

Grau en Estadística

Títol: Estudi de la fiabilitat d'ela escintigrafia miocardial ^{123}I – MIBG en pacients caucàsics amb malaltia de Parkinson

Autor: Maria Vivó Pascual

Director: Josep Lluís Carrasco

Departament: Fonaments Clínics, Estadística Aplicada

Convocatòria: Juny 2020



Resum

El propòsit d'aquest treball és l'estudi de les mesures preses per un conjunt de metges per a avaluar la fiabilitat de la tècnica d'imatge $^{123}I - MIBG$ i concloure si és reproducible. Els índexs estimats i interpretats per a l'obtenció dels resultats són els índex de fiabilitat o Intraclass Correlation Coefficient (ICC) i l'índex de concordança o Concordance Correlation Coefficient (CCC), ambdues mesures basades en l'estudi de la variabilitat de les dades. La mostra d'estudi consta d'un total de 27 individus i està conformada per un conjunt de pacients patològics i un altre de pacients control. Sobre aquests i a partir de la tècnica d'imatges s'extreuen mesures sobre un índex en dos moments temporals diferents i seràn les variables d'estudi al llarg del treball.

Un cop finalitzat l'anàlisi de la variabilitat, es comparen els resultats obtinguts amb un anàlisi que es va realitzar sobre les mateixes dades i amb el mateix objectiu però fent ús d'una metodologia subòptima. Finalment, s'avaluen les conseqüències d'una metodologia inadequada.

Paraules clau

Error de mesura, Model Lineal Mixt, Fiabilitat, Concordança, $^{123}I - MIBG$

Classificació AMS

62P10 Applications to biology and medical sciences; **62J10** Analysis of variance and covariance; **62J05** Linear regression; **62-07** Data analysis.

Reliability study of $^{123}\text{I} - \text{MIBG}$ myocardial scintigraphy in caucasian patients with Parkinson disease

Abstract

The purpose of this paper is to study the measurements taken by a group of doctors in order to analyze the reliability of the image technique $^{123}\text{I} - \text{MIBG}$ and conclude whether if it is a reproducible methodology. The estimated indices to obtain results are the Intraclass Correlation Coefficient (ICC) and the Concordance Correlation Coefficient (CCC), both of them based on the study of the variability within and between clusters of the data. The sample of study is composed of 27 caucasian individuals and can be segregated in diagnosed group and control group. Measurements are taken from the images obtained using the technique of study in two different moments, 15 minutes and 4 hours post-injection. This information is taken by three doctors, one of them realizes the same measurement 3 times.

Once the variability study is finished, the results of the investigation are compared with those of an original study made years ago using a different methodology, therefore reporting different insights. Lastly, the negative effects of the application of a suboptimal methodology are exposed.

To carry out the analysis, the statistical software "R" and the packages related to modelling and visualization of data will be used.

Key words

Measurement error, Linear Mixed Models, Reliability, Concordance, $^{123}\text{I} - \text{MIBG}$

Índex

1	Introducció	7
2	Metodologia	9
2.1	Disseny de l'experiment	9
2.1.1	Selecció de la mostra	9
2.1.2	Observadors i mesures	10
2.2	Tècnica d'imatge	11
2.3	Modelització estadística	12
2.3.1	Comparació de models	13
2.4	Índex de fiabilitat	14
2.4.1	Definició	14
2.4.2	Procediment per l'estimació de ρ	15
2.5	Índex de concordança	18
2.5.1	Definició	18
2.5.2	Procediment per l'estimació de ρ_c	18
2.6	Càlcul dels intervals de confiança	19
2.7	Determinació de l'excel·lència d'un índex	19
3	Determinació del model general i estudi de la fiabilitat	21

3.1	Especificació del model	21
3.2	Estudi de la fiabilitat al mesurar <i>index4h</i>	22
3.3	Estudi de la fiabilitat al mesurar <i>index15mins</i>	24
4	Estudi de models més complexos i estudi de la fiabilitat	26
4.1	Especificació del model	26
4.2	Estudi de l'error aleatori heteroscedàstic segons observador i diagnòstic al mesurar <i>index4h</i>	28
4.3	Estudi de l'error aleatori heteroscedàstic segons observador i diagnòstic al mesurar <i>index15mins</i>	32
5	Estudi de la concordança	36
5.1	Estudi de les diferències sistemàtiques entre observadors	36
5.1.1	Diferències sistemàtiques per a la variable <i>index4h</i>	36
5.1.2	Diferències sistemàtiques per a la variable <i>index15mins</i>	38
5.2	Càlcul índex de concordança global	39
5.3	Càlcul índex de concordança per parelles	40
6	Conclusions de l'anàlisi	42
7	Limitacions en l'anàlisi de l'article clínic original	43
8	Comparació entre l'anàlisi original l'actual	44

A	Apèndix	46
A.1	Verificació i estudi descriptiu de les dades	46
A.1.1	Exploració multivariant de les dades	47
A.2	Implementació de les funcions en R	51
A.3	Validació dels resultats/funcions	53
A.4	Variàncies associades als càlculs dels índexs	54
A.5	Estudi per a la validació de funcions a partir de bootstrap	59
A.6	Codi R	61

Agraïments

Abans de procedir amb el treball m'agradaria agrair a totes aquelles persones que m'han ajudat tant en la realització del treball com en el curs de la carrera en general.

Pel que fa a aquest últim curs i en concret al treball en qüestió, l'ajuda i suport del meu tutor Josep Lluís Carrasco (JL) ha estat fonamental i molt apreciada. Al teu costat ha estat una experiència que podria repetir cada curs!

D'altra banda la carrera no hauria estat igual de meravellosa sense els companys amb els quals he tingut la sort de compartir aquests anys que tant ens han ensenyat. A ells també els hi dedico aquest treball i els hi desitjo el millor.

Finalment i igual d'important es troba la meva família, la que sempre m'ha animat a donar el millor de mi en la vida. Està dedicat a ells, per saber que podia quan ni jo mateixa ho sospitava.

1 Introducció

El Departament de Fonaments Clínics de la Facultat de Medicina de la Universitat de Barcelona està interessat en la optimització d'un estudi que es va realitzar fa 10 anys sobre la validesa sobre una tècnica d'imatge pel mesurament d'un índex en pacients diagnosticats de Parkinson. L'interès de re-analitzar aquestes dades es troba en que la metodologia utilitzada aleshores es considera subòptima [6]. De manera que un dels aspectes clau en la realització del treball és remarcar la importància de la metodologia en el camp de l'estadística per a poder assegurar la validesa i eficiència dels resultats. Una metodologia estadística inapropiada és un defecte que es pot presentar en estudis de recerca per desconeixement de les metodologies òptimes que potencialment es poden aplicar. Conseqüentment quan això passa, la validesa i per tant fiabilitat, d'aquests tipus d'estudis no és pot assegurar i les conclusions que se n'extreuen poden estar distorsionades.

Les dades que s'analitzen en aquest treball consisteixen en mesures repetides d'un conjunt de pacients, s'obtenen a partir de la realització d'una tècnica d'imatge anomenada $^{123}\text{I} - \text{MIBG} - \text{I}$ -metaiodobenzylguanidine – sobre la qual se'n extreu la informació referent a *heart-to-metastinum ratio* als 15 minuts i 4 hores posteriors a la injecció d'un medicament. Aquesta informació és extreta partir de la mateixa imatge per 3 observadors diferents, diversos cops en el cas d'un d'ells.

L'objectiu de l'estudi és l'anàlisi de la reproductibilitat de la tècnica d'imatge $^{123}\text{I} - \text{MIBG}$. Amb aquest propòsit és determinarà la fiabilitat dels observadors per separat, i si n'és possible s'estimarà un índex de fiabilitat global. Així mateix, s'avaluarà la concordança, és a dir, la intercanviabilitat de les mesures de diferents observadors per a finalment contrastar si aquesta tècnica és reproducible.

Tant l'índex de fiabilitat com l'índex de concordança són mesures agregades de concordança per a variables contínues, fent ús de la descomposició de la variància en entre i intra-individus, sent la font de discordança la variabilitat intra-individus. Aquests dos estadístics conformen les principals eines d'anàlisi utilitzades per a extreure resultats i conclusions. En aquest treball, per a validar les formules i algorismes relacionats amb l'error estàndard que s'han programat expressament per al treball, s'ha realitzat un estudi a partir de bootstrap, la informació referent a aquest procés es troba a l'apèndix del treball, concretament a l'apartat A.3.

Per a realitzar aquest estudi de fiabilitat i concordança del conjunt de mesures primer és porta a terme un estudi de validació i exploració de les dades. Aquest és troba adjunt a l'apartat A.1 de l'apèndix donat que no és d'interès directe. Seguidament, s'analitza quina és l'estructura de la variància de l'error aleatori més adequada per a cada model, tant el que treballa amb *index4h* com a variable resposta com el que ho fa amb *index15mins*. Un cop definit el model per les dues variables es procedirà a l'estimació dels diferents índexs. És en el desenvolupament d'aquest procés on s'hi troba la principal diferència entre l'anàlisi realitzat a aquest estudi respecte l'article esmentat. S'explicaran ambdós procediments i s'exposaran les raons per les que l'estudi original és subòptim.

Finalment i de gran importància, s'agraeix a les doctores que ens han permès fer les anàlisis amb les seves dades. Són la Dra. A. Muxí i la Dra. P. Paredes del departament de Medicina Nuclear de l'Hospital Clínic de Barcelona.

2 Metodologia

En aquest apartat es tracta quin ha estat el disseny de l'experiment, l'estructura de les dades i és profunditza més en la tècnica d'estudi que s'avalua en aquest treball, i de la que se'n extreuen les mesures que conformen les variables d'estudi. Posteriorment es procedeix a l'explicació del tipus de model amb els que es treballa de manera exclusiva i quina ha estat la metodologia a l'hora de contrastar diferents models. Un altre aspecte en el que es profunditza és en els conceptes de fiabilitat i concordança, la seva expressió i importància així com el mètode d'estimació. Finalment s'aclara com s'ha procedit en la construcció dels intervals de confiança a partir de la transformació Z de Fisher i el mètode per a la determinació de l'excel·lència d'un índex.

2.1 Disseny de l'experiment

2.1.1 Selecció de la mostra

La població de l'estudi és una mostra de 27 individus caucàsics, la meitat d'aquests són pacients patològics i és caracteritzen per ser diagnosticats de Parkinson, amb una durada mitja de la malaltia de 10 anys. L'altre meitat són pacients control, és a dir, individus no diagnosticats de Parkinson ni altres patologies segons el seu historial mèdic, així com examinacions físiques i neurològiques. De manera que al llarg de l'estudi es fa referència al grup control com a grup de pacients saludables, en contrast al grup de pacients malalts. A l'hora de seleccionar els individus de cada grup s'ha forçat un disseny balancejat pel que fa als trets sexe i edat.

A l'hora de seleccionar els pacients malalts s'ha mesurat la posició que ocupa cada un d'ells en l'escala de Hoehn and Yahr (*HY*), mesura que és troba dins el rang de l'1 al 5 –essent 1 malaltia lleu i 5 malaltia molt severa- aquesta informació s'ha enregistrat en la variable *NYON*. Dels 13 pacients malalts 7 van ser diagnosticats *HY* de primer nivell, 6 de segon nivell i 1 de tercer.

2.1.2 Observadors i mesures

Per a realitzar l'estudi es van seleccionar 3 metges amb diferents trets en relació a la seva formació com a radiòlegs en tècniques d'imatge per prendre les mesures. Les seves característiques i la codificació que van rebre són:

- Metge especialista que va rebre una formació específica sobre la tècnica. Realitza 1 mesura a cada pacient (*observador 1*)
- Metge especialista que no va rebre una formació específica sobre la tècnica. Realitza 3 mesures a cada pacient (*observador 2*)
- Metge resident a l'hospital on es realitza l'estudi, realitza 1 mesura a cada pacient (*observador 3*)

Els professionals encarregats de dirigir el projecte van decidir que l'observador 2 prendria 3 mesures de cada pacient, amb la finalitat de poder calcular posteriorment la variància intra-observador utilitzant únicament les mesures realitzades per el metge especialista no entrenat. Aquest és un dels aspectes que caracteritza l'anàlisi estadística del treball com a subòptima. L'esquema resultant de l'experiment s'ha reproduït en la Figura 2.1.

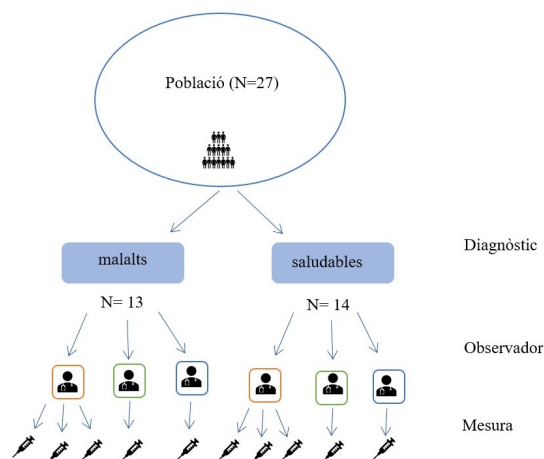


Figura 2.1: Esquema del disseny de l'experiment clínic

2.2 Tècnica d'imatge

La variable que és mesura és heart-to-mediastinum ratio (H/M) a partir de la tècnica d'imatge $^{123}\text{I} - \text{MIBG}$, procediment de creació d'imatges mèdiques nuclear no invasiu que fa ús d'un yode radioactiu anomenat ^{123}I per investigar el subministrament de sang al cor. Les imatges que s'obtenen són com la que està representada a la Figura 2.2. La valoració d'aquestes imatges és d'importància decisiva en l'avaluació i en la presa de decisions sobre el diagnòstic d'un pacient.

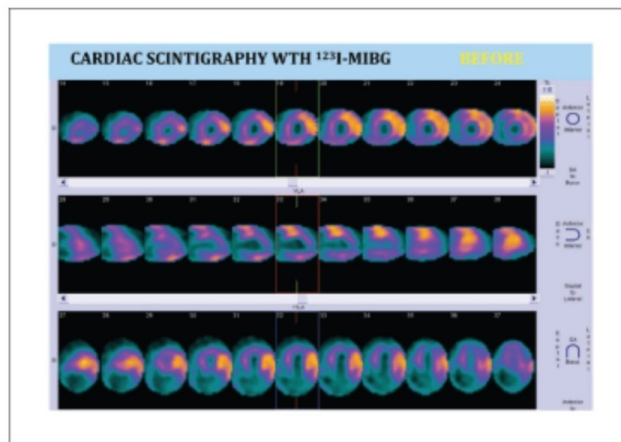


Figura 2.2: Tècnica d'imatge $^{123}\text{I} - \text{MIBG}$

S'obtenen aquestes imatges en dos moments temporals diferents: als 15 minuts i 4 hores posteriors a la injecció de ^{123}I . Seguidament és quantifica la captació miocardiàca de $^{123}\text{I} - \text{MIBG}$ a partir de l'índex H/M, aleshores s'utilitza aquest indicador pel càlcul de la taxa de dissolució, anomenada *washout rate* (*WR*) com la diferència percentual entre la diferència early H/M i delayed H/M. El rati entre la mesura H/M i el WR associat s'enregistra a la variable *aclaram*. De manera que la informació que contenen les variables *WR* i *aclaram* són operacions derivades de les variables *index15mins* i *index4h*.

