STA 303/1002-Methods of Data Analysis II

Sections L0101& L0201, Winter 2019

Dr. Shivon Sue-Chee

UNIVERSITY OF TORONTO

February 14, 2019

STA 303/1002: Week 6- Case Study III Inference

Binary Logistic Regression Example

- Case Study III Inference: The Donner Party Example
  - Confidence interval for Odds Ratio
  - Testing/comparing models
  - Wald vs Likelihood Ratio Tests
  - Other Model Fit Statistics
- In R:
  - Effect Plots
  - Related R packages and functions
- Joke: *"I asked a statistician for her phone number... and she gave me an estimate."*(www.workjoke.com)

WALD CHI-SQUARE PROCEDURES

Logistic Regression: Inference on a single β

Test

- Hypotheses: $H_0 : \beta_j = 0$ (X_j has no effect on log-odds)
  $H_a : \beta_j \neq 0$
- Test Statistic: $\boxed{z = \frac{\widehat{\beta}_j}{SE(\widehat{\beta}_j)}}$

$$\boxed{z^2 \sim \chi^2_1}$$

  where
  - $\widehat{\beta}_j$- maximum likelihood (ML) estimate and
  - $SE(\widehat{\beta}_j)$- estimated standard error from the numerical procedure that generated the MLE.
  - By standard large-sample results, MLE's are normally distributed. Thus, for large n, under H_0, z is an observation from an approx. $\mathcal{N}(0, 1)$ distribution.

C.I

- 95% Confidence interval: $\boxed{\widehat{\beta}_j \pm 1.96 SE(\widehat{\beta}_j)}$

Examples: Inference on single β's

$\text{logit}(\pi) = \mu$

Using R output ('coefficients'):

	Age	*Sex*
Test statistic	$(-0.078/0.0373)^2$	
P-value	0.036	
CI for β	$-0.078 \pm 1.96(0.0373)$	
	=(-0.15, -0.0049)	
CI for Odds ratio	$(e^{-0.15}, e^{-0.0049})$=(0.86,0.995)	
Conclusion	For the same sex, the odds ratio for a 1-year increase in age is between .86 and 0.995.	

$\hat{\beta}_j \pm 1.96\, SE(\hat{\beta}_j)$

Recall the relationship between $\mathcal{N}(0,1)$ and Chi-square distribution:

Examples: Inference on single β's

Using R output:

	$-ve$ *Age*	$+ve$ *Sex*
Test statistic	$(-0.078/0.0373)^2$	$4.47 = 2.114^2$
P-value	0.036	0.0345
95% CI for β	$-0.078 \pm 1.96(0.0373)$ $=(-0.15, -0.0055)$	(0.117, 3.078)
CI for Odds ratio	$(e^{-0.15}, e^{-0.0055}) = \boxed{(0.86, 0.995)}$	$\boxed{(1.124, 21.72)}$
Conclusion	For the same sex, the odds ratio for a 1-year increase in age is between .86 and 0.995.	

- Note: Both marginal p-values are less that 0.05 and the confidence intervals for the odds ratios do not include 1.
- Hence, we have moderate evidence that both *Age* and *Sex* have an effect on survival over and above each other.
- Recall: If $Z \sim \mathcal{N}(0, 1)$, then $Z^2 \sim \chi_1$.

Additional CI Examples

41 vs 21

10 vs 30

(15 — 65)

Using R output:

MLE

$\downarrow$

$\hat{\beta}$

- Q: Find a 95% CI for the change in odds of survival for a 40-yr old to 20-yr old of the same sex.
- A:
  - The log odds change by -0.078*(40-20)=-1.56.
  - 95% CI for the change in log odds is 20*(-0.15, -0.00~~55~~ 49) =(-3.0, -0.11).
  - 95% CI for the odds ratio is (0.05, 0.896).
  - The odds of survival of a 40-yr old woman were $e^{-1.56} = 0.21$ times the odds of survival for a 20-yr old.

e^x – monotone

- Note that it is not appropriate to compute CI for π since $0 \leq \pi \leq 1$ and it is not normally distributed.

