrix(c(-1, 0.4, 0.6), nrow = 3)
x <- sample(c(0,1,2), size = n, replace = TRUE)
X <- matrix(cbind(1, as.numeric(x==1), as.numeric(x==2)), ncol = 3)
y <- rbinom(n, 1, expit(X %*% beta))
```

Simulated example: Fit GLM (unrestricted)

```
## Initialize and run GLM
betaj <- c(0,0,0)
tol <- 1
iter <- 1

while(tol > 1e-15 & iter < 50)
{
  betaj.prior <- betaj
  etaj <- c(X %*% betaj)
  Gn <- t(X) %*% (y - expit(etaj))
  An <- t(X) %*% diag(expit(etaj)*(1 - expit(etaj))) %*% X
  betaj <- betaj + solve(An) %*% Gn
  tol <- sum((betaj - betaj.prior)^2)
  iter <- iter + 1
}
bhat <- betaj
Vhat <- solve(An)

> c(bhat)
[1] -1.0726368  0.2841794  0.7159619
```

Test 1: $H_0 : \beta_1 = 0$ (groups 0 and 1 the same).

- Model: $Y \sim \text{Bernoulli}(p = \text{expit}(\beta_0 + \beta_1 1(x = 1) + \beta_2 1(x = 2)))$.
- Constrained model: $Y \sim \text{Bernoulli}(p = \text{expit}(\beta_0 + \beta_2 (x = 2)))$.

Simulated example: Fit GLM (constrained)

```
X0 <- cbind(X[,1], X[,3])
betaj <- c(0,0)
tol <- 1
iter <- 1

while(tol > 1e-15 & iter < 50)
{
  betaj.prior <- betaj
  etaj <- c(X0 %*% betaj)
  Gn <- t(X0) %*% (y - expit(etaj))
  An <- t(X0) %*% diag(expit(etaj)*(1 - expit(etaj))) %*% X0
  betaj <- betaj + solve(An) %*% Gn
  tol <- sum((betaj - betaj.prior)^2)
  iter <- iter + 1
}

bhat.0 <- cbind(c(betaj[1], 0, betaj[2]))

> c(bhat.0)
[1] -0.9304754  0.0000000  0.5738004
```

Simulated example: Likelihood ratio test

```
loglik <- sum(log(dbinom(y, size = 1, prob = expit(X %*% bhat))))
loglik.0 <- sum(log(dbinom(y, size = 1, prob = expit(X %*% bhat.0))))
LR <- -2*(loglik.0 - loglik)
pLR <- 1 - pchisq(LR, df = 1)
```

Simulated example: Score test

```
Gn0 <- t(X) %*% (y - expit(X %*% bhat.0))
An0 <- t(X) %*% diag(c(expit(X %*% bhat.0)*
                     (1 - expit(X %*% bhat.0)))) %*% X
S <- t(Gn0) %*% solve(An0) %*% Gn0
pS <- 1 - pchisq(S, df = 1)
```

Simulated example: Wald test

```
C <- matrix(0, nrow = 1, ncol = 3)
C[1,2] <- 1

> C
     [,1] [,2] [,3]
[1,]    0    1    0


W <- t(C %*% bhat) %*% solve(C %*% Vhat %*% t(C)) %*% (C %*% bhat)
pW <- 1 - pchisq(W, df = 1)
```

Simulated example: Report results of tests

```
> pLR
[1] 0.5247948

> pS
          [,1]
[1,] 0.5248092

> pW
          [,1]
[1,] 0.5253544
```

Examples: Logistic regression

Simulated example: Compare to ANOVA function

```
> pLR
[1] 0.5247948

> pS
          [,1]
[1,] 0.5248092

zz.Full <- glm(y ~ X[,2] + X[,3], family = binomial)
zz.Reduced <- glm(y ~ X[,3], family = binomial)

> anova(zz.Full, zz.Reduced, test = "LRT")$Pr[2]
[1] 0.5247948

> anova(zz.Full, zz.Reduced, test = "Rao")$Pr[2]
[1] 0.5248089
```

Test 2: $H_0 : \beta_2 - \beta_1 = 0$ (groups 1 and 2 the same).

- Model: $Y \sim \text{Bernoulli}(p = \text{expit}(\beta_0 + \beta_1 1(x = 1) + \beta_2 1(x = 2)))$.
- Constrained model: $Y \sim \text{Bernoulli}(p = \text{expit}(\beta_0 + \beta_1 1(x > 0)))$.

Simulated example: Fit GLM (constrained)

