CHAPTER 17 - Robustness of procedures

17.7 – Robustness of the t-procedure

A **confidence interval** or significance test is called **robust** if the **confidence level** or ***P*-value** does not change very much when the conditions for use of the procedure are violated.

- Except in the case of small **samples**, the condition that the data are an SRS from the **population** of interest is more important than the condition that the **population** **distribution** is Normal.
- ***Sample*** *size less than 15:* Use *t* procedures if the data appear close to Normal (roughly symmetric, single peak, no **outliers**). If the data are clearly skewed or if **outliers** are present, do not use *t*.
- ***Sample*** *size at least 15:* The *t* procedures can be used except in the presence of **outliers** or strong skewness.
- *Large* ***samples****:* The *t* procedures can be used even for clearly **skewed distributions** when the **sample** is large, roughly $n \geq 40$.

Theoretical conditions of t procedures:

- SRS (this is the most important)
- Population must be normally distributed

If these two conditions are met, then t confidence intervals and t tests are exact

IN PRACTICE

No population is exactly normal

Thus, no t-procedures has an exact distribution

Explore the Shape of the sample:

- Outliers are always a problem. We can't use t-procedures in this case
- Strong skewness are a problem when $n < 40$
- Strong skewness is not a problem when $n = 40$ or $n > 40$ because the Central Limit theorem applies

Theoretical Conditions of *t* procedures:

- SRS.
- Population must be normally distributed.

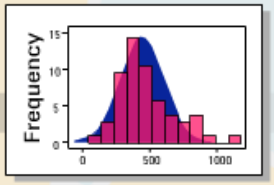
If these two conditions are met, then *t* confidence intervals and *t* tests are exact.

But what should we check in practice?

In Practice

No population is exactly normal:

- Thus, no *t*-procedure has an exact *t*-distribution.



Questions:

How are *t*-procedures affected by not having an SRS?
Confidence levels and *P*-values are off, but we don't know how much nor in what direction.

How are *t*-procedures affected by lack of normality?
Confidence levels and *P*-values are approximately correct provided data are not strongly skewed and have no outliers!

Definition of Robust

A statistical procedure (confidence interval or significance test) is *robust* if: Confidence level or *P*-value does not change very much when conditions of procedure are not met.

Are *t* procedures *robust* with respect to:

- Lack of SRS (randomization)? **Never**
  - **Check:** Method of data collection for use of probability selection.
    - Ask: Can data be regarded as an SRS?
- Lack of normality? **Yes, but only IF data are not strongly skewed and have no outliers.**
  - **Check:** Plot of data:
    - Look for: Reasonable symmetry and single-peaked appearance
    - Look out for: Strong skewness, outliers

Sample size less than 40:
OK to use t procedures unless outliers or strong skewness are present.
Sample size at least 40:
OK to use t procedures even when data are strongly skewed due to CLT.

Can a *t* Procedure Be Used?

Case 1

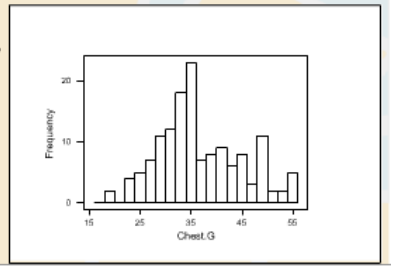
Researchers collected data on n = 143 bears in order to estimate the chest girth for all bears. Is using a t confidence interval procedure appropriate? Why or why not?

SRS? **Seems reasonable to consider data as SRS.**

Is t robust with respect to normality? Examine plot of data.

Yes, because sample size is sufficiently large to apply Central Limit Theorem.

Use of t procedure is appropriate.



Can a *t* Procedure Be Used?

Case 2

Researchers at an agricultural station want to determine if the average yield of a variety of alfalfa exceeds three tons per acre; they randomly selected sixteen fields from those available for use in the experiment. Is using a t significance test procedure appropriate? Why or why not?

SRS? **Fields were randomly selected.**

Is t robust with respect to normality?

Examine plot of data.

Yes, because there are no outliers or strong skewness in the data.

Use of t procedure is appropriate.

Leaf Unit = 0.010

29	3
29	9
30	4
30	69
31	1
31	79
32	023
32	567
33	1
33	
34	0

Can a *t* Procedure Be Used?

Case 3

A manufacturing engineer wants to estimate the elasticity of a product using a newly developed additive; eleven units were produced and tested. Is using a t confidence interval procedure appropriate? Why or why not?

SRS? **Seems reasonable to consider data as SRS.**

Is t robust with respect to normality? Examine plot of data.

No, because there is an outlier in the data and n < 40.

Use of t procedure is NOT appropriate.