Aquests atributs junt als que s'han referenciat a apartats previs, *HY* i *NYON*, conformen la base de dades amb la que és treballa en l'estudi. Així i tot, les úniques variables d'interès donat l'objectiu establert són *index15mins* i *index4h*.

2.3 Modelització estadística

En primer lloc cal definir quina és la notació que s'utilitzarà en aquest estudi a l'hora de especificar un model. Es procedeix de la següent manera: per a indicar que un efecte del model és fix, s'utilitza una lletra grega (α, β, \dots), altrament, si l'efecte és aleatori, és refereix a aquest amb una lletra llatina (a, b, \dots).

A l'hora de plantejar quina és la modelització més adequada, cal tenir en compte que es tracta d'un estudi de mesures repetides, a cada pacient se li realitzen 5 mesures. S'han de modelar explícitament dos fonts de variabilitat, aquestes són la variància entre mesures del mateix individu i la variància entre diferents individus. D'aquesta manera, els pacients, variable *ID*, han de ser considerats en el model com un efecte aleatori donat que l'objecte d'estudi es la població, no els 27 pacients concretament. Com és vol tenir en compte aquesta variabilitat entre individus i a l'hora el propòsit de l'estudi és l'anàlisi de la fiabilitat i concordança dels observadors, s'utilitza un *Model Lineal Mixte* - conegut com *LMM* de l'anglès - en que l'observador és un efecte fix i l'individu és un efecte aleatori. L'expressió matricial resultat és la següent:

$$Y = X\beta + Zb + \varepsilon$$

La variable *Y* representa el valor de la variable dependent/resposta. Les matrius *X* i *Z* són les matrius de disseny que corresponen als efectes fixes i aleatoris respectivament. El vector β descriu l'efecte del factor observador en l'expectació de la variable resposta. Els efectes aleatoris *b* s'assumeix que segueixen una distribució normal amb mitjana zero. Finalment ε és el vector d'errors aleatoris, representa la variabilitat intra-individu, també s'assumeix que segueix una distribució normal de mitjana zero i variància σ^2 .

Els estimadors REML (Restricted/Residual Maximum Likelihood), procediment proposat per *Patterson y Thompson* (1971), consisteixen en la factorització de la versemblança completa en dos parts independents, una d'elles no conté la mitjana, assumint que al utilitzar aquesta part de la informació no és perd informació respecte la versemblança completa. Així doncs, el procés d'estimació REML consisteix amb l'aplicació del procés de màxima versemblança a una combinació lineal de les observacions:

$$CY = CX\beta + CZb + C\varepsilon \Rightarrow E(CY) = 0$$

Cal trobar C a partir de la formulació de $n - p$ contrastes que *eliminen* els efectes fixes, $CX = 0$. Aleshores la versemblança de σ_b^2 i σ^2 donat CY és la versemblança restringida i proporciona estimacions no esbiaixades pels components de la variància.

El procediment REML és el que el paquet utilitzat per a modelitzar, anomenat nlme, aplica per defecte a l'ajustar els LMM.

2.3.1 Comparació de models

Una de les conseqüències de l'aplicació del procediment REML és que a l'hora de comparar models aquests han de estar conformats per els mateixos efectes fixes. El problema es presenta a l'hora de comparar la versemblança de models jeràrquics. Si els models es comparessin utilitzant una mesura d'entropia, com l'AIC per exemple, no hi hauria problema. En els casos en que aquesta condició no es satisfà els models són especificats sota una metodologia d'estimació *Maximum Likelihood (ML)*. Així i tot, al escollir el model final per a l'estudi per cada variable és procedirà a la seva estimació per REML [5].

Aleshores, el mètode utilitzat per a comparar models és la comparació de la versemblança dels models a partir del **test de raó de versemblança**. Aquest procediment consisteix en l'avaluació de l'equivalència de dos models, un model anomenat model complert o sencer i un altre amb alguna restricció en algun dels paràmetres del model sencer. Si la restricció, i per tant la hipòtesis nul·la, no presenta evidència en les dades, les versemblances d'ambdós models no haurien de diferir més que en l'error aleatori. L'estadístic de contrast és el rati de les versemblances i avalua si aquest valor és significativament diferent a 1, sota la hipòtesis nul·la aquest estadístic segueix una distribució χ^2 amb graus de llibertat la diferència de paràmetres dels models.

2.4 Índex de fiabilitat

2.4.1 Definició

El mesurament és un procés essencial en el camp de l'estadística, en aquest cas aplicat a la investigació mèdica. Especialment en aquest tipus d'estudis les variables a mesurar poden resultar complicades de quantificar i conseqüentment suposa la presència d'un cert grau d'error de mesura. La qualitat de la mesura depèn tant de la seva validesa com de la seva fiabilitat i no menys important, de la inexistència d'errors sistemàtics.

El concepte de fiabilitat pot ser definit com el nivell de confiança amb el que pot afirmar-se que s'obtenen els mateixos valors al efectuar diversos mesuraments d'una mateixa variable baix condicions similars. L'estudi de la fiabilitat és basa en la mesura de la concordança en termes de variabilitat en les mesures preses per un o més observadors o mètodes. En aquest cas s'avalua la variància d'un mateix observador al realitzar les mesures de les variables referents a l'índex H/W i entre ells. Es realitzarà un conjunt de contrastos per determinar-ne l'equivalència dels índexs i en cas afirmatiu se'n calcularà l'índex de fiabilitat global pel conjunt d'observadors de la variable d'estudi. Finalment s'estudiarà com afecta el diagnòstic del pacient en la capacitat de mesurament de l'observador en la variable d'estudi.

Amb aquesta finalitat i donat que les variables d'estudi són quantitatives, s'estimarà el valor del Coeficient de Correlació Intraclasse (ICC) per a dur a terme l'estudi de la fiabilitat. Aquest coeficient realitza el càlcul de la mitjana de les correlacions entre les dades d'un mateix clúster. Es basa en la idea que la variabilitat pot ser descomposta en la variabilitat deguda a les diferències entre els individus (σ_{α}^2) i a la variabilitat entre les mesures de cada individu (σ_{ϵ}^2), també anomenada variabilitat residual. L'ICC pot definir-se com la proporció de variabilitat total que és deu a la variabilitat entre subjectes.

La fórmula matemàtica de l'índex de fiabilitat s'expressa en l'equació (1) i pren domini de 0 - situació que és dona si no hi ha clúster i per tant les dades d'un mateix individu són independents o bé si la variància corresponent a l'error residual tendeix a infinit- a 1, que és la situació òptima, en la que la variància total és variabilitat entre individus i per tant la variància de l'error aleatori és nul·la.

$$\rho = \frac{\sigma_{\alpha}^2}{\sigma_{\alpha}^2 + \sigma_e^2} \in [0, 1] \quad (1)$$

Amb relació a la interpretació precisa del valor de l'ICC en l'àmbit clínic, no és gaire clara, donat que no hi ha cap referent científic que indiqui a partir de quin valor pot considerar-se una bona fiabilitat. Aquest estudi s'ha basat en articles científics que tracten de manera exhaustiva aquest aspecte [7] i en el criteri i experiència dels professionals en el camp d'estadística clínica del Departament de Fonaments Clínics de la Universitat de Barcelona. Aleshores, s'ha considerat a partir d'un coeficient de 0.7 una bona fiabilitat, mentre que a partir d'un valor de $\rho \in [0.9, 1]$ s'interpretarà com una fiabilitat excel·lent.

2.4.2 Procediment per l'estimació de ρ

Ens proposem estimar el valor de l'Intraclass Correlation Coeficient i l'interval de confiança del 95%, $IC(\rho)_{95\%}$, associat per a cada observador i variable.

El primer aspecte que cal subratllar és el funcionament del paquet que s'utilitzarà per a modelitzar, `nlme`. En la situació d'un model amb error aleatori heterocedàstic, cas que satisfan tots els models amb els que es treballarà en aquest estudi. El procediment que es segueix és el càlcul de la desviació estàndard residual basal - en aquest cas per a l'observador 1 - i uns pesos associats, θ_j . El producte dels pesos al quadrat i la variància residual basal conforma la variància residual dels altres observadors, essent el valor del pes per el primer observador $\theta_1 = 1$. Conseqüentment, el càlcul de ρ consisteix en aquest cas en l'operació de l'equació (2).

$$\rho_{mj} = \frac{\sigma_{\alpha}^2}{\sigma_{\alpha}^2 + \theta_j^2 \sigma_e^2} \Rightarrow \rho_{mj} \begin{cases} j = 1, 2, 3 \text{ observador} \\ m = 1, 2 \text{ (index4h, index15mins)} \end{cases} \quad (2)$$

Un altre aspecte rellevant al treballar amb el paquet `nlme` és la matriu de les estimacions dels components de la variància, Σ_M , la que és calcula en escala logarítmica, de manera que Σ s'expressa com s'observa en l'equació (3).

$$\Sigma_M = \begin{pmatrix} Var(\tau) & cov(\tau, \eta_1) & cov(\tau, \eta_2) & cov(\tau, \tau_e) \\ cov(\tau, \eta_1) & Var(\eta_1) & cov(\eta_2, \eta_1) & cov(\tau_e, \eta_1) \\ cov(\tau, \eta_1) & cov(\eta_2, \eta_1) & Var(\eta_2) & cov(\tau_e, \eta_2) \\ cov(\tau, \tau_e) & cov(\tau_e, \eta_1) & cov(\tau_e, \eta_2) & Var(\tau_e) \end{pmatrix} \quad (3)$$

$$\tau = \log(\sigma_\alpha^2) \quad \eta_1 = \log(\theta_1) \quad \eta_2 = \log(\theta_2) \quad \tau_e = \log(\sigma_e^2)$$

Un cop s'obté el valor de $\hat{\rho}$ caldrà procedir al càlcul de l'interval de confiança, el que requereix de $\hat{\rho}$ i del valor de la variància de $\hat{\rho}$, aquesta pot aproximar-se utilitzant el **mètode delta**, procediment expressat en l'equació (4).

$$\hat{\sigma}(\rho_{mj}) = \Delta' \Sigma_\rho \Delta \quad (4)$$

On Σ_ρ representa la matriu de variàncies i covariàncies associada a les components de la variància i es defineix a l'equació (5).

$$\Sigma_{\rho_{obs_i}} = \begin{pmatrix} Var(\sigma_\alpha^2) & Cov(\sigma_\alpha^2, \sigma_{e_i}^2) \\ Cov(\sigma_\alpha^2, \sigma_{e_i}^2) & Var(\sigma_{e_i}^2) \end{pmatrix} \quad (5)$$

Com que l'output de la funció `lme` proporciona la matriu d'aquestes components en escala logarítmica és procedeix a la utilització del mètode delta [2]. Aquest procediment consisteix en l'aproximació de l'error estàndard de transformacions de variables aleatòries a partir de l'aproximació de la sèrie de Taylor de primer ordre. L'expressió general resultant és troba expressada en l'equació (6).

$$Var(G(X)) \approx \nabla G(X)' Cov(X) \nabla G(X) \quad (6)$$

Posteriorment és realitzen les operacions del conjunt d'equacions (7),(8) i (9) per a l'obtenció dels elements de la matriu en el cas del primer observador.

$$\hat{\sigma}_\alpha^2 = e^{2\tau} \quad \hat{\sigma}_e^2 = e^{2\tau_e}$$

$$\text{Var}(\hat{\sigma}_\alpha^2) \approx \left(\frac{\delta e^{2\tau}}{\delta \tau}\right)^2 \text{Var}(\tau) = 4(\hat{\sigma}_\alpha^2)^2 \text{Var}(\tau) \quad (7)$$

$$\text{Var}(\hat{\sigma}_e^2) \approx \left(\frac{\delta e^{2\tau_e}}{\delta \tau_e}\right)^2 \text{Var}(\tau_e) = 4(\hat{\sigma}_e^2)^2 \text{Var}(\tau_e) \quad (8)$$

$$\text{cov}(\hat{\sigma}_\alpha^2, \hat{\sigma}_e^2) \approx \left(\frac{\delta e^{2\tau}}{\delta \tau}\right) \left(\frac{\delta e^{2\tau_e}}{\delta \tau_e}\right) \text{cov}(\tau, \tau_e) = 4\hat{\sigma}_\alpha^2 \hat{\sigma}_e^2 \text{cov}(\tau, \tau_e) \quad (9)$$

Pel que fa als altres observadors, cal tenir en compte que la variància residual és el producte entre la variància residual del primer observador i un pes al quadrat, θ_j^2 , que el en cas de l'observador 1 pren valor 1. Per a l'aproximació dels components de la matriu de covariàncies dels altres dos observadors cal tenir en compte el quadrat dels pesos θ_2, θ_3 . Les expressions resultants per al càlcul de les estimacions són:

$$\hat{\sigma}_{e_j}^2 = \theta_j^2 \cdot \sigma_e^2 \Rightarrow \log(\sigma_{e_j}^2) = 2\eta_1 + 2\tau_e = \gamma_k \Rightarrow$$

$$\boxed{\hat{\sigma}_{e_j}^2 = e^{2\gamma_k}} \quad (10)$$

$$\text{Var}(\hat{\sigma}_{e_j}^2) = \left(\frac{\delta e^{2\tau_k}}{\delta \tau_k}\right)^2 \text{Var}(\gamma_k) = (\hat{\sigma}_j^2)^2 \text{Var}(\gamma_k) \quad (11)$$

$$\text{Var}(\gamma_k) = \text{Var}(\eta_j + 2\tau_e) = \text{Var}(\eta_j) + 4\text{Var}(\tau_e) + 4\text{cov}(\eta_j, \tau_e) \quad (12)$$

$$j = 2, 3 \quad k = 1, 2$$

$$\begin{aligned} \text{cov}(\hat{\sigma}_\alpha^2, \hat{\sigma}_{e_j}^2) &= \text{cov}(e^{\log(\sigma_\alpha^2)}, e^{\log(\sigma_{e_j}^2)}) = \left(\frac{\delta e^{\log(\sigma_\alpha^2)}}{\delta \log(\sigma_\alpha^2)}\right) \cdot \left(\frac{\delta e^{\log(\sigma_{e_j}^2)}}{\delta \log(\sigma_{e_j}^2)}\right) \cdot \text{cov}(\log(\sigma_\alpha^2), \log(\sigma_{e_j}^2)) = \\ &= \hat{\sigma}_\alpha^2 \hat{\sigma}_{e_j}^2 [2\text{cov}(\tau, \eta_j) + 4\text{cov}(\tau, \tau_e)] \quad (13) \end{aligned}$$

Per altra banda, per obtenir Δ cal realitzar les derivades parcials de $\hat{\rho}$, aquestes és troben en les equacions (14) i (15).

$$\frac{\delta \rho_i}{\delta \sigma_\alpha^2} = \frac{1 - \rho_i}{\sigma_\alpha^2 + \sigma_e^2} \quad (14)$$

$$\frac{\delta \rho_i}{\delta \sigma_e^2} = -\frac{\rho_i}{\sigma_\alpha^2 + \sigma_e^2} \quad (15)$$

2.5 Índex de concordança

2.5.1 Definició

L'índex de concordança (ρ_c) es diferencia de l'índex de fiabilitat (ρ) al tenir en compte l'efecte observador, mesurant la falta d'exactitud o en altres paraules, la diferència entre les mitjanes dels observadors. Valors alts d'aquest índex indiquen intercanviabilitat entre les mesures dels diferents observadors. Se'n calcula el seu valor a partir de la fórmula en l'equació (16).

$$\rho_c = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\beta^2 + \sigma_e^2} \begin{cases} \sigma_\alpha^2 \text{ variància entre individus} \\ \sigma_\beta^2 \text{ variabilitat entre mitjanes dels observadors} \\ \sigma_e^2 \text{ variància residual o intra-individus} \end{cases} \quad (16)$$

2.5.2 Procediment per l'estimació de ρ_c

L'índex de concordança o Concordance Correlation Coefficient (CCC), pot ser estimat a partir dels components de la variància del model d'efectes mixtes amb els individus com efecte aleatori i els observadors com efecte fix. Aquest procediment permet ser generalitzat a la situació de més de dos observadors.