```
X0 <- cbind(X[,1], X[,2] + X[,3])
betaj <- c(0,0)
tol <- 1
iter <- 1

while(tol > 1e-15 & iter < 50)
{
  betaj.prior <- betaj
  etaj <- c(X0 %*% betaj)
  Gn <- t(X0) %*% (y - expit(etaj))
  An <- t(X0) %*% diag(expit(etaj)*(1 - expit(etaj))) %*% X0
  betaj <- betaj + solve(An) %*% Gn
  tol <- sum((betaj - betaj.prior)^2)
  iter <- iter + 1
}

bhat.0 <- cbind(c(betaj[1], betaj[2], betaj[2]))

> c(bhat.0)
[1] -1.072637  0.513021  0.513021
```

Simulated example: Wald test

```
C <- matrix(0, nrow = 1, ncol = 3)
C[1,2] <- -1
C[1,3] <- 1

> C
     [,1] [,2] [,3]
[1,]    0   -1    1


W <- t(C %*% bhat) %*% solve(C %*% Vhat %*% t(C)) %*% (C %*% bhat)
pW <- 1 - pchisq(W, df = 1)
```

Simulated example: Report results of tests

```
## The code for pS and pLR are the same as previously shown
> pLR
[1] 0.3039759

> pS
          [,1]
[1,] 0.3048367

## The code for pW relied on another definition of C
> pW
          [,1]
[1,] 0.3060022
```

Simulated example: Compare to ANOVA function

```
> pLR
[1] 0.3039759

> pS
          [,1]
[1,] 0.3048367

zz.Full <- glm(y ~ X[,2] + X[,3], family = binomial)
zz.Reduced <- glm(y ~ I(X[,2] + X[,3]), family = binomial)

> anova(zz.Full, zz.Reduced, test = "LRT")$Pr[2]
[1] 0.3039759

> anova(zz.Full, zz.Reduced, test = "Rao")$Pr[2]
[1] 0.3048367
```

Test 3: $H_0 : \beta_1 = \beta_2 = 0$ (all groups the same).

- Model: $Y \sim \text{Bernoulli}(p = \text{expit}(\beta_0 + \beta_1 1(x = 1) + \beta_2 1(x = 2)))$.
- Constrained model: $Y \sim \text{Bernoulli}(p = \text{expit}(\beta_0))$.

Examples: Logistic regression

Simulated example: Fit GLM (constrained)

```
X0 <- cbind(X[,1])
betaj <- c(0)
tol <- 1
iter <- 1

while(tol > 1e-15 & iter < 50)
{
  betaj.prior <- betaj
  etaj <- c(X0 %*% betaj)
  Gn <- t(X0) %*% (y - expit(etaj))
  An <- t(X0) %*% diag(expit(etaj)*(1 - expit(etaj))) %*% X0
  betaj <- betaj + solve(An) %*% Gn
  tol <- sum((betaj - betaj.prior)^2)
  iter <- iter + 1
}

bhat.0 <- cbind(c(betaj[1], 0, 0))

> c(bhat.0)
[1] -0.7233002  0.0000000  0.0000000
```

Simulated example: Wald test

```
C <- matrix(0, nrow = 2, ncol = 3)
C[1,2] <- 1
C[2,3] <- 1

> C
     [,1] [,2] [,3]
[1,]    0    1    0
[2,]    0    0    1
```

```
W <- t(C %*% bhat) %*% solve(C %*% Vhat %*% t(C)) %*% (C %*% bhat)
pW <- 1 - pchisq(W, df = 2)
```

Simulated example: Report results of tests

```
## The code for pS and pLR are the same as previously shown
## except now with two degrees of freedom!
> pLR
[1] 0.2336163

> pS
          [,1]
[1,] 0.232581

## The code for pW relied on another definition of C
> pW
          [,1]
[1,] 0.2370055
```

Simulated example: Compare to ANOVA function

```
> pLR
[1] 0.2370055

> pS
          [,1]
[1,] 0.2325809

zz.Full <- glm(y ~ X[,2] + X[,3], family = binomial)
zz.Reduced <- glm(y ~ 1, family = binomial)

