BASICS OF
STATISTICAL RISK ANALYSIS

BACKGROUND

- **STATISTICS:** set of logical and mathematical-probabilistic tools for the study of real phenomena that occur with repeated determinations characterized by ***variability***.

- Its division into two branches (*i.e.*, DESCRIPTIVE, INFERENTIAL) responds more to schematization needs: in real applications there are no such clear demarcations.

- **DESCRIPTIVE STATISTICS**: data collection and analysis by means of graphs and summary indices (position, variability and shape).

- **INFERENTIAL STATISTICS:** set of methods that allow to generalize results based on a partial observation (**sample**) : process in **inductive inference !**

BACKGROUND (*cont.*)

- **INFERENCE:**
  - is the process of reaching a conclusion from a given set of statements (or *premises*)
  - it is of two types: **deductive** and **inductive**

- **Example 1:** **Deductive Argument** (from general to the particular)

  Premises: Socrates is a man
  All men are mortal
  Conclusion: Socrates is mortal **VALID ARGUMENT**

- **Example 2:** **Inductive Argument** (from particular to the general)

  Premises: Last September was the rainiest on record
  John's birthday is in September
  Conclusion: It rained on John's last birthday **PLAUSIBLE ARGUMENT**

- ***A basic problem in inductive inference is to devise ways of measuring the strength of an inductive argument!***

BACKGROUND (*cont.*)

- **Statistical INFERENCE** uses two methodologies: **hypothesis testing** and **parameter estimation.**
- **Statistical hypothesis**: an assertion regarding the parameters of one or more populations that we want to test or investigate.
- **Hypothesis testing** : the procedure that leads to a decision concerning a particular hypothesis. It is based on a random sample extracted from the population of interest (survey).
- **Null Hypothesis**: identified as H_0, is the "default hypothesis", the "thing that is accepted", the "currently accepted value for a certain parameter".
- **Alternative Hypothesis**: identified as H_a or H_1 it also called, in some books, "the research hypothesis". It involves the assertion to be tested.

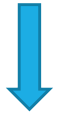
BACKGROUND (*cont.*)

STATISTICAL HYPOTHESIS TESTING is useful in many cases, *e.g.*:

- **check if a certain value lies within the confidence interval**
  (typical application: determining if a result is an OOS)
- **compare two datasets to see if they are really different or belong to the same population**
  (typical applications of this are in: suppliers validation, comparison of analytical data generated by different methods, *etc.*)
- **check the strength of the correlation between one or more causes and the undesirable effect**

BACKGROUND (*cont.*)

- **Example**: Within a Company it is believed that, on the average, a given chemical process leads to 100 kg of API. A QA Officer claims that, after the last change to the equipment, the **average yield** is no longer 100 kg.



Statistical hypothesis: H_0: $\mu = 100$ kg (**Null hypothesis**)

H_1: $\mu \neq 100$ kg (**Alternative hypothesis**)

} *two-tails test*

Note :

- Hypothesis are always statements about the population or distribution being studied, NOT about the sample.
- ***H_0 and H_1 are mathematical opposites of one another and together they cover all possibilities !***

BACKGROUND (*cont.*)

- There are just two possible outcomes:
  - **Reject the Null Hypothesis**: we then believe H_1 to be the case
  - **Fail to reject the Null Hypothesis** : we basically keep H_0

- ***How can we do the testing ? How can we reject H_0 or not?***

  We use a ***test statistic***, a parameter calculated from a sample (*e.g.*, mean) and the concrete way, the process, by which we can decide if we reject the null hypothesis or not is the ***hypothesis test***.

BACKGROUND (*cont.*)

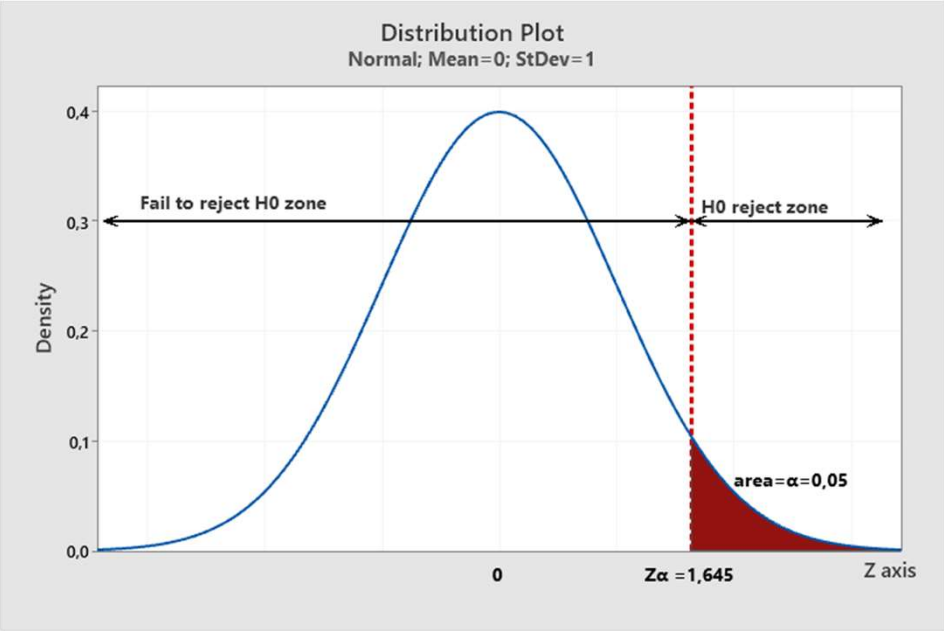
- Since we are taking a decision (*i.e.*, to reject H_0 or fail to reject H_0) there will be a ***level of confidence*** (indicated with ***C*** and typically: 95% or 99% or 0.95 and 0.99) ***which tells us how sure we are that we have made the right decision or choice.***