$0 < \pi < 1$

$\hat{\pi} = \frac{e^{M}}{1+e^{M}}$

Model Assumptions for Binary Logistic Regression

1. Underlying probability model for response is Bernoulli. $\pi(1-\pi)$
2. Observations are independent.
3. The form of the model is correct.
   - Linear relationship between logits and explanatory variables
   - All relevant variables are included; irrelevant ones excluded
4. Sample size is large enough for valid inference-tests and CIs. (Recall large-sample properties of MLEs.)

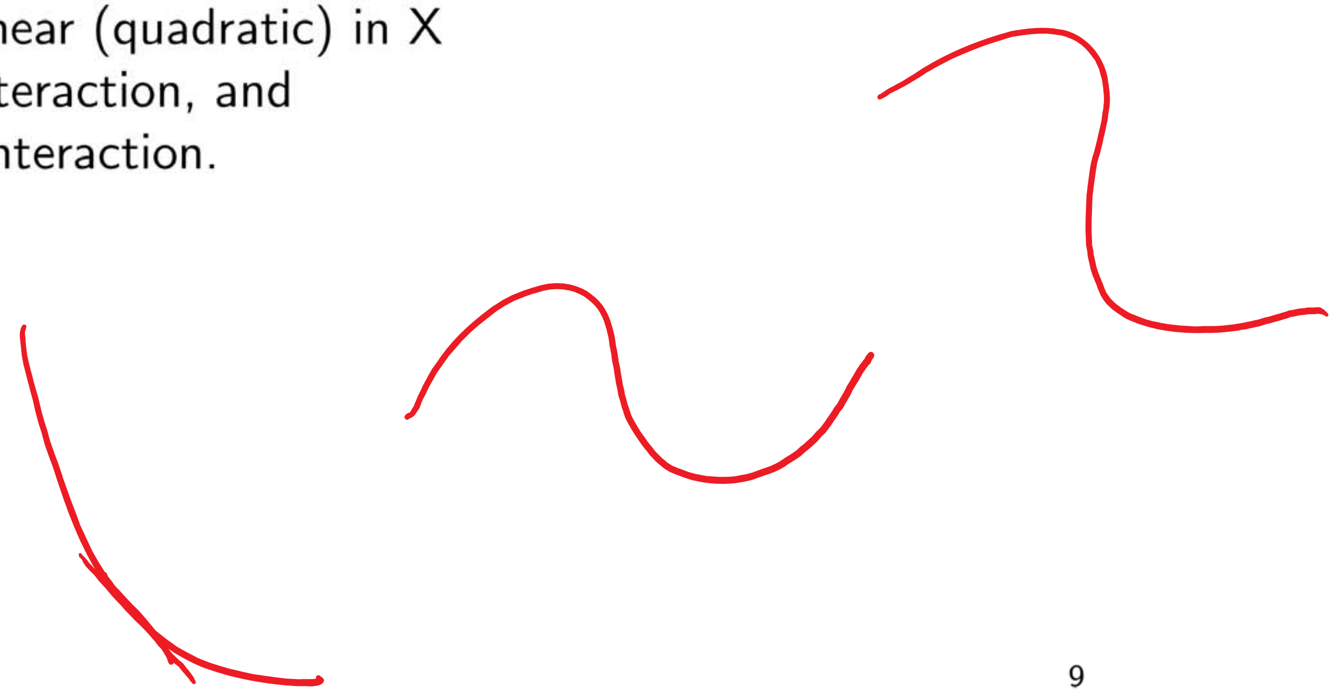
Binary Logistic Regression vs Linear Regression

- Both utilize MLE's for the β's
- *Less assumptions to check for than in linear (least squares) regression*
  - No need to check for outliers since Y is either 0 or 1.
  - No residual plots; No meaning can be inferred from residuals
  - Variance is not constant

Case Study III: Testing model assumptions

other factors:
health status

- Design — **Independence**: We know that there were families within Donner's party, so we have concerns that the observations were not independent!
- Inference — **Form of the model**: Test higher-order terms such as
  - Age^2- non-linear (quadratic) in X
  - $Sex * Age$ interaction, and
  - $Age^2 * Sex$ interaction.



