PRINCIPALS PROCEDIMENTS PER AVALUAR LA CONCORDANÇA ENTRE OBSERVADORS AMB MESURES D'ESCALA QUANTITATIVA

Josep Lluís Carrasco
Departament de Salut Pública

UNIVERSITAT DE BARCELONA

Continguts

1. INTRODUCCIÓ
2. COMPONENTS DE LA DISCORDANÇA
3. ANÀLISI DESAGREGAT
4. COEFICIENT DE CONCORDANÇA
5. COEFICIENT DE CORRELACIÓ INTRACLASSE
6. MÈTODE BLAND-ALTMAN
7. TOTAL DEVIATION INDEX
8. COMENTARIS FINALS

Introducció

- Dos mètodes de mesura (observadors, avaluadors,…) amb resposta quantitativa mesuren la mateixa característica.

- Es desitja mesurar el grau d'acord entre els dos mètodes:
  - Són els resultats d'un mètode reproduïbles pel segon?
  - Comparació d'un mètode amb un altre de referència
  - Realment mesuren la mateixa característica?

- La qüestió no és si concorden o no, sinó en quin grau discorden i quines implicacions te aquesta discordança.

Exemples:

- Dues balances de bany. Es vol substituir una per una altra. Ens engreixarem uns Kg sobtadament?

- Dos analitzadors automàtics donant valors de paràmetres bioquímics (concentració en sang de glucosa, colesterol,…)

- Es desitja substituir un esfigmomanòmetre antic per un de nou. Obtindrem les mateixes mesures de PA?

- Bioequivalència individual. Comparació de les biodisponibilitats de dos productes farmacèutics.

Exemple PA

- Instruments: Un esfigmomanòmetre manual de mercuri i un automàtic OMRON 711
- Mostra: 384 individus
- Es va mesurar simultàniament la pressió arterial (sistòlica i diastòlica) dues vegades amb cada aparell.
- Es desitja avaluar el impacte que pot tenir substituir els aparells manuals pels automàtics

Components de la Discordança

Model de mesura dels dos mètodes (X i Y):

$$X_i = \alpha_X + \beta_X \tau_i + e_{X\,i}$$

$$Y_i = \alpha_Y + \beta_Y \tau_i + e_{Y\,i}$$

α_X, α_Y Error sistemàtic constant

β_X, β_Y Error sistemàtic proporcional

$\tau_i \sim N(0, \sigma_\tau^2)$ Valor real

$e_{Xi} \sim N(0, \sigma_{eX}^2)$, $e_{Yi} \sim N(0, \sigma_{eY}^2)$ Errors aleatoris

Concordança perfecta, $Y_i = X_i$ implica que $\alpha_Y = \alpha_X;\ \beta_Y = \beta_X;\ e_{Y\,i} = e_{X\,i} = 0$

En termes de les mitjanes, variàncies i covariància de les dues variables

$$\mu_X = E(X_i) = \alpha_X \,;\, \mu_Y = E(Y_i) = \alpha_Y$$

$$\sigma_X^2 = Var(X_i) = \beta_X^2\sigma_\tau^2 + \sigma_{eX}^2;\, \sigma_Y^2 = Var(Y_i) = \beta_Y^2\sigma_\tau^2 + \sigma_{eY}^2$$

$$Cov(X_i, Y_i) = Cov\left(\alpha_X + \beta_X\tau_i + e_{X_i}; \alpha_Y + \beta_Y\tau_i + e_{Y_i}\right) = \beta_X\beta_Y\sigma_\tau^2$$

$$\rho_{XY} = Corr(X_i; Y_i) = \frac{(\beta_X\beta_Y\sigma_\tau^2)}{\sqrt{(\beta_X^2\sigma_\tau^2 + \sigma_{eX}^2)(\beta_Y^2\sigma_\tau^2 + \sigma_{eY}^2)}}$$

