# Wald tests

# Where we're going

So far:

1. How can we estimate parameters / fit a model?
    - Maximum likelihood estimation
    - Fisher scoring
2. Was our fitted model good?
    - Logistic regression diagnostics

Currently:

3. How can we use our fitted model for inference?
    - Convergence of MLEs
    - Wald tests

Coming up:

4. How else can we test hypotheses? What about confidence intervals?
5. Why did we focus so much on MLEs?

# Formal definition

Wald test for one parameter

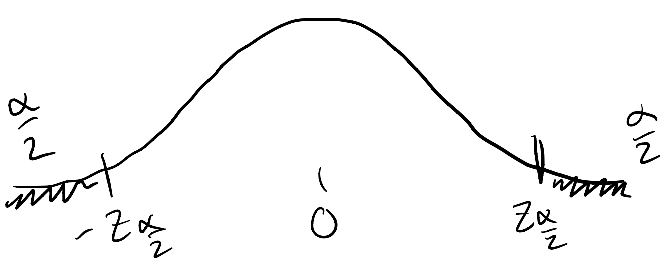
Let $\theta \in \mathbb{R}$ be a parameter of interest, and let $\hat{\theta}_n$ be an estimator for which $\frac{\hat{\theta}_n - \theta}{S_n} \xrightarrow{d} N(0,1)$, for some sequence $S_n$

$(S_n^2 \approx Var(\hat{\theta}_n))$

To test $H_0: \theta = \theta_0$ vs. $H_A: \theta \neq \theta_0$

Let $Z_n = \frac{\hat{\theta}_n - \theta_0}{S_n}$. The Wald test rejects $H_0$

when $|Z_n| > z_{\frac{\alpha}{2}}$ where $z_{\frac{\alpha}{2}}$ is the upper $\frac{\alpha}{2}$ quantile of $N(0,1)$



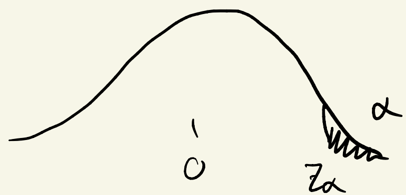
## Comments

- Under $H_0$ $(\theta = \theta_0)$, $Z_n \approx N(0,1)$ (if $n$ is sufficiently large)

So $P(\text{reject } H_0 \mid H_0 \text{ is true}) \approx P(|Z| > z_{\frac{\alpha}{2}})$ $\quad Z \sim N(0,1)$

$= \alpha$

- To test $H_A: \theta > \theta_0$, reject when $Z_n > z_\alpha$




$\Rightarrow P(\text{reject } H_0 \mid H_0 \text{ is true})$

$= P(Z > z_\alpha) = \alpha$

To test $H_A: \theta < \theta_0$, reject when $Z_n < -z_\alpha$

- Any asymptotically normal statistic can be used to construct a Wald test

# Hypothesis tests for a population mean

Let $Y_1, Y_2, \ldots, Y_n$ be an iid sample from a population with mean $\mu$ and variance $\sigma^2$. We want to test

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

CLT: $\quad \dfrac{\bar{Y}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1)$

Wald test: $\quad Z_n = \dfrac{\bar{Y}_n - \mu_0}{\sigma/\sqrt{n}}$

what if $\sigma$ is unknown?

$$Z_n = \frac{\bar{Y}_n - \mu_0}{\hat{\sigma}/\sqrt{n}} \xrightarrow{d} N(0,1) \quad (\text{under } H_0)$$

$$\hat{\sigma} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n} (Y_i - \bar{Y}_n)^2}$$

or $\quad \hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (Y_i - \bar{Y}_n)^2}$

Proof: Slutsky's

$\left( \dfrac{\sigma}{\hat{\sigma}} \xrightarrow{p} 1 \right)$

# Hypothesis tests for a population proportion

Let $Y_1, Y_2, \ldots, Y_n \overset{iid}{\sim} Bernoulli(p)$. We want to test

$$H_0 : p = p_0 \qquad H_A : p \neq p_0$$

What is our Wald test statistic?

$$Z_n = \frac{\hat{p} - p_0}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

$$\hat{p} = \frac{1}{n}\hat{\sum_{i=1}} Y_i = \bar{Y}_n$$

$$Var(Y_i) = p(1-p) \quad \Rightarrow \text{CLT:} \quad \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \xrightarrow{d} N(0,1)$$

Alternatively: $\frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$ under $H_0$, both $\xrightarrow{d} N(0,1)$

# Testing multiple parameters

Logistic regression model for the dengue data:

$$Y_i \sim Bernoulli(p_i)$$

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 WBC_i + \beta_2 PLT_i$$

> Researchers want to know if there is any relationship between white blood cell count or platelet count, and the probability a patient has dengue. What hypotheses should they test?

$H_0: \beta_1 = \beta_2 = 0$

$H_A$: at least one of $\beta_1, \beta_2 \neq 0$

# Testing multiple parameters

```
m1 <- glm(Dengue ~ WBC + PLT, data = dengue,
          family = binomial)
summary(m1)
```

```
...
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.6415063  0.1213233   21.77   <2e-16 ***
## WBC         -0.2892904  0.0134349  -21.53   <2e-16 ***
## PLT         -0.0065615  0.0005932  -11.06   <2e-16 ***
## ---
...
```

> Can the researchers test their hypotheses using this output?

## Wald tests for multiple parameters

$$\mathcal{I}^{-1}(\beta) = \begin{bmatrix} Var(\hat{\beta}_0) & \cdots & \cdots \\ \vdots & \boxed{\begin{matrix} Var(\hat{\beta}_1) & \cdots \\ \vdots & \ddots \end{matrix}} \end{bmatrix}$$

$C\mathcal{I}^{-1}(\beta)C^T$

We know $\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} \approx N(\beta, \mathcal{I}^{-1}(\beta))$

We want to test $\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$

Notice that $\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{C} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = C\beta$

$$\Rightarrow \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = C\hat{\beta}$$

$$\underset{\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}}{\overset{\shortparallel}{C\hat{\beta}}} \approx N(\underset{\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}}{\overset{\shortparallel}{C\beta}}, \; C\mathcal{I}^{-1}(\beta)C^T)$$

If $Z \sim N(0,1)$, then $Z^2 \sim \chi^2_1$ (HW)

$\Rightarrow$ $Z_n = \dfrac{\hat{\theta} - \theta_0}{S_n}$, rejecting when $|Z_n| > z_{\frac{\alpha}{2}}$ is equivalent to rejecting when

$$Z_n^2 > \chi^2_{1,\alpha}$$

where $\chi^2_{1,\alpha}$ = upper $\alpha$ quantile of $\chi^2_1$

If $Z \in \mathbb{R}^q$ and $Z \sim N(0, I)$, then $Z^T Z \sim \chi^2_q$

# Class activity

https://sta711-s23.github.io/class_activities/ca_lecture_17.html

- Wald tests for the dengue data