A l'hora de calcular els paràmetres de la variància, concretament σ_β^2 , cal recordar que l'efecte observador s'inclou en el model com un efecte fix, de manera que el component σ_β^2 indica la variabilitat entre les mitjanes dels observadors. En aquest cas l'estimador apropiat s'expressa en l'equació (17), on k fa referència al nombre d'observadors, n als nombre d'individus i m el nombre de rèpliques d'una mesura d'un observador en un mateix pacient [2].

$$\hat{\sigma}_\beta^2 = \frac{1}{k(k-1)} \sum_{i=1}^{k-1} \sum_{j=i+1}^k (\bar{Y}_i - \bar{Y}_j)^2 - \frac{\hat{\sigma}_e^2}{n \cdot m} \quad (17)$$

Els components de variància i covariància són estimats a partir de la inversa de la matriu d'informació de Fisher, excepte en el cas del càlcul de la variància del component σ_β^2 que serà

aproximada a partir de l'equació (18). [2]

$$Var(\hat{\sigma}_\beta^2) = \frac{4}{k^2(k-1)^2} \sum_{i=1}^{k-1} \sum_{j=i+1}^k \{(\bar{Y}_i - \bar{Y}_j)^2 \cdot Var(\bar{Y}_i - \bar{Y}_j)\} + \frac{Var(\hat{\sigma}_e^2)}{(n \cdot m)^2} \quad (18)$$

Si l'error aleatori és heteroscedàstic és procedirà considerant $\sigma_e^2 = \frac{1}{k} \sum_{i=1}^k \sigma_{e_i}^2$.

2.6 Càlcul dels intervals de confiança

Pel càlcul dels intervals de confiança de les estimacions de ρ i ρ_c , cal remarcar dos aspectes claus, en primer lloc que el nivell de significació s'ha fixat per defecte al 5%, de manera que els intervals són del 95% de confiança. Per altra banda en lloc de utilitzar el valor de l'estimació de ρ pel càlcul directe dels extrems de l'interval, s'ha fet ús d'una transformació que accelera la convergència a la distribució normal anomenada transformació Z de Fisher [3], consistent en la següent alteració:

$$\hat{Z} = \frac{1}{2} \log\left(\frac{1 + \hat{\rho}_i}{1 - \hat{\rho}_i}\right) \sim N(0, V(\hat{Z}))$$

$$\hat{V}(\hat{Z}) = \frac{\hat{V}(\hat{\rho}_i)}{(1 - \hat{\rho}_i^2)^2}$$

2.7 Determinació de l'excel·lència d'un índex

L'objectiu d'aquest treball és l'optimització de l'article clínic esmentat anteriorment a partir de l'avaluació de la fiabilitat i concordança de 3 observadors. No obstant, no hi ha cap mena d'estàndard científic que defineixi a partir de quin punt el valor de l'índex de fiabilitat/concordança és excel·lent. És per això que s'ha definit un punt de tall, a partir d'estudis previs, és considerarà que valors d'aquests índexs per sobre de 0.9 són excel·lents [7].

Com s'ha hagut de fer ús d'aquest test recurrentment al llarg de l'estudi, s'ha decidit crear una funció que avalua el contrast desenvolupat a continuació.

$$H_0 : \rho \leq 0.9$$

$$H_1 : \rho > 0.9$$

Per avaluar-lo s'ha fet ús de la transformació Z de Fisher tant per el valor de l'estimació de ρ com per el valor del valor contrastat, 0.9, que un cop transformat pren valor 1.472. La raó de la transformació és una ràpida convergència a una distribució $N(0, V(\hat{Z}))$ de manera que el test resultant és:

$$T = \frac{\hat{Z} - 1.472}{\sqrt{\text{Var}(\hat{Z})}} \sim N(0, 1)$$

Sota el supòsit que estableix H_0 l'estadístic T segueix una distribució Normal d'esperança 0 i variància 1. Si el valor de l'estadístic de contrast és major a $z_{0.95}$, és a dir, 1.96, és rebutjarà H_0 i s'assumirà que l'índex fiabilitat/concordança és excel·lent.

3 Determinació del model general i estudi de la fiabilitat

En una primera aproximació de l'estudi de la fiabilitat s'analitza com és la fiabilitat de manera global segons els diferents observadors per als dos índexs d'estudi *index4h* i *index15mins*. Si la fiabilitat és equivalent per els diferents observadors es procedirà al càlcul de l'índex de fiabilitat global.

3.1 Especificació del model

A l'hora d'ajustar el model més adequat per a la situació d'estudi cal tenir en compte que aquest experiment és un estudi clínic amb mesures repetides sobre un determinat nombre de pacients. A la Figura 3.1 s'hi troben representats gràficament els valors de la variable *index15mins* per cada pacient, pot observar-se una certa variabilitat entre ells. Per tal de tenir en compte aquesta font de variabilitat, la variable referent al pacient, *ID*, ha de ser modelitzada com un efecte aleatori.

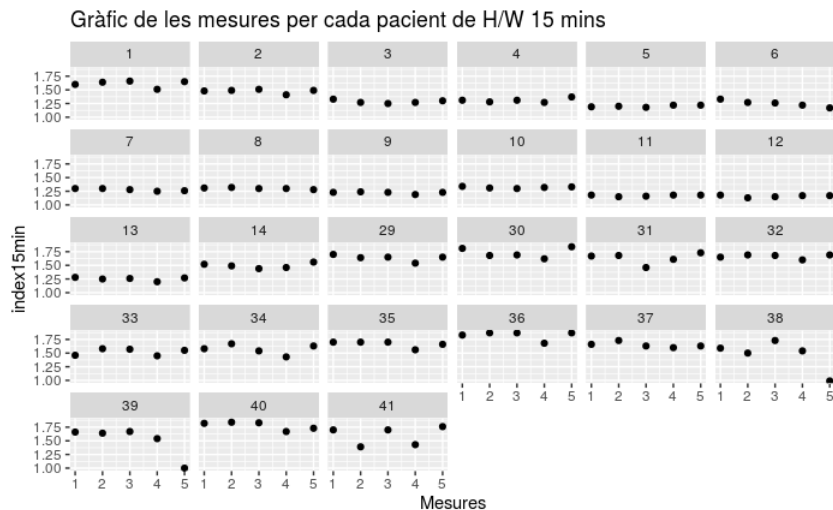


Figura 3.1: Facet plot de la variable *index15mins* per a les diferents mesures de cada individu

La variable observador és considera com un efecte fix, de manera que els resultats només faran referència a aquests tres observadors. El model resultant es troba expressat en l'equació (19).

$$Y_{ijk} = \mu + \alpha_i + B_j + e_{ijk} \begin{cases} i = 1, 2, 3 & \text{Observador} \\ j = 1, 2, \dots, 27 & \text{Pacient} \\ k = 1, 2, 3 & \text{Rèplica} \end{cases} \quad (19)$$

Restriccions : $\sum_{j=1}^3 \alpha_i = 0$

Suposicions : $B_j \sim N(0, \sigma_B^2)$ $e_{ijk} \sim N(0, \sigma_{e_j}^2)$

En el model especificat prèviament, Y_{ijk} és el valor de la variable resposta, *index4h* o *index15mins*, en la rèplica k de l'individu i realitzada per a l'observador j . α_i representa l'efecte de l'observador, B_j representa l'efecte individu i e_{ijk} representa l'error aleatori. En un principi s'assumeix una estructura per la variància de l'error aleatori homoscedàstica, $e_{ijk} \sim N(0, \sigma_e^2)$, al llarg de l'estudi es contrastarà aquesta amb altres d'heteroscedàstiques: $\sigma_e^2 = \sigma_{e_j}^2 I_N$.

3.2 Estudi de la fiabilitat al mesurar *index4h*

En primer lloc s'estudien els índex de fiabilitat per observador de la variable *index4h*.

	$\hat{\rho}$	$\hat{\sigma}_{\hat{\rho}}$	$IC_{inf}(\rho)$	$IC_{sup}(\rho)$
Observador 1	0.946	0.0175	0.898	0.971
Observador 2	0.911	0.0336	0.817	0.958
Observador 3	0.945	0.0239	0.873	0.976

Taula 3.1: Estimació de l'índex de fiabilitat ρ per cada observador i l'error estàndard i interval de confiança del 95% associats

De manera general sembla que els índexs de fiabilitat per els diferents observadors són molt elevats i prenen valors relativament similars, així i tot l'error estàndard associat a l'índex de fiabilitat del segon observador pren un valor bastant més elevat als altres. Cal aleshores contrastar de manera formal si els ρ_i poden considerar-se equivalents, per tant si les diferents mesures preses per els observadors presenten la mateixa variabilitat. Per a contrastar-ho s'ha plantejat el següent contrast d'hipòtesis:

$$\begin{aligned}
H_0 : \rho_1 = \rho_2 = \rho_3 = \rho \\
H_1 : \rho_i \neq \rho_j \\
i \neq j \\
i, j = 1, 2, 3 \text{ (observador)}
\end{aligned}$$

Per a resoldre aquest tipus de contrastos, és realitza la comparació de models – Contrastant el model amb una estructura de variàncies homoscedàstica respecte una estructura heteroscedàstica determinada pels observadors - a partir del test de raó de versemblança. El resultat del contrast del model nul - Model amb estructura de variàncies homogènia- i el model amb una estructura de variàncies heterogènia, presenta un p valor associat de 0.382, valor superior al nivell de significació fixat, $\alpha = 0.05$. Conseqüentment, no podem afirmar que els models siguin significativament diferents i ens quedem amb aquell que suposa una única variància. Aquests resultats ens permeten realitzar aleshores el càlcul d'un índex de fiabilitat global (ρ_{global}) i la posterior avaluació de l'excel·lència d'aquest.

$\hat{\rho}$	$\hat{\sigma}_{\hat{\rho}}$	$IC_{inf}(\rho)$	$IC_{sup}(\rho)$
0.937	0.0187	0.888	0.965

Taula 3.2: Estimació ρ_{global} per la variable *index4h*, l'error estàndard i interval de confiança associats

$$\begin{aligned}
H_0 : \rho_{global} \leq 0.9 \\
H_1 : \rho_{global} > 0.9
\end{aligned}$$

El contrast d'excel·lència del 95% de confiança, que considera un índex excel·lent si el valor d'aquest es estrictament superior a 0.9, presenta un p valor de $0.0577 > 0.05$, H_0 no pot rebutjar-se H_0 a favor de H_1 . Conseqüentment no pot afirmar-se que l'índex de fiabilitat global és excel·lent, tot i que convé tenir en compte que el p valor és troba al límit del contrast ja que el valor de l'índex de fiabilitat és molt elevat. A mode de conclusió les mesures preses pels observadors són molt precises.

3.3 Estudi de la fiabilitat al mesurar *index15mins*

Realitzem el mateix procediment per la variable *index15mins* i obtenim els resultats representats a la Taula 3.3.

	$\hat{\rho}$	$\hat{\sigma}_{\hat{\rho}}$	$IC_{inf}(\hat{\rho})$	$IC_{sup}(\hat{\rho})$
Observador 1	0.915	0.0297	0.834	0.957
Observador 2	0.946	0.0248	0.876	0.964
Observador 3	0.540	0.1006	0.314	0.7072

Taula 3.3: Estimació ρ per cada observador, l'error estàndard i interval de confiança associats

Pel que fa a aquesta variable i índex, els resultats no són tant bons com en el cas estudiat prèviament. El metge codificat com a observador 3 al realitzar les mesures de l'índex H/W als 15 minuts pren les mesures amb molta variabilitat, $\hat{\rho}_3 = 0.54$, comparat amb el valor que pren l'índex de fiabilitat per la variable *index4h* i amb els altres dos metges per aquesta variable.

Es contrasta l'equivalència dels valors d'aquests índexs i s'obté un valor del *p valor* associat al contrast menor a 0.01, per consegüent els models comparats no són equivalents i s'han de considerar variàncies diferents. Per aquest fet no és coherent realitzar el càlcul de l'índex de fiabilitat global per aquesta variable, així i tot, es contrasta si la fiabilitat dels observadors 1 i 2 és equivalent.

$$H_0 : \rho_1 = \rho_2 = \rho$$

$$H_1 : \rho_i \neq \rho_j$$

$$i \neq j$$

$$i = j = 1, 2 \text{ (observadors)}$$

Efectivament el *p valor* resultant al comparar el model de variàncies heterogènies per als 3 observadors respecte el que considera una única variància pels observadors 1 i 2 és major al nivell de significació 0.05. Aleshores poden considerar-se dues variàncies diferents, una pels observadors 1 i 2 i una altre diferent per l'observador 3.

És realitza el càlcul de $\hat{\rho}_{1U2}$ i $\hat{\rho}_3$ i és verifica l'equivalència entre aquest model i el model amb una estructura de variàncies totalment heteroscedàstica pels observadors. Amb un p valor de $0.3426 \gg 0.05$ no és pot rebutjar H_0 .

	$\hat{\rho}$	$\hat{\sigma}_{\hat{\rho}}$	$IC_{inf}(\hat{\rho})$	$IC_{sup}(\hat{\rho})$
Observadors 1 U 2	0.927	0.0221	0.869	0.960
Observador 3	0.552	0.0990	0.329	0.716

Taula 3.4: Estimació de ρ per els observadors 1 i 2 conjuntament i 3 de la variable *index15mins*

Aleshores s'avalua l'excel·lència dels dos índexs de fiabilitat i s'obtenen p valors associats al contrast de 0.148 i 0.999, respectivament per $\hat{\rho}_{1U2}$ i $\hat{\rho}_3$, ambdós valors molt superiors a α . Conseqüentment no pot assumir-se que la fiabilitat dels observadors és excel·lent tot i ser molt elevada per al conjunt d'observadors 1 i 2.

Un últim comentari que val la pena dir és que si s'exclouen del model les dues mesures que pren l'observador 3 que s'allunyen tant del conjunt de mesures i s'estima el valor de l'índex de fiabilitat s'obté un valor de 0.922 associat a aquest observador, considerablement elevat, al nivell dels altres dos observadors.

4 Estudi de models més complexos i estudi de la fiabilitat

4.1 Especificació del model

En l'anàlisi descriptiu de les dades, referent a l'apartat A.1 de l'apèndix, s'ha observat com en funció del diagnòstic del pacient el rang de valors que prenen les variables d'estudi és considerablement diferent, de manera que la variable referent al diagnòstic del pacient, *estatus*, sembla tenir un efecte significatiu en el resultat de les mesures. Tanmateix, a l'hora d'estudiar com afecta al model el diagnòstic del pacient és molt important tenir en compte el context de l'estudi, en concret l'objectiu d'aquest que es la avaluació de la fiabilitat dels observadors en el moment de mesurar un determinat índex d'un pacient. No és convenient introduir la variable referent al diagnòstic en el model com un efecte fix ja que aquesta variable explica variabilitat entre individus i aquesta es veuria reduïda. Aleshores el valor del ICC/CCC no seria comparable a aquells estimats globalment sense l'efecte diagnòstic en el model per ser estimats amb una variància entre individus diferents [1, 4]. Així i tot, pot contrastar-se si la variància de l'error aleatori difereix significativament en funció del diagnòstic o bé segons una combinació del diagnòstic i l'observador que realitza la mesura. De manera que s'estudia si una estructura de la variància de l'error aleatori heteroscedàstica suposa una millora significativa en el model.