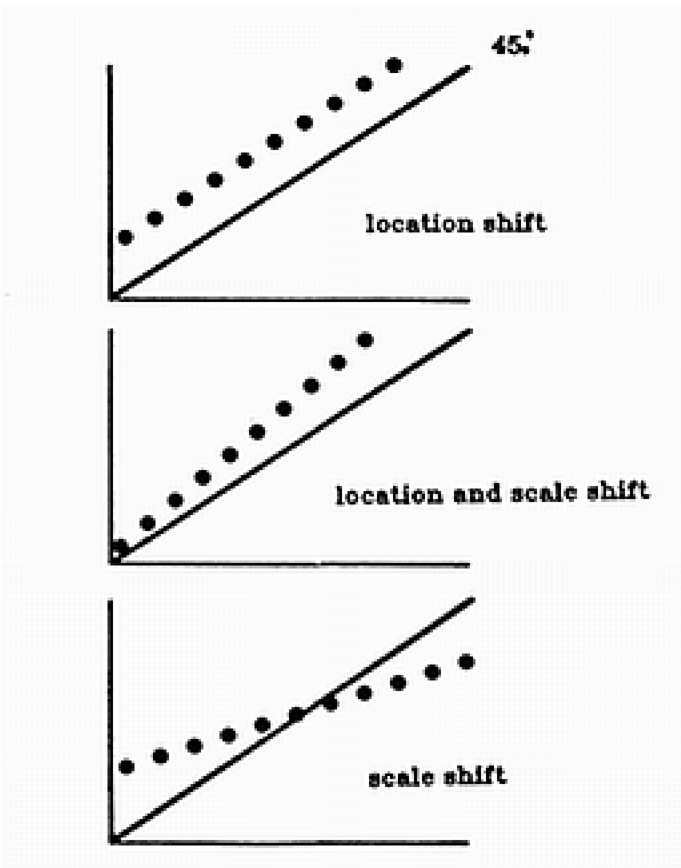
Concordança perfecta

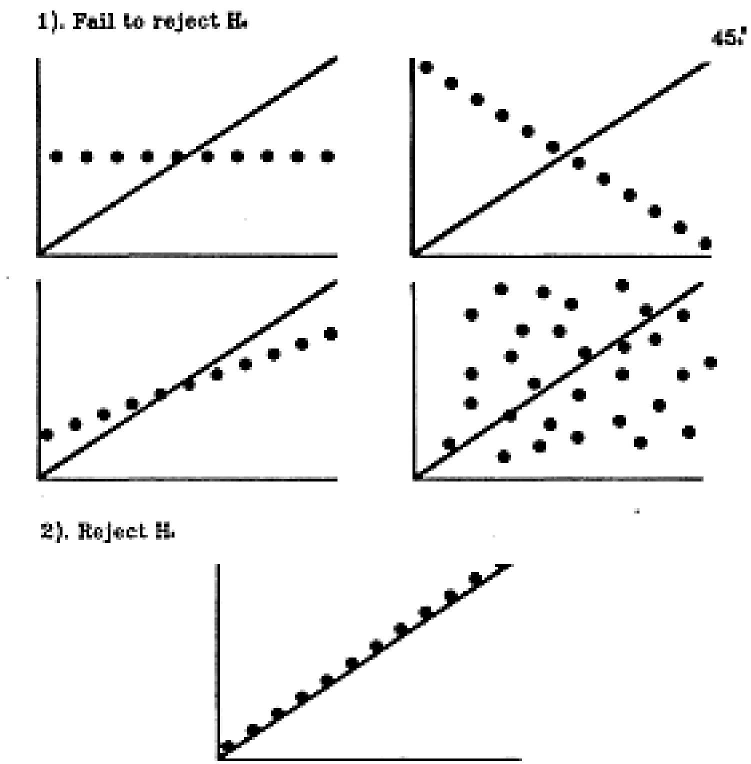
1) $\alpha_Y = \alpha_X \Rightarrow \mu_X = \mu_Y$. Igualtat de mitjanes