nested

Comparing models: Likelihood Ratio Test

- Idea: Compare likelihood of data under FULL (F) model, $\mathcal{L}_F$ to likelihood under REDUCED (R) model, $\mathcal{L}_R$ of same data.

$$\text{Likelihood ratio} : \frac{\mathcal{L}_R}{\mathcal{L}_F}, \quad \text{where } \mathcal{L}_R \leq \mathcal{L}_F$$

$\text{logit}(\pi) = \beta_0$ — Null

$\text{logit}(\pi) = \beta_0 + \beta_1 Age + \beta_2 Sex + \beta_3 Age \times Sex + \beta_4 Age^2 + \beta_5 Age^2 \times Sex$

- Hypotheses: $H_0 : \beta_1 = \cdots = \beta_k = 0$

  (Reduced model is appropriate; fits data as well as Full model)

  H_a: at least one $\beta_1, \cdots, \beta_k \neq 0$

  (Full model is better)

- Test Statistic: Deviance (residual),

$$G^2 = -2 \log \mathcal{L}_R - (-2 \log \mathcal{L}_F) = -2 \log \left(\frac{\mathcal{L}_R}{\mathcal{L}_F} \right) \sim \chi^2_k$$

- For large n, under H_0, G^2 is an observation from a chi-square distribution with k df.

SATURATED

Case Study III Exercise: Comparing models

Using R output,

Q: Determine whether a model with the 3 higher-order polynomial terms and/or interaction terms is an improvement over the additive model.

Model 2

Model 1

- Hypotheses:

  H_0: Additive model is better, $\text{logit}(\pi) = \alpha_0 + \alpha_1 Age + \alpha_2 1_F$ $p=2$

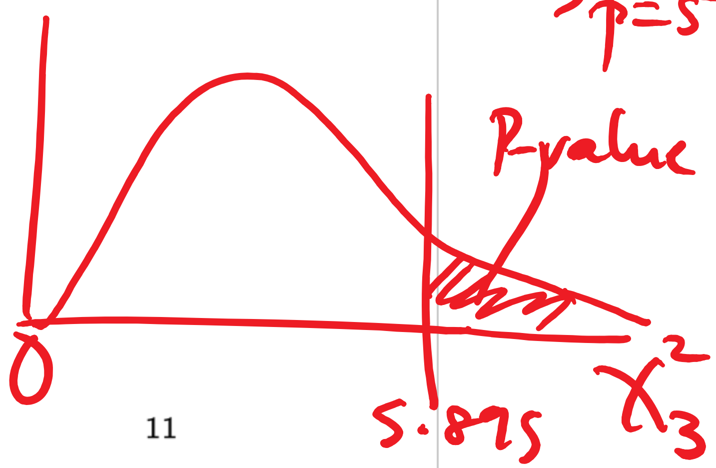
  H_a: Model 2 is better, $\text{logit}(\pi) = \beta_0 + \beta_1 Age + \beta_2 1_F + \boxed{\beta_3 Age \times 1_F + \beta_4 Age^2 + \beta_5 Age^2 \times 1_F}$ $\rightarrow p=5$

- Test Statistic:

$G^2 = \text{Deviance} = 51.256 - 45.361 = 5.895 \sim \chi^2_3$

- Distribution of TS:
- P-value: $P(\chi^2_3 \geq 5.895) = 0.1168$
- Conclusion:

Evidence that the additive model is a better fit compared to the higher-order model.



Testing β's: Wald versus LRT test

	Wald	LRT
Testing whether a single β=0	✓	✓
Comparing nested models		✓
Small to moderate sample sizes β near boundary of parameter space		✓

Can compare

MLE's

Beware of regularity conditions

$0 < \hat{\pi} < 1$

Case Study III Exercise: Comparing models

Using R output,

Q: Determine whether the effect of *Age* on the odds of survival differ with *Sex*.

