Lecture 2

Brennan Becerra

2024-01-11

Data Input

1.

```
>x <- c(1,2,3,4,5)
>x <- scan()
1:1
2:2
3:3
4:4
5:5
6:
>
```

Codeblock 1 (language-r)

2.

```
>x <- scan(file="path //datafilename")
```

Codeblock 2 (language-r)

- To read text-formatted data file

3. R Dataframe

```
>c1 <- c(1,2,3,4)
>c2 <- c(0.1,0.2,0.3,0.6)
>D1 <- data.frame(c1,c2)
>D1
```

Codeblock 3 (language-r)

General Definitions and Examples

Estimation

$100 * (1 - \alpha)\%$ confidence interval for unknown θ.
Test of Hypothesis: Have a claim about unknown θ.

$$H_0 : \text{Null Hypothesis}$$
$$H_1 : \text{Alternative Hypothesis}$$

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_0 \end{cases} \text{ or } \begin{cases} H_0 : \theta \geq \theta_0 \\ H_1 : \theta < \theta_0 \end{cases}$$

Ex1. $\theta =$ Average Height of BSU students.

- $\theta =$ Average height of BSU students is larger than 5.6 ft.

$$H_0 : \theta \leq 5.6$$
$$H_1 : \theta > 5.6$$

Ex2. $\theta =$ Average Height of BSU students.

- $\theta =$ Average height of BSU students is less than 5.6 ft.

$$H_0 : \theta \geq 5.6$$
$$H_1 : \theta < 5.6$$

Ex3. $\theta =$ Average Height of BSU students.

- $\theta =$ Average height of BSU students is 5.6 ft.

$$H_0 : \theta \neq 5.6$$
$$H_1 : \theta = 5.6$$

Recap of steps:

1. Test of Hypothesis: Have a claim about unknown θ.

$$H_0 : \text{Null Hypothesis}$$
$$H_1 : \text{Alternative Hypothesis}$$

2. Select Significant Level, I.E. the value of α, in $100 * (1 - \alpha)\%$.
3. Under the assumption that H_0 is true, calculate *Test Statistic*(TS).

> **Test Statistic**
>
> 1. A random variable
> 2. Has a probability distribution

4. Based on α and TS, and type of $\underbrace{H_1}_{\neq,<,>}$, most statistical programs yield a "p-value," p.
5. Decision: If $p < \alpha$, then REJECT H_0.

> **Population**
>
> A collection of individuals or objects, that are of interest in quantitative study.

> **Sample**
>
> We can define a (random) *sample* to be a subset of the population of interest.

> **Parameter**
>
> We can define a *parameter* as a fixed number giving descriptive characteristics to an entire *population*.

TRUTH (or Nature)

	H_1 **is true**	H_0 **is true**
H_0 is True	Correct	Type II Error $P(H_0$ is true when H_1 True) $= \beta$
H_1 is True	Type I Error $= P(H_1$ is True when H_0 is true) $= \alpha$	Correct

Fix α to minimize β.

Read Chapter 3 before next Tuesday.