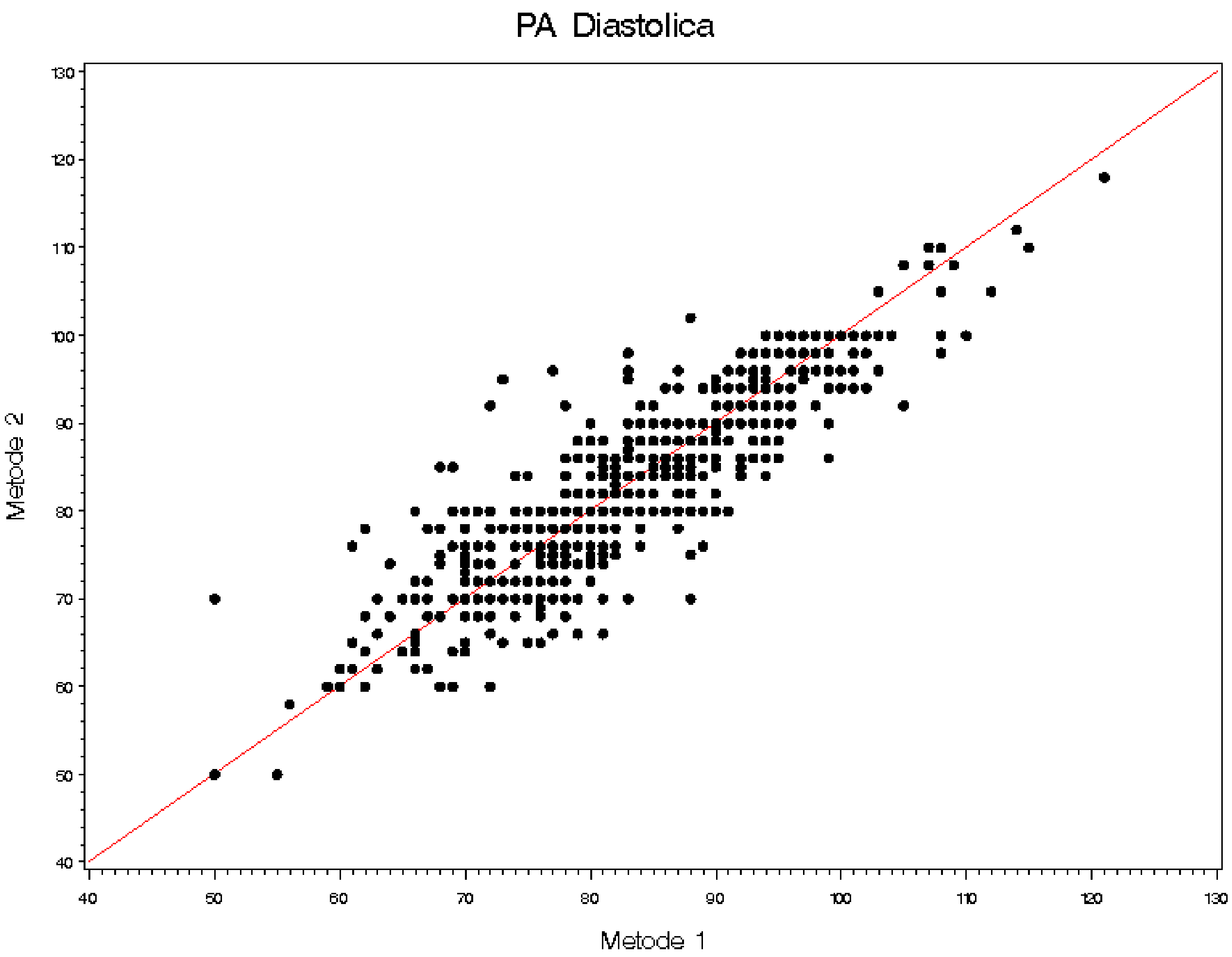
2) $e_{Y_i} = e_{X_i} = 0 \Rightarrow \sigma_{eY}^2 = \sigma_{eX}^2 = 0 \Rightarrow \rho_{XY} = 1$ Associació lineal perfecta

3) $\beta_Y = \beta_X = \beta \text{ i } \sigma_{eY}^2 = \sigma_{eX}^2 \Rightarrow \sigma_X^2 = \sigma_Y^2$ Igualtat de variàncies

Lin (1989). *Biometrics* **45**(1), 255.






PA Diastolica
Metode 2
Metode 1
130
120
110
100
90
80
70
60
50
40
40
50
60
70
80
90
100
110
120
130

Anàlisi Desagregat

Es tracta d'avaluar les tres components per separat:

1) Diferència de mitjanes $d = \mu_X - \mu_Y$

2) Diferència o quocient de variàncies $q = \frac{\sigma_Y^2}{\sigma_X^2}$

3) Associació lineal ρ_{XY}

Concordança perfecta: s'hauria d'avaluar les següents hipòtesis

$$\begin{array}{lll} H_{01}: d = 0 & H_{02}: q = 1 & H_{03}: \rho_{XY} = 1 \\ H_{A1}: d \neq 0 & H_{A2}: q \neq 1 & H_{A3}: \rho_{XY} \neq 1 \end{array}$$

Però no te sentit ja que avaluar la concordança perfecta no és una qüestió probabilística sinó determinista: Si alguna mesura és diferent ja no hi haurà concordança perfecta.

La pregunta adequada és saber si cadascun dels paràmetres es manté dins d'uns límits

$$d \in [-\theta_1, \theta_1];\, q \in [\theta_2^{-1}, \theta_2];\, \rho_{XY} > \theta_3$$

Hipòtesis associades

$$\begin{array}{lllll} H_{01}: d = -\theta_1 & H_{02}: d = \theta_1 & H_{03}: q = \theta_2^{-1} & H_{04}: q = \theta_2 & H_{05}: \rho_{XY} = \theta_3 \\ H_{A1}: d < -\theta_1 & H_{A2}: d > \theta_1 & H_{A3}: q < \theta_2^{-1} & H_{A4}: q > \theta_2 & H_{A5}: \rho_{XY} < \theta_3 \end{array}$$

Problema: La hipòtesi nul·la és la de concordança!!!

Per exemple, si no rebutgéssim totes les H_0 per un problema de potència, podríem dir que els dos mètodes concorden?

Es tracta d'un problema d'equivalència. Posem allò que volem demostrar a la hipòtesi alternativa.

