

# TIPS BIOESTADÍSTICOS

## Tema 13: El coeficiente de correlación de concordancia de Lin

PETER B. MANDEVILLE

*The most exciting phrase to hear in science, the one that heralds discoveries, is not "Eureka!" but "Now that's funny..."*

Isaac Asimov

En cualquier validación de un ensayo (*assay*) o proceso de validación de un instrumento, la reproducibilidad de mediciones de ensayo a ensayo es de interés. Además, cuando un nuevo ensayo o instrumento está desarrollado, es de interés evaluar si el nuevo puede reproducir los mismos resultados que un ensayo tradicional o estándar de oro. Tales procesos de validación frecuentemente se evalúan utilizando el coeficiente de correlación de Pearson,<sup>a</sup> la prueba de t pareada, el análisis por cuadrados mínimos con pendiente igual a 1 y el intercepto igual a 0, el coeficiente de variación, o el coeficiente de correlación intraclase.<sup>b</sup> Existen desventajas en el uso de to-

<sup>a</sup> De acuerdo con Zar,<sup>5</sup> se utiliza el coeficiente de correlación de Pearson cuando es posible designar una variable como X y otra variable como Y.

<sup>b</sup> De acuerdo con Zar,<sup>5</sup> se utiliza el coeficiente de correlación intraclase cuando tienen pares de observaciones, pero es imposible decir que las observaciones en la primera columna tienen algo sistemático, que es diferente al que tienen las observaciones en la segunda columna.

das estas estadísticas, ya que no se puede evaluar la reproducibilidad utilizando solamente una de éstas.<sup>2,3,4</sup>

La correlación y regresión lineal son suficientes para evaluar el acuerdo entre dos métodos, pero existe la necesidad de un índice de resumen para evaluar la reproducibilidad de las mediciones. Este índice fue desarrollado por Lin (1989) y se denomina el *coeficiente de correlación de concordancia* y está dado por Shoukri.

Por ejemplo, un laboratorio analítico puede tener interés en saber si las mediciones de una sustancia son las mismas utilizando dos diferentes instrumentos, o cuándo son efectuados por dos diferentes técnicos.<sup>5</sup>

Se grafica el resultado obtenido con el método 1 contra el resultado obtenido con el método 2 para cada muestra, con el fin de observar si los puntos están sobre una línea de 45 grados que pasa por el origen, esto es la línea de concordancia.<sup>2</sup> Reproducibilidad perfecta del ensayo será cuando los puntos caen sobre la línea de concordancia.<sup>5</sup> El resultado

es el grado de reproducibilidad.<sup>2</sup>

Lin (1989) ha mostrado que el método de evaluar reproducibilidad es superior al de comparación de coeficientes de variación, prueba de t pareada, regresión, correlación de Pearson y correlación intraclase.<sup>5</sup> Además ha mostrado que la prueba es robusta con n tan pequeño como 10.<sup>2,5</sup>

Se encuentra una tabla para la determinación del tamaño de la muestra en su artículo.<sup>6</sup>

El coeficiente de correlación de concordancia,  $r_c$ , puede tener valores desde -1 a +1 y sus valores absolutos no pueden ser mayores que el coeficiente de correlación de Pearson,  $r$ , por lo cual se puede decir que  $-1 \leq -|r| \leq r_c \leq |r| \leq 1$  y que

$r_c = 0$  solamente si  $r = 0$ .<sup>2,5</sup>

<sup>c</sup> Agradezco a la Q.F.B. Esperanza de la Cruz Mendoza y a la Dra. Celia Aradillas García por los datos.