- Hypotheses:

  $H_0: \text{logit}(\hat{\pi}) = \hat{\alpha}_0 + \hat{\alpha}_1 Age + \hat{\alpha}_2 1_F$

  $H_a: \text{logit}(\hat{\pi}) = \hat{\beta}_0 + \hat{\beta}_1 Age + \hat{\beta}_2 1_F + \hat{\beta}_3 Age \times 1_F$

- Test Statistic:

  Wald: $Z^2 = 2.94$

  LRT: $G^2 = 51.256 - 47.346 = 3.91$

- Distribution of TS:

  Wald: $\sim \chi^2_1$

  LRT: $\sim \chi^2_1$

- P-value:

  Wald: $P(\chi^2_1 > 2.94) = 0.0865$

  LRT: $P(\chi^2_1 > 3.91) = 0.048$

- Conclusion:

Inconclusive evidence that the Additive model is better.

Comparing models: 'Global' LRT

do not need any of the predictors.

- Idea: Compares Fitted model to NULL [$\text{logit}(\pi) = \beta_0$] model
- Hypotheses: $H_0 : \beta_1 = \cdots = \beta_p = 0$

  (NULL model is appropriate)

  H_a: at least one $\beta_1, \cdots, \beta_p \neq 0$

  (Fitted model is better)

$$\text{logit}(\pi) = \beta_0 + \underbrace{\beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p}_{p}.$$

Case Study III Exercise: 'Global' LRT

Using R output,
Q: Determine whether or not the additive model fits better than the Null model.