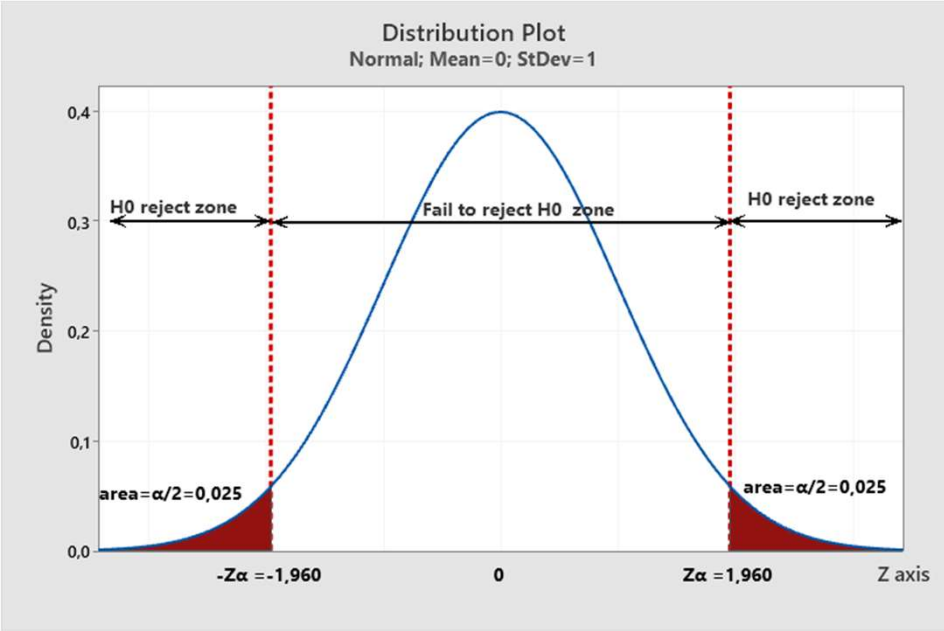
  The complement to one (1) of ***C*** is the so-called ***level of significance***, indicated with α (= 1- ***C***) and equal to 0.05 or 0.01.

  Practically, ***level of confidence*** and ***level of significance*** tell us the same: how sure we are that we are making the right decision or not !

BACKGROUND (*cont.*)

Practically, the ***level of significance*** is an area defined by a t_{α} value that represent the corresponding ***test statistic*** value printout in standard tables.





BACKGROUND (*cont.*)

- The most commonly used α-level is 0.05. *At this level, the chance of finding an effect that does not really exist is only 5%.*
- **The smaller the α value, the less likely you are to incorrectly reject the null hypothesis.** However, a smaller value for α also means a decreased chance of detecting an effect if one truly exists (lower power)
- **Sometimes it may be better to choose a smaller value for α :** testing samples from a new machine to decide whether to purchase a dozen. Since the cost of purchasing and installing a dozen machines is very high, you want to be sure that the new machine is more accurate before making the purchase.

BACKGROUND (*cont.*)

P-value:

- ***corresponds to the probability of making a mistake rejecting the null hypothesis (H_0)***
- in other words, this means what follows:
  - accepting the null hypothesis is equivalent to believing that there are no differences between the two quantities subject to hypothesis verification, *i.e.*, $\mu_1 = \mu_2$, $\sigma_1 = \sigma_2$, *etc.*
  - considering the null hypothesis as untrue is therefore equivalent to believing that there are differences between the two quantities subject to hypothesis verification
  - the *P-value* indicates the probability that one has to make a mistake by drawing the latter conclusion, *i.e.*, that there are differences between the two quantities subject to hypothesis verification.