En el siguiente ejemplo se determina el grado de reproducibilidad de insulina en plasma a través de dos métodos: RIA y QML con 172 pacientes.<sup>c</sup> Primero se efectuó el análisis propuesto por Mandeville.<sup>7</sup>

El primer paso es leer los datos que fueron de formato csv de Excel,

```
> dat <-
read.csv("celia.csv",header=T)
```

verificar los nombres de las variables

```
> names(dat)
[1] "FOLIO" "RIA" "QLM"
```

y el número de repeticiones.

```
> length(dat$RIA)
[1] 172
```

El paso siguiente es para converger la matriz en una con las variables respuesta (Resp), tratamiento (Trat) y folio (Folio).

```
> attach(dat)
> Resp <- c(RIA,QLM)
> Trat <-
as.factor(c(rep("RIA",172),rep("QLM",172)))
> Folio <-
as.factor(c(FOLIO,FOLIO))
> detach(dat)
> Dat <-
data.frame(Resp,Trat,Folio)
```

Entonces se efectuó ANOVA

```
> res <-
lm(Resp~Trat+Folio,data=Dat)
> anova(res)
Analysis of Variance Table
```

Response: Resp

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Trat	1	64	64	4.66	0.032
Folio	171	53898	315	23.04	<2e-16
Residuals	171	2339	14		

Existe una diferencia significativa y se calcula las medias y la diferencia entre las medias.

```
> tmp <-
tapply(Dat$Resp,Dat$Trat,mean)
> tmp
   QLM  RIA
11.05 11.91
> tmp[[2]]-tmp[[1]]
[1] 0.8608
```

La media de RIA es mayor que la media de QLM y en promedio la diferencia es de 0.8608 unidades.

Se calcula el coeficiente de correlación intraclass y los límites de confianza

```
> ICC31(res)
```

Intraclass Correlation Coefficient (ICC)

Shrout-Fleiss 3,1 (raters fixed, patients random)

```
ICC: 0.92
LCL: 0.89
UCL: 0.94
```

y el coeficiente de correlación de Pearson y los límites de confianza.

```
> attach(dat)
> cor.test(RIA,QLM)
```

Pearson's product-moment correlation

data: RIA and QLM  
t = 37.93, df = 170, p-value < 2.2e-16

alternative hypothesis: true correlation is not equal to 0  
95 percent confidence interval:  
0.9273 0.9595  
sample estimates:  
cor  
0.9457

El código siguiente calcula el coeficiente de correlación de concordancia de Lin y es una modificación del código de Ziegler<sup>8</sup> que incluye las correcciones de los errores tipográficos en Lin<sup>2</sup> y Lin<sup>6</sup> que fueron publicadas posteriormente.<sup>9</sup>

```
Lincc <- function(x,y){
  mx <- mean(x); my <- mean(y)
  rc <- (2*cov(x,y))/(
  (var(x)+var(y)+((mx-my) ^ 2))
  r <- cor(x,y); n <- length(x)
  u <- ((n-1)*((mx-my) ^ 2))/
  sqrt(sum((x-mx) ^ 2)*sum((y-my) ^ 2))
  sigmazc <- sqrt((((1-(r ^ 2))*rc ^ 2)/
  ((1-(rc ^ 2))*r ^ 2)+(2*(rc ^ 3)*(1-rc)*u)/
  (r*(1-(rc ^ 2)) ^ 2)-((rc ^ 4)*(u ^ 2))/
  (2*(r ^ 2)*(1-(rc ^ 2)) ^ 2))/(n-2))
  delta <- qnorm(1-.05/2)*sigmazc
  zc <- 0.5*log((1+rc)/(1-rc))
  li <- zc-delta; ls <- zc+delta
  LI <- (exp(2*li)-1)/(exp(2*li)+1)
  LS <- (exp(2*ls)-1)/(exp(2*ls)+1)
  cat("Lin's Concordance
  Coefficient:»,rc,»\n»)
  cat("límites:»,LI,»y»,LS,»\n»)
}
> Lincc(QLM,RIA)
```

Lin's Concordance Coefficient: 0.9147  
límites: 0.893 y 0.9322

La tabla presenta los resultados y se nota que en este caso los de los tres métodos son muy parecidos, pero pueden tener resultados muy diferentes uno del otro.

