Confidence intervals

Last time: Wald confidence intervals

Method: $\hat{\beta} \stackrel{\sim}{\approx} N(\beta, \mathcal{I}^{-1}(\beta)) \Rightarrow \hat{\beta}_i \stackrel{\sim}{\approx} N(\beta_i, \underbrace{\mathcal{I}^{-1}(\beta)_{ii}}_{i^{th} \text{ diagonal entry}})$

Wald CI: $\hat{\beta}_i \pm \underbrace{z_{1-\frac{\alpha}{2}}}_{1-\frac{\alpha}{2} \text{ quantile of } N(0,1)} \sqrt{\mathcal{I}^{-1}(\hat{\beta})_{ii}}$

e.x. 95% CI $\Rightarrow$ 1.96

Example: Titanic data

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 Sex_i + \beta_2 Age_i + \beta_3 SecondClass_i + \beta_4 FirstClass_i + \beta_5 Sex_i \cdot Age_i$$

```
...
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.408232   0.330916   1.234 0.217337
## Sexmale     -1.163444   0.437622  -2.659 0.007848 **
## Age         -0.007186   0.011684  -0.615 0.538522
## Pclass2      1.191858   0.243233   4.900 9.58e-07 ***
## Pclass1      2.697561   0.295822   9.119  < 2e-16 ***
## Sexmale:Age -0.049851   0.014782  -3.373 0.000745 ***
...
```

95% CI for β_3: $1.192 \pm 1.96(0.243) = (0.716, 1.668)$

Test $H_0: \beta_3 = b$ $\quad$ $H_A: \beta_3 \neq b$

$1-\alpha$ CI for β_3 $= \{ b : \text{fail to reject } H_0: \beta_3 = b \text{ at level } \alpha \}$

95% CI for β_3: $[0.716, 1.668]$

$H_0: \beta_3 = 0.716$

$H_A: \beta_3 \neq 0.716$

$$z = \frac{1.192 - 0.716}{0.243} = 1.96$$

p-value $= 0.05$

$H_0: \beta_3 = 1.668$

$H_A: \beta_3 \neq 1.668$

$z = -1.96$ $\quad$ p-value $= 0.05$

$H_0: \beta_3 = 1$

$H_A: \beta_3 \neq 1$

$$z = \frac{1.192 - 1}{0.243} = 0.192$$

p-value $= 0.848 > 0.05$

$H_0: \beta_3 = \hat{\beta}_3$

$H_A: \beta_3 \neq \hat{\beta}_3$

$$Z = \frac{0}{SE(\hat{\beta})} = 0$$

p-value = 1

Confidence intervals for linear combinations

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be an observed set of data

Model: $Y_i \sim \text{Bernoulli}(p_i)$

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta^T X_i \qquad \beta \in \mathbb{R}^{k+1}$$

Let $\theta = a^T \beta \qquad a \in \mathbb{R}^{k+1}$

we want $1-\alpha$ CI for θ

1) $\hat{\theta} = a^T \hat{\beta}$

2) $\hat{\beta} \approx N(\beta, \mathcal{I}^{-1}(\beta)) \implies \hat{\theta} \approx N(\underbrace{a^T\beta}_{\theta}, a^T \mathcal{I}^{-1}(\beta) a)$

3) $1-\alpha$ Wald CI: $\hat{\theta} \pm z_{1-\frac{\alpha}{2}} \sqrt{\text{Var}(\hat{\theta})}$

$$= a^T \hat{\beta} \pm z_{1-\frac{\alpha}{2}} \sqrt{a^T \mathcal{I}^{-1}(\hat{\beta}) a}$$

contrasts: special case where $\sum a_i = 0$

Class activity

https://sta712-f22.github.io/class_activities/ca_lecture_16.html

95% CI for $\beta_4 - \beta_3$: $(0.913, 2.098)$

we are 95% confident that the true difference in log odds of survival between first & second class passengers is between 0.913 and 2.098, holding sex & age fixed

odds scale : $(e^{0.913}, e^{2.098}) = (2.492, 8.150)$

95% CI for log odds of survival for a female, 20 yr old, 2^{nd} class passenger

est. log odds $= a^T \hat{\beta}$ $\qquad a^T = (1, 0, 20, 1, 0, 0)$

Interval: $(0.982, 1.931)$

what about the probability?

$$\left(\frac{e^{0.982}}{1+e^{0.982}}, \frac{e^{1.931}}{1+e^{1.931}} \right) = (0.728, 0.873)$$

