drp winter 25

bao han ngo

e-values in hypothesis testing

what is an p-value?

- a random variable P with $P(P \leq \alpha) \leq \alpha$ under the null
- a smaller p-value represents more evidence against the null
- disadvantages of p-values
  - pre-specify α
  - pre-specify sample size
  - difficult to combine dependent p-values from different experiments
  - need to know data collection procedure

what is an e-value?

- a random variable E with an expected value of at most 1 under the null, $E_{H_0}[E] \leq 1$
- a larger e-value represents more evidence against the null
- 1/e is a valid p-value
- advantages of e-values
  - can specify α after the fact
  - can peek at the data/stop early
  - can easily combine e-values from different experiments
  - can be used even when data collection procedure is unknown

Employment status [reference: unemployed]		
Employed	−0.569	1.442
Not in the labor force	−0.344	−0.012
Education level [reference: no high school]		
High school graduate	0.129	−0.012
Some college	1.217**	1.442*
College graduate	2.270**	2.391**

Note: All models adjust for survey weight.
Coefficients for age reflect a 10 year change.
* $p < 0.05$.
** $p < 0.01$.

likelihood ratio as an e-value

- $H_0\colon \theta = \theta_0$
- $H_1\colon \theta = \theta_1$
- likelihood ratio: $\Lambda = \frac{L(\theta_0|x)}{L(\theta_1|x)}$
- inverse of likelihood ratio: $\frac{1}{\Lambda} = \frac{L(\theta_1|x)}{L(\theta_0|x)}$ has an expected value of 1 → $\frac{1}{\Lambda}$ is an e-value

simulating p- and e-value validity

- t samples from Bernoulli(p = 0.5)
- want to test (and maybe reject) the null hypothesis $H_0: p = 0.5$ using t samples

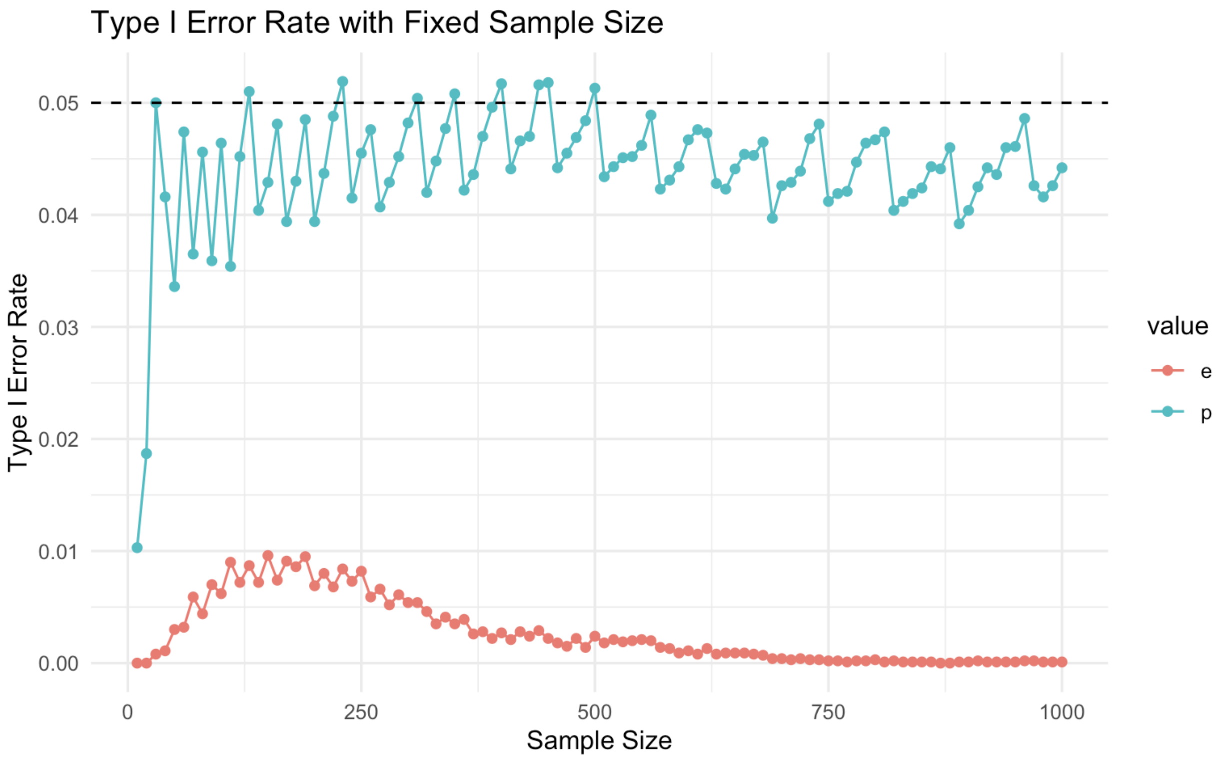
p-value

- $H_0: p = 0.5$
- $H_1: p > 0.5$
- p based on binomial tail probability
- reject if $p < 0.05$

e-value

- $H_0: p = 0.5$
- $H_1: p = 0.6$
- e = $\frac{0.6^x \times (1-0.6)^{t-x}}{0.5^x \times 0.5^{t-x}}$
- reject if e > $\frac{1}{0.05}$

p- and e-values are valid with pre-specified sample size t



simulating p- and e-value validity with early stopping

- 1000 samples from Bernoulli(p = 0.5)
- testing after every 10 samples

p-value

- $H_0 : p = 0.5$
- $H_1 : p > 0.5$
- p based on binomial tail probability
- reject if $p < 0.05$
- rejection in 26.2% of samples

e-value

- $H_0 : p = 0.5$
- $H_1 : p = 0.6$
- e = $\frac{0.6^x \times (1-0.6)^{t-x}}{0.5^x \times 0.5^{t-x}}$
- reject if e > $\frac{1}{0.05}$
- rejection in 3.7% of samples

an example

- help !!! how much data should I collect ?????
- n = 200 observations from Bernoulli(p=0.6)

p-value

- $H_0: p = 0.5$
- $H_1: p > 0.5$
- $p_{n=200} = 0.01$
    - ($p_{n=93} = 0.048$)

e-value

- $H_0: p = 0.5$
- $H_1: p = 0.6$
- $e_{n=157} = 21.8 > \frac{1}{0.05}$

- we could have stopped collecting data after 157 points !!

conclusions

- p-values are being abused
- e-values could lead to time and money saved in online experiments