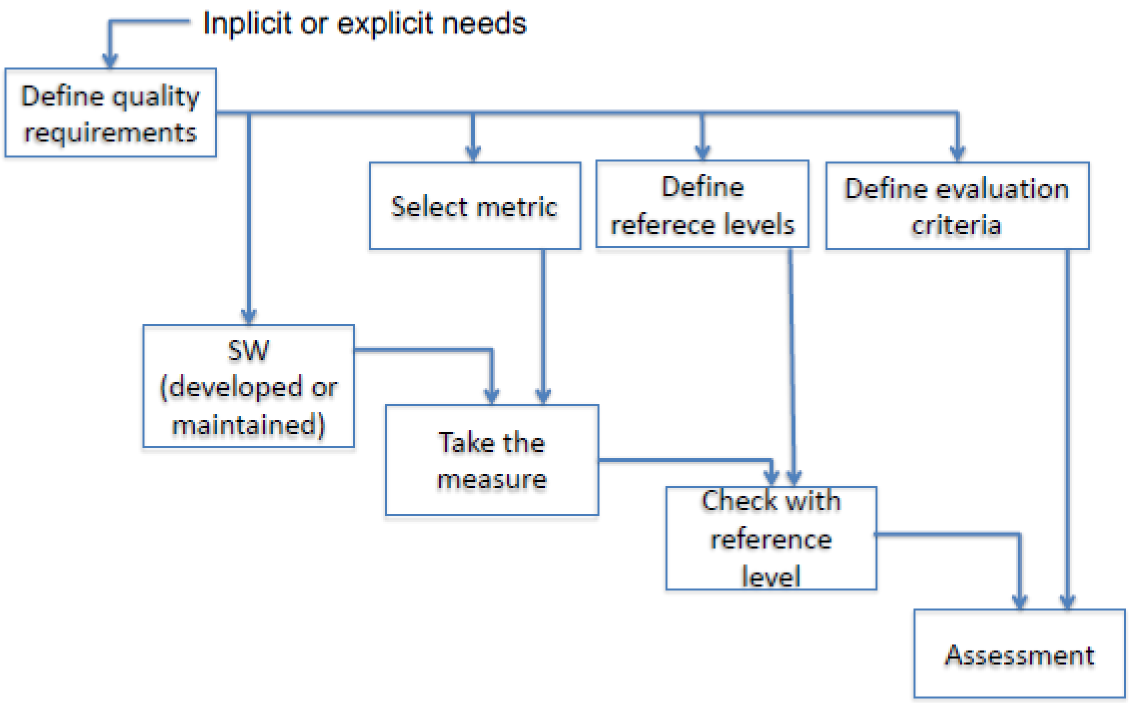
8. Measures and Statistics

8.1 Measures

We cannot govern what cannot measure. We measure for monitoring and making decisions.

Approach

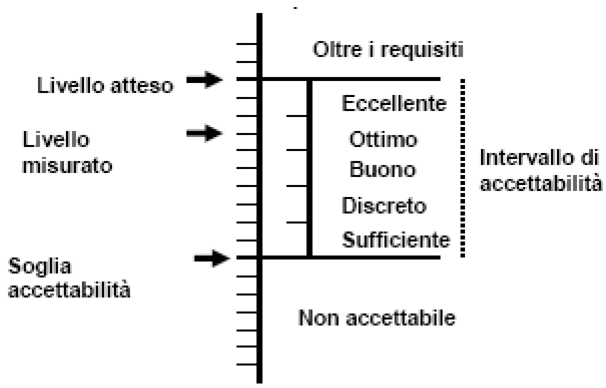


Process

We need some rating: define reference level

Metrics: provide quantitative values

We have to map quantitative data to a qualitative scale



8.2 Scales

1)Nominal Scale

classifies persons or objects into categories with common characteristics

A pre-defined non ordered set of distinct values (e.g. possible types of programming errors)

the average makes sense only to check the frequency

2)Ordinal Scale

rank the order of the items we measure in terms of which has less or more of the quality represented by the variable (e.g. CMMI levels). Possibility of use of operators (=, ≠, <, >).

