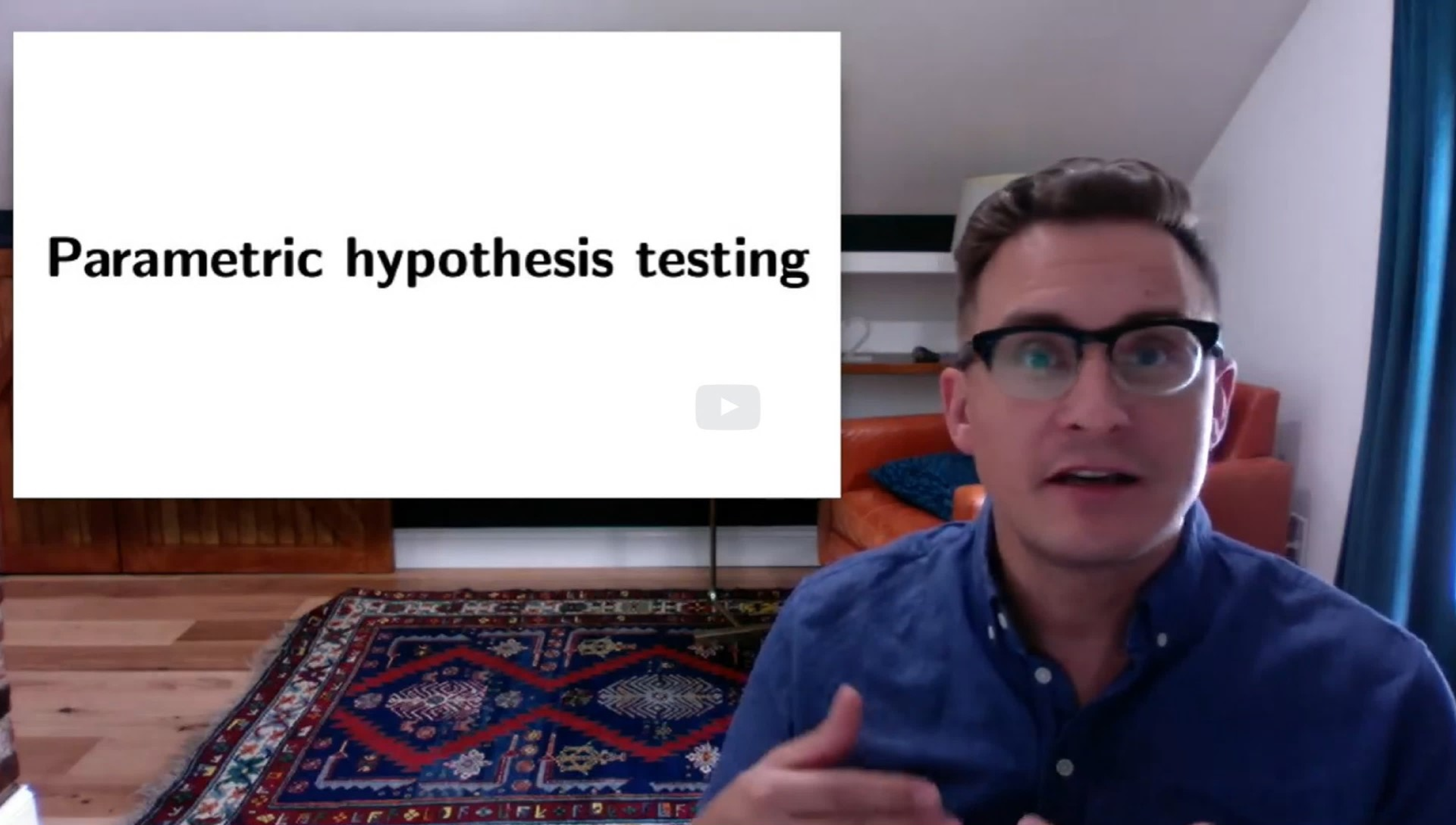


0:23 / 3:39

Speed 1.0x

- Given the duality between confidence intervals (CI) and tests, it is not surprising that the same tools will be used.
- A simple approach: first build CI, then deduce a test is nice but *limited*: one/two-sided, two sample tests are more common than confidence intervals for (say) $\mu_{\mathrm{d}} > \mu_{\mathrm{c}}$.
- Easier to unfold the same machinery: this is the principle behind the WALD TEST
- Wald's test only guarantees *asymptotic* level. An alternative is the T - test