Les hipòtesis adequades a contrastar són:

$$
\begin{array}{lllll}
H_{01}: d = -\theta_1 & H_{02}: d = \theta_1 & H_{03}: q = \theta_2^{-1} & H_{04}: q = \theta_2 & H_{05}: \rho_{XY} = \theta_3 \\
H_{A1}: d > -\theta_1 & H_{A2}: d < \theta_1 & H_{A3}: q > \theta_2^{-1} & H_{A4}: q < \theta_2 & H_{A5}: \rho_{XY} > \theta_3
\end{array}
$$

Problema de multiplicitat de l'error de tipus I? Principi d'unió-intersecció:

1) H_0 global com a unió de totes les H_0
2) H_A global com a intersecció de totes les H_A

$$H_0 = H_{01} \cup H_{02} \cup H_{03} \cup H_{04} \cup H_{05}$$

$$H_A = H_{A1} \cap H_{A2} \cap H_{A3} \cap H_{A4} \cap H_5$$

H_0 es rebutja amb un nivell de significació global $\leq \alpha$ **si totes** les H_0 són rebutjades al nivell de significació α.

Handicaps del procediment unió-intersecció:

1) Garanteix l'estabilitat del nivell de significació però pot ser poc potent.

2) S'han de definir almenys 3 límits (mitjanes, variàncies, correlació)

En l'exemple de pressió arterial es podrien definir els següents límits

$$\theta_1 = 10 ; \theta_2 = 1.25; \theta_3 = 0.9$$

$$\hat{d} = 0.49; \quad \hat{q} = 1.00; \qquad \hat{\rho}_{XY} = 0.82$$

Un cop triats els límits es duen a terme els tests apropiats:

Choudhary PK, Nagaraja HN. (2005) *Biometrical Journal* **47**(4), 1-8

Carrasco JL; Jover. (2005). *Biometrical Journal* **47**(5), 623-634

En aquest exemple es rebutjaren totes les hipòtesis excepte la cinquena (correlació).

El Coeficient de Concordança

Lin (1989). *Biometrics* **45**, 255; Lin (1992) *Biometrics* **48**, 599

Mesura la distància mitjana dels parells de dades (X,Y) a la recta de 45º

Mean Squared Deviation (MSD)

$$D = Y - X$$

$$MSD = \epsilon^2 = E(D^2) = E[(Y - X)^2] = (\mu_X - \mu_Y)^2 + \sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}$$

MSD amb X,Y incorrelacionades $\epsilon_{\rho=0}^2 = (\mu_X - \mu_Y)^2 + \sigma_X^2 + \sigma_Y^2$

Coeficient de concordança

$$\rho_{CCC} = 1 - \frac{\epsilon^2}{\epsilon_{\rho=0}^2} = \frac{2\rho_{XY}\sigma_X\sigma_Y}{(\mu_X - \mu_Y)^2 + \sigma_X^2 + \sigma_Y^2}$$

Pren valors en el interval [-1 , 1]

1) $\rho_{CCC} = 1$ en el cas de concordança perfecta $\mu_X = \mu_Y; \sigma_X^2 = \sigma_Y^2; \rho_{XY} = 1$

2) $\rho_{CCC} = 0$ implica independència $\rho_{XY} = 0$

3) $\rho_{CCC} = -1$ significa discordança perfecta. Valors negatius tenen poc sentit

La hipòtesis a testar és $\begin{array}{l} H_0: \rho_{CCC} = \theta \\ H_A: \rho_{CCC} > \theta \end{array}$

Un límit per a considerar intercanviables dos mètodes podria ser θ=0.9

En el nostre exemple l'estimació del CCC va ser $\hat{\rho}_{CCC} = 0.8188$; $\hat{\rho}_{CCC,5\%} = 0.7962$

Per tant no es pot concloure que siguin intercanviables

El Coeficient de Correlació Intraclasse

$$Y_{ij} = \mu + \alpha_i + \beta_j + e_{ij}$$

Y_{ij} mesura de 'observador *j* en l'individu *i*
$\alpha_i \sim N(0, \sigma_\alpha^2)$ efecte individu; *i*=1,…,n;
β_j efecte mètode; *j*=1,…,k;
$e_{ij} \sim N(0, \sigma_e^2)$ error aleatori;

$$\rho_{ICC} = \frac{Cov(Y_{ij}, Y_{ij'})}{Var(Y_{ij})}$$

L'efecte mètode ha de ser fix o aleatori?

- Si tan sols es vol avaluar la concordança entre un nombre concret (fix) d'observadors l'efecte hauria de ser fix.
- Si el que tenim és una mostra aleatòria d'una població d'observadors i es desitja inferir el resultats a aquesta població l'efecte hauria de ser aleatori

- Observador fix

$$Cov(Y_{ij}, Y_{ij'}) = \sigma_\alpha^2$$

$$Var(X_{ij}) = \sigma_\alpha^2 + \sigma_e^2$$

$$\rho_{ICC} = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_e^2}$$

Aquest índex no mesura concordança, mesura **consistència**!!!

Els mètodes són **consistents** si, independentment d'a on es centren (mitjanes), l'ordre (correlació) i distància entre les mesures (variància) és la mateixa entre els observadors

- Observador aleatori $\beta_j \sim N(0, \sigma_\beta^2)$

$$Cov(Y_{ij}, Y_{ij'}) = \sigma_\alpha^2$$

$$Var(X_{ij}) = \sigma_\alpha^2 + \sigma_\beta^2 + \sigma_e^2$$

$$\rho_{ICC} = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\beta^2 + \sigma_e^2}$$

Aquest índex si que mesura concordança. Per tant per a que l'ICC mesuri concordança s'ha de considerar l'observador com a aleatori.