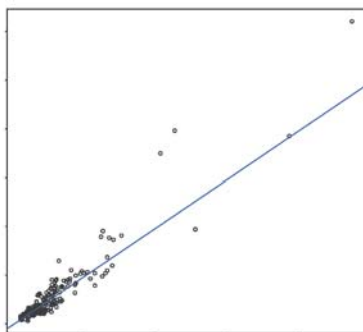
Estadística	Límite inferior	límite estimado	Límite superior
Pearson	0.945	0.946	0.9595
ICC	0.890	0.920	0.94
Lin	0.893	0.915	0.9322

Para examinar el rechazo, se levantó una gráfica de concordancia entre los resultados de los métodos QML y RIA para la medición de insulina.

```
plot(QLM,RIA,xlab="QLM",ylab="RIA")
abline(col=blue,0,1)
```

Se observa que hay concordancia hasta aproximadamente las 30 unidades.

Otra manera para mostrar el acuerdo gráficamente es con una gráfica Bland-Altman, donde se muestran las diferencias entre las mediciones y establecen cortes que son calculados al



multiplicar el error estándar por 2, y suma y resta este producto de la media de las diferencias. Estos cortes son aproximadamente los límites de confianza a 95% para las diferencias.<sup>3,4</sup>

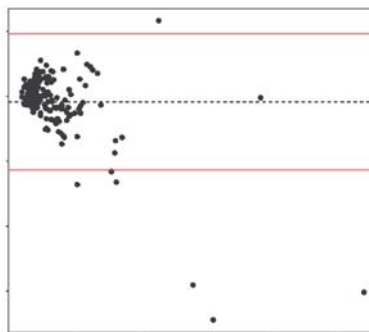
```
BAplot <- function(x,y,titl){
  mn <- 0.5*(x+y); diff <- x-y
  md <- mean(diff); vd <- var(diff)
  plot(mn,diff,pch=16,xlab="Media",
  ylab="Diferencia",main=titl,
  sub=paste("Media = ",md))
  abline(h=md,lty=2);
  abline(h=md+c(2,-
  2)*sqrt(vd),col="red")
  cat("Cortes:",md+2*sqrt(vd),"y",md-
  2*sqrt(vd),"\n")
}
```

```
> BAplot(QLM,RIA,"QLM y RIA")
Cortes: 9.6 y -11.32
```

Otra vez muestra que hay concordancia hasta aproximadamente las 30 unidades.

Para ver si hay concordancia, y es menor de 30 unidades, se repite el análisis, omitiendo los valores iguales o mayores que 30.

```
> res <-
```



```
lm(Resp~Trat+Folio,subset=Resp<30,
data=Dat) > anova(res)
Analysis of Variance Table
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Trat	1	64	64	4.66	0.032
Folio	171	53898	315	23.04	<2e-16
Residuals	171	2339	14		

Todavía existe una diferencia significativa entre los resultados de los dos métodos, por lo que es necesario evaluar si las diferencias son clínicamente importantes o no, como se expuso antes en Mandeville.<sup>7</sup>

Referencias

- George E. P. Box, J. Stuart Hunter y William G. Hunter. (2005). *Statistics for Experimenters: Design, Innovation, and Discovery*. Second edition. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., New York, NY, USA.
- Lawrence I-Kuei Lin. (1989). A Concordance Correlation Coefficient to Evaluate Reproducibility. *Biometrics* 45, 255-268 (March).
- Mohamed M. Shoukri. (2004). *Measures of Interobserver*
- J. Martin Bland and Douglas G. Altman. (1986). *Statistical Methods for Assessing Agreement Between Two Methods of Clinical Measurement*. *Lancet*, I, 307-310.
- Jerrold H. Zar. (1996). *Biostatistical Analysis*. Third edition. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.
- Lawrence I-Kuei Lin. (1992). *Assay validation using the concordance correlation*

- coefficient. *Biometrics* 48, 599-604 (June).
7. Peter B. Mandeville. (2005). Tips Bioestadísticos. Tema 9: El coeficiente de correlación intraclase (ICC). *CienciaUANL*, VOL. VIII, No. 3, julio-septiembre 2005: 414-416.
8. Craig H. Ziegler. (08nov2003). Annotating graphs using text. R-help.
9. Lawrence I-Kuei Lin. (2000). A Note on the Concordance Correlation Coefficient. *Biometrics* 56, 324-325 (March).