The Wald test (1)

- Statistical model $(E, \{\mathbb{P}_\theta\}_{\theta \in \Theta})$
- Estimator $\hat{\theta}$ such that $\boxed{\dfrac{\hat{\theta} - \theta}{\sqrt{\widehat{\mathrm{var}}(\hat{\theta})}} \xrightarrow[n \to \infty]{(d)} \mathcal{N}(0,1)}$

  where $\widehat{\mathrm{var}}(\hat{\theta})$ is an estimator of the variance of $\hat{\theta}$
- For example, in the Bernoulli case, $\widehat{\mathrm{var}}(\hat{p}) = \dfrac{\hat{p}(1-\hat{p})}{n}$.

p

$\hat{p} = \overline{X}_n \qquad \widehat{\mathrm{var}}(\hat{p}) = \dfrac{\hat{p}(1-\hat{p})}{n} \qquad \dfrac{\hat{p} - p}{\sqrt{\widehat{\mathrm{var}}(\hat{p})}} = \sqrt{n}\,\dfrac{\hat{p} - p}{\sqrt{\hat{p}(1-\hat{p})}} \xrightarrow[n\to\infty]{(d)} \mathcal{N}(0,1)$

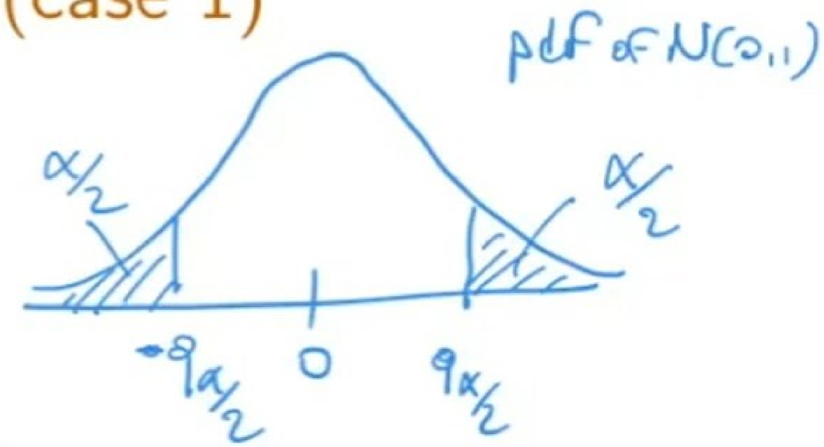
3:14 / 6:32 Speed 1.0x

The Wald test (2)

	$H_0: \theta = \theta_0$ $H_1: \theta \neq \theta_0$	$H_0: \theta \leq \theta_0$ $H_1: \theta > \theta_0$	$H_0: \theta \geq \theta_0$ $H_1: \theta < \theta_0$
Wald Test ψ	$\mathbb{1}\{\lvert W \rvert > q_{\alpha/2}\}$	$\mathbb{1}\{W > q_{\alpha}\}$	$\mathbb{1}\{W < -q_{\alpha}\}$

$$W := \frac{\hat{\theta} - \theta_0}{\sqrt{\widehat{\mathsf{var}}(\hat{\theta})}}$$

Asymptotic level of the Wald test (case 1)



If

$$\begin{aligned} H_0 &: \quad \theta = \theta_0 \\ H_1 &: \quad \theta \neq \theta_0 \end{aligned}$$

Then, for any $\theta = \theta_0$,

$$\lim_{n\to\infty} \mathbb{P}_{\theta_0}[\psi = 1] = \lim_{n\to\infty} \mathbb{P}_{\theta_0}\left[|W| > q_{\alpha/2}\right] = \mathbb{P}\left[|Z| > q_{\alpha/2}\right] = \alpha$$

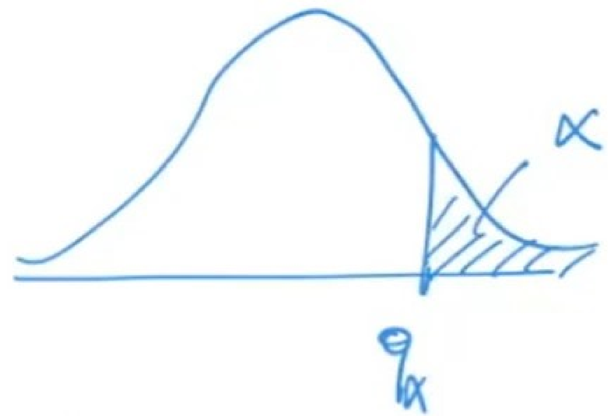
Note that it is important to take the same θ_0 in $\mathbb{P}_{\theta_0}$ and W!