El model que es procedirà a estudiar seguidament és el corresponent a l'expressió de l'equació (20), el canvi que és realitza respecte el model estudiat prèviament (equació (19)) es troba en l'estructura de la variància de l'error aleatori del model. En aquest model pot observar-se com l'error segueix una distribució normal en funció de l'observador i , on D és una variable binària que pren valor 1 quan el pacient forma part de la mostra de malalts i 0 altrament.

$$Y_{ijk} = \mu + \alpha_i + B_j + e_{ijk} \begin{cases} i = 1, 2, 3 & \text{Observador} \\ j = 1, 2, \dots, 27 & \text{Pacient} \\ k = 1, 2, 3 & \text{Rèplica} \end{cases} \quad (20)$$

Restriccions : $\sum_{j=1}^3 \alpha_i = 0$

Suposicions : $B_j \sim N(0, \sigma_B^2)$ $e_{ijk} \sim N(0, \sigma_{e_{i,D}})$

Així que es contrastarà l'estructura de la variància de l'error aleatori homoscedàstica respecte una estructura heteroscedàstica segons l'observador i d'altra banda respecte una estructura heteroscedàstica segons el diagnòstic. Finalment és contrastarà l'estructura homoscedàstica respecte una estructura de la variància de l'error aleatori heteroscedàstica segons la combinació de observador i diagnòstic.

Seguidament s'enumeren i descriuen les quatre estructures que seran contrastades.

1. Homoscedàstica: $\sigma_e^2 I_n$
2. Heteroscedàstica segons observador: dependent de l'observador que hagi pres la mesura aquesta presentarà una variabilitat associada diferent.

$$\sigma_{e_i}^2 I_n \Rightarrow \sigma_{e_i}^2 = \{ \sigma_{e_1}^2, \sigma_{e_2}^2, \sigma_{e_3}^2 \}$$

3. Heteroscedàstica segons observador i diagnòstic: dependent de l'observador i el diagnòstic del pacient en la mesura, aquesta presentarà una variabilitat associada diferent. Hi ha 3 metges, representats per el subíndex i , i 2 possibles diagnòstics, representats per el subíndex j , de manera que en aquest cas la variància de l'error aleatori presentarà 6 nivells.

$$\sigma_{e_{ij}}^2 I_n \Rightarrow \sigma_{e_{ij}}^2 = \{ \sigma_{e_{11}}^2, \sigma_{e_{31}}^2, \sigma_{e_{12}}^2, \sigma_{e_{32}}^2, \sigma_{e_{13}}^2, \sigma_{e_{33}}^2 \}$$

4. Heteroscedàstica segons observador i diagnòstic determinats: a partir de l'estudi de l'estructura de la variància de l'error aleatori és pot determinar si és completament heteroscedàstic segons observador i diagnòstic o si ho és parcialment, de manera que algun/s nivells presenten la mateixa variabilitat associada. En aquest cas l'estructura seguirà essent la de una matriu diagonal d'estructura $\sigma_{e_{ij}}^2 I_n$, els nivells seran determinats a partir dels resultats extrets al comparar models.

És fa ús del test de raó de versemblança per a realitzar els contrastos i per als dos primers, referents al contrast d'una estructura homoscedàstica respecte una d'heteroscedàstica segons observador o diagnòstic, no s'obté suficient evidència per a rebutjar la hipòtesis nul·la d'homogeneïtat en la variància, amb p valors respectivament de 0.385 i 0.0697. Ara bé, al realitzar la comparació

amb una estructura heterogènia segons la combinació d'observador i diagnòstic s'obté un *p* valor associat de 0.0015, conseqüentment és pot afirmar que hi ha suficient evidència per a rebutjar H_0 a favor del model que presenta una estructura de la variància de l'error aleatori heteroscedàstic segons la observador i diagnòstic. Tot seguit s'estudia la fiabilitat per a les diferents combinacions d'observador i diagnòstic i s'analitza quines d'aquestes són significativament diferents i de quina manera.

4.2 Estudi de l'error aleatori heteroscedàstic segons observador i diagnòstic al mesurar *index4h*

Aquesta subsecció consisteix en l'estudi de la fiabilitat dels observadors però aquest cop tenint en compte una estructura heteroscedàstica de la variància de l'error aleatori. L'objectiu és esbrinar si hi ha una combinació dels factors observador i diagnòstic que suposin un model més simple però equivalent al que treballa amb totes les combinacions en l'estructura de l'error aleatori a partir de l'estudi de la fiabilitat de les diferents combinacions.

En aquesta primera part s'analitzen les mesures de la variable *index4h*, més endavant és realitzarà el mateix procediment per a l'altre variable d'estudi *index15mins*.

Tots els resultats es troben continguts en la Taula 4.1, on s'hi troba indicada en primer lloc l'estimació de l'índex de fiabilitat i seguidament la variància associada al índex i els intervals del 95% de confiança. Primerament és destacable el fet que el valor dels índexs és molt alt en tots els casos exceptuant l'observador 2 al mesurar pacients saludables. Així mateix, si s'analitza la segona columna, on es troba la variància del valor comentat, pot observar-se una alta variabilitat entre els diferents índexs, a més, mentre que els observadors 1 i 3 presenten una menor variabilitat al realitzar les mesures en els pacients saludables, l'observador 2 presenta la situació contrària amb una variància per el primer cas considerablement superior que els altres observadors.

D'entre totes les possibles comparacions entre les nou mesures, aquelles que es de major interès són les que fan referència a les següents situacions:

1. Diferència entre les mesures dels observadors per un mateix diagnòstic

	$\hat{\rho}$	$\hat{\sigma}_{\hat{\rho}}$	$IC_{inf}(\hat{\rho})$	$IC_{sup}(\hat{\rho})$
Diagnòstic saludable				
Observador 1	0.974	0.00980	0.946	0.987
Observador 2	0.868	0.0555	0.708	0.943
Observador 3	0.979	0.0127	0.932	0.994
Diagnòstic malalt				
Observador 1	0.918	0.0294	0.837	0.960
Observador 2	0.960	0.0264	0.857	0.989
Observador 3	0.909	0.0458	0.763	0.966

Taula 4.1: Estimacions per la interacció entre observador i diagnòstic en la variable *index4h*

2. Diferència entre les mesures de diferents diagnòstics per un mateix observador

De manera que es contrasten aquestes nou diferències a partir de la realització de comparacions múltiples. El procediment que s'ha seguit ha estat el disseny de la matriu de diferències i el càlcul de l'estadístic de contrast E , equació (21), i a partir d'aquest s'han extret els p valors que posteriorment han estat ajustat pel **mètode de Holm**. Per a realitzar les comparacions múltiples cal recordar que al treballar amb Models Lineals Mixtes la variància de l'error aleatori es defineix com $\hat{\sigma}_{e_{ij}}^2 = \hat{\sigma}_e^2 \hat{\theta}_{ij}$, així doncs el contrast es formula com:

$$\begin{cases} H_0 : \sigma_{e_{ij}}^2 = \sigma_{e_{i'j'}}^2 \\ H_1 : \sigma_{e_{ij}}^2 \neq \sigma_{e_{i'j'}}^2 \end{cases} \Rightarrow \begin{cases} H_0 : \theta_{ij} = \theta_{i'j'} \\ H_1 : \theta_{ij} \neq \theta_{i'j'} \end{cases}$$

L'estadístic associat al contrast s'ha calculat com:

$$E = \frac{\hat{\theta}_{ij} - \hat{\theta}_{i'j'}}{\sqrt{\text{Var}(\hat{\theta}_{ij} - \hat{\theta}_{i'j'})}} \Rightarrow \begin{cases} i = 1, 2, 3 & (\text{observador}) \\ j = 1, 2 & (\text{variable}) \end{cases} \quad (21)$$

Aleshores els p valors associats als contrastos es troben en la Taula 4.2, aquells que prenen un valor menor al nivell de significació 0.05.

Contrast	p valor associat
Comparació de l'error aleatori entre observadors al mesurar pacients saludables	
Observador 1 vs Observador 2	0.0014
Observador 1 vs Observador 3	0.6955
Observador 2 vs Observador 3	0.0194
Comparació de l'error aleatori entre observadors al mesurar pacients malalts	
Observador 1 vs Observador 2	0.5814
Observador 1 vs Observador 3	0.6955
Observador 2 vs Observador 3	0.5814
Comparació de l'error aleatori d'un observador al mesurar pacients saludables i malalts	
Observador 1	0.0048
Observador 2	0.2238
Observador 3	0.0931

Taula 4.2: Comparacions múltiples d'interès per a la variable *index4h*

Tenint en consideració els *p valors* resultants, Taula 4.2, pot afirmar-se amb un nivell de confiança del 95% que els observadors 1 i 3 al mesurar la variable *index4h* en pacients saludables i malalts no presenten diferències significatives. Altrament l'observador 2 presenta diferències significatives respecte els altres dos al mesurar a pacients saludables. En conseqüència, l'estructura de la variància de l'error aleatori que sembla més adequada és heteroscedàstica per les següents combinacions:

1. Observadors $1 \cup 3$ al mesurar pacients saludables
2. Observadors $1 \cup 3$ al mesurar pacients malalts
3. Observador 2

S'està reduint d'un total de sis a la meitat de combinacions diferents en l'estructura de la variància, fet que podria resultar en una millora significativa del model. Per a comprovar-ho es comparen el model que presenta la estructura que s'ha exposat prèviament i el model amb totes les combinacions mitjançant el test de raó de versemblança. S'obté un *p valor* associat de 0.3506, de

manera que no hi ha suficient evidència per a rebutjar la hipòtesis nul·la, H_0 , d'equivalència dels models. En altres paraules, els models no són significativament diferents i seguint el principi de parsimònia escollim el més simple, el que defineix l'estructura de la variància com a l'estructura determinada prèviament.

Cal estudiar aleshores els índexs de fiabilitat del model escollit i la realització del contrast d'excel·lència per a cada estimació. A la Taula 4.3 es troben els resultats per al model d'estructura de la variància final. Al observar la Taula 4.3, sembla que ρ associat a l'observador 2 i als observadors $1 \cup 3$ al mesurar malalts son molt similars, convé aleshores contrastar si és possible trobar una estructura més simple.

	$\hat{\rho}$	$\hat{\sigma}(\hat{\rho})$	$IC_{inf}(\rho)$	$IC_{sup}(\rho)$
Observadors $1 \cup 3$ saludables	0.975	0.0089	0.950	0.987
Observadors $1 \cup 3$ malalts	0.906	0.0348	0.809	0.955
Observador 2	0.919	0.0271	0.846	0.958

Taula 4.3: Estimació ρ per cada observador i l'error estàndard i interval de confiança associats per al model que explica *index4h*

Al comparar el model d'estructura de 3 nivells amb un model que només considera dos nivells - Observadors $1 \cup 3$ saludables i Observadors $1 \cup 3$ malalts $\cup 2$ - el *p valor* associat és 0.667, de manera que els models són equivalents i el model més simple és l'escollit com a model final. Amb aquest model s'obtenen els resultats de la Taula 4.4.

	$\hat{\rho}$	$\hat{\sigma}(\hat{\rho})$	$IC_{inf}(\rho)$	$IC_{sup}(\rho)$
Observadors $1 \cup 3$ saludables	0.975	0.0089	0.950	0.987
Observadors $1 \cup 3$ malalts $\cup 2$	0.914	0.0262	0.845	0.953

Taula 4.4: Estimació ρ per cada observador i l'error estàndard i interval de confiança associats per al model que explica *index4h*

Finalment és realitza el contrast d'excel·lència als índexs de fiabilitat associats a l'estructura final del model per a la variable *index4h*. En el cas de les mesures dels observadors $1 \cup 3$ en pacients saludables s'obté un *p valor* associat menor a 0.0001, de manera que hi ha una forta evidència per rebutjar H_0 , confirmant que l'índex és excel·lent. Contràriament l'altre índex presenta un *p valors*

associats al contrast de 0.313, per tant no pot afirmar-se que siguin excel·lents. Així i tot els valors que pren l'índex és considerablement elevat i consegüentment la fiabilitat és bona.

A tall de conclusió, pot afirmar-se que els observadors 1 i 3 prenen les mesures amb la mateixa fiabilitat per la variable *index4h* i que aquesta és molt alta tant al mesurar pacients saludables com malalts. L'observador 2 al mesurar ambdós tipus de pacients presenta la mateixa fiabilitat que els observadors 1 i 3 al mesurar a pacients malalts. De manera general la fiabilitat és considerablement elevada.

4.3 Estudi de l'error aleatori heteroscedàstic segons observador i diagnòstic al mesurar *index15mins*

El següent pas és la realització del mateix anàlisi per el model amb variable resposta *index15mins*. Els resultats associats a l'estudi és troben a les taules 4.5, 4.6, 4.7.

En primer lloc s'ha procedit a comparar l'estructura de la variància de l'error aleatori homogenia respecte els models amb una estructura heterogenia segons observador, diagnòstic i la combinació d'ambdós. Els resultats obtinguts han estat els mateixos que al tractar la variable *index4h*, el model que presenta una estructura de la variància heteroscedàstica segons totes les combinacions d'observador i diagnòstic és significativament diferent respecte el model d'estructura homoscedàstica i per tant conté més informació, així que és l'escollit per a continuar l'estudi.

A la Taula 4.5 s'hi troben expressades les estimacions dels índexs de fiabilitat per a les diferents combinacions d'observador i diagnòstic. En aquest cas, de manera general per a tots els observadors, la fiabilitat al mesurar pacients saludables és major respecta els pacients malalts, amb una variabilitat associada a les estimacions menor. L'observador 3 presenta una fiabilitat molt baixa al mesurar pacients malalts, la causa d'aquest fenomen són dues de les mesures que difereix molt respecte les altres preses per el mateix observador, com pot observar-se al realitzar l'estudi multivariant a l'apartat A.1 de l'apèndix, concretament la Figura (9(d)).

Seguidament es realitza l'estudi de les diferències entre les combinacions d'estudi. Contràriament a la situació observada en la variable *index4h*, en aquest cas la mesurar pacients saludables els ob-

	$\hat{\rho}$	$\hat{\sigma}(\hat{\rho})$	$IC_{inf}(\rho)$	$IC_{sup}(\rho)$
Diagnòstic saludable				
Observador 1	0.985	0.0061	0.967	0.993
Observador 2	0.953	0.0260	0.864	0.985
Observador 3	0.953	0.0232	0.878	0.982
Diagnòstic malalt				
Observador 1	0.846	0.0611	0.675	0.930
Observador 2	0.936	0.0576	0.661	0.989
Observador 3	0.376	0.1160	0.130	0.578

Taula 4.5: Estimacions per la interacció entre observador i diagnòstic en la variable *index15mins*

servadors no presenten diferències significatives. Així i tot, si que difereix l'observador 3 amb els altres al mesurar pacients malalts. A més, els observadors 1 i 3 presenten diferències molt significatives al realitzar mesures de pacients amb diferents diagnòstics, efectivament l'error comés és significativament major al diagnosticar malalts.