Però, i si per disseny aquest efecte ha de ser fix?

La quantitat σ_β^2 serà una suma de quadrats enlloc d'una variància

$$\sigma_\beta^2 = \frac{1}{k-1}\sum_{j=1}^{k}\beta_j^2$$

CCC vs ICC

Carrasco, Jover (2003). *Biometrics*, **59**, 849

Cas k=2 mètodes

1) $\sigma_\alpha^2 = \sigma_{12}$;
2) $\sigma_\beta^2 = \frac{1}{2}(\mu_1 - \mu_2)^2$;
3) $\sigma_e^2 = \frac{1}{2}(\sigma_1^2 + \sigma_2^2) - \sigma_{12}$

$$\rho_{ICC} = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\beta^2 + \sigma_e^2} = \frac{2\sigma_{12}}{(\mu_1 - \mu_2)^2 + \sigma_1^2 + \sigma_2^2} = \rho_{CCC}$$

Aquesta igualtat també es demostra per k>2 mètodes

A l'exemple de PA

$$\hat{\sigma}_\alpha^2 = 77.82; \hat{\sigma}_\beta^2 = 0.10 ; \hat{\sigma}_e^2 = 17.12$$
$$\hat{\rho}_{ICC}=0.8188, \hat{\rho}_{ICC,5\%} = 0.7962$$

La principal font de discordança és l'error aleatori.

- Avantatges de l'ICC
  1) Generalització senzilla per a incloure rèpliques i extensió a *k>2* observadors.
  2) El mètode d'estimació és Màxim Versemblant (REML): bon comportament asimptòtic de les estimacions (si la versemblança és certa).
  3) Permet adequar l'índex a situacions no estàndard:
     - Count data (Carrasco, Jover (2005). *Statistics in Medicine*, **24**, 4021)
       Dues tècniques de recompte de CD34.
     - Mesures repetides. (Carrasco, King, Chinchilli (2009). *Journal of Biopharmaceutical Statistics*, **19**, 90).
       Dos mètodes per quantificar la concentració de cortisol en sang en asmàtics.
       Es va mesurar la concentració de cada pacient en quatre visites.

- Handicap: És una mesura dependent de la covariància i per tant del rang de mesura de la mostra: Cal determinar prèviament quin és el rang real de la variable analitzada o en el que es vol inferir el grau de concordança i comprovar si la mostra ho compleix

Escala CCC (ICC)

Agreement	**Almost perfect**	**>0.9**
	Good	0.7-0.9
	Fair	0.5-0.7
Disagreement	Poor	0.3-0.5
	Bad	0.1-0.3
	Independence	<0.1

- Valors negatius:
  - En l'ICC no són possibles (components de la variància positives) però sí en d'altres estimadors del CCC.
  - Valors propers a zero: pot ser degut a variabilitat mostral en un cas d'independència.
  - Valors allunyats de zero i negatius: revisar les dades.

Mètode Bland-Altman

Bland, Altman (1986). *Lancet*, **i** (8476), 307
Bland, Altman (1995). *Lancet*, **346** (8982), 1085
Bland, Altman (1999). *Statistical Methods in Medical Research*, **8**, 135

Y_{i1} mesures del primer observador
Y_{i2} mesures del segon observador

amb *i=1,..,n* individus

Es defineix la diferència entre els dos observadors com $d_i = Y_{i1} - Y_{i2}$

Es calcula la mitjana i la variància de la variable diferència

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i$$

$$s_d^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(d_i - \bar{d}\right)^2$$

Limits of agreement

$[L_1, L_2] = \bar{d} \pm 2s_d$. Dins d'aquests límits es troben aproximadament el 95% de les diferències.

Procediment: Comparar aquests límits amb uns especificats prèviament (habitualment simètrics) $[-\theta, \theta]$. Si $[L_1, L_2] \in [-\theta, \theta]$ es conclou que els mètodes concorden

Es tracta d'un procediment escalat (unitats de mesura = variable original)

Què succeeix si es tenen rèpliques? Es modelen les dades mitjançant el model d'efectes mixtes.

$$Y_{ijl} = \mu + \alpha_i + \beta_j + e_{ijl}$$

$$d_i \sim N(\beta_1 - \beta_2, 2\sigma_e^2)$$

$$[L_1, L_2] = \hat{\beta}_1 - \hat{\beta}_2 \pm 2\sqrt{2\hat{\sigma}_e^2}$$

Intervals de Tolerància

Els límits de concordança pretenen estimar un interval de probabilitat

Estimar un interval de probabilitat amb una certa confiança es coneix com a **interval de tolerància**.