$$W = \frac{\hat{\theta} - \theta_0}{\sqrt{\widehat{\mathrm{Var}}(\hat{\theta})}} \xrightarrow[n\to\infty]{(d)} \overset{Z}{\mathcal{N}(0,1)}$$

Asymptotic level of the Wald test (case 2 & 3)

If

$$H_0: \quad \theta \leq \theta_0$$
$$H_1: \quad \theta > \theta_0$$



Then, for any $\theta \leq \theta_0$,

$$\lim_{n\to\infty} \mathbb{P}_\theta[\psi = 1] = \lim_{n\to\infty} \mathbb{P}_\theta[W > q_\alpha]$$

$$= \lim_{n\to\infty} \mathbb{P}_\theta\left[\frac{\hat\theta - \theta_0}{\sqrt{\widehat{\mathrm{var}}(\hat\theta)}} > q_\alpha\right]$$

$$\leq \lim_{n\to\infty} \mathbb{P}_\theta\left[\frac{\hat\theta - \theta}{\sqrt{\widehat{\mathrm{var}}(\hat\theta)}} > q_\alpha\right]$$

$$= \mathbb{P}[Z > q_\alpha] = \alpha$$

$$\frac{\hat\theta - \theta_0}{\sqrt{\widehat{\mathrm{var}}(\hat\theta)}} = \frac{\hat\theta - \theta}{\sqrt{\widehat{\mathrm{var}}(\hat\theta)}} + \underbrace{\boxed{\frac{\theta - \theta_0}{\sqrt{\widehat{\mathrm{var}}(\hat\theta)}}}}_{\leq 0}$$

$$\underbrace{\frac{\hat\theta - \theta_0}{\sqrt{\widehat{\mathrm{var}}(\hat\theta)}}}_{\geq q_\alpha} \leq \frac{\hat\theta - \theta}{\sqrt{\widehat{\mathrm{var}}(\hat\theta)}}$$

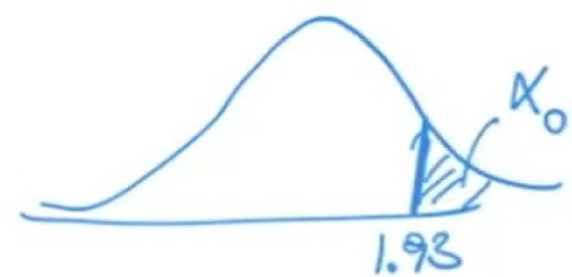
Example 1: News

More than 2/3 of Americans get news on social media

Is this quote from a 2018 Pew Research Center study justified?

$X_1, \ldots, X_n \overset{iid}{\sim} \mathsf{Ber}(p)$, $p \in [0,1]$,

$$H_0 : \quad p \leq 2/3$$
$$H_1 : \quad p > 2/3$$



This claim is based on $n = 4{,}581$ randomly sampled U.S., $\hat{p} = .68. = \overline{X_n}$

X_n
$\shortparallel$

95%

$$W^{\text{obs}} = \sqrt{4{,}581}\,\frac{.68 - 2/3}{\sqrt{.68(1-.68)}} = 1.93 > 1.645 \text{ so Reject at asymptotic level } 5\%$$

The p-value is α_0 such that

$$q_{\alpha_0} = 1.93 \iff \alpha_0 = \mathbb{P}(Z > 1.93) = 1 - 0.9732 = 2.68\%$$

Fail to reject at asymptotic level $\underline{\alpha = 1\%}$.

p-values for the Wald test

- Recall that $W := \dfrac{\hat{\theta} - \theta_0}{\sqrt{\widehat{\mathsf{var}}(\hat{\theta})}}$.
- Denote by W^{obs} the realization (observed value) of W in a given example. For the News example, $W^{\text{obs}} = 1.93$
- Then p-values and asymptotic p-values are given by