Prenent en consideració els resultats observats a les taules 4.5 i 4.6, la conclusió que se'n extreu és que l'estructura de la variància de l'error aleatori sembla ser pràcticament heteroscedàstica per a totes les combinacions d'observador i diagnòstic. Concretament ve determinada per les següents combinacions.

1. Observador 1 al mesurar pacients saludables
2. Observador 1 al mesurar pacients malalts
3. Observador 3 al mesurar pacients saludables
4. Observador 3 al mesurar pacients malalts
5. Observador 2

Contrast	p valor associat
Comparació de l'error aleatori entre observadors al mesurar pacients saludables	
Observador 1 vs Observador 2	0.122
Observador 1 vs Observador 3	0.071
Observador 2 vs Observador 3	0.775
Comparació de l'error aleatori entre observadors al mesurar pacients malalts	
Observador 1 vs Observador 2	0.586
Observador 1 vs Observador 3	0.0001
Observador 2 vs Observador 3	0.008
Comparació de l'error aleatori d'un observador al mesurar pacients saludables i malalts	
Observador 1	\ll 0.0001
Observador 2	0.775
Observador 3	\ll 0.0001

Taula 4.6: Comparacions múltiples d'interès per a la variable *index15mins*

Al realitzar el test basat en la raó de versemblança per comparar els models de diferents estructures heteroscedàstiques s'obté un *p valor* associat de 0.789, consegüentment no és pot rebutjar que els models són equivalents i a partir de la llei de parsimònia s'escull el model amb una estructura depenent de 5 combinacions en lloc de les 6 combinacions possibles. La taula de l'estimació dels índex de fiabilitat estimats per al model finalment escollit, Taula 4.7, presenta resultats molt similars a la Taula 4.5.

Es porta a terme aleshores el contrast d'excel·lència per cada un dels ρ estimats i s'obtenen *p valors* associats menors a 0.05 per tots menys l'últim dels contrastes, l'observador 3 al mesurar pacients diagnosticats malalts.

Per concloure, en el cas de la variable *index15mins* la fiabilitat dels observadors és considerablement elevada en les diferents situacions excepte en el cas de l'observador 3 al mesurar pacients

	$\hat{\rho}$	$\hat{\sigma}(\hat{\rho})$	$IC_{inf}(\rho)$	$IC_{sup}(\rho)$
Diagnòstic saludable				
Observador 1	0.985	0.0061	0.967	0.993
Observador 3	0.953	0.0232	0.878	0.982
Diagnòstic malalt				
Observador 1	0.846	0.0611	0.675	0.930
Observador 3	0.376	0.1160	0.130	0.578
Obs 2				
Observador 2	0.936	0.0576	0.661	0.989

Taula 4.7: Estimacions per la interacció entre observador i diagnòstic en la variable *index15mins*

malalts, en que presenta una fiabilitat molt baixa deguda a una alta variabilitat en els mesuraments.

5 Estudi de la concordança

Un cop especificada l'estructura de la variància de l'error aleatori més adient pels diferents models i mesurada la concordança de les mesures d'un mateix observador en diferents situacions (de manera global i segons diagnòstic), el següent pas és l'estudi de la concordança de les mesures **entre** els diferents observadors. Per a mesurar aquest tipus de concordança s'utilitza el *Concordance Correlation Coefficient* (CCC) que és un cas particular de l'ICC.

5.1 Estudi de les diferències sistemàtiques entre observadors

A mode de pas previ al càlcul dels índexs CCC, és realitza l'estudi de la mitjana de les mesures en funció dels observadors. La finalitat d'aquest procediment és la de contrastar si l'efecte dels observadors és significatiu respecte la mitjana global del model. Per a contrastar formalment l'efecte fix referent als observadors s'utilitza el test de raó de versemblança, contrastant el model amb l'efecte dels observadors respecte el model nul. Les hipòtesis confrontades seguidament contrasten el valor que pren el paràmetre associat a la variable *observador*.

$$H_0 : \beta_i = 0$$

$$H_1 : \beta_i \neq 0 \quad \forall i = 1, 2, 3$$

Per ambdós models, el que pren com a variable dependent *index15mins* i el que pren *index4h*, el *p valor* associat al test és menor a 0.001, de manera que l'efecte dels observadors és significativament diferent a zero en ambdós casos.

5.1.1 Diferències sistemàtiques per a la variable *index4h*

A continuació s'estudia la mitjana segons observador, la variància associada i l'interval de confiança del 95% en el model que estudia la variable *index4h*, elements presentats a la Taula 5.1. Pot observar-se que l'observador 2 presenta una esperança que sembla ser significativament menor a la dels observadors 1 i 3, la que és similar.

	$\hat{\mu}$	$\hat{\sigma}(\hat{\mu})$	$IC_{inf}(\mu)$	$IC_{sup}(\mu)$
Observador 1	1.43	0.229	1.32	1.54
Observador 2	1.34	0.232	1.23	1.45
Observador 3	1.43	0.230	1.33	1.54

Taula 5.1: Estimació μ per cada observador i l'error estàndard i interval de confiança associats per al model que explica *index4h*

A partir de comparacions múltiples es verifiquen les conclusions extretes a partir de la Taula 5.1. En efecte, els observadors 1 i 3 no presenten en les mesures diferències significatives. En canvi, l'observador 2 és significativament diferent als altres dos metges. Així doncs, pot considerar-se que l'observador 2 pren les mesures significativament més baixes que els altres observadors.

	$\hat{\mu}$	$\hat{\sigma}(\hat{\mu})$	$IC_{inf}(\mu)$	$IC_{sup}(\mu)$
Diagnòstic saludable				
Observador 1	1.21	0.191	1.13	1.28
Observador 2	1.14	0.211	1.05	1.23
Observador 3	1.21	0.193	1.14	1.29
Diagnòstic malalt				
Observador 1	1.68	0.198	1.60	1.76
Observador 2	1.57	0.198	1.49	1.65
Observador 3	1.68	0.211	1.59	1.77

Taula 5.2: Estimació μ pels diferents nivells de la interacció entre observador i diagnòstic i l'error estàndard i interval de confiança associats per al model que explica *index4h*

A l'estudiar prèviament l'estructura de la variància de l'error aleatori s'ha identificat un clar efecte de la combinació entre observador i diagnòstic. Aquest efecte pot ser observat també a la Taula 5.2, en que s'ha estimat el valor de la esperança de la variable *index4h* segons les dife-

rents combinacions d'observador i diagnòstic. Semblen observar-se diferències significatives entre l'observador 2 i els altres observadors, els que presenten el mateix valor de l'esperança.

Al estudiar les esperances segons el diagnòstic del pacient s'observa el mateix patró per ambdós diagnòstics, observadors 1 i 3 presenten els mateixos valors amb errors estàndards associats a les estimacions similars i l'observador 2 presenta valors menors.

5.1.2 Diferències sistemàtiques per a la variable *index15mins*

En el cas del model que explica la variable *index15mins*, de la mateixa manera que ha passat en el cas estudiat prèviament, l'observador 2 presenta de manera significativa - basat en el test de comparacions múltiples - mesures menors als altres observadors, els que en aquest cas també presenten un valor idèntic, tot i que l'error estàndard associat difereix lleugerament. Cal subratllar que en aquest cas tant les diferències com l'error estàndard associat a les estimacions prenen valors menors. Els *p* valors associats al test de comparacions per a cada parella d'observadors són 0.0001, 0.8939, 0.0022, de manera que l'observador 2 presenta un valor significativament menor als altres observadors.

	$\hat{\mu}$	$\hat{\sigma}(\hat{\mu})$	$IC_{inf}(\mu)$	$IC_{sup}(\mu)$
Observador 1	1.47	0.195	1.39	1.55
Observador 2	1.42	0.196	1.34	1.50
Observador 3	1.47	0.199	1.39	1.55

Taula 5.3: Estimació μ per cada observador i l'error estàndard i interval de confiança associats per al model que explica *index15mins*

Al realitzar l'estudi de les mitjanes segons diagnòstic i observador, expressat a la Taula 5.4, sembla ser que a l'hora de mesurar als pacients malalts els 3 observadors reporten mesures centrades en valors significativament diferents, essent l'observador 1 el que presenta una major desviació respecte els altres. Cal tenir en compte també que l'error estàndard associat a l'observador 3 al realitzar les mesures en els pacients malalts és molt superior respecte els altres casos.

	$\hat{\mu}$	$\hat{\sigma}(\hat{\mu})$	$IC_{inf}(\mu)$	$IC_{sup}(\mu)$
Diagnòstic saludable				
Observador 1	1.31	0.171	1.25	1.37
Observador 2	1.28	0.175	1.22	1.35
Observador 3	1.32	0.177	1.26	1.38
Diagnòstic malalt				
Observador 1	1.67	0.181	1.60	1.74
Observador 2	1.56	0.175	1.50	1.62
Observador 3	1.59	0.280	1.43	1.76

Taula 5.4: Estimació μ pels diferents nivells de la interacció entre observador i diagnòstic i l'error estàndard i interval de confiança associats per al model que explica *index15mins*

5.2 Càlcul índex de concordança global

Un cop estudiat l'efecte dels observadors i les seves diferències en l'esperança de les mesures, es procedeix a l'estimació de l'índex CCC, el que aportarà informació sobre la intercanviabilitat dels observadors. Aquesta és una propietat molt valuosa en proves clíniques ja que corrobora si el mètode, la tècnica d'imatge en aquest cas, és reproduïble. Hi ha un paquet existent, *cccrm*, que conté les funcions necessàries pel càlcul de l'índex de concordança, ρ_c , però requereix que l'estructura de variància del model donat sigui homoscedàstica, condició que els models d'estudi no satisfan. Per aquesta raó s'ha creat una funció que donat un model amb una estructura de variàncies heteroscedàstica i un determinat nombre d'observadors és capaç de calcular-ne l'índex de concordança associat. Els resultats per al conjunt d'observadors en les variables *index4h* i *index15mins* poden trobar-se en la Taula 5.5.

En primer lloc s'estudien els resultats de les mesures de la variable *index4h*. Malgrat l'estimació de l'índex de concordança és un valor elevat, $\hat{\rho}_c = 0.888$, a l'aplicar el contrast d'excel·lència s'obté un *p valor* associat de 0.53, de manera que l'índex resulta no ser excel·lent (valor estrictament superior a 0.9). Cal fer referència a l'estudi previ de les diferències sistemàtiques en el que

Variable	$\hat{\rho}_c$	$\hat{\sigma}(\hat{\rho}_c)$	$IC_{inf}(\rho_c)$	$IC_{sup}(\rho_c)$
index4h	0.888	0.0317	0.807	0.936
index15mins	0.884	0.0335	0.798	0.935

Taula 5.5: Estimació ρ_c i l'error estàndard i interval de confiança associats per al model que explica les variables *index4h* i *index15mins*

s'ha identificat l'observador 2 com a possible individu no intercanviable amb els altres dos. Per a comprovar aquesta hipòtesis s'avalua la concordança dels observadors per parelles.

5.3 Càlcul índex de concordança per parelles

Els resultats de l'estimació de l'índex de concordança per les diferents parelles d'observadors és troben representats a la Taula 5.6. Els observadors 1 i 3 són intercanviables, presentant el valor per l'índex de concordança més elevat, això es deu a que com s'ha observat abans al presentar la mateixa esperança la variabilitat entre observadors estimada és nul·la. Així i tot, els observadors 2 i 3 també presenten una concordança considerablement elevada. El valor pels observadors 1 i 2 és el més baix, no obstant això no significa que no presentin una bona concordança.

Parella d'observadors	$\hat{\rho}_c$	$\hat{V}ar(\hat{\rho}_c)$	$IC_{inf}(\rho_c)$	$IC_{sup}(\rho_c)$
Observadors 1 \cup 2	0.872	0.0371	0.777	0.928
Observadors 1 \cup 3	0.966	0.00012	0.937	0.982
Observadors 2 \cup 3	0.931	0.0338	0.824	0.974

Taula 5.6: Estimació ρ_c i l'error estàndard i interval de confiança associats per al model que explica la variable *index4h* per cada parell d'observadors

Es procedeix a realitzar el contrast d'excel·lència pels tres índexs i només el referent a la concordança entre els observadors 1 i 3 resulta prendre un valor significativament superior a 0.9. Així i tot els resultats són molt elevats i de manera general podria acceptar-se la intercanviabilitat

entre mesures de diferents observadors en el cas de la variable *index4h*.

D'altra banda, els resultats referents a la variable *index15mins* són similars en el marc general al estimar l'índex de concordança pels tres observadors, pren valor 0.884, l'índex no és excel·lent tot i prendre un valor elevat. S'estudia aleshores la concordança per parelles d'observadors, els resultats és troben expressats a la Taula 5.7. Com pot observar-se, al analitzar les mesures per parelles de metges, l'índex pren un valor molt elevat en tots els casos, els índex no són excel·lents però molt elevats.

Parella d'observadors	$\hat{\rho}_c$	$\hat{Var}(\hat{\rho}_c)$	$IC_{inf}(\rho_c)$	$IC_{sup}(\rho_c)$
Observadors 1 \cup 2	0.911	0.0275	0.839	0.952
Observadors 1 \cup 3	0.902	0.001	0.822	0.947
Observadors 2 \cup 3	0.981	0.0124	0.824	0.974

Taula 5.7: Estimació ρ_c i l'error estàndard i interval de confiança associats per al model que explica la variable *index15mins* per cada parell d'observadors

Així que de la mateixa manera que a l'estudiar la variable *index4h* pot afirmar-se que tot i que l'índex de concordança no és excel·lent, els valors estimats són molt elevats, de manera que els 3 observadors podrien considerar-se intercanviables.

6 Conclusions de l'anàlisi

Un cop analitzats i exposats els resultats pels diferents índexs cal extreure'n conclusions. Primerament, a l'estudiar la fiabilitat dels diferents observadors s'ha observat que de manera general (al no tenir en compte el diagnòstic del pacient), al mesurar ambdues variables, *index4h* i *index15mins*, la precisió dels observadors és molt elevada, menys l'observador 3 al mesurar *index15mins*, que presenta una fiabilitat considerablement baixa, de 0.54. S'ha identificat la causa d'aquest valor en un parell de les mesures realitzades significativament més baixes a la resta, a més al realitzar l'aproximació de l'ICC sense aquestes mesures l'índex presenta un valor considerablement elevat i similar als altres observadors, seria aconsellable revisar aquestes avaluacions per entendre la raó de la discordança. Si és té en compte ara que l'observador 3 és l'únic metge no especialitzat i a més no va rebre cap formació específica sobre la tècnica, podria ser un d'aquests factors el que suposa aquesta falta de precisió en les seves mesures. A més, al estudiar més a fons la fiabilitat dels observadors, tenint en compte aquest cop el diagnòstic del pacient, en el cas de la variable *index15mins* s'ha identificat una estructura bastant més complexa. D'aquesta manera s'ha identificat que les mesures en que l'observador 3 presenta una fiabilitat molt baixa és en pacients malalts, mentre en pacients saludables l'índex de fiabilitat associat és quasi excel·lent. Globalment la fiabilitat dels observadors al realitzar mesures en pacients malalts és menor respecte pacients saludables.