Un interval de tolerància bilateral que conté el p% de les diferències entre el primer i segon mètodes amb una confiança $(1-\alpha)\%$ ve donat per

$$\bar{d} \pm g(p; n; 1-\alpha) s_d$$

On

$$g(p; n; 1-\alpha) = Z_{\frac{(1+p)}{2}} \left(1 + \frac{1}{2n}\right) \sqrt{\frac{n-1}{\chi^2_{\alpha; n-1}}}$$

Hahn, G.J. and Meeker, W.Q. (1991), *Statistical Intervals: A Guide for Practitioners*, New York: John Wiley & Sons, Inc.

Exemple Dades Pressió

- Límits previs: ± 10 mmHg

- p=0.95; α=0.05;

$$\hat{\beta}_1 - \hat{\beta}_2 = 0.49\ ;\quad \hat{\sigma}_e^2 = 17.12;\quad g(p; n; 1-\alpha)\text{=}2.087$$

$$[L_1, L_2] = [-11.72\ ; 12.70]$$

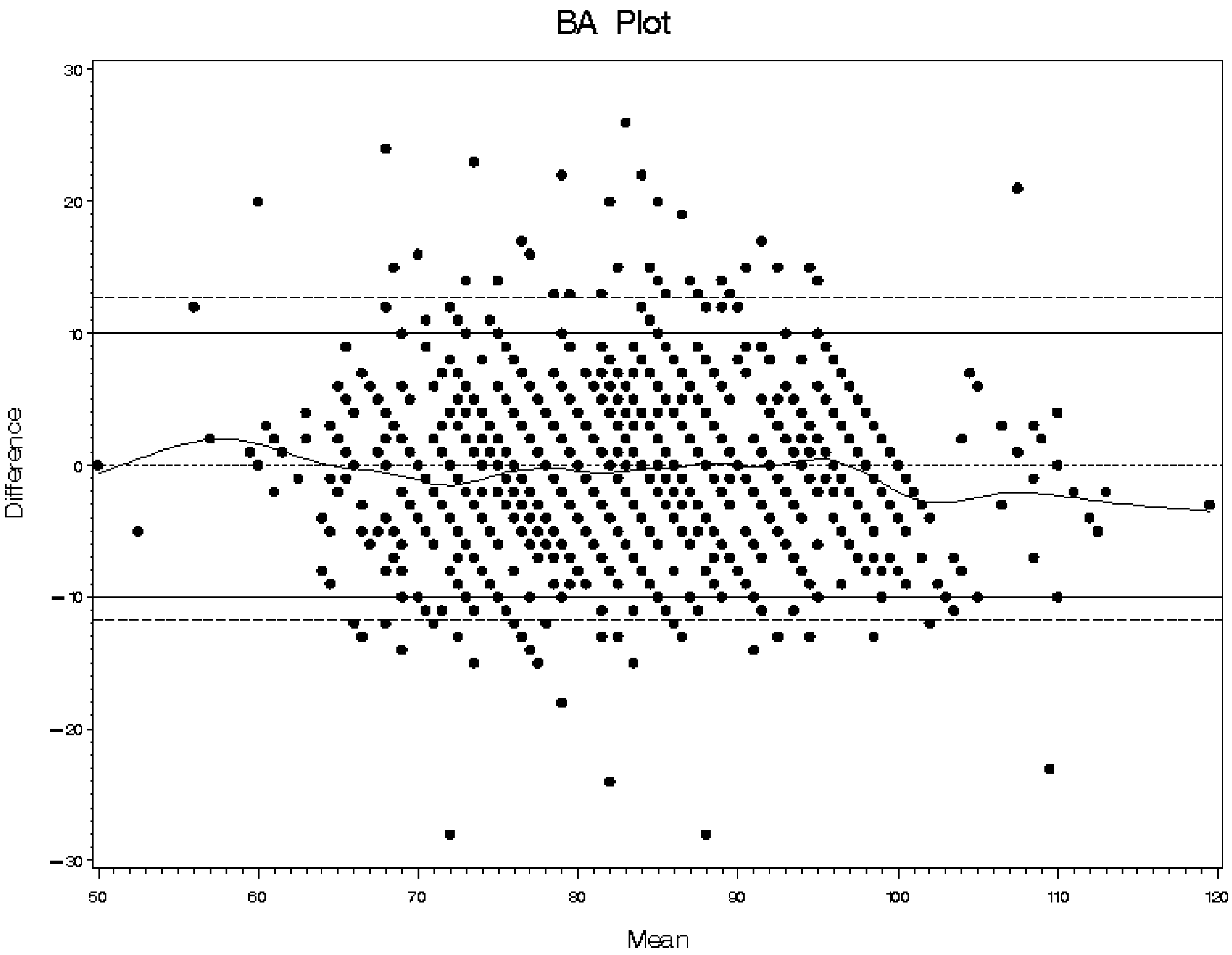
- p=0.9; α=0.05;

$$g(p; n; 1-\alpha)\text{=}1.75$$

$$[L_1, L_2] = [-9.76\ ; 10.74]$$

No es compleix el criteri per a considerar els mètodes concordants.

