Confidence intervals for the mean, when variance is unknown.

$$X_1, \ldots, X_n \sim N(\theta, \sigma^2)$$

$$Z = \frac{\mu - \bar{X}}{\sigma/\sqrt{n}} \sim N(0,1)$$

what if σ^2 is unknown?

$$T = \frac{\mu - \bar{x}}{s/\sqrt{n}}$$ where s^2 is sample variance

<u>Example 3.4.7</u>

$X_1, \ldots, X_n \sim N(\theta, \sigma^2)$

σ^2 is unknown

BUT for a sample $x_1, \ldots, x_n$ where

$n = 12 \quad s^2 = 7$

Compute confidence interval for mean θ with $\alpha = 0.05$

ie. 95% confidence interval.

$$T = \frac{\theta - \bar{x}}{s/\sqrt{n}}$$

$$P(-t_{\nu, 1-\alpha/2} < T < t_{\nu, 1-\alpha/2}) = 1 - \alpha$$

ν

Greek letter "nu" "degrees of freedom" of Student's t-distribution

χ^2_k - k - degrees of freedom

$\nu = n-1$

$n = 12 \Rightarrow \nu = 11$

$t_{\nu, 1-\alpha/2} = 2.201$

Boxplot - outliers.

median

$q_{0.25}$ - lower quartile
$q_{0.75}$ - upper quartile

$\Big\}\ q_{0.75} - q_{0.25}$

max value

min value.

Tukey's rule:

$x \notin [\,q_{0.25} - 1.5\,IQR,\ q_{0.75} + 1.5\,IQR\,]$

$\Rightarrow$ x is an outlier.

Hypothesis testing

Definition 4.1:

A hypothesis is a statement about a parameter (or parameters) of interest.

Null hypothesis H_0

"The default position"

→ gives us our assumptions about random variables

→ gives us distribution of our test statistic

Alternative hypothesis H_1

A "complementary" position

Example:

$$X_1, \ldots, X_n \sim N(\theta, \sigma^2)$$

$H_0 : \theta = 0$ ← this is our assumption

$H_1 : \theta \neq 0$ ← this is what we want to show

Example 4.1.2. Suppose a vaccine is developed to prevent infection of a particular disease. The 'vaccine efficacy' is defined as the proportionate reduction in infection rate between vaccinated and unvaccinated individuals. Denoting the vaccine efficacy as VE, one possible null hypothesis (and alternative hypothesis) is

$$H_0 : VE \leq 0.3,$$
$$H_1 : VE > 0.3.$$

This choice reflects the conservative view that a vaccine will only be considered effective if the efficacy is greater than 30%. △

p-values

- H_0 is made (null hypothesis)
- Gives us assumptions about random variables $X_1, X_2, \ldots, X_n$
- Allows us to derive a distribution for a test statistic $T = t(X)$, distribution F_T

Then: observe data $x = (x_1, x_2, \ldots, x_n)$

Compute test statistic $t = t(x)$

$p = 1 - F_T(T)$: random variable

cdf

Using observations $p = 1 - F_T(t(x))$

number, score

$$F_T(t(x)) \in [0,1]$$

$$\Rightarrow p = 1 - F_T(t(x)) \in [0,1]$$

Values close to 0 : indicate data "far" from H_0

	Random variable	Realisation
Statistic	$\bar{X}$	$\bar{x}$
Confidence interval	$[L(X), U(X)]$	(l, u)
p-value	$\cancel{P}$	p

Example

$X_1, X_2, \ldots, X_n \sim N(\theta, \sigma^2)$ $\quad n = 100$

Assume $\sigma^2 = 1$

θ : value is unknown.

We want to show that θ is not 0.

$H_0 : \theta = 0$

$H_1 : \theta \neq 0$

Should be 0.15

observe $x_1, \ldots, x_{100}$ $\bar{x} = 1.5$ $n = 100$

$$\boxed{z = \frac{\bar{x}}{\sigma / \sqrt{n}} = \frac{0.15}{1/10} = 1.5}$$

$$\bar{X} \sim N\left(\theta, \frac{\sigma^2}{n}\right) = N\left(\theta, \frac{1}{n}\right)$$
$$= N\left(\theta, \frac{1}{100}\right)$$

$$F_Z(1.5) = 0.9332$$

$$p = 1 - F_Z(1.5) = 0.0668$$

Specify significance threshold α

<u>IN ADVANCE</u>

$\alpha = 0.05$

$0.0668 \nless 0.05$

eg. $\alpha = 0.05$

$\alpha = 0.01$

$\alpha = 0.001$

Then compare p to α

if $p < \alpha$: REJECT H_0

$p \nless \alpha$: FAIL TO REJECT H_0