We are able to classifies subjects or rank them. (possiblity highest to lowest). Have to determine the distance between 2 levels.

Careful for the average!!

3)Interval Scale

allow to ranke the order of the items that are measured, and to quantify and compare the sizes of differences between them.

Has equal intervals. (e.g. exams mark)

requires precise definition of the unite (e.g. Celsius, Farheneit)

The arbitrary of the zero implies not comparability.

4)Ratio Scale

Similar to interval, but there is an absolute zero (not arbitrary)

e.g.: height, weight, time etc. Phisical principle for 0.

not used in educational research and testing.

5)Absolute Scale

a Ratio scal with non negative integers

e.g. LOC,

Choose the scale

depends on the attributes, must correspond to the set of relations we need.

Synopsis of measure scales

Scale types	Admissible transformations	Basic empirical operation	Appropriate statistical indexes	Appropriate statistical tests	EXAMPLES
NOMINAL	any one-to-one transformation	equality test	Mode Frequency	not parametric	labeling classify
ORDINAL	$M(x) \geq M(Y)$ implies that $M'(x) \geq M'(Y)$	equality test and > <	Median Percentiles Spearman r Kendall W Kendall T	not parametric	preferences ordering di entità
INTERVALS	$M'=aM+n (a>0)$ [*positive, linear*]	equality test and > < + and -	Aritmetic mean Standard deviation Pearson correlation Multiple correlation	not parametric	Fahrenheit o Celsius date time
RATIO	$M'=aM$ $(a>0)$ [*similarity transformation*]	equality test and > < + and - * and /	Geometric mean Armonic mean Coefficiente di variazione Percentage variation Correlation index	not parametrico and parametric	time intervals Kelvin lenghts
ABSOLUTE	$M'=M$ [*identity*]				entity count

8.3 Types of measures

Ratio and proportion

Ratio: result of division between two different and disjoint domains.

It's not a percentage.

Proportion: result of a division between 2 values where the dividend contributes to the divisor (e.g. number of satisfied users/number of users)

assume values between 0 and 1.

Percentage: proportion normalized to 100

Rate: value associated with the dynamics of a phenomenon

change of a quantity with respect to another wuantity (x) on which it depends.

See example at pg.22

Definitions

Dev process: reqs, analysis, design, integration, tests….

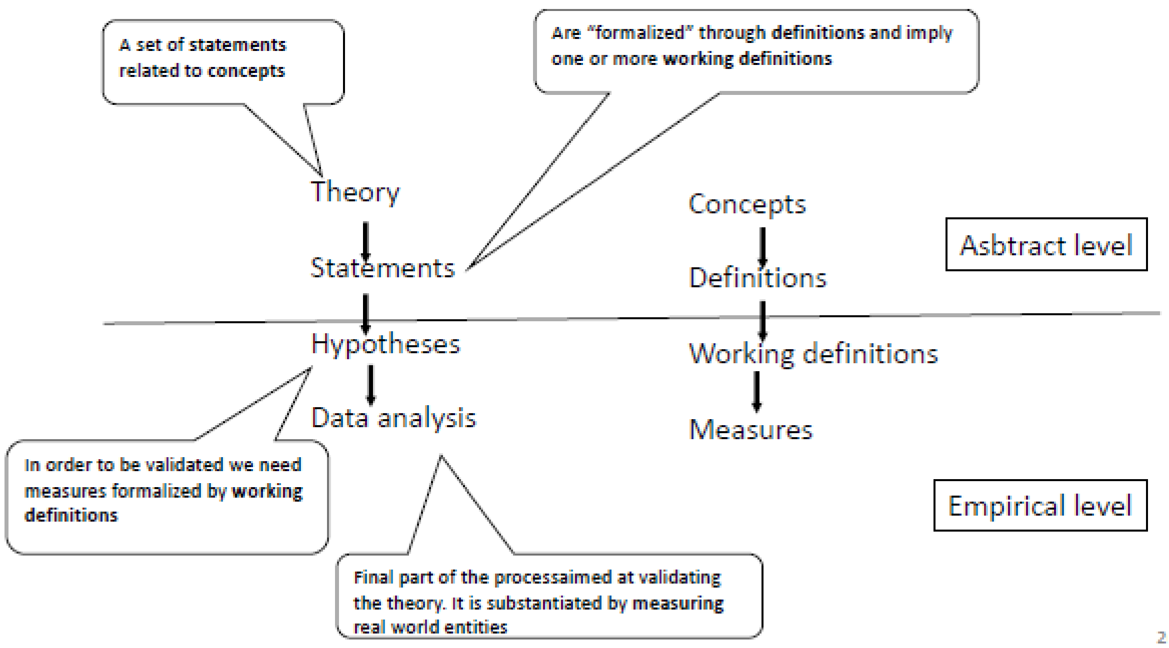
Final part of the dev process: integration and associated testing

