19th March '2024

Continuing with Confidence intervals for σ^2 :

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$$ ⟵ Pivotal quantity for σ^2

$100(1-\alpha)\%$ Confidence interval :

Want area under upper tail = area under lower tail = $\alpha/2$.

Let's look at Figure 6.4 :→

We want to find all dobs (σ_0^2) s.t :

1) $\mathbb{P}(\chi^2_{(n-1)} \geq b) \geq \alpha/2$; $b = \chi^2_{1-\alpha/2, (n-1)}$

2) $\mathbb{P}(\chi^2_{(n-1)} \leq a) \leq \alpha/2$; $a = \chi^2_{\alpha/2, (n-1)}$

$$\boxed{a \leq \frac{(n-1)S^2}{\sigma_0^2} \leq b}$$

$$\Rightarrow \left\{ \frac{1}{a} \geq \frac{\sigma_0^2}{(n-1)S^2} \geq 1/b \right\} \Rightarrow \left\{ 1/b \leq \frac{\sigma_0^2}{(n-1)S^2} \leq 1/a \right\}$$

$$\Rightarrow \left\{ \frac{(n-1)S^2}{b} \leq \sigma_0^2 \leq \frac{(n-1)S^2}{a} \right\}$$ or,

$\left\{ \frac{(n-1)S^2}{b}, \frac{(n-1)S^2}{a} \right\}$ is our $100(1-\alpha)\%$ C.I with

$a = \chi^2_{\alpha/2, (n-1)}$, $b = \chi^2_{1-\alpha/2, (n-1)}$

Using R and our data,

$n = 352$, $S^2 = 780.087^2$

$a = 300.99$, $b = 404.797$

95% C.I for δ^2 is $[726^2, 842^2]$ $H_0: \delta^2 = 750^2$

750^2 is here

[See R code for this test and C.I Section 6.2.3]

Section 6.3: Two Sample Models.

In this Section, we consider two Samples.

Y_{1i} = i^{th} Salary in 1^{st} Sample (Work term # 1)

Y_{2j} = j^{th} Salary in 2^{nd} Sample (Work term # 2)

$i = 1, 2, \cdots, n_1$; $j = 1, 2, \cdots, n_2$

We assume the samples are independent.

$Y_{1i} \overset{iid}{\sim} N(\mu_1, \delta_1^2)$; $Y_{2j} \overset{iid}{\sim} N(\mu_2, \delta_2^2)$

We want to test, $H_0: \mu_1 = \mu_2$ i.e., $\mu_1 - \mu_2 = 0$

(Case 1): $\rightarrow$ We assume that $\delta_1^2 = \delta_2^2 = \delta^2$, known.

L.R.S ; $D = Z^2$, where $Z = \dfrac{\bar{Y}_1 - \bar{Y}_2}{\delta\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim N(0,1)$

(Case 2): $\rightarrow$ We assume that $\delta_1^2 = \delta_2^2 = \delta^2$ but, unknown

L.R.S ; $D = g(T^2)$, where $T = \dfrac{\bar{Y}_1 - \bar{Y}_2}{S_{pool}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$

monotone

$D \approx \chi^2_{k-q}$.

$$S_{pool} = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

(Case 3) :→ We can't assume that σ_1^2, σ_2^2 are known nor equal.

$$L.R.S\ ;\quad D = g(T^2),\ \text{where}\quad T = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim t_{(\nu)}$$

ν → Satterthwaite Approximation to the degree of freedom..!

C.I's for the Cases :→

1) $(\bar{Y}_1 - \bar{Y}_2) \pm Z_{1-\alpha/2}\, \sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

2) $(\bar{Y}_1 - \bar{Y}_2) \pm t_{1-\alpha/2,\, (n_1+n_2-2)} \cdot S_{pool} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

3) $(\bar{Y}_1 - \bar{Y}_2) \pm t_{1-\alpha/2,\, \nu} \cdot \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$

Commenting on plots : (You will do this in Assignment 6)

A Comment on a graph has two Components :

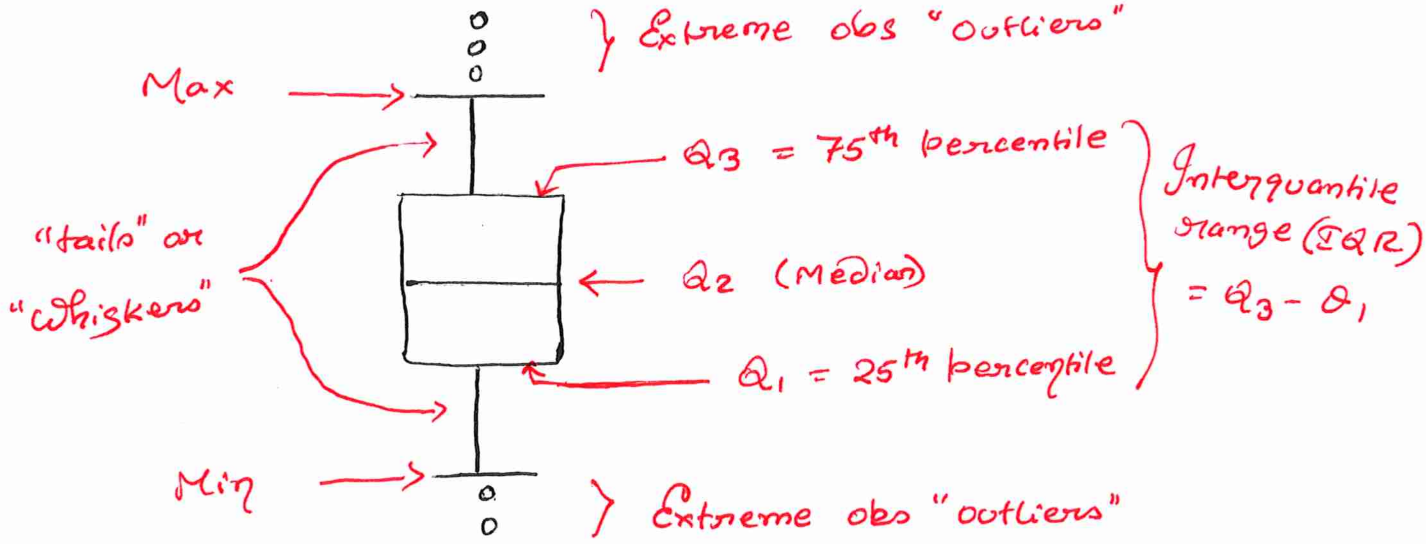
1) pointing out a characteristics of the plot (data).
2) Gives a reason/interpretation of said characteristics.

Both are required for an adequate Commentary.

(1) requires that You understand the purpose of a given graph.

Boxplots, for Example, demonstrate data distribution according to the "5-number summary" :

Min, Q_1, Median, Q_3, Max



☆ Always plot Your Data & look at Summary Statistics before testing !! ☆

Side by Side boxplots can demonstrate the similarities and differences between two distributions.

Examples (1)

- ↳ is the spread similar? i.e length of whiskers/ length of box.
- ↳ Do either have outliers?
- ↳ Do their boxes overlap?
- ↳ Is there any differences in their medians?
- ↳ Any skew present? (Median/whiskers may not be Symmetric)

☆ What do all these Characteristics mean? ☆ Still need (2)

You do not need to comment on all these things every single time. Rather, find something meaningful that correlates to your hypothesis of interest/assumptions on populations.

—☆—