# Lecture 17: Wald Tests

# Last time

Under regularity conditions, if $\hat{\theta}_n$ is a MLE of $\theta$ then

- $\hat{\theta}_n \xrightarrow{p} \theta$
- $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, I_1^{-1}(\theta))$

Counterexample: Suppose we measure some outcome for $n$ individuals, and we take two measurements for each individual

$$Y_{11}, Y_{12} \sim N(\mu_1, \sigma^2)$$
$$Y_{21}, Y_{22} \sim N(\mu_2, \sigma^2)$$
$$Y_{31}, Y_{32} \sim N(\mu_3, \sigma^2)$$
$$\vdots \qquad \vdots$$
$$Y_{n1}, Y_{n2} \sim N(\mu_n, \sigma^2)$$

(different means for each individual)

We want $\hat{\sigma}^2$. Some math: $\hat{\sigma}^2 = \frac{1}{4n} \sum_{i=1}^{n} (Y_{i1} - Y_{i2})^2 \xrightarrow{p} \frac{\sigma^2}{2}$

not to $\sigma^2$. Don't have consistency, because # parameters are growing with $n$

# Where we are going

So far:

- How can we estimate parameters / fit a model?
  - MLE
  - Fisher scoring
- Asymptotic properties of MLEs

Next:

- How can we use our estimates for inference?
  - Wald tests

Future:

- General hypothesis testing framework
- Other ways of testing hypotheses
- Confidence intervals
- GLM diagnostics

# Hypothesis test for a population mean

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

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

value under $H_0$

Test statistic:

$$Z_n = \frac{\bar{Y}_n - \mu_0}{\sigma/\sqrt{n}} \quad (\text{if } \sigma \text{ is known})$$

$$Z_n = \frac{\bar{Y}_n - \mu_0}{s/\sqrt{n}} \quad \text{or } t = \quad (\text{if } \sigma \text{ is unknown})$$

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

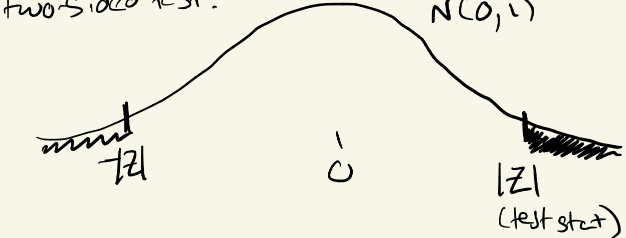
$$Z_n = \frac{\bar{Y}_n - \mu_0}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1)$$ (by CLT, under mild assumptions) if $H_0$ is true

CLT: $$\frac{\bar{Y}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1)$$ ; if $\mu = \mu_0$ ($H_0$ is true)

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

Test stat: $Z_n = \dfrac{\bar{X}_n - \mu_0}{\sigma / \sqrt{n}}$

two-sided test:




if $H_0$ is true: $Z_n \approx N(0,1)$

<u>p-value</u>: (probability of "observed data or more extreme", if $H_0$ is true)

observed test statistic (from data) is $z$

$$\text{p-value} = P(|Z_n| \geq |z| \mid H_0 \text{ is true})$$
$$= P(|Z_n| \geq |z| \mid \mu = \mu_0)$$

if $\mu = \mu_0$, then $Z_n \approx N(0,1)$

$$\text{p-value} \approx 2\,\Phi(-|z|)$$

cdf of $N(0,1)$

what if $\sigma$ is unknown?

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

$$Z_n = \frac{\bar{Y}_n - \mu_0}{S/\sqrt{n}}$$

(typically use a t-distribution, but for $n$ sufficiently large $n$, $N(0,1)$ will work)

know that $\frac{\bar{Y}_n - \mu_0}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1)$

want to show $\frac{\bar{Y}_n - \mu_0}{S/\sqrt{n}} \xrightarrow{d} N(0,1)$ also

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

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i^2 - 2Y_i\bar{Y}_n + \bar{Y}_n^2) = \frac{n}{n-1} \cdot \frac{1}{n} \sum_{i=1}^{n} (Y_i^2 - 2Y_i\bar{Y}_n + \bar{Y}_n^2)$$

$$\frac{1}{n} \sum_{i=1}^{n} Y_i^2 \xrightarrow{P} \mathbb{E}[Y_i^2] \quad \text{(WLLN)}$$

$$-\frac{2}{n} \sum_i Y_i\bar{Y}_n = -2\bar{Y}_n \cdot \left(\frac{1}{n}\sum_i Y_i\right) = -2\bar{Y}_n^2$$

$$\frac{1}{n} \sum_i \bar{Y}_n^2 = \bar{Y}_n^2$$

$$\mathbb{E}[S^2] = \sigma^2$$

$$\mathbb{E}\left[\frac{n-1}{n} S^2\right] = \sigma^2\left(\frac{n-1}{n}\right)$$

(slightly biased)

$$S^2 = \frac{n}{n-1}\Big( \underbrace{\frac{1}{n}\sum_i Y_i^2}_{\xrightarrow{p} \mathbb{E}[Y_i^2] \text{ (WLLN)}} - \underbrace{\bar{Y}_n^2}_{\xrightarrow{p} (\mathbb{E}[Y_i])^2 \text{ (WLLN + CMT)}} \Big)$$

$\bar{Y}_n \xrightarrow{p} \mathbb{E}[Y_i]$ (WLLN)

$\bar{Y}_n^2 \xrightarrow{p} (\mathbb{E}[Y_i])^2$ (continuous mapping theorem)

$$\xrightarrow{p} \mathbb{E}[Y_i^2] - (\mathbb{E}[Y_i])^2 = \sigma^2 \quad \text{(Slutsky's theorem)}$$

$$S^2 = \underbrace{\frac{n}{n-1}}_{\to 1}\Big( \xrightarrow{p} \sigma^2 \Big)$$

$$S^2 \xrightarrow{p} 1 \cdot \sigma^2 = \sigma^2$$

$$S \xrightarrow{p} \sigma \quad \text{(CMT)}$$

$$\frac{\bar{Y}_n - \mu_0}{S/\sqrt{n}} = \underbrace{\frac{\bar{Y}_n - \mu_0}{\sigma/\sqrt{n}}}_{\xrightarrow{d} N(0,1) \text{ under } H_0} \cdot \overbrace{\frac{\sigma}{S}}^{\xrightarrow{p} 1} \xrightarrow{d} N(0,1) \cdot 1 = N(0,1)$$

(Slutsky's again!)

(under $H_0$)

# Hypothesis test for a population proportion

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

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

# Wald test for one parameter