Rigorous: the adheres to the process documentation

Quality of released sw: # of errors per KLOC discovered during testing

1) The greate the percentage of KLOC, the lesser the number of KLOC

2) The greater the effectiveness of the inspection the lesser is the number of errors per KLOC discovered during the system testing
3. The greater the efficacy of tests, in terms of discovered errors, the lesser the number of errors per KLOC discovered during the system test.



8.4 Descriptive statistics

We define the following parameters:

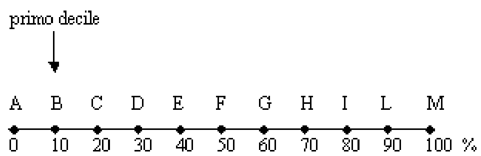
- Mean $\mu = (x_1 + x_2 + \dots + x_n)/n$
- Variance **var**$=[(x_1-\mu)^2 + (x_2-\mu)^2 + \dots (x_n-\mu)^2]/n$
- Standard deviation $\sigma = \text{var}^{1/2}$

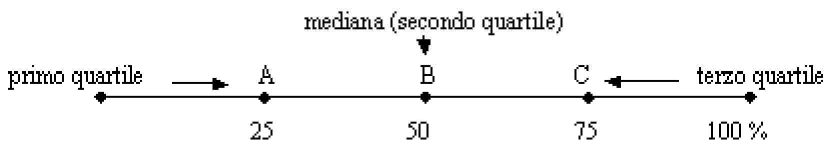
Usually the variance is indicated by σ^2

Median: middle data point in an ordered set

Percentile: value of a variable below which a certain percent of observations fall

Quartile: any of the three values which divide the sorted data set into four equal parts, so that each part represents one fourth of the sampled population.



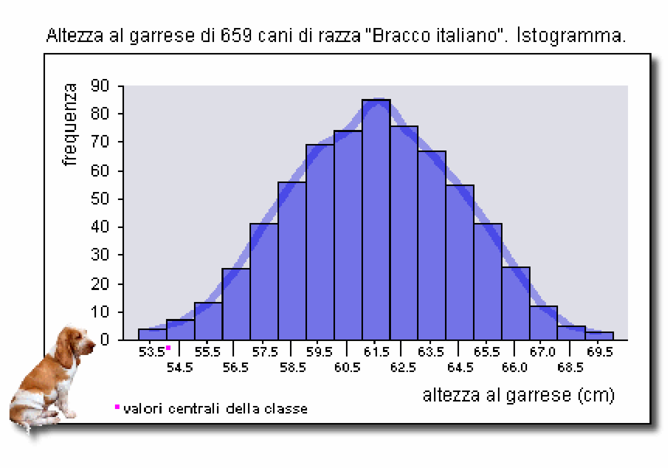


Mode: the most frequently occurring data point

Range: the distance between the smallest and largest data point.