	$H_0: \theta = \theta_0$ $H_1: \theta \neq \theta_0$	$H_0: \theta \leq \theta_0$ $H_1: \theta > \theta_0$	$H_0: \theta \geq \theta_0$ $H_1: \theta < \theta_0$
Wald test	$\lvert W \rvert > q_{\alpha/2}$	$W > q_\alpha$	$W < -q_\alpha$
p-value	$\mathbb{P}(\lvert W \rvert > \lvert W^{obs} \rvert)$	$\mathbb{P}(W > W^{obs})$	$\mathbb{P}(W < W^{obs})$
asymp. p-value	$\mathbb{P}(\lvert Z \rvert > \lvert W^{obs} \rvert)$	$\mathbb{P}(Z > W^{obs})$	$\mathbb{P}(Z < W^{obs})$

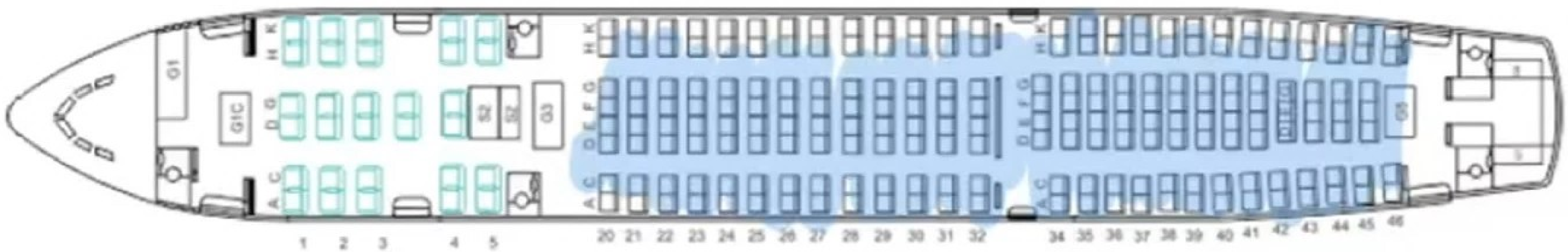
where $Z \sim \mathcal{N}(0,1)$

Example 2: How to board a plane?

What is the fastest method to board a plane?

R2F or WiIMA?

- R2F= Rear to Front (JetBlue)



- WiIMA=Window, Middle, Aisle (United)

Model and Assumptions

R2F

- X: boarding time of a random JetBlue flight.

$$\mathbb{E}[X] = \mu_1, \qquad \text{var}[X] = \sigma_1^2$$

WilMA

- Y: boarding time of a random United flight.

$$\mathbb{E}[Y] = \mu_2, \qquad \text{var}[Y] = \sigma_2^2$$

- We have $X_1, \ldots, X_n$ independent copies of X and $Y_1, \ldots Y_m$ independent copies of Y.
- We further assume that the two samples are independent

Is there a difference between the two boarding methods:

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$

Equivalently, write $\theta = \mu_1 - \mu_2$, we get

$$H_0: \theta = 0$$
$$H_1: \theta \neq 0$$

We have two samples: this is a two-sample testing problem.

Asymptotically normal estimator for θ

- Define the estimator $\hat{\theta} = \bar{X}_n - \bar{Y}_m$
- We have by the CLT:

$$\frac{\hat{\theta} - \theta}{\sqrt{\mathrm{var}(\hat{\theta})}} \xrightarrow[n \to \infty]{(d)} \mathcal{N}(0, 1)$$

- But: $\mathrm{var}(\hat{\theta}) = \mathrm{var}(\bar{X}_n) + \mathrm{var}(\bar{Y}_m) = \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}$
- We can estimate σ_1^2 by $\hat{\sigma}_1^2$ and σ_2^2 by $\hat{\sigma}_2^2$ where

$$\hat{\sigma}_1^2 := \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2 \qquad \hat{\sigma}_2^2 := \frac{1}{m} \sum_{i=1}^{m} (Y_i - \bar{Y}_m)^2$$

$\sigma_1^2 = \mathbb{E}[(X - \mathbb{E}[X])^2]$

- Both estimators are consistent so by Slutsky

$\widehat{\mathrm{var}}(\hat{\theta}) = \frac{\hat{\sigma}_1^2}{n} + \frac{\hat{\sigma}_2^2}{m}$

$$\frac{\hat{\theta} - \theta}{\sqrt{\frac{\hat{\sigma}_1^2}{n} + \frac{\hat{\sigma}_2^2}{m}}} \xrightarrow[\substack{n \to \infty \\ m \to \infty}]{(d)} \mathcal{N}(0, 1)$$

