4.1 CONCEPTES BÀSICS D'ESTADÍSTICA

DISTRIBUCIÓ DELS RESULTATS D'ANÀLISIS REPLICATS

Població de resultats

Amb una anàlisi repetida no trobo el mateix valor perquè hi ha variabilitat de resultat. Existeixen diferents valors amb freqüència alta. Tinc variabilitat que conforma una població de resultats de dades.

Reference (true) value: μ

Absolute error: $E_a = x_i - \mu$

Una població de dades ve caracteritzada per uns trets distintius:

1. **Tipus de distribució** que segueix aquesta població de dades. Marcada pel patró de comportament de la freqüència d'aparició duna dada donada.
2. **Valor central**: valor numèric que representa tota la població de dades. Pot ser el valor del mig o no.
3. **La dispersió de les dades**: com s'organitzen al voltant d'aquest valor central amb una certa variabilitat.

ESTADÍSTICS PARAMÈTRICS

Valor central

- **Mitjana aritmètica (AM)**

$$\boxed{\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}}$$

Dispersió

- **Desviació estàndard (SD)**
- **Variancia (s^2)**

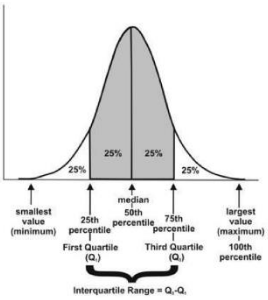
$$\boxed{s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}}$$

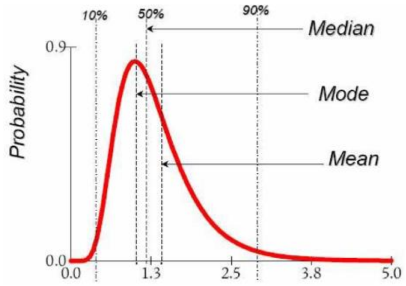
- Caracteritzen la distribució de resultats.
- Són molt sensibles a la presència de valors anòmals: valors que no haurien de pertànyer a la població de dades.

ESTADÍSTICS ROBUSTOS/ NO PARAMÈTRIQUES

Estadístics que no estan associats a una distribució estadística donada (distribució desconeguda) i són menys sensibles als anòmals, inclús per a distribucions esbiaixades (*skewed)*: població no simètrica amb pics desviats dreta/esquerra.

- **Mediana(Q_2)**: és el nombre que separa la meitat superior i inferior. Punt central de la dreta i de l'esquerra, estima el valor central d'una distribució de dades. Pot ser igual que la mitjana si tinc distribució simètrica. Es un paràmetre robust: no es basa en la distribució que tenim sinó que és el punt central.
- **Moda** és el valor que apareix més vegades.
- **Interval interquartil(IQR):** és la diferencia entre Q_3-Q_1 per tal de mesurar la dispersió estadística de la població de dades.





QUALITAT DELS RESULTATS ANALÍTICS: ERRORS EXPERIMENTALS

ERROR SISTEMÀTIC (DETERMINAT) → BIAIX

Bias estimation: $B = \bar{x} - \mu$

Relative bias: $B_r\,(\%) = \frac{\bar{x} - \mu}{\mu} \cdot 100$

- Defineix un patró diferenciat entre la referència i els valors replicats.
- Si provenen de la mateixa font tenen el mateix signe: per tant, existeixen valors per sobre o per sota del valor de referència.
- Quan es coneixen s'han d'evitar i corregir. Errors inacceptables

→ **VERACITAT**-Absència de biaix: grau d'acceptació de concordança entre els valors obtinguts i el valor de referència. Un 5-10% biaix molt alts! Han de ser inferiors al 5% com a màxim. **Fonts**:

- Presa de mostra incorrecta
- Tractament de mostra incorrecte
- Contaminació de la mostra
- Pèrdua d'anàlit
- Reaccions no controlades
- Selecció del mètode incorrecte
- Presència d'interferències
- Calibratge incorrecte

ERROR ALEATORI (indeterminat) → PRECISIÓ

- Petites diferències entre replicats, sense un patró definit respecte el valor de referència (diferències poden ser positives o negatives). També es pot fer respecte la mitjana o mediana obtinguda. No cal μ.
- Ha de ser minimitzat
- Descrit estadísticament i quantificat per la desviació estàndard.

→ **PRECISIÓ:** grau de concordança entre els replicats i la mitjana.

Standard deviation (SD):

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

Relative standard deviation (RSD):

$$S_r\,(\text{RSD}) = \frac{s}{\bar{x}}$$

Variance: s^2

ERROR TOTAL

→ **EXACTITUT**: els resultats són exactes quan són precisos i veraços alhora (sense biaix). No està afectat per errors aleatoris o sistemàtics. Per estimar l'error global tenim la incertesa. Sempre: Ens basem en la hipòtesi que no cometem errors sistemàtics.