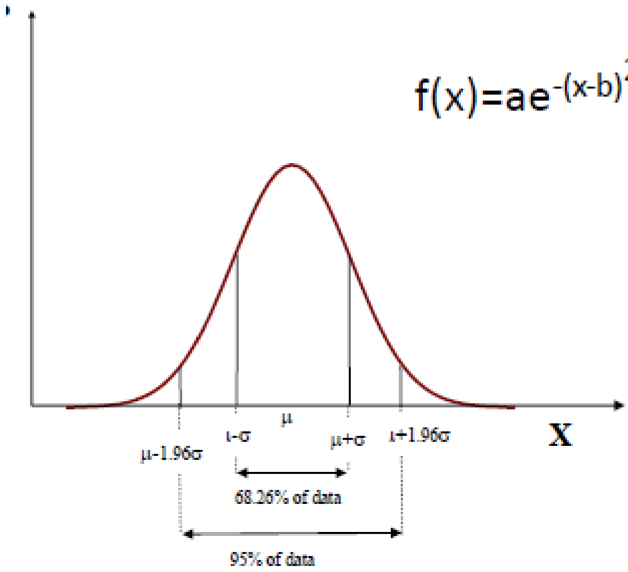
Frequency distribution

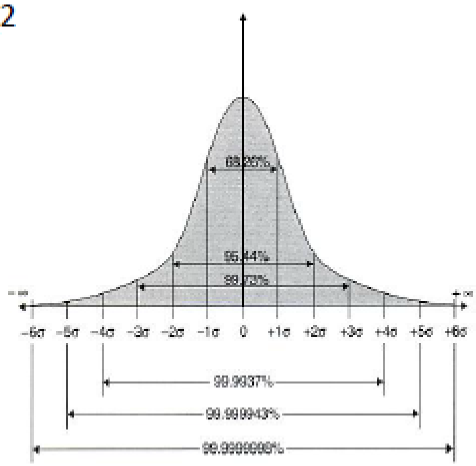
It is obtained by splitting the values observed and by indicating, for each of them, the corresponding frequency.