Applying the Wald test

$$W = \frac{\hat{\theta} - 0}{\sqrt{\frac{\hat{\sigma}_1^2}{n} + \frac{\hat{\sigma}_2^2}{m}}} \qquad \psi = \mathbb{1}\{ |W| > q_{\alpha/2} \}$$

Data from JetBlue (R2F) and United (WiIMA):

	R2F	WiIMA
Average (mins)	24.2	25.9
Std. Dev (mins)	5.1	4.3
Sample size	72	56

$$W = \frac{24.2 - 25.9}{\sqrt{\frac{5.1^2}{72} + \frac{4.3^2}{56}}} = -2.04 \Rightarrow$$

Reject at asymp. level 5%

The asymp. p-value is given by

$$\alpha_0 = \mathbb{P}[|Z| > |-2.04|] = 2\mathbb{P}[Z < -2.04] = \underline{4.14\%}$$

Example 3: Waiting for the T

Waiting times for the T: $X_1, \dots, X_n \stackrel{iid}{\sim} \mathsf{Exp}(\lambda)$.

$$\begin{aligned} H_0 &: \quad \lambda \geq 1 \\ H_1 &: \quad \lambda < 1 \end{aligned}$$

- Recall that using the Delta-method, we got for $\hat{\lambda} = 1/\bar{X}_n$,

$$\sqrt{n}(\hat{\lambda} - \lambda) \xrightarrow[n \to \infty]{(d)} \mathcal{N}(0, \lambda^2)$$

- Therefore, by Slutsky

$$\sqrt{n}\,\frac{\hat{\lambda} - \lambda}{\hat{\lambda}} \xrightarrow[n \to \infty]{(d)} \mathcal{N}(0, 1)$$

MAIN INGREDIENT FOR WALD TEST

- Test statistic

$$W = \sqrt{n}\frac{1}{\hat{\lambda}}(\hat{\lambda} - 1)$$

- Reject at 5% if $W < -q_\alpha = -1.645$

Example 4: MLE and the Wald test

- Recall that under some regularity conditions, we have:

$$\sqrt{n}(\hat{\theta}^{\mathsf{MLE}} - \theta) \xrightarrow[n\to\infty]{(d)} \mathcal{N}(0, \frac{1}{I(\theta)})$$