BACKGROUND (*cont.*)

- Closely connected with the P-value is, **α**, the ***level of significance.***

- Used in hypothesis testing, **α is the maximum acceptable level of risk for rejecting a true null hypothesis (type I error) and is expressed as a probability ranging between 0 and 1**

- Before beginning any analysis, **α** should be set and then compared with *P-values* as follows:
  - if *P-value* ≤ α-level ⇨ reject H_0 in favor of H_1
  - if *P-value* > α-level ⇨ fail to reject H_0

BACKGROUND (*cont.*)

- **Sometimes choosing a larger value for α is better:** a jet engine manufacturer testing the strength of cheaper ball bearings. Saving a small amount of money does not outweigh the potentially disastrous effects if the bearings are weaker. Therefore, you might want to be more likely to detect a real weakness in the cheaper bearings.

IN GENERAL, TO MAKE IT SIMPLE:

- Since a probability of making a mistake of less than or equal to 5% (α = 0.05) is *normally accepted*:
  - **P-value > 0.05 ⇨ the differences between samples are not statistically significant: *the Null hypothesis fails to be rejected***
  - **P-value ≤ 0.05 ⇨ the differences between samples are statistically significant: *the Null hypothesis can be rejected***

BACKGROUND (*cont.*)

From a wider point of view, the hypothesis test can be considered as a «statistical criminal hearing» in which a defendant is innocent until proven otherwise !

It is based on the verification of two opposing hypotheses:

- The **null hypothesis H_0** (*defendant innocent*): it is considered true as long as there is no evidence to the contrary
- The **alternative hypothesis H_1** (*defendant guilty*): it is considered confirmed when the null hypothesis can be rejected

BACKGROUND (*cont.*)

Two types of errors may be committed when testing hypothesis:

- **Type I error (**or **risk α)** : the null hypothesis is rejected when it is true
- **Type II error (** or **risk β)** : the null hypothesis is not rejected when it is false

D.C. Montgomery, Statistical Quality Control: A Modern Introduction – J. Wiley (2013)

BACKGROUND (*cont.*)

The probabilities of these two types of error are denoted as:

$$\boldsymbol{\alpha} = P\{\text{type I error}\} = P\{\text{ reject } H_0 \mid H_0 \text{ is true }\}$$

$$\boldsymbol{\beta} = P\{\text{type II error}\} = P\{\text{ fail to reject } H_0 \mid H_0 \text{ is false }\}$$

Power of a statistical test = 1 – β = $P\{\text{ reject } H_0 \mid H_0 \text{ is false }\}$

D.C. Montgomery, Statistical Quality Control: A Modern Introduction – J. Wiley (2013)

STATISTICAL RISK ANALYSIS

In Quality Control:

- **risk α,** or just **α,** is called **PRODUCER'S RISK** because it denotes the probability that a good lot will be rejected, or the probability that a process producing acceptable values of a particular quality characteristic will be rejected as performing unsatisfactorily.

- **risk β,** or just **β**, is called **CONSUMER'S RISK** because it denotes the probability of accepting a lot of poor quality, or allowing a process that is operating in an unsatisfactory manner relative to some quality characteristic to continue in operation.

D.C. Montgomery, Statistical Quality Control: A Modern Introduction – J. Wiley (2013)

STATISTICAL RISK ANALYSIS

- The **PRODUCER'S RISK** (or **risk α**) is stated in conjunction with the numerical definition of the *maximum quality level that may be acceptable* or **ACCEPTABLE QUALITY LEVEL** (AQL).

  This quantity, as defined by MIL-STD-105D (1963), represents the maximum number of defects (or defective parts) per hundred of units that, for the purpose of sampling inspection, can be considered satisfactory as a process average.

  A sampling plan should have a low **PRODUCER'S RISK** for quality which is equal to or better than the AQL.

  In sampling plans, **risk α** is usually fixed at 0.05 (*i.e.*, ***C*** = 95%), but it could vary from 0.01 to 0.10 (*i.e.*, ***C*** from 99% to 90%).

J.M. Juran and F.M. Gryna, Juran's Quality Control Handbook, 4th Edition, McGraw-Hill (1988)

STATISTICAL RISK ANALYSIS

- The **CONSUMER'S RISK** (or **risk β**) is stated in conjunction with the numerical definition of *rejectable quality* or *the level of quality that is unsatisfactory* and therefore should be rejected by the sampling plan.

  This "rejectable quality", indicated as **lot tolerance percent defective** (LTPD), is usually set equal to 0.10. A CONSUMER'S RISK of 0.10 represents the lot quality for which the probability of acceptance is 0.10, *i.e.*, only 10% of such lots will be accepted.