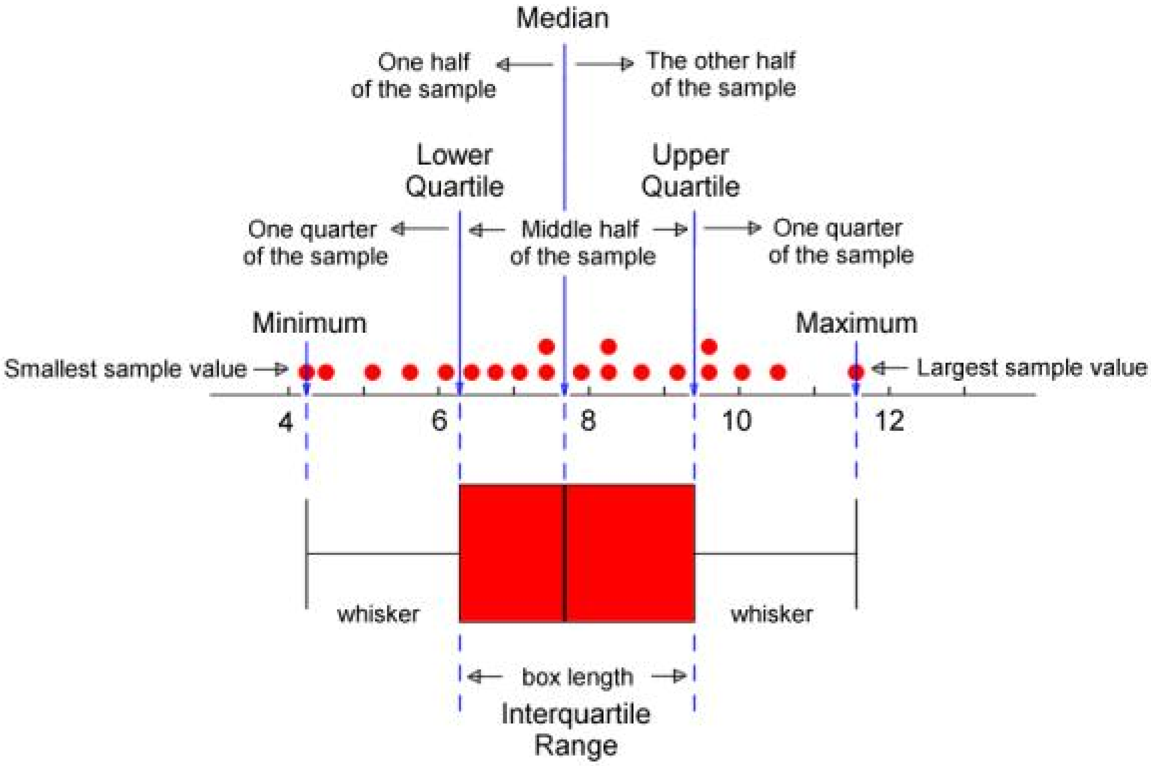
Normal

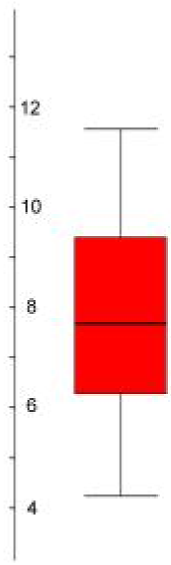
$$f(x)=ae^{-(x-b)^2/c^2}$$



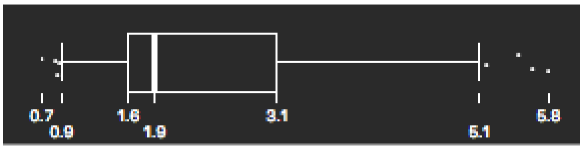


Boxplot



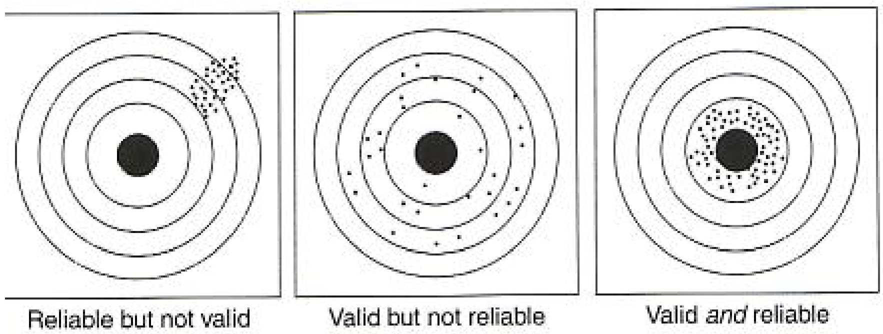


Boxplot and outliers



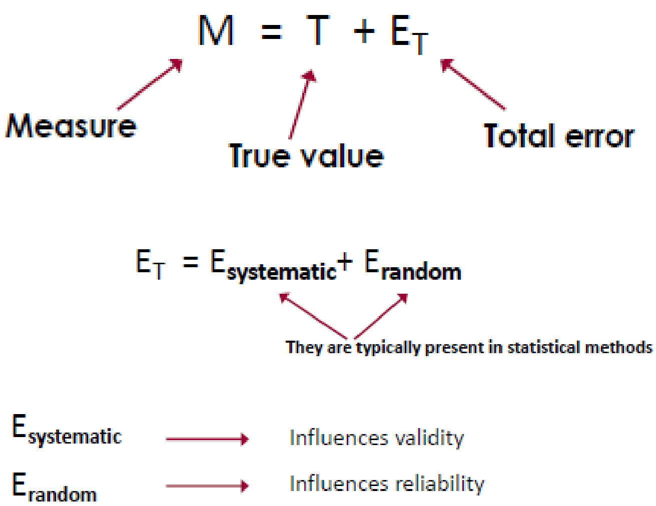
Quality of a measure

- Reliability: consistency of a measure, also the repeatability of a measurement; The smaller sigma^2 the more reliable the measure.
- Validity: strength of our conclusions, inferences or propositions; is the measure measuring what we actually are looing for? (similar to accuracy)



Errors in measuring

The difference between the measured value and the true one is called total error (ET)



Systematic errors influence validity because they occur constantly

If we assume that there can be only Er → its contribution on the averaged can be ignored.

