5.1 Hypothesis Testing

Tuesday, September 22, 2020 11:30 AM

Hypothesis is a prediction based on theory, research, or observation, that is being tested

Hypothesis testing tests whether the outcome of a study supports or refutes a hypothesis

Logic of hypothesis testing

- backwards way based on what we do know

Hypothesis Testing Procedure

- create two mutually exclusive hypotheses (Null and Alternative (research) hypothesis)
- comparing hypotheses in light of sample evidence is a statistical test
- Research question: question we want answered
- Research (Alternative) hypothesis (H_1): formal (statistical) hypothesis to be tested based on the research question
- Null hypothesis: the opposite of the research hypothesis
- Null and research hypothesis must be mutually exclusive + exhaustive

$z_i = x_i - \bar{x} / s$

- alpha (significance) level: probability we're willing to accept that the observed result is due to chance

Type I and II errors

- retain null hypothesis when null is true Type II
- reject null hypothesis when null is false Type I

- Type I has α (alpha) level → probability of making said type of error
- Type II has β (beta) level

- Power $(1-\beta)$: probability of rejecting null when it is false

$z = \bar{x} - \mu_{\bar{x}} / \sigma_{\bar{x}} = \bar{x} - \mu / \frac{\sigma}{\sqrt{n}}$

$t = \bar{x} - \mu_{\bar{x}} / s_{\bar{x}} = \bar{x} - \mu / \frac{s}{\sqrt{n}}$

Testing a sample mean when σ is unknown

- use a t test instead of a z test

Example

Pop size $N = 4$ random var X is age $X = \{18, 20, 22, 24\}$

$\mu = \sum x_i / N = \frac{18+20+22+24}{4} = 21$

$\sigma = \sqrt{\sum (x_i - \mu)^2 / N} = 2.236$

Sampling Distribution of s^2

- s^2 is more likely to underestimate σ^2 than to overestimate
- corresponding t-value will be larger than z-value esp. is N is small
- As $N \uparrow$, t approaches z

Degrees of freedom (df)

- df is # of independent pieces of information free to vary
- in one-sample t test we estimate a sample mean so we have n-1 df
- there is a t distribution for every number of df

t distribution

- larger tails than normal (z) distribution
- small df = large tails, large df = small tails
- as df becomes larger (>120), t approaches z

Conditions for using t

- independence of observations
- observations should be nearly normal

Confidence Intervals

1) $t = (\bar{x} - \mu) / (s / \sqrt{n})$
2) substitute critical value for $t \Rightarrow \pm 2.365 = (\bar{x} - \mu) / (s / \sqrt{n})$ any number
3) substitute statistics
4) solve for μ
5) calculate upper + lower limits