J.M. Juran and F.M. Gryna, Juran's Quality Control Handbook, 4th Edition, McGraw-Hill (1988)

STATISTICAL RISK ANALYSIS

The **PRODUCER'S** and **CONSUMER'S RISKS** and associated AQL and LTPD are graphically summarized by an operating characteristic curve such as that shown here on the right. The distribution of individual measurements is assumed normal and the plan desired such that the curve passes through the two points: (p_1, 1-α) and (p2, β), where:

- p_1 = acceptable proportion defective
- 1-α = probability of acceptance at p_1
- p_2 = rejectable proportion defective
- β = probability of acceptance at p_2

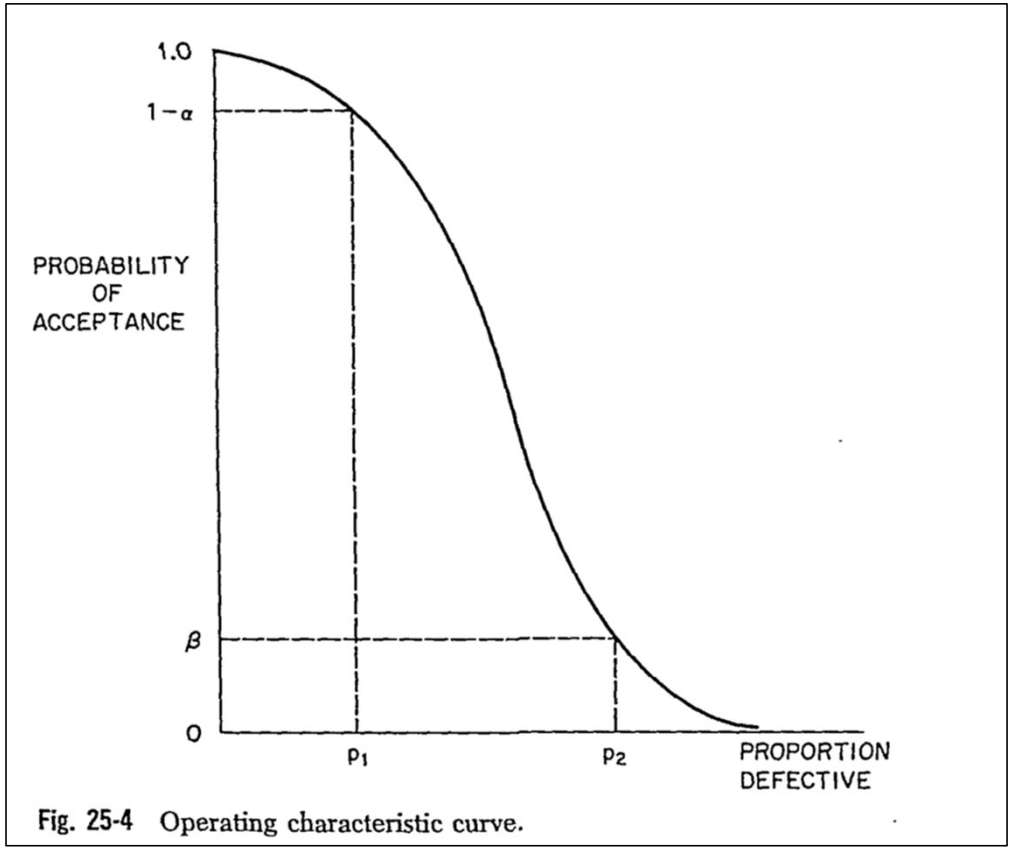


Fig. 25-4 Operating characteristic curve.

J.M. Juran and F.M. Gryna, Juran's Quality Control Handbook, 4th Edition, McGraw-Hill (1988)

STATISTICAL RISK ANALYSIS *(cont.)*

In summary:

	We Reject H_0. (accept H_a)	We Fail to Reject H_0 (not enough evidence to accept H_a)
H_0 is true.	Type I Error	Correct Decision
H_0 is false. (H_a is true)	Correct Decision	Type II Error

What just seen represents the basis of the so-called: **STATISTICAL RISK ANALYSIS**

STATISTICAL RISK ANALYSIS *(cont.)*

To easily visualize and quickly remember the two types of error, consider the so called **TABLE OF TRUTH**

		True situation of the defendant	
		Innocent	**Guilty**
Verdict	**Innocent**	**Correct**	**Incorrect verdict**: the court finds someone who is guilty innocent **Error type II = β**
	Guilty	**Incorrect verdict**: the court found one innocent guilty **Error type I = α**	**Correct**

CONCLUSIVE SUMMARY

- Risks **α** and **β** risks are related to each other
- Usually risk α = 5%. This corresponds to a 95% confidence level (***C***) of not making type I errors
- Given a phenomenon that has a certain variability (σ) and a certain sample size (n), the risk β is calculated accordingly
- ***Other things being equal, if we choose a smaller α the risk β increases: requiring clear evidence to reject the null hypothesis (H_0) increases the probability of incurring in a type II error***
- ***To reduce α and keep β constant (or vice versa), it is necessary to increase the sample size (n)***