In this case: E(M) = T

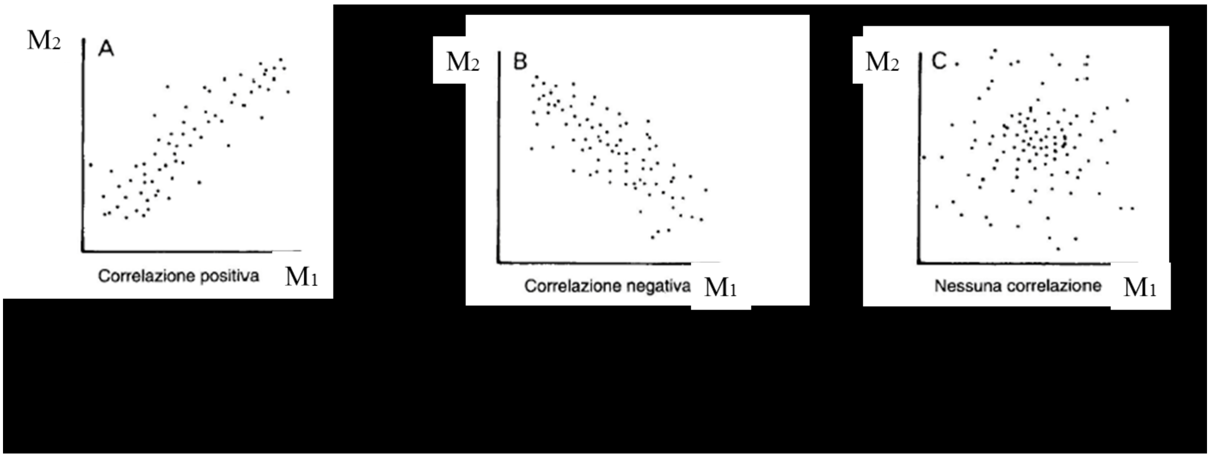
the smaller the error the lesser the influence
M = T + Er → var(M) = var(T) + var(Er)

The reliability of a measure is the ratio between the variance of the measured quantity and the variance of the metric
rhom=var(T)/var(M) = [var(M)-var(Er)]/var(M)= 1- [var(Er)/var(M)]

Reliability is between 0 and 1

Correlation

relationship between two variables holds (+1,-1,0)



8.5 Inferential statistics

We identify the following parameters
– Mean (sample) μ= (x1+ x2+... +xN)/n
– Variance (sample) var=[(x1-μ)2+ (x2-μ)2 +...(xN-μ)2]/(n-1)
– Standard deviation (sample) s=var1/2
– Percentile /median / mode etc. (sample)

Are random variables

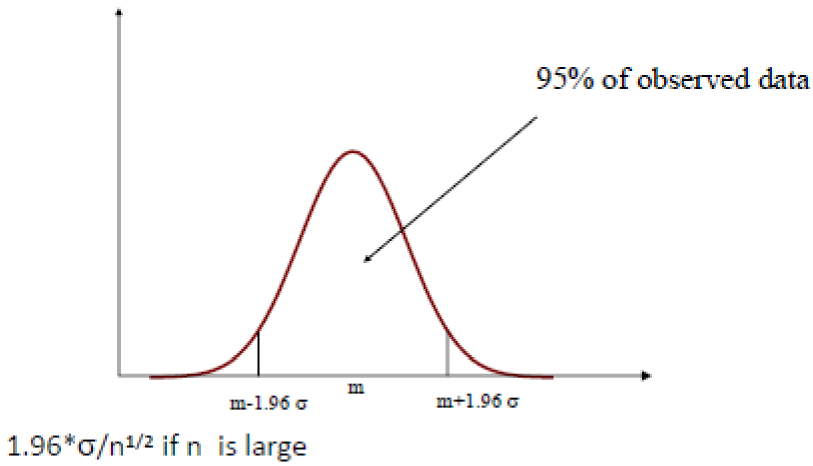
Confidence interval

We can estimate the probability that the mean of a population M is within an interval centered on the mean of a sample of n elements of such population.