> anova(zz.Full, zz.Reduced, test = "LRT")$Pr[2]
[1] 0.3039759

> anova(zz.Full, zz.Reduced, test = "Rao")$Pr[2]
[1] 0.3048367
```

Notes:

- In this example, all tests should have been valid because the model was correctly specified.
- This was exemplified (though certainly not proven) by the fact that the resulting p-values were so similar.
- Let's try an example where the model is *not* fully correct.

Table of Contents

Examples: Gaussian regression

Model setup: For simulation

- Exposure: $X \sim \text{Uniform}(1, 10)$.
- Outcome: $Y \sim \mathcal{N}(\beta_0 + \beta_1 x + \beta_2 x^2, \sigma^2_X = 0.25x^2)$.
- Let $\beta_0 = 50$, $\beta_1 = 0$ and $\beta_2 = 0$.
- Let $N = 100$.
- Consider the hypothesis test: $H_0 : \beta_1 = \beta_2 = 0$
  - No quadratic association between X and mean Y.
- OLS: correctly specified mean model, but model-based variance wrong.
- Consider LR test, score test, Wald test (model-based variance), and Wald test (sandwich-based variance).

Examples: Gaussian regression

Simulated example: Generate data

```
set.seed(7345)
n <- 1000
beta <- matrix(c(50, 0, 0), nrow = 3)
x <- runif(n, 1, 10)
X <- matrix(cbind(1, x, x^2), ncol = 3)
y <- rnorm(n, X %*% beta, x/2)
```

Simulated example: Gaussian GLM

```
## Gaussian GLM
bhat <- solve(t(X) %*% X) %*% t(X) %*% y
phi.hat <- t(y - X %*% bhat) %*% (y - X %*% bhat)/(n - 3)
An <- t(X) %*% X
Bn <- t(X) %*% diag(c(y - X %*% bhat)^2) %*% X
```

Simulated example: Variance estimators (unrestricted)

```
## Model-based variance
V1 <- solve(An)*as.numeric(phi.hat)

## Sandwich variance
V2 <- solve(An) %*% Bn %*% solve(An)
```

Simulated example: Variance estimators (unrestricted)

```
## Restricted model
X0 <- X[,c(-2,-3)]
bhat.0 <- solve(t(X0) %*% X0) %*% t(X0) %*% y
bhat.0 <- cbind(c(bhat.0, 0, 0))
phi.hat0 <- t(y - X %*% bhat.0) %*% (y - X %*% bhat.0)/(n - 1)
```

Simulated example: Likelihood ratio test

```
## Likelihood ratio test
loglik <- sum(log(dnorm(y, X %*% bhat, sqrt(phi.hat*(n - 3)/n))))
loglik.0 <- sum(log(dnorm(y, X %*% bhat.0, sqrt(phi.hat0*(n - 1)/n))))
LR <- -2*(loglik.0 - loglik)
pLR <- 1 - pchisq(LR, df = 2)
```

Simulated example: Score test

```
## Score test
Gn0 <- t(X) %*% (y - X %*% bhat.0)/as.numeric(phi.hat)
An0 <- t(X) %*% X/as.numeric(phi.hat)
S <- t(Gn0) %*% solve(An0) %*% Gn0
pS <- 1 - pchisq(S, df = 2)
```

Simulated example: Wald tests

```
## Wald tests
C <- matrix(0, nrow = 2, ncol = 3)
C[1,2] <- 1
C[2,3] <- 1

## Model-based Wald test
W1 <- t(C %*% bhat) %*% solve(C %*% V1 %*% t(C)) %*% (C %*% bhat)
pW1 <- 1 - pchisq(W1, df = 2)

## Sandwich-based Wald test
W2 <- t(C %*% bhat) %*% solve(C %*% V2 %*% t(C)) %*% (C %*% bhat)
pW2 <- 1 - pchisq(W2, df = 2)
```

Simulated example: Results

```
> pLR
[1] 0.05747821

> pS
           [,1]
[1,] 0.05750231

> pW1
           [,1]
[1,] 0.05750231

> pW2
          [,1]
[1,] 0.1752994
```

Discussion questions:

- Which p-values (if any) are valid for the test of interest?
  - Note nuances regarding sample size. . .
- Can the likelihood ratio and score tests be modified to be valid?

Summary

So far:

- Hypothesis testing for GLMs.

Summary

Up next:

- Diagnostics.