Representació gràfica



Total Deviation Index

- En el mètode BA el interval es troba centrat a la mitjana de les diferències i els seus límits depenen de com es realitzen aquestes diferències: $Y_1 - Y_2$ o $Y_2 - Y_1$

  Suposem que $d_i = Y_{i1} - Y_{i2}$ ens dona el interval $[L_1, L_2]$. Si fem $d_i^* = Y_{i2} - Y_{i1}$, aleshores $\bar{d}^* = -\bar{d}$ i $[L_1^*, L_2^*] = [-L_2, -L_1]$

- Si $abs(L_2) > abs(L_1)$ la diferència més gran que es pot donar és $abs(L_2)$, amb independència del sentit de la diferència.

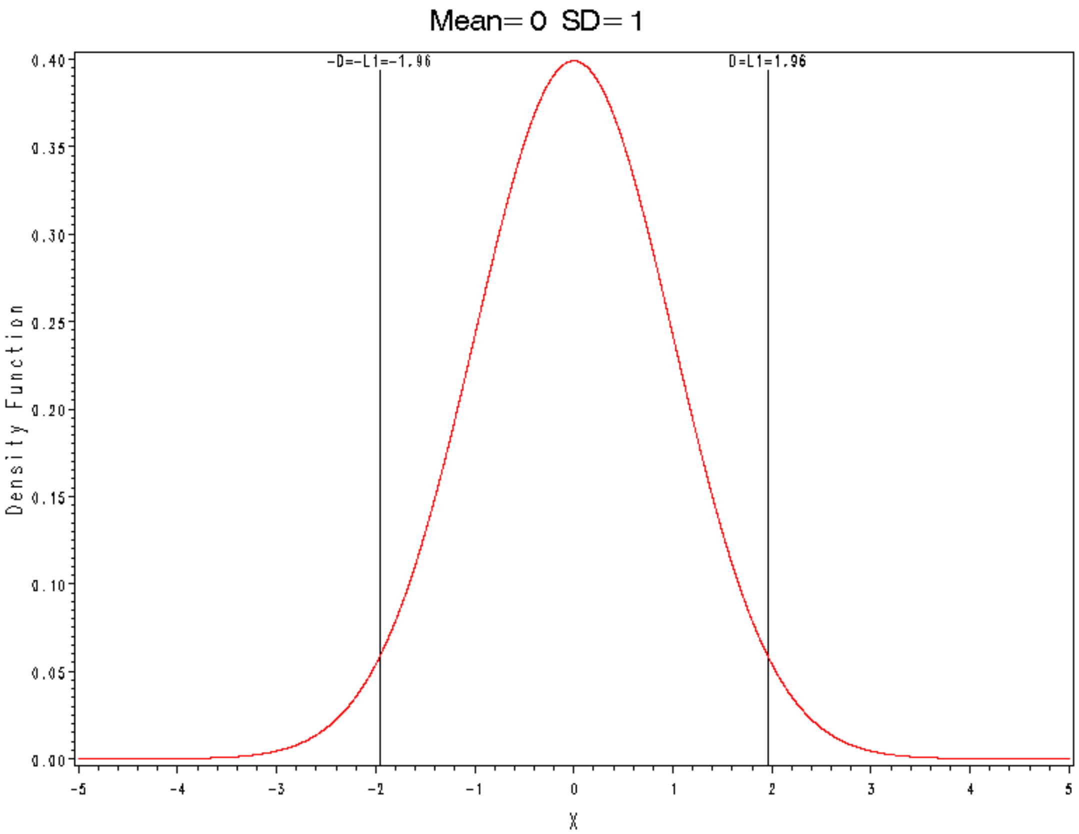
- En l'exemple hem vist que $[L_1, L_2] = [-11.72\ ; 12.70]$, per tant la diferència més gran entre els dos mètodes és de 12.70 mmHG.

  Això és equivalent a donar un interval del tipus $[-\Delta, \Delta]$ on $\Delta = \max(abs[L_1, L_2])$

- Problema: Pr $(-\Delta < d < \Delta) \geq p$; i només serà igual a p si $\bar{d}$=0.