The error probability is proportional to the standard deviation of the sample and inversely proportional to the size of the sample

Confidence interval size indicates the reliability of an estimate. A confidence level refers to the percentage of all possible samples that can be expected to include the true population parameter.

If the value of a parameter using a sample is x, with confidence interval [x-d, x+d] at confidence level P, then the actual population M parameter will be in [x-d, x+d] with a P probability.

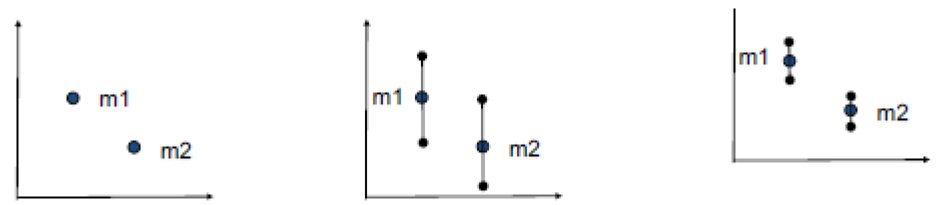


Ex. p.63 SE_07

8.6 Hypotheses Verification

A statistic test consist of challenging the hypothesis that the means of different samples are the same, the hypothesis that all true means are equal indicates that we assume that all observed differences are random: this hypothesis is called null hypothesis.

We fix a priori the probability of having an error (alfa).



Two means coming from two samples

Two confidence intervals that likely share the same mean

Two confidence intervals that likely DO NOT share the same mean hypotesis is a statement on the distribution of variables (H)

H0 nulle hypothesis → mua=mub

H1 (alternative hypotesis) → mua≠mub

We define a formula on the sample means capturing data differences. The formula generate a random variable and we compute the p-value. We select a priori alfa (significance level)

if:

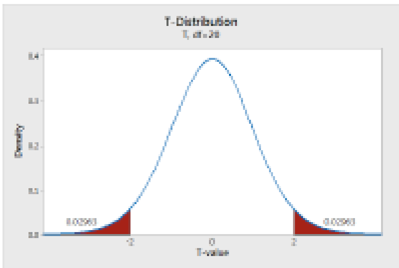
p>alfa → accept H0

p≤alfa → accept H1

t-test

allows to compare the difference between the means values of 2 samples

$$t = \frac{\mu_a - \mu_b}{\sqrt{\frac{\sigma_a^{2} + \sigma_b^{2}}{n}}}$$



assuming a and b from the same popuation we can compute the probability density of t nd the probability of t being equal or greater than a value

Indicates the probability that t≥X

p-value and alfa

A value of p >=α indicates that the difference between the observed means is "random" and we have to select the null hypothesis.

Probability of error

H0 is true:

reject H0 (alfa) type I error (false positive)

accept H0 (1-alfa)

H0 is false:

reject H0 (1-beta)

accept H0 (beta) type II error (false negative)

Type II errors arise frequently when the sample sizes are too small.

8.7 ANOVA

For n>2 samples we use ANOVA

ANOVA → Analysis of Variance

allows to analyze 2+ samples comparing the internal intra-variability with the inter-variability between groups.

F: Snedecor Variable

ANOVA produce same p-value as t-test

The hypotheses are the followings:
– H0: μ1 = μ2 …=μI
– H1: at least two among the means are different
I samples of J items

I > 2 different samples: {C1, … ,CI }

Yij is the j-th observation on the i-th sample

Where:

- Mean of sample *i* : $\mu_i = (\sum_{j=1}^{J} Y_{ij}) / J$
- General mean: $\mu = (\sum_{i=1}^{I} \mu_i) / I$

I samples of J items

F-test (Fisher test)

The **random** Snedecor variable F :

$$F = \frac{SS_B / (I-1)}{SS_W / [I(J-1)]}$$

$$SS_B = J \sum_{i=1}^{I} (\mu_i - \mu)^2$$

$$SS_W = \sum_{i=1}^{I} \sum_{j=1}^{J} (Y_{ij} - \mu_i)^2$$

This test tells us whether to:
- accept H0: p>a (F<F-crit)