Tipus d'error		**Propietat qualitativa**		**Mesura quantitativa**
Errors sistemàtics	→	**VERACITAT**	→	**Biaix**
↓		↓		↓
Error total	→	**EXACTITUD**	→	**Incertesa**
↑		↑		↑
Errors aleatoris	→	**PRECISIÓ**	→	**Desviació estàndard**

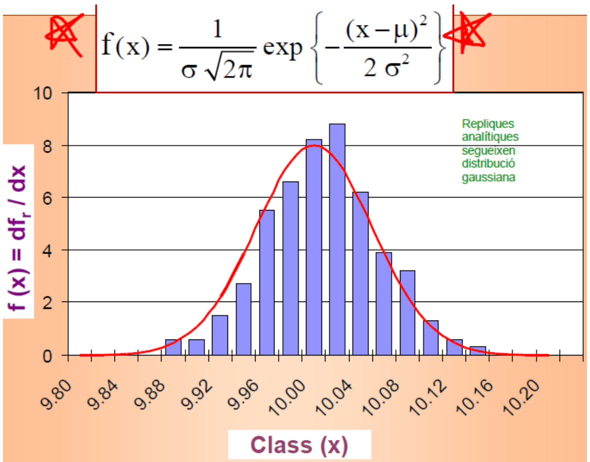
DISTRIBUCIÓ GAUSSIANA

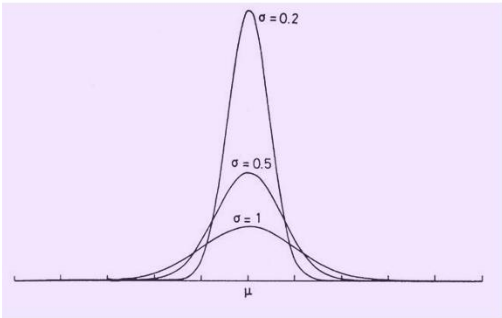
Descric la variabilitat inherent a les dades amb una distribució gaussiana. L' àrea total de la funció és 1. Cobreix totes les dades del −∞ a +∞. La dispersió es calcula amb la desviació estàndard. **σ:** paràmetre teòric de la funció de Gauss. Si **σ** és petita, la distribució és estreta al voltant del valor de referencia.

Hipòtesi inicial: cap dada té probabilitat nul·la (a priori). No hi ha error sistemàtic.

Histograma: represento dades agrupades per classes d'una certa amplada amb intervals 0,02 g o mL i mesuro el nombre de vegades que surt dins d'aquest grup. A mesura que fem més mesures obtenim una distribució normal. Les repliques analítiques segueixen una distribució gaussiana.

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$



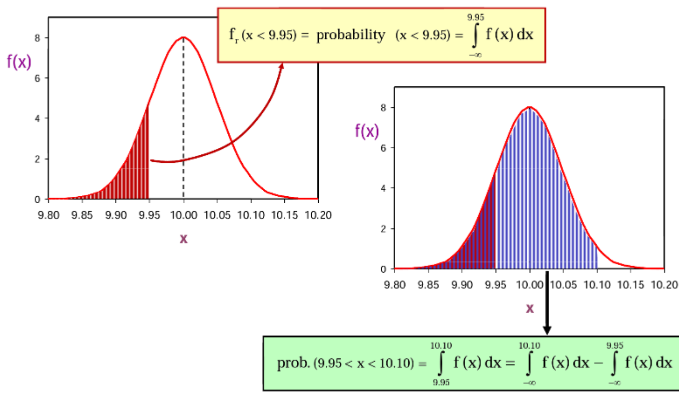


$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{n}}$$

Probabilitat d'obtenir un valor donat o un interval de valors: perquè puc calcular l'àrea parcial integrant, segons els valors de x que m'interessen.

Integral normalitzada que depèn de l'amplada de la corba que m'interessa de major freqüència més pròxima al valor real. Vull saber la probabilitat d'obtenir un valor concret i que no depengui del valor de la variable. Normalitzem la funció de gauss, estandarditzarla amb la variable **z** que em permet igualar totes les dades independentment del nombre de repliques.



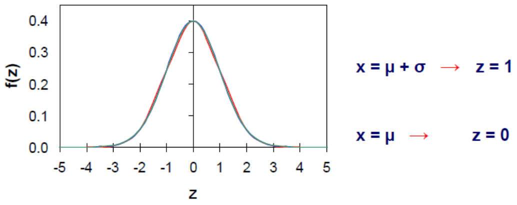
DISTRIBUCIÓ NORMAL ESTANDARDITZADA: Quan n>30 i/o coneixem σ: sigma depèn del nº d'observacions.
Canvi de variable: aplica la condició obligatòria: sigma igual a la unitat.

$$\boxed{z = \frac{(x-\mu)}{\sigma}}$$

Gràcies al canvi de variable no cal integrar, tenim totes les probabilitats **tabulades**.