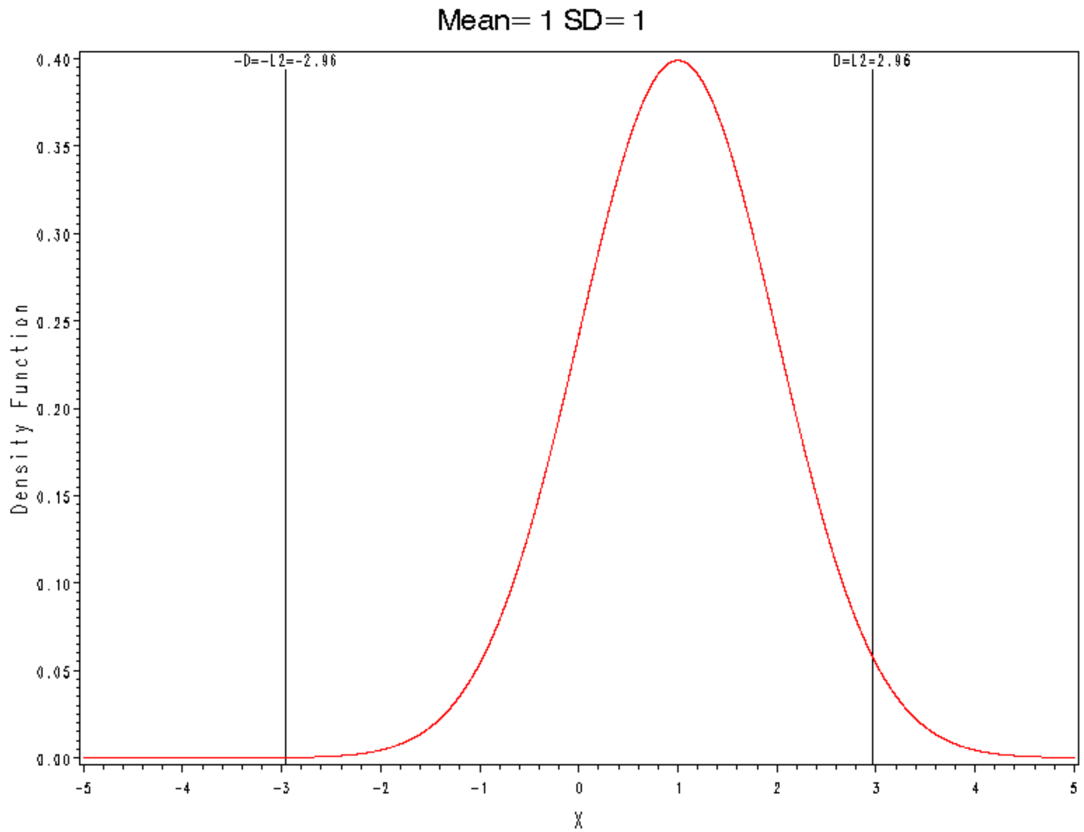
  Mitjana=0 Desviació estàndard=1. El interval de probabilitat del 95% és

$$[L_1, L_2] = [-1.96\,;1.96] \text{ i per tant } [-\Delta, \Delta] = [-1.96\,;1.96]$$



Mitjana=1 Desviació estàndard=1. El interval de probabilitat del 95% és

$$[L_1, L_2] = [-0.96\,;2.96] \text{ i per tant } [-\Delta, \Delta] = [-2.96\,;2.96]$$



Però el interval $= [-2.96\,;2.96]$ conté més del 95% de la probabilitat!!!

El TDI per una determinada probabilitat p es defineix com aquell $\kappa > 0$ tal que

$$\Pr(|D| < \kappa) = p$$

La inferència amb la distribució de $|D|$ pot ser complicada

- Lin proposa una aproximació basada en el MSD (Lin (2000). *Statistics in Medicine*, **19**, 255)

$$TDI_p = \Phi^{-1}\left(\frac{1+p}{2}\right)\sqrt{MSD}$$

Malauradament l'aproximació és acceptable per valors petits de μ_d^2/σ_d^2. Per exemple, amb p= 0.9 si $\frac{\mu_d^2}{\sigma_d^2} < 1$.

Per valors superiors el TDI és sensiblement sobreestimat.

- Choudary i Nagaraja ((2007). *Journal of Statistical Planning and Inference*, **137**, 279) proposen una aproximació basada en una distribució khi-quadrat no centrada.

  L'aproximació és correcta per $0.8 \leq p \leq 0.95$ i $n \leq 30$.

  Per mides mostrals superiors l'aproximació tendeix a sobreestimar: es proposa utilitzar un bootstrap-t.

- Les funcions es troben disponibles a:

  http://www.utdallas.edu/~pankaj/prob_criteria/r_code.html

- En l'exemple de PA $\frac{\mu_d^2}{\sigma_d^2} = \frac{0.49}{34.24} \cong 0$ i n=384

  El TDI del 95% per un nivell de significació del 5% és

  - Aproximació Lin. 11.86
  - Bootstrap-t Choudary. 12.23

Comentaris Finals

1) La qüestió d’interès no és avaluar si els mètodes concorden perfectament, sinó avaluar el grau de discordança.

2) Els diferents procediments són compatibles en un mateix anàlisi.

3) Previ a l’anàlisi s’han d’establir uns límits a partir dels quals la discordança no sigui acceptable.

4) En un anàlisi de “intercanviabilitat” de mètodes una discordança superior a la desitjada no implica necessàriament que no es pugui substituir un mètode per un altre, altrament pot ser una raó per fer-ho.

5) Els mètodes exposats també són aplicables a estudis de fiabilitat (concordança entre repliques d’un sol mètode).