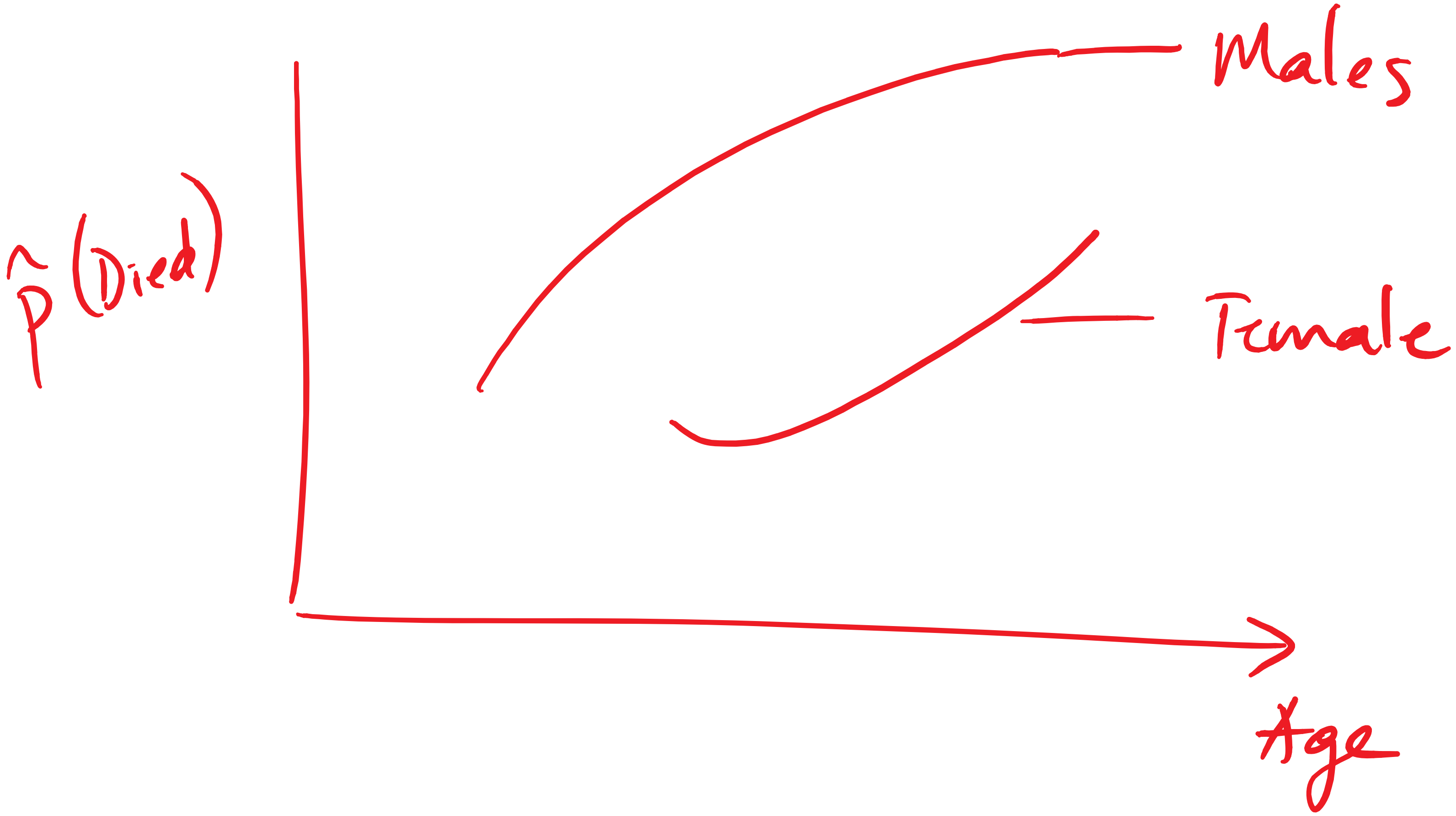
H_0: Null is better

- Hypotheses: H_a: Additive is better fit
- Test Statistic: $G^2 = 61.827 - 51.256 = 10.571$
- Distribution of TS: χ^2_2
- P-value: $P(\chi^2_2 > 10.571) = 0.005$ (In R).
- Conclusion:

Strong evidence that the fitted model is better than the NULL model

Plot

Q: How would the plot of estimated probabilities change if we modelled probability of death rather than survival?



Over 50yrs

Q: Should one be reluctant to draw conclusions about the ratio of male and female odds of survival for the Donner Party members over 50?

Yes; no females older than 50.

Other Model Fit Statistics

- Two popular fit statistics: AIC and BIC; combines log-likelihood with a penalty.
- Useful for comparing models with same response and same data
- Extends from normal regression to GLMs
  1. Akaike's Information Criterion (AIC)

$$\boxed{AIC = -2 \log \mathcal{L} + 2(p+1)}$$

  2. Schwarz's (Bayesian Information) Criterion (BIC)

$$BIC = -2 \log \mathcal{L} + (p+1) \log N$$

where

- p-number of explanatory variables, and
- N=sample size

- smaller is better

When is AIC = BIC?

If $\log N = 2$

$\Rightarrow N = e^2$
$= 7.3$

O.W: BIC > AIC

Model Fit Statistics: AIC and BIC

- Smaller is better!
- BIC applies stronger penalty for model complexity than AIC
- AIC Rule of Thumb:
  - One model fits **better** than another if difference in AIC's > 10
  - One model model is essentially **equivalent** to another if the difference in AIC's < 2

In R: BIC("fitted model").
AIC("fitted").

Using AIC: Case Study III Example

- Fitted models are based on same response and data.
- Based on AIC, choose a 'best' model.

Model	Variables	AIC	BIC
1	{age,sex}	57.256	62.676
2	{age,sex,age*sex,age^2,age^2*sex}	57.361	68.201
3	{age,sex,age*sex,age^2 }	55.830	64.863
4	{age,sex,age*sex}	$\boxed{55.346}$	$\boxed{62.573}$

Results:

- Difference in AIC between 1 and 3 is within 2
- There is some indication that 2 is worse than 3 and 4.
- Choose Model 1 (the simplest)

Related R packages and functions

- Packages:
  - `aod`: analysis of over-dispersed data
  - `ggplot2`: graphics
  - `Sleuth3`: data sets for Ramsey and Schafer's text
  - `effects`: effects displays for GLM and other models
- Functions:
  - create a factor: `as.factor()`
  - cross Tabulations: `xtabs()`
  - specifying the reference level: `relevel()`
  - generalized linear models: `glm()`
  - find deviance: `deviance()`
  - confidence interval: `confint()`
  - model coefficients: `coef()`
  - variance-covariance matrix: `vcov()`
  - `wald.test()`
  - `AIC()`
  - `BIC()`