A l'estudiar la concordança entre els observadors, s'ha pogut comprovar que els observadors són intercanviables en un índex elevat, encara que no excel·lent. Per consegüent pot concloure's que la tècnica d'imatge és altament reproducible, ja que tot i que només s'han estudiat tres metges, aquests representen un ampli rang d'experiència, des de un metge resident fins a un especialista format en la tècnica d'imatge.

7 Limitacions en l'anàlisi de l'article clínic original

L'anàlisi original presenta dos aspectes de gran rellevància que provoquen que l'estudi sigui qualificat de subòptim. El primer d'ells és el el procediment per a l'**estimació de la variància intra-observador i interobservador**. A l'hora de calcular l'aproximació de la variància intraobservador per a la posterior estimació de l'ICC, els investigadors van fer ús únicament de les mesures dels observadors que presenten mesures repetides per cada pacient, és a dir, utilitzant exclusivament les mesures de l'observador 1. Aquest procediment no invalida el resultat però suposa una pèrdua de precisió al no estar treballant amb totes les dades disponibles. Conseqüentment, suposant que els observadors presentaven la mateixa fiabilitat, és va estimar un únic índex global, ρ_g , per als 3 observadors i basat únicament en les dades del primer observador.

D'altra banda, per a realitzar el càlcul de l'estimació de la variància interobservador, es va fer ús de les mesures dels observadors 2 i 3, (Y_{i12}, Y_{i13}) , mentre que les mesures de l'observador 1 van ser agregades a partir del càlcul de la mitjana per treballar amb una sola dada. De manera que és va considerar $(\bar{Y}_{i.1} = Y_{i11}^*)$, empobrint l'anàlisi. La variància residual resultant de procedir d'aquesta manera correspon a la formula (23).

$$\bar{Y}_{i.1} = \frac{1}{k} \sum_{i=1}^3 (\mu + \alpha_i + \beta_i + e_{i1k}) = \mu + \alpha_i + \beta_i + \frac{1}{k} e_{i1k} = \mu + \alpha_i + \beta_i + e_{i1k}^* \quad (22)$$

$$Var(e_{i1}^*) = Var\left(\frac{1}{k} e_{i1k}\right) = \frac{1}{k^2} [k \cdot \sigma_e^2] = \frac{1}{k} \sigma_e^2 \quad (23)$$

A causa del pas previ realitzat a l'equació (22) la variància residual resultant és subestima en la k -èsima¹ part del seu vertader valor, on k és el nombre de rèpliques, en aquest cas $k = 3$. Els resultats associats a l'article original són molt breus i consisteixen simplement en la taula adjunta a la Figura 7.1.

¹ k fa referència al nombre d'observadors, en aquest estudi aleshores la variància serà subestimada a una tercera part del valor real

Table 3 Intraobserver and interobserver variability

	Intra (same observer)		Inter (3 different observers)	
	ICC	CV%	ICC	CV%
H/M15m	0.97	2.1%	0.96	3.8%
H/M4h	0.97	3.1%	0.95	3.2%

H/M15m heart to mediastinum ratio at 15 min post-injection, *H/M4h* heart to mediastinum ratio at 4 h post-injection, *ICC* interclass correlation coefficient, *CV* coefficient of variability

Figura 7.1: Taula referent als ICC de l'article original [6]

8 Comparació entre l'anàlisi original i l'actual

Els canvis realitzats en la metodologia de l'anàlisi i les conclusions de manera global han estat les mateixes: la tècnica d'imatge és altament reproduïble. Així i tot, al treballar amb una estructura de la variància de l'error aleatori més complexa s'han arribat a conclusions parcials diferents en les anàlisis. Així com l'anàlisi original suposa que els diferents observadors presenten la mateixa fiabilitat, en l'estudi que s'ha portat a terme s'ha descobert una estructura heteroscedàstica que s'ajusta molt millor a les dades i que ens aporta un major grau de coneixement. S'ha observat per exemple com la fiabilitat és diferent entre alguns observadors i en alguns casos també segons el diagnòstic del pacient. Aquesta informació pot ser d'utilitat als metges que hagin de fer ús de la tècnica d'imatge. Un altre aspecte que difereix entre els estudis és la especificació d'un valor a partir del qual es considera un índex com a excel·lent. En l'anàlisi portat a terme en aquest estudi pocs dels índexs han resultat ser excel·lents, tot i que donat valors considerablement elevats s'ha considerat que eren prou bons com per verificar la tècnica. En canvi, l'anàlisi original, a l'utilitzar únicament les mesures d'un observador per calcular la variància intra-observador i subestimar la variància inter-observador, els valors resultants dels índexs són més elevats.

A tall de conclusió, la idea central d'aquest estudi és posar en manifest la importància de la metodologia en l'estadística, específicament en els estudis estadístics clínics. Una metodologia subòptima o directament equivocada pot distorsionar greument els resultats i conclusions dels estudis estadístics, els que per a obtenir millors resultats o per falta de suficients coneixements, no apliquen de manera adequada els mètodes estadístics.

Referències

- [1] Josep L Carrasco, Alejandro Caceres, Georgia Escaramis, and Lluís Jover. Distinguishability and agreement with continuous data. *Statistics in medicine*, 33(1):117–128, 2014.
- [2] Josep L Carrasco and Lluís Jover. Estimating the generalized concordance correlation coefficient through variance components. *Biometrics*, 59(4):849–858, 2003.
- [3] Josep L Carrasco, Tonya S King, and Vernon M Chinchilli. The concordance correlation coefficient for repeated measures estimated by variance components. *Journal of Biopharmaceutical Statistics*, 19(1):90–105, 2009.
- [4] I Lawrence and Vernon Chinchilli. Rejoinder to the letter to the editor from atkinson and nevell. 53(2):777–778, 1997.
- [5] Charles E McCulloch and John M Neuhaus. Generalized linear mixed models. *Encyclopedia of biostatistics*, 4, 2005.
- [6] África Muxí, Pilar Paredes, Ignacio Navales, Francesc Valldeoriola, Carles Gaig, Francisco Lomeña, Andres de la Cerda, Oriol Solà, Beatriz Domenech, Eduardo Tolosa, et al. Diagnostic cutoff points for 123 i-mibg myocardial scintigraphy in a caucasian population with parkinson's disease. *European journal of nuclear medicine and molecular imaging*, 38(6):1139–1146, 2011.
- [7] Salvador Pita Fernández, Sonia Pértega Díaz, and E Rodríguez Maseda. La fiabilidad de las mediciones clínicas: el análisis de concordancia para variables numéricas. *Cad Aten Primaria*, 10(4):290–6, 2003.

A Apèndix

A.1 Verificació i estudi descriptiu de les dades

El tractament de les dades previ a l'estudi és un dels escenaris més importants del projecte, anomenat *preprocessing*. En aquesta secció és pren un primer contacte amb les dades, el significat dels atributs i la seva naturalesa (tipus de dada). També es realitzen tasques com és la imputació de dades mancants i transformacions si s'escauen.

Les dades que es van proporcionar per a la realització del treball es trobaven en molt bones condicions, per aquest motiu no s'han hagut de fer modificacions alguna per a treballar amb elles. Pel que fa a la detecció i tractament de dades mancants no s'ha hagut de realitzar cap mena de procediment donat que no n'hi han.

Un altre punt d'estudi en aquest apartat és la detecció d'outliers. S'ha realitzat una anàlisi gràfica de les variables d'interès segregant segons diagnòstic, ja que com és detecta més endavant en l'exploració multivariant de les dades, el rang de valors que prenen les variables segons el diagnòstic difereix significativament, tal com s'observa també als boxplots.

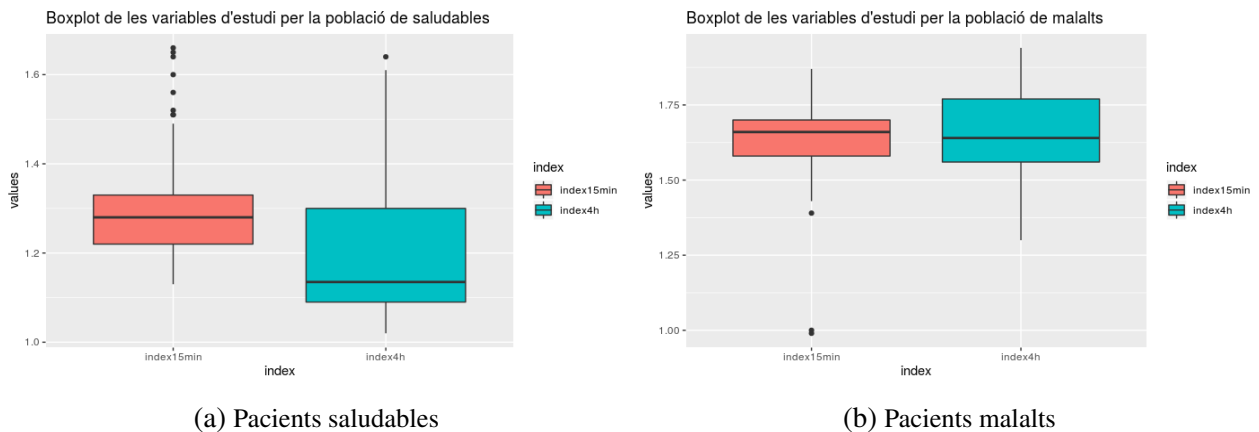


Figura A.1: Boxplots de les variables d'estudi *index4h* i *index15mins* segregant pel diagnòstic dels pacients

La variable *index4h* sembla no presentar outliers en la població de malalts, mentre que per la població de pacients saludables presenta una cua significativament més allargada en la cota superior que en la inferior, de manera que pren algun valor bastant distant de la resta. Per altra

banda la variable *index15mins* presenta outliers en els dos casos, tot i que amb diferents matisos: a l'avaluar-se en els pacients malalts presenta molt poca dispersió i dos outliers molt allunyats de la resta de observacions. En la població de saludables també s'observa la cua de la cota superior significativament més allargada i bastants outliers, tenint en compte que el total d'individus es conforma per només 14 observacions.

A.1.1 Exploració multivariant de les dades

El primer pas ha estat la visualització tant gràfica com numèrica univariant i bivariant de les principals variables d'estudi. En primer lloc s'han estudiat les variables *index15mins* i *index4h* de manera aïllada i posteriorment s'ha estudiat l'efecte del diagnòstic i els observadors en aquestes a partir de tècniques de visualització multivariant.

Anàlisi descriptiu univariant

En primer lloc es realitza l'anàlisi descriptiu numèric i gràfic de les variables *index4h* i *index15mins*. Com pot observar-se en la Taula A.1, ambdues variables prenen pràcticament el mateix rang de valors com a domini i la mitjana difereix de manera molt lleu, en 5 centèsimes. Aquestes variables mesuren el mateix índex en dos moments temporals diferents, de manera que no resulta molt sorprenent que el rang de valors sigui el mateix, a més s'identifica als gràfics de les distribucions de les variables un clara bimodalitat, segurament degut a que s'hi troben barrejades dues poblacions diferents, pel context en el que s'esta treballant, tot sembla indicar que el diagnòstic dels pacients és el causant d'aquest tret. Tot i haver-hi un nombre de similituds, també es destaca el fet que els valors en el que es centren les dues potencials poblacions per ambdues variables difereixen.

Nom variable	Tipus	Estadístics Descriptius					
		Min	Q_1	Median	Mean	Q_3	Max
index15mins	Continua	0.99	1.27	1.49	1.47	1.66	1.87
index4h	Continua	1.02	1.13	1.40	1.42	1.64	1.94

Taula A.1: Taula descriptiva de les variables numèriques d'estudi *index4h* i *index15mins*

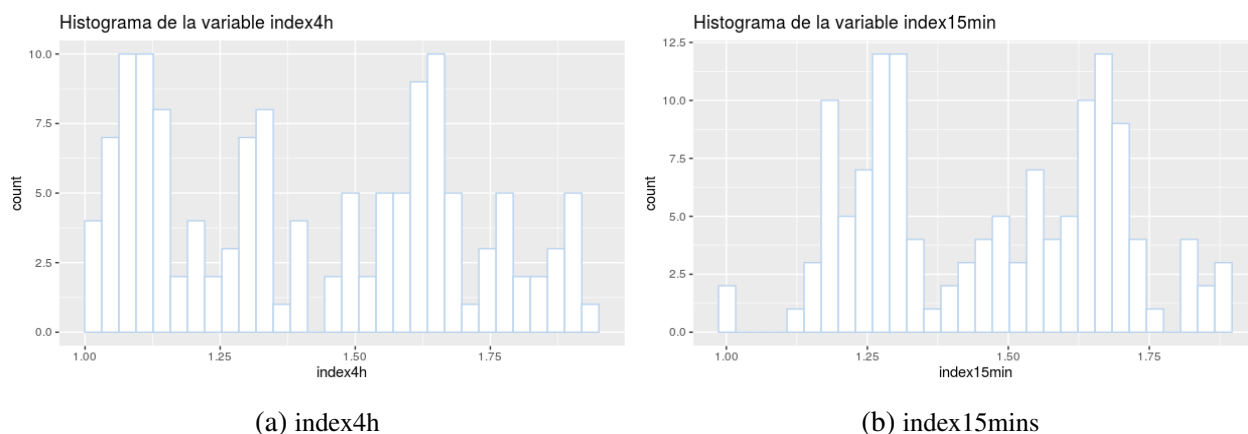


Figura A.2: Histograma de les variables d'estudi *index4h* i *index15mins*

Anàlisi descriptiu bivariant

- Variable diagnòstic

Seguidament s'efectua l'estudi de visualització multivariant de les variables *index4h* i *index15mins* amb la finalitat d'analitzar el comportament d'aquestes variables sota l'efecte dels factors referents al diagnòstic i l'observador. Al realitzar el gràfic d'un histograma representant les variables d'estudi en funció del diagnòstic dels pacients, representat en la Figura A.3, pot observar-se que la distribució dels valors és pràcticament discriminable segons el diagnòstic del pacient. Els pacients saludables prenen valors d'H/W de manera general més baixos en els dos moments temporals respecte els pacients malalts, a més la seva distribució és més platocúrtica. Un altre tret destacable que s'identifica clarament en la Taula A.2, és el decrement en mitjana en el valor de H/W pels pacients saludables mentre que els pacients malalts presenten un lleuger increment del

valor amb el pas del temps. Més encara, sembla haver un outlier en la variable *index15mins* per un dels pacients malalts que es troba molt allunyat de la resta de pacients amb el mateix diagnòstic, prenent un valor significativament baix.