This works b/c $\frac{e^x}{1+e^x}$ (or odds, just e^x) is bijective on the right range

Inverting the likelihood ratio test

Example: Normal data

$$X_1, \ldots, X_n \sim N(\mu, 1)$$

$$H_0: \mu = \mu_0$$
$$H_A: \mu \neq \mu_0$$

$$L(\mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(X_i-\mu)^2\right\}$$

$$\Rightarrow \ell(\mu) = n \log\left(\frac{1}{\sqrt{2\pi}}\right) - \frac{1}{2}\sum_{i=1}^{n}(X_i-\mu)^2$$

LRT: $2\ell(\bar{X}) - 2\ell(\mu_0) \sim \chi^2_1$

$$2\ell(\bar{X}) - 2\ell(\mu_0) = \sum_{i=1}^{n}(X_i-\mu_0)^2 - \sum_{i=1}^{n}(X_i-\bar{X})^2$$

$$\sum_{i=1}^{n}(X_i-\bar{X})^2 = \sum_{i=1}^{n}X_i^2 - 2\bar{X}\sum_{i=1}^{n}X_i + \sum_{i=1}^{n}(\bar{X})^2 = \sum_{i=1}^{n}X_i^2 - n(\bar{X})^2$$

$$\sum_{i=1}^{n}(X_i-\mu_0)^2 = \sum_{i=1}^{n}X_i^2 - 2n\bar{X}\mu + n\mu^2$$

$$\Rightarrow \quad 2\ell(\bar{X}) - 2\ell(\mu_0) = n(\mu_0^2 - 2\mu\bar{X} + (\bar{X})^2)$$

$$= n(\bar{X} - \mu_0)^2$$

$\Rightarrow$ reject when $n(\bar{X} - \mu_0)^2$ is large

or, reject when $\left|\sqrt{n}(\bar{X} - \mu_0)\right|$ is large

$$\sqrt{n}(\bar{X} - \mu_0) \sim N(0,1)$$

$\Rightarrow$ reject when $\sqrt{n}(\bar{X} - \mu_0) > z_{1-\frac{\alpha}{2}}$ or $< -z_{1-\frac{\alpha}{2}}$

So, inverting the LRT for a normal distribution is equivalent to the Wald confidence interval!

Example: Exponential data

$X_1, \dots, X_n \sim \text{Exponential}(\lambda)$ $\qquad f(x) = \lambda e^{-\lambda x}$

$$L(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} \implies \ell(\lambda) = \sum_{i=1}^{n} (\log \lambda - \lambda x_i)$$

$$\implies \frac{d}{d\lambda} \ell(\lambda) = \sum_{i=1}^{n} \left(\frac{1}{\lambda} - x_i \right) \overset{\text{set}}{=} 0 \implies \hat{\lambda} = \frac{1}{\sum_i x_i}$$

Test: $H_0 : \lambda = \lambda_0$

$H_A : \lambda \neq \lambda_0$

$$\text{LRT: } 2\ell(\hat{\lambda}) - 2\ell(\lambda_0)$$

$$= 2(n \log \hat{\lambda} - \hat{\lambda} \Sigma x_i) - 2(n \log \lambda_0 - \lambda_0 \Sigma x_i)$$

$$= 2\left(n \log\left(\frac{1}{\bar{x}}\right) - 1\right) - 2(n \log \lambda_0 - \lambda_0 \Sigma x_i)$$

$$= 2(-n \log(\bar{x}) - 1 - n \log \lambda_0 + \lambda_0 \Sigma x_i)$$

reject when $2(-n \log(\bar{x}) - 1 - n \log \lambda_0 + \lambda_0 \Sigma x_i) > \chi^2_{1, 1-\alpha}$

We can numerically determine values λ_0 for which we reject, but there isn't a nice closed-form solution

Types of research questions

So far, we have learned how to answer the following questions:

- What is the relationship between the explanatory variable(s) and the response?
- What is a "reasonable range" for a parameter in this relationship?
- Do we have strong evidence for a relationship between these variables?

> What other kinds of research questions might we ask?

Making predictions

- For each passenger, we calculate $\hat{p}_i$ (estimated probability of survival)
- But, we want to predict *which* passengers actually survive

> How do we turn $\hat{p}_i$ into a binary prediction of survival / no survival?

Confusion matrix

		Actual	
		$Y = 0$	$Y = 1$
Predicted	$\widehat{Y} = 0$	344	70
	$\widehat{Y} = 1$	80	220

> Did we do a good job predicting survival?