where $I(\theta)$ is the Fisher information

- Using Slutsky, we get

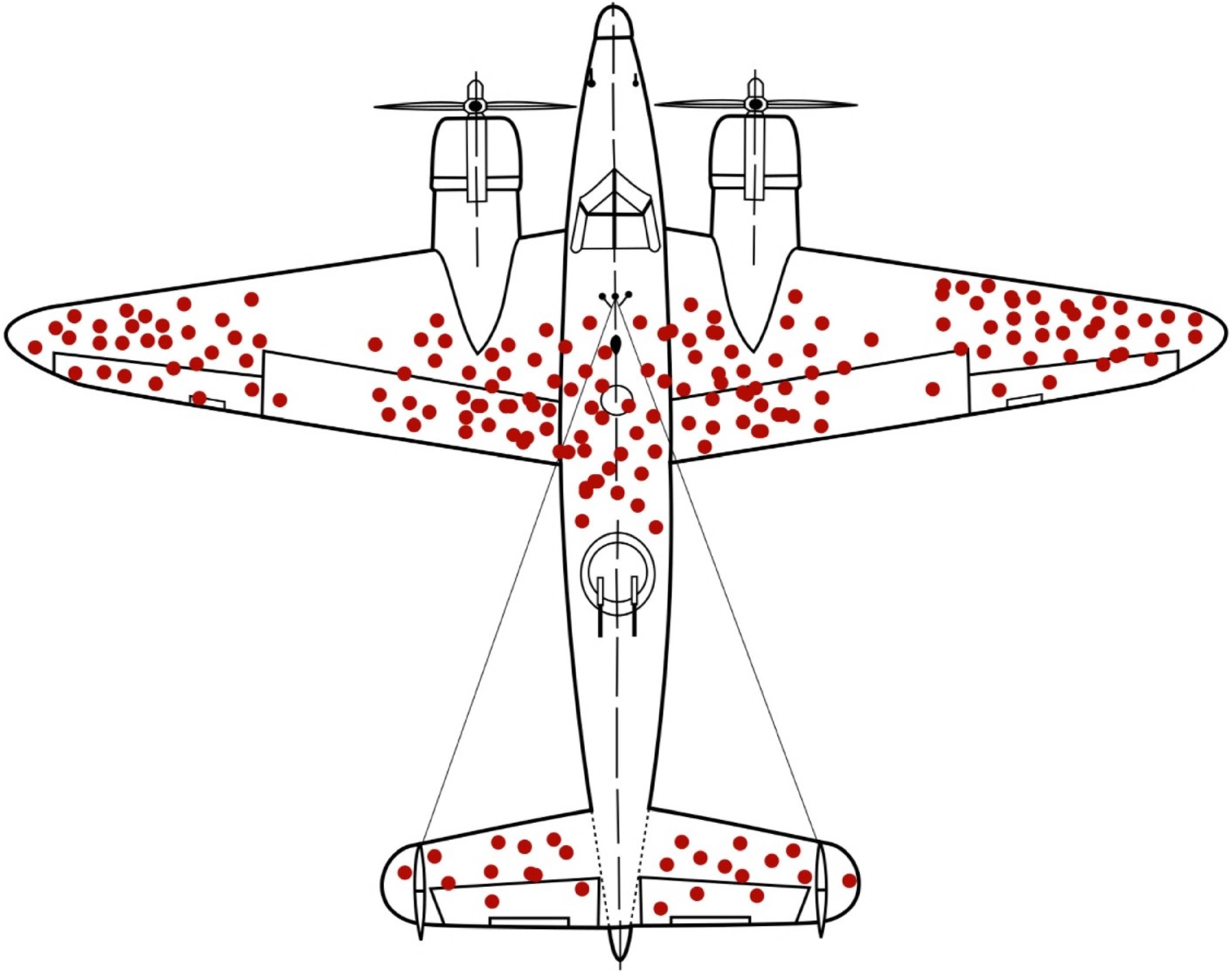
+ CMT

$$\sqrt{n \; I(\hat{\theta})}\,(\hat{\theta}^{\mathsf{MLE}} - \theta) \xrightarrow[n\to\infty]{(d)} \mathcal{N}(0, 1)$$

- Therefore, we can use the Wald test with test statistic given by

$$W = \sqrt{n \; I(\hat{\theta}^{MLE})}\;(\hat{\theta}^{MLE} - \theta_0)$$

5:38 / 5:52 Speed 1.0x



A test based on the log-likelihood

- Consider an i.i.d. sample $X_1, \ldots, X_n$ with statistical model $(E, (\mathbb{P}_\theta)_{\theta \in \Theta})$, where $\Theta \subseteq \mathbb{R}^d$ $(d \geq 1)$.

- Suppose the null hypothesis has the form

$$H_0 : (\theta_{r+1}, \ldots, \theta_d) = (\theta_{r+1}^{(0)}, \ldots, \theta_d^{(0)}),$$

  for some fixed and given numbers $\theta_{r+1}^{(0)}, \ldots, \theta_d^{(0)}$.

- Let

$$\hat{\theta}_n = \underset{\theta \in \Theta}{\operatorname{argmax}} \ \ell_n(\theta) \quad \text{(MLE)}$$

  and

$$\hat{\theta}_n^c = \underset{\theta \in \Theta_0}{\operatorname{argmax}} \ \ell_n(\theta) \quad \text{("constrained MLE")}$$

  where $\Theta_0 = \{\theta \in \Theta : (\theta_{r+1}, \ldots, \theta_d) = (\theta_{r+1}^{(0)}, \ldots, \theta_d^{(0)})\}$

Likelihood ratio test

Test statistic:

$$T_n = 2\left(\ell_n(\hat{\theta}_n) - \ell_n(\hat{\theta}_n^c)\right).$$

Wilks' Theorem

Assume H_0 is true and the MLE technical conditions are satisfied.
Then,

$$T_n \xrightarrow[n\to\infty]{(d)} \chi^2_{d-r}$$

Likelihood ratio test with asymptotic level $\alpha \in (0,1)$:

$$\psi = \mathbb{1}\{T_n > q_\alpha\},$$

where q_α is the $(1-\alpha)$-quantile of χ^2_{d-r} (see tables).