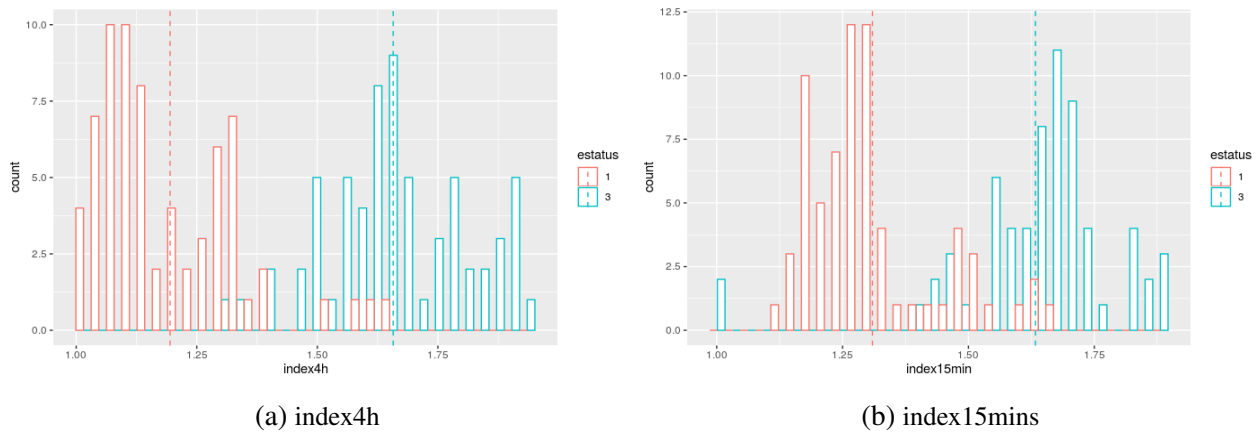


Figura A.3: Histograma de la variable d'estudi *index4h* i *index15mins* en funció del diagnòstic

Diagnòstic	H/W als	
	15 minuts	4 hores
Saludables	1.31	1.19
Malalts	1.63	1.66

Taula A.2: Taula descriptiva de les variables *index15mins* i *index4h* en funció de *estatus*

- Variable observador

Pel que fa a la distribució del les variables d'estudi en funció de l'observador que pren les mesures, els gràfics de la Figura A.5 presenten les distribucions per cada un dels observadors i variable. No resulta gaire fàcil extreure conclusions donada la bimodalitat, la que té raó de ser del diagnòstic. De manera gràfica sembla que els observadors presenten una variabilitat diferent per les variables, dona la impressió que són més precisos al mesurar la variable *index15mins*. Al llarg

de l'estudi s'intenta esbrinar si aquestes variàncies que s'observen en els observadors i entre ells són significativament diferents.

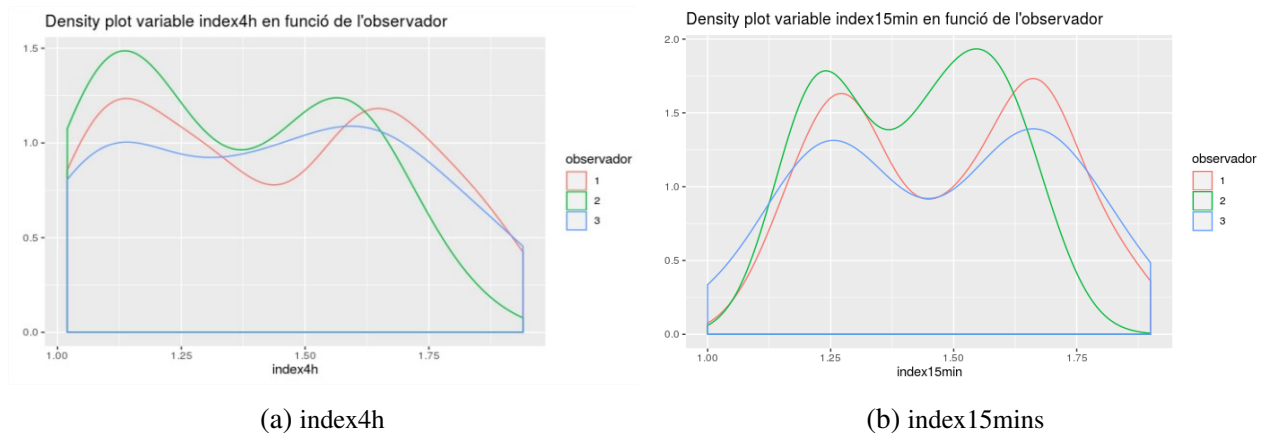
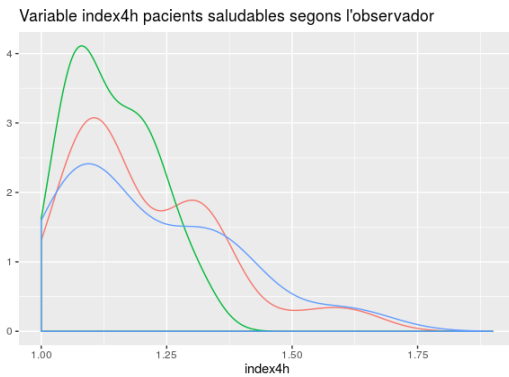


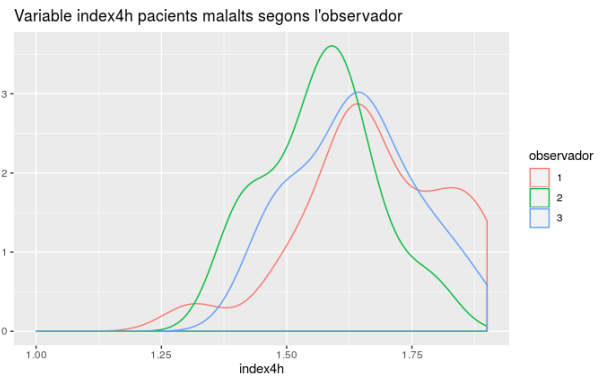
Figura A.4: Density plot de la variable d'estudi index4h i index15mins en funció de l'observador

- Variable observador i diagnòstic

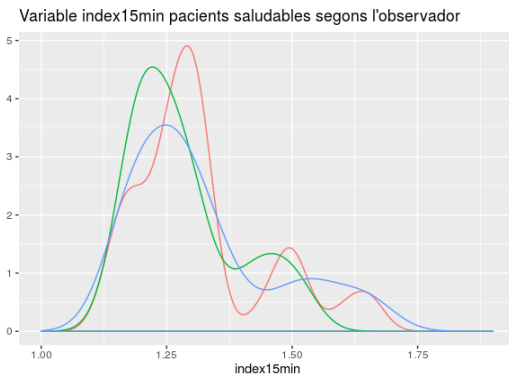
De manera més exhaustiva, es realitza l'estudi gràfic dels diferents observadors en funció del diagnòstic del pacient per a les variables d'estudi. Els gràfics que es troben a la Figura A.5 són elements clau donat que són gràfics molt informatius sobre les discrepàncies entre els valors de les mesures dels diferents observadors. Així com els tres primers gràfics semblen presentar densitats similars pels diferents observadors, val la pena comentar que no és el cas de l'últim gràfic, que reflecteix les mesures preses per els metges per la variable *index15mins* en els pacients malalts. En altres paraules, hi ha una major variabilitat entre el valor de les mesures dels metges. D'altra banda també cal destacar com alguns gràfics presenten distribucions més leptocúrtiques, com n'és el cas dels gràfics (c) i (d) respecte els dos primers. Aquest aspecte aporta informació sobre la variabilitat intra-observador, concepte en el que es profunditzarà al llarg del treball.



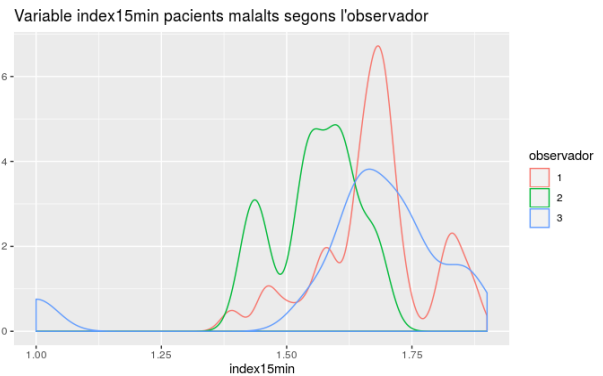
(a) index4h pacients saludables



(b) index4h pacients malalts



(c) index15mins pacients saludables



(d) index15mins pacients malalts

Figura A.5: Density plot de la variable d'estudi index4h i index15mins en funció del diagnòstic i l'observador

A.2 Implementació de les funcions en R

Tots els algorismes utilitzats en aquest treball han estat creats i implementats amb el llenguatge de programació R, amb els paquets addicionals:

1. Per a l'anàlisi exploratori de les dades

- readxl
- dplyr
- tidyr
- summarytools
- ggplot2
- car
- MASS

2. Per a la modelització

- nlme
- lsmeans

La part tècnica del treball ha consistit en la creació e implementació del codi necessari per poder dur a terme l'anàlisi de la tècnica d'imatge, el codi és troba dividit en dos scripts, un referent a les funcions que s'han desenvolupat en els recents apartats, `Funcions.R`, i l'altre referent a l'execució d'aquestes funcions i la consegüent obtenció dels resultats que s'aniran expressant al document en els propers apartats, `Codi.TFG.Rmd`. Tot el codi pot trobar-se en l'últim apartat de l'apèndix, A.7.

Les diferents funcions implementades al codi poden diferenciar-se principalment en:

1. Funcions pel càlcul dels elements de la variància:

- `dev.theta` que desenvolupa el mètode deltha fent ús de la funció preexistent `Deriv`, facilitant molts càlculs descrits a l'apartat (2.4.2).
- `Lmat` encarregada de la creació de la matriu de disseny dels diferents contrastos que s'estudiaran.
- `sa`, `se`, `theta` per a extreure del model les estimacions dels paràmetres $\sigma_\alpha, \sigma_e, \theta$
- `sb`, `var.sb` pel càlcul de $\hat{\sigma}_b, \hat{Var}(\sigma_b)$

2. Funcions pel càlcul de l'índex de fiabilitat/concordança

- `càlcul.output` conté el codi per al càlcul dels índex de fiabilitat i concordança i els errors estàndards associats
- `intervals` és la funció que s'utilitza cada cop que és vol obtenir un interval de confiança per a l'índex d'estudi.
- `contr.excel` pel contrast formal de l'excel·lència d'un índex.

Les funcions que s'han anat codificant per a l'estudi, s'han aplicat en el cas concret de 3 observadors i una variància heteroscedàstica, així i tot, són generalitzables al un nombre d'observadors n , essent n un nombre finit entre 1 i qualsevol real enter.

La funció que s'ha utilitzat per a la validació de les altres operacions ha estat `boot.model`, consistent en la creació d'unes noves dades equivalents a les originals, a partir de la distribució dels diferents efectes, individus i observadors. En l'script principal, `Codi.TFG.Rmd`, s'aplica la funció per a validar els resultats de cada apartat.

A.3 Validació dels resultats/funcions

La validació dels resultats és un aspecte fonamental, en aquest treball s'ha realitzat una rigorosa validació de tots els resultats i funcions abans de prendre conclusions. La validació de resultats ha estat especialment necessària a l'hora de fer ús del mètode delta, donat que les estimacions realitzades a partir d'aquest procediment no és consideren aproximacions gaire potents amb mostres petites comparat amb altres procediments com bootstrap paramètric. A partir del model i les distribucions de les diferents variables s'han generat 1000 cops unes noves dades (equivalents a les originals) i s'ha calculat ρ i ρ_c .

Per crear els conjunts de dades equivalents en primer lloc s'han generat de manera aleatòria els valors dels efectes aleatoris A_i, e_{ijk} a partir de les distribucions:

$$A_i \sim N(0, \sigma_\alpha^2)$$
$$e_{ijk} \sim N(0, \sigma_\alpha^2)$$

Per altra banda s'obté el valor de β a partir del model, donat que es un efecte fix. Aleshores es realitza el sumatori dels diferents valors per a cada observació, de manera que l'individu y_{ijk} pren valor $\mu + A_i + \beta_j + e_{ijk}$.

Aleshores per a validar les funcions s'ha realitzat el càlcul dels paràmetres i del valor de l'índex d'estudi per bootstrap un cop per cada funció. Els resultats es troben en les taules de l'apartat 2 del apèndix, les diferències entre els dos procediments són d'ordre de mil·lèsimes, de manera que els resultats s'han donat per vàlids i les funcions per efectives. Cal comentar també que en el futur aquestes funcions s'inclouran en el paquet `cccrm`. S'han recopilat totes les taules comparatives al següent apartat de l'apèndix.

A.4 Variàncies associades als càlculs dels índexs

- Variàncies associades al càlcul de l'índex de fiabilitat

1. Observadors al mesurar *index4h*

$\hat{\sigma}_\alpha^2$	$\hat{\sigma}_{e_1}^2$	$\hat{\sigma}_{e_2}^2$	$\hat{\sigma}_{e_3}^2$
0.0733	0.00423	0.00549	0.00427

Taula A.3: Estimació de les variàncies entre individus i la variància residual per a cada observador al realitzar les mesures de la variable *index4h*

2. Global al mesurar *index4h*

$\hat{\sigma}_\alpha^2$	$\hat{\sigma}_{e_1}^2$
0.07141	0.00484

Taula A.4: Estimació de les variàncies entre individus i la variància residual global al realitzar les mesures de l'*index4h*

3. Observadors al mesurar *index15mins*

$\hat{\sigma}_\alpha^2$	$\hat{\sigma}_{e_1}^2$	$\hat{\sigma}_{e_2}^2$	$\hat{\sigma}_{e_3}^2$
0.0381	0.00355	0.00277	0.01074

Taula A.5: Estimació de les variàncies entre individus i la variància residual per a cada observador al realitzar les mesures de l'*index15min*

4. Observadors 1 ∪ 2 i 3 al mesurar *index15mins*

σ_{α}^2	$\hat{\sigma}_{e_{1\cup 2}}^2$	$\hat{\sigma}_{e_3}^2$
0.040	0.00316	0.0101

Taula A.6: Estimació de les variàncies entre individus i la variància residual per al conjunt d'observadors 1 ∪ 2 i 3 al realitzar les mesures de l'*index15min*

5. Interacció observador i diagnòstic al mesurar *index4h*

Interacció Observador:Diagnòstic	$\hat{\sigma}_{\alpha}^2$	$\hat{\sigma}_e^2$
Observador 1 al mesurar saludables	0.0736	0.00199
Observador 2 al mesurar saludables	0.0736	0.00472
Observador 3 al mesurar saludables	0.0736	0.00175
Observador 1 al mesurar malalts	0.0736	0.00362
Observador 2 al mesurar malalts	0.0736	0.00246
Observador 3 al mesurar malalts	0.0736	0.00384

Taula A.7: Estimacions per la interacció entre observador i diagnòstic en la variable *index4h*

6. Interacció observador i diagnòstic al mesurar *index15mins*

Interacció Observador:Diagnòstic	$\hat{\sigma}_\alpha^2$	$\hat{\sigma}_e^2$
Observador 1 al mesurar saludables	0.0389	0.00060
Observador 2 al mesurar saludables	0.0389	0.00107
Observador 3 al mesurar saludables	0.0389	0.00107
Observador 1 al mesurar malalts	0.0389	0.00207
Observador 2 al mesurar malalts	0.0389	0.00127
Observador 3 al mesurar malalts	0.0389	0.00624

Taula A.8: Estimacions per la interacció entre observador i diagnòstic en la variable *index15mins*

Estructura proposada combinacions observador i diagnòstic al mesurar *index4h*

Interacció Observador:Diagnòstic	$\hat{\sigma}_\alpha^2$	$\hat{\sigma}_e^2$
Observador 1 \cup 3 al mesurar saludables	0.07413	0.00191
Observador 1 \cup 3 al mesurar malalts	0.07413	0.00773
Observador 2	0.07413	0.00655

Taula A.9: Estimacions per la interacció entre observador i diagnòstic en la variable *index4h*

Estructura final combinacions observador i diagnòstic al mesurar *index4h*

Interacció Observador:Diagnòstic	$\hat{\sigma}_\alpha^2$	$\hat{\sigma}_e^2$
Observador 1 \cup 3 al mesurar saludables	0.0737	0.00192
Observador 1 \cup 3 al mesurar malalts \cup 2	0.0737	0.0069