Z SCORE- PARÀMETRE Z

Mitjana o valor de replicats és igual al valor real si la z es igual a zero. Així tinc una distribució centrada sobre 0. I la sigma pren el valor de la unitat. Per tenir una sigma per dalt o per baix, la probabilitat de tenir un valor experimental al voltant de 1 sigma es de 68,2%. 2 sigmes estem pròxims al 95% i una mica més. Excloem dades molt altes o molt baixes. 1,96 vegades la sigma.



TAULES:

DISTRIBUCIÓ NORMAL STANDARDITZADA: Taula única estandarditzada, és una distribució universal per a qualsevol població de dades. Representa la probabilitat associada amb un valor de la z (alfa o nivell de significació). Hi ha d'una cua només ens interessa una frontera. La probabilitat de z és 0,5 perquè està al mig. L'àrea cap a la dreta es la meitat. A mesura que la z augmenta la alfa va disminuint.
Si tenim de dues cues la probabilitat de z és 1. La probabilitat de l'àrea que queda exclosa (α) és l'àrea de la dreta de la z. Hi ha una relació entre els valors d'una taula d'una i de dues cues:

Z (α, 1 cua) = z (2 α, 2 cues)
Z (α, 2 cues) = z (α/2 , 1

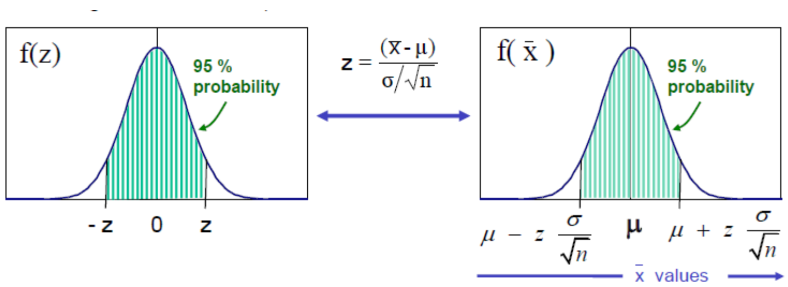
DISTRIBUCIÓ T-STUDENT: Quan tenim baix nombre d'observacions **n<30** i/o desconeixem **σ. t depèn de s i dof**
És una millor estimació de σ per a poblacions amb un nombre de dades reduït. Evoluciona a una distribució normal quan s'incrementa el nombre de graus de llibertat: nombre de desviacions independents. **Dof = n-1**
Darrera de qualsevol variable hi ha distribució normal però com calculo la sigma? S'ha d'estimar a partir de s.
Accepto que les dades són normals però no puc aplicar la funció gaussiana amb una determinada sigma.
Quin valor de t tinc en funció de probabilitat: àrea que excloc i graus de llibertat.

DISTRIBUCIÓ DE MITJANES
Objectiu anàlisi: estimar valor real de referència. ~~Distribució de replicats~~ NO Distribució de mitjanes.
Distribució normal de replicats: mesures individuals de la variable: μ ± σ (SD)

$$\boxed{\sigma_x = \frac{\sigma}{\sqrt{n}}}$$

Distribució normal de mitjanes: calculada per n repliques individuals μ ± σ_x (SD)
La sigma de la distribució de mitjanes es igual a la sigma de rèpliques entre l'arrel de n.
TEOREMA DE LÍMIT CENTRAL: la distribució de mitjanes evoluciona a una distribució normal quan el nombre de dades augmenta. Qualsevol variable si la promitjo és una distribució gaussiana.

Interval de confiança del valor mitjà
El valor de referència es pot predir de la mitjana i σ/s (només considerant que existeix un error aleatori). Podem escollir els valors de probabilitat que acceptem: amb un 95% de probabilitat, per exemple.



si conec la sigma d'estudis previs:

$$\bar{x} = \mu \pm z\frac{\sigma}{\sqrt{n}} \Rightarrow \boxed{\mu = \bar{x} \pm z\frac{\sigma}{\sqrt{n}}}$$

Si la mitjana pertany a una mitjana gaussiana:

$$\boxed{\mu = \bar{x} \pm t_{(n-1,\, d.o.f.)}^{2}\frac{s}{\sqrt{n}}}$$

INTERVAL DE CONFIANÇA BASAT EN LA INCERTESA
El valor de referència es pot predir de la mitjana i de l'estimació de l'error total: incertesa.

Incertesa estàndard(u_c) **$\mu = \bar{x} \pm u_c$** expressada com una SD global
Incertesa expandida: **$U = u_c \cdot k$** **$\mu = \bar{x} \pm U$** corregida amb un factor de cobertura