- reject H0: p<=a (F>=F-crit)

Example

Example: I=3 samples, J=5 elements each

$$F = \frac{SS_B/(I-1)}{SS_W/[I(J-1)]}$$

I-1=2
I(J-1)=12

The probability that F>=3.89=0.05

numerator

denominator

nu \ nu	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161	200	216	225	230	234	237	239	241	242	244	246	248	249	250	251	252	253	254
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,41	19,43	19,45	19,45	19,46	19,47	19,48	19,49	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,74	8,70	8,66	8,64	8,62	8,59	8,57	8,55	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,91	5,86	5,80	5,77	5,75	5,72	5,69	5,66	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,68	4,62	4,56	4,53	4,50	4,46	4,43	4,40	4,37
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00	3,94	3,87	3,84	3,81	3,77	3,74	3,70	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,57	3,51	3,44	3,41	3,38	3,34	3,30	3,27	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,28	3,22	3,15	3,12	3,08	3,04	3,01	2,97	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,07	3,01	2,94	2,90	2,86	2,83	2,79	2,75	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,91	2,85	2,77	2,74	2,70	2,66	2,62	2,58	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,79	2,72	2,65	2,61	2,57	2,53	2,49	2,45	2,40
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,69	2,62	2,54	2,51	2,47	2,43	2,38	2,34	2,30
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,60	2,53	2,46	2,42	2,38	2,34	2,30	2,25	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,53	2,46	2,39	2,35	2,31	2,27	2,22	2,18	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,48	2,40	2,33	2,29	2,25	2,20	2,16	2,11	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,35	2,28	2,24	2,19	2,15	2,11	2,06	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,38	2,31	2,23	2,19	2,15	2,10	2,06	2,01	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,27	2,19	2,15	2,11	2,06	2,02	1,97	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,31	2,23	2,16	2,11	2,07	2,03	1,98	1,93	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,28	2,20	2,12	2,08	2,04	1,99	1,95	1,90	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,25	2,18	2,10	2,05	2,01	1,96	1,92	1,87	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,23	2,15	2,07	2,03	1,98	1,94	1,89	1,84	1,78

F-critical values at α = 0.05

I=2
J=7

$$\mu_a = (1+6+1+1+6+6+2)/7 = 3{,}28$$

$$\mu_b = (5+3+1+6+2+4+1)/7 = 3{,}14$$

$$\mu = (1+6+1+1+6+6+2+5+3+1+5+2+4+2)/14 = 3{,}21$$

$$SS_B = 7[(3{,}28-3{,}21)^2 + (3{,}14-3{,}21)^2] = 0{,}0714$$

$$SS_W = (1-3{,}28)^2 + (6-3{,}28)^2 + (1-3{,}28)^2 + (1-3{,}28)^2 + (6-3{,}28)^2 + (6-3{,}28)^2 + (2-3{,}28)^2 + \\ + (5-3{,}14)^2 + (3-3{,}14)^2 + (1-3{,}14)^2 + (6-3{,}14)^2 + (2-3{,}14)^2 + (4-3{,}14)^2 + (1-3{,}14)^2 = 62{,}29$$

$$F = \frac{0{,}0714/(2-1)}{62{,}29/[2(7-1)]} = 0{,}01376 \quad << \quad \text{F-crit}_{1,12} = 4.75$$

→ Accept H_0

Analysis of variance on tested KLOC and trend of defects

Most often the attempt of demonstrating an hypothesis substantiates in searching a relation between two variables

e.g.:

- Proportion KT=(tested KLOC)/total KLOC
- Defect rate in the first year DR=(D/KLOC)*k (let k be 1)

ANOVA allows us to evaluate the probability that the differences among the means are random.

To see if 2 of 3 samples belong to the same population It is necessary to compare the single pairs n*(n-1)/2=3

See exercise at p.94 of SE_07_MeasuresAndStatistics