Taula A.10: Estimacions per la interacció entre observador i diagnòstic en la variable *index4h***Estructura final combinacions observador i diagnòstic al mesurar *index15mins***

Interacció Observador:Diagnòstic	$\hat{\sigma}_\alpha^2$	$\hat{\sigma}_e^2$
Observador 1 al mesurar saludables	0.0389	0.00060
Observador 3 al mesurar saludables	0.0389	0.00206
Observador 1 al mesurar malalts	0.0389	0.00192
Observador 3 al mesurar malalts	0.0389	0.00731
Observador 2	0.0389	0.0650

Taula A.11: Estimacions per la interacció entre observador i diagnòstic en la variable *index15mins*

- Variàncies associades al càlcul de l'índex de concordança

1. Paràmetres pel càlcul de l'índex de concordança global

Variable	$\hat{\sigma}_\alpha^2$	$\hat{\sigma}_{e_1}^2$	$\hat{\sigma}_{e_2}^2$	$\hat{\sigma}_{e_3}^2$	$\hat{\sigma}_B^2$
index4h	0.0708	0.001985	0.01098	0.00427	0.002939
index15mins	0.0372	0.0006042	0.001738	0.00176	0.000783

Taula A.12: Estimació de les variàncies entre individus i la variància residual per a cada observador al realitzar les mesures de l'*índex4h*

2. Paràmetres pel càlcul de l'índex de concordança dels observadors 2 a 2 per *índex4h*

Variable	$\hat{\sigma}_\alpha^2$	$\hat{\sigma}_{e_1}^2$	$\hat{\sigma}_{e_2}^2$	$\hat{\sigma}_B^2$
Observadors 1U2	0.0711	0.00209	0.01030	0.00405
Observadors 2U3	0.0570	1.6520e-11	0.0143	0.00420
Observadors 1U3	0.0447	0.0006042	0.0020	0.00000

Taula A.13: Estimació de les variàncies entre individus i la variància residual per a cada observador al realitzar les mesures 2 a 2 per *índex4h*

3. Paràmetres pel càlcul de l'índex de concordança dels observadors 2 a 2 per *index15mins*

Variable	$\hat{\sigma}_\alpha^2$	$\hat{\sigma}_{e_1}^2$	$\hat{\sigma}_{e_2}^2$	$\hat{\sigma}_B^2$
Observadors 1U2	0.0389	0.000593	0.00203	0.01327
Observadors 2U3	0.0286	6.009e-12	0.00291	0.00056
Observadors 1U3	0.0447	0.000583	0.00199	0.00000

Taula A.14: Estimació de les variàncies entre individus i la variància residual per a cada observador al realitzar les mesures 2 a 2 per *index4h*

A.5 Estudi per a la validació de funcions a partir de bootstrap

(a) Paràmetres al mesurar *index4h* per bootstrap

	Funcions	Bootstrap
$\hat{\sigma}_\alpha^2$	0.07330	0.07370
$\hat{\sigma}_{e_1}^2$	0.00423	0.00420
$\hat{\sigma}_{e_2}^2$	0.00549	0.00474
$\hat{\sigma}_{e_3}^2$	0.00427	0.00412

Taula A.15: Estimació de les variàncies entre individus i la variància residual per a cada observador al realitzar les mesures de l'*index4h* a partir de 500 repliques de bootstrap parametric

(b) ICC i CCC observadors al mesurar *index4h* per bootstrap

	Funcions	Bootstrap
Index de fiabilitat		
$\hat{\rho}_{Observador_1}$	0.945	0.943
$\hat{\rho}_{Observador_2}$	0.930	0.935
$\hat{\rho}_{Observador_3}$	0.945	0.943
Index de concordança		
$\hat{\rho}_{Observador_1}$	0.945	0.943
$\hat{\rho}_{Observador_2}$	0.930	0.935
$\hat{\rho}_{Observador_3}$	0.945	0.943

Taula A.16: Estimació de les variàncies entre individus i la variància residual per a cada observador al realitzar les mesures de l'*index4h* a partir de 500 repliques de bootstrap parametric

A.6 Codi R

- Funcions utilitzades pel càlcul de paràmetres ICC

```
1
2# Complementary functions recurrently used in more complex functions (
   regarding the clculations of ICC, CCC and confidence intervals)
3sa<-function(tau) exp(2*(tau)) # Function to obtain within individuals
   variance
4ltheta<-function(nobs,b) c(0,if(nobs > 1){b[2:(nobs)]}) # logarithm of
   parameter theta (weights of the residual variance for the observers i =
   2,3, ...)
5se<-function(ltheta,log_s) ifelse(ltheta == 0,exp(2*log_s),exp(2*ltheta+2*log
   _s)) # Function to obtain between individuals variance
6icc<-function(tau,ltheta,log_s) sa(tau)/(sa(tau)+se(ltheta,log_s)) # Function
   to clculate ICC index
7# Functions to clculate partial derivates of the function ICC
8d1_icc<-Deriv(icc,"tau")
9d2_icc<-Deriv(icc,"ltheta")
10d3_icc<-Deriv(icc,"log_s")
11
12
13dev.theta<-function(lt,i,vars,ntotal){
14  elements<-NULL
15  if((ntotal-1) > 1){
16    for(k in 2:(ntotal-1)){
17      elements<- c(elements,ifelse(i != k,0,d2_icc(vars[1],lt[i],vars[ntotal]))
18        )
19    }
20    m<-c(d1_icc(vars[1],lt[i],vars[ntotal]),
21      elements,
22      d3<-icc(vars[1],lt[i],vars[ntotal]))
23    matrix(m,nrow= ntotal)
```

24

25 }

- Funció pel càlcul d'interval de confiança i contrast d'excel·lència

```
1
2 #Function to calculate confidence intervals (CI) using Z-Fisher transformation
3 intervals<-function(index,var_index){
4   z<-0.5*log((1+index)/(1-index))
5   sez<-var_index/(((1+index)**2)*((1-index)**2))
6   llz<-z+qnorm(0.025)*sqrt(sez)
7   ulz<-z+qnorm(0.025,lower.tail = F)*sqrt(sez)
8   ll95 = (exp(2*llz)-1)/(exp(2*llz)+1);
9   ul95 = (exp(2*ulz)-1)/(exp(2*ulz)+1);
10  data.frame(int_inf=ll95,int_sup=ul95)
11 }
12
13 #Function for contrasting whether if a value is excellent (> 0.9)
14
15 contr.excel<-function(estim,var_e,pt_ex){
16   z<-0.5*log((1+estim)/(1-estim))
17   valor<-0.5*log((1+pt_ex)/(1-pt_ex))
18   sez<-sqrt(var_e/((1-estim**2)**2))
19   est_contrast<-(z-valor)/sez
20   pval<-pnorm(est_contrast,lower.tail = FALSE)
21   ret<-paste("El p valor associat al contrast d'excel·lència s de",round(pval
22     ,4))
23   return(ret)
24 }
```

- Funció pel càlcul de l'índex de fiabilitat

```

1# Function that recompiles all the related to ICC index, CI and parameters
2
3output.icc<-function(model,boot=FALSE,params= FALSE){
4  vars<-attr(model$apVar,"Pars")
5  ntotal<-length(vars)
6  nobs<-ntotal-1
7  var_ind<-sa(vars[1])
8  lt<-ltheta(nobs,vars)
9  var_e<- se(lt,vars[ntotal])
10 iccs<-icc(vars[1],lt,vars[ntotal])
11
12 if(params == TRUE){
13   params<-c(var_ind,var_e)
14   noms<-c("var ind")
15   for(i in 1:nobs){noms<-c(noms,paste("var residual obs",i))}
16   names(params)<-noms
17   return(params)
18 }
19
20 if(boot == FALSE){
21  var.icc<-noms<-NULL
22  for(i in 1:nobs){
23   var.icc<-c(var.icc,t(dev.theta(lt,i,vars,ntotal))%%model$apVar%%dev.
24     theta(lt,i,vars,ntotal))
25  }
26  interval<-intervals(iccs,var.icc)
27  out<-data.frame(icc= iccs,std_e= var.icc, interval_inferior= interval
28    [,1],interval_superior= interval[,2])
29  row.names(out)<- for(i in 1:nobs){noms<-c(noms,paste("observador",i))}
30  return(out)
31 }else{
32  df.boot=c(var_ind,var_e,iccs)

```



```

31   return(df.boot)
32 }
33
34 }

```

- Funcions utilitzades pel càlcul de paràmetres CCC

```

1
2 # Functions to calculate CCC index
3
4 varB<-function(model,nobs)model$varFix[1:nobs,1:nobs] # Components of
   observer variance
5 seh<-function(ltheta,log_s,nobs) sum(exp(ltheta+2*log_s))/nobs # Mean of the
   residuals
6
7 # Function to create de design matrix
8
9 Lmat<-function(k){
10  aux1<-diag(k)*0
11  aux1[1,1:k]<-1
12
13  for (i in 1:(k-1)) aux1[1+i,1+i]=1
14  L=array(0,dim=c(k,k*(k-1)/2))
15  cont=0
16  for (i in 1:k){
17    for (j in 1:(k-1)){
18      if (i>j){
19        cont=cont+1
20        L[,cont]=aux1[,i]-aux1[,j]
21      }
22    }
23  }
24  return(L)
25 }

```

```

26
27 #Computes variability between observers' means
28 sb<-function(nobs,b,varB){
29   L<-Lmat(nobs) # Building L matrix
30   difmed=t(L)%*%b # clculating observers sum of squares
31   #vardifmed=t(L)%*%varB%*%L
32   A<-L%*%t(L)
33   aux2<-(t(difmed)%*%difmed)-sum(diag(A%*%varB))
34   sb<-max(aux2/(nobs*(nobs-1)),0)
35   return(sb)
36 }
37
38 # Variance of sb
39 var.sb<-function(k,b,varB){
40   L<-Lmat(k) # Building L matrix
41   A<-L%*%t(L)
42   xx<- ( 2*sum(diag( ( A%*%varB)**2 ) ) ) + 4*( t(b)%*%A%*%varB%*%A%*%b ) ) /
         ((k*(k-1))**2)
43   return(xx)
44 }

```

- Funcions utilitzades pel càlcul de l'índex de concordança i l'error estàndard associat a l'índex

```

1
2 ccc<-function(tau,ltheta,log_s,k,b,varB) sa(tau)/(sa(tau)+sb(k,b,varB)+seh(
   ltheta,log_s,k))
3
4 #Standard error of CCC index
5
6 #Partial derivates of CCC function
7 d1_ccc<-Deriv(ccc,"tau")
8 d2_ccc<-Deriv(ccc,"ltheta")
9 d3_ccc<-Deriv(ccc,"log_s")

```

```

10 d4_ccc<-Deriv(seh,"ltheta")
11 d5_ccc<-Deriv(seh,"log_s")
12
13 # Standard error
14 var_ccc<- function(model){
15   n<-model$dims$ngrps[1]
16   vars<-attr(model$apVar,"Pars")
17   ntotal<-length(vars)
18   b<-model$coef$fixed
19   nobs<-length(b)
20   varB<-varB(model,nobs)
21   # nh<-length(unique(dades.rel$obs_est)) # Number of levels heterocedastic
      error
22   var_ind<-sa(vars[1])
23   lt<-ltheta((ntotal-1),vars)
24   var_e<- mean(se(lt,vars[ntotal]))
25   ccc1<-ccc(vars[1],lt,vars[ntotal],nobs,b,varB)
26   dtau<-d1_ccc(vars[1],lt,vars[ntotal],nobs,b,varB)
27   dlt<-NULL
28   for (i in 2:(ntotal-1)){
29     dlt<-c(dlt,d2_ccc(vars[1],lt[i],vars[ntotal],nobs,b,varB))
30   }
31   dls<-d3_ccc(vars[1],lt,vars[ntotal],nobs,b,varB)
32   # Derivative sb
33   den<-var_ind+sb(nobs,b,varB)+var_e
34   dsb<-(-1)*ccc1/den
35   if (sb(nobs,b,varB)==0) dev.sb<-0
36   dev<-matrix(c(dtau,dlt,dls,dsb),ncol=1)
37   # Covariance respect seh
38   dev5<-NULL
39   for (i in 2:(ntotal-1)){
40     dev5<-c(dev5,d5_ccc(lt[i],vars[ntotal],nobs))
41   }
42   dev4<-d4_ccc(lt,vars[ntotal],nobs)

```

```

43 # Creation of sigma matrix
44 sigma<-array(0,c(ntotal+1,ntotal+1))
45 sigma[1:ntotal,1:ntotal]<-model$apVar
46 sigma[ntotal+1,ntotal+1]<-var.sb(nobs,b,varB)
47 sigma[1:ntotal,ntotal+1]<-c(0,dev5,dev4)%*%model$apVar*(-1/n)
48 sigma[ntotal+1,1:ntotal]<-c(0,dev5,dev4)%*%model$apVar*(-1/n)
49
50 var.ccc<-t(dev)%*%sigma%*%dev
51 return(var.ccc)
52 }
53
54
55 # Function that recompiles all the related to ICC index, CI and parameters
56
57 output.ccc<-function(model){
58   vars<-attr(model$apVar,"Pars")
59   ntotal<-length(vars)
60   b<-model$coef$fixed
61   nobs<-length(b)
62
63   iccc<- ccc(vars[1],ltheta((ntotal-1),vars),vars[ntotal],nobs,b,varB(model,
64     nobs))
65   var_iccc<-var_ccc(model)
66   out1<-data.frame(CCC= iccc, Estd=sqrt(var_iccc) ,Int_inf= intervals(iccc,
67     var_iccc)[,1], Int_sup= intervals(iccc,var_iccc)[,2])
68   out2<-data.frame(var_ind = sa(vars[1]), var_e = se(ltheta(nobs,vars),vars[
69     ntotal]), var_dif_obs= sb(nobs,b,varB(model,nobs)))
70   out<-list(out1,out2)
71   return(out)
72 }

```

- Algoritme que genera noves dades equivalents a les originals per aplicar les funcions creades (Bootstrap paramètric)

```

1
2 # Function to validate the other functions (Parametric Bootstrap)
3
4 boot.model<-function(dd,var_ind,var_e,beta){
5   n_Alpha<-length(unique(dd$ID))
6   n_e= c(n_Alpha*3,n_Alpha,n_Alpha)
7
8   alph= rnorm(n_Alpha,0,sqrt(var_ind))
9
10  error_i1k = rnorm(n_e[1],0,sqrt(var_e[1]))
11  error_i2k = rnorm(n_e[2],0,sqrt(var_e[2]))
12  error_i3k = rnorm(n_e[3],0,sqrt(var_e[3]))
13
14  y1<-as.numeric(rep(beta[1],n_e[1])+rep(alph,3)+error_i1k)
15  y2<-as.numeric(rep(beta[1]+beta[2],n_e[2])+alph+error_i2k)
16  y3<-as.numeric(rep(beta[1]+beta[3],n_e[3])+alph+error_i3k)
17  y<-c(y1,y2,y3)
18
19  ind<-as.factor(rep(1:27,5))
20  obs<-as.factor(c(rep(1,27*3),rep(2,27),rep(3,27)))
21  data.boot<-data.frame(y,ind,obs)
22  m.boot<-lme(y~obs,random= ~ 1|ind,weights = varIdent(form= ~1|obs),data=
      data.boot)
23  clcul_output(m.boot,boot = TRUE)
24 }

```