

Meta-análisis en R

Guía básica

Juan David Leongómez^{1,*}

04 septiembre, 2021

¹ Laboratorio de Análisis del Comportamiento Humano (LACH), Facultad de Psicología, Universidad El Bosque, Bogotá, Colombia. jdleongomez.info

* Correspondencia: [Juan David Leongómez <jleongomez@unbosque.edu.co>](mailto:jleongomez@unbosque.edu.co)

Descripción

Este documento contiene todo el código explicaciones básicas, paso a paso, para hacer un meta-análisis en R, usando los paquetes [metafor](#) (Viechtbauer, 2010) y [robumeta](#) (Fisher & Tipton, 2015). Está basado, y siguiendo [este video](#), creado por Daniel S. Quintana (2021).

Índice

1. Meta-análisis de correlación	1
1.1. Base de datos de ejemplo	1
1.2. Transformación de r de Pearson a z de Fisher	2
1.3. Hacer el meta-análisis	3
1.3.1. Más información sobre heterogeneidad	4
1.3.2. Diagnóstico de influencia	4
1.3.3. <i>Forest plot</i> (diagrama de bosque)	6
Referencias	9

1. Meta-análisis de correlación

1.1. Base de datos de ejemplo

La base de datos `dat.molloy2014`, tomada de Molloy et al. (2013) viene incluida con `{metafor}`. Básicamente, estudia si existe una asociación entre la diligencia (*conscientiousness*) y la adherencia a la medicación. En otras palabras, ¿las personas más diligentes son más propensas a cumplir con la medicación prescrita?

```
dat <- get(data(dat.molloy2014))
dat <- mutate(dat, study_id = 1:16) #add study_id column
dat <- dat %>% select(study_id, authors:quality) #bring study_id to first column
```

La base da datos, que he asignado a un objeto llamado `dat`, tienen ahora la siguiente estructura:

Tabla 1

Estructura de la base de datos de la base de datos

study_id	authors	year	ni	ri	controls	design	a_measure	c_measure	meanage	quality
1	Axelsson et al.	2009	109	0.187	none	cross-sectional	self-report	other	22.00	1
2	Axelsson et al.	2011	749	0.162	none	cross-sectional	self-report	NEO	53.59	1
3	Bruce et al.	2010	55	0.340	none	prospective	other	NEO	43.36	2
4	Christensen et al.	1999	107	0.320	none	cross-sectional	self-report	other	41.70	1
5	Christensen & Smith	1995	72	0.270	none	prospective	other	NEO	46.39	2
6	Cohen et al.	2004	65	0.000	none	prospective	other	NEO	41.20	2
7	Dobbels et al.	2005	174	0.175	none	cross-sectional	self-report	NEO	52.30	1
8	Ediger et al.	2007	326	0.050	multiple	prospective	self-report	NEO	41.00	3
9	Insel et al.	2006	58	0.260	none	prospective	other	other	77.00	2
10	Jerant et al.	2011	771	0.010	multiple	prospective	other	NEO	78.60	3
11	Moran et al.	1997	56	-0.090	multiple	prospective	other	NEO	57.20	2
12	O'Cleirigh et al.	2007	91	0.370	none	prospective	self-report	NEO	37.90	2
13	Penedo et al.	2003	116	0.000	none	cross-sectional	self-report	NEO	39.20	1
14	Quine et al.	2012	537	0.150	none	prospective	self-report	other	69.00	2
15	Stilley et al.	2004	158	0.240	none	prospective	other	NEO	46.20	3
16	Wiebe & Christensen	1997	65	0.040	none	prospective	other	NEO	56.00	1

Nota: Datos tomados de Molloy et al. (2013).

La columna **ri** contiene los coeficientes de correlación de Pearson (la columna **ni** contiene los tamaños de muestra de cada estudio). Dado que los coeficientes de Pearson no tienen una distribución normal, esto podría llevar a calcular varianzas incorrectas, especialmente cuando se trata de correlaciones con tamaños de muestra pequeños.

Adicionalmente, en este ejemplo tenemos una serie de moderadores:

- **controls:** número de variables controladas
- **design:** si se utilizó un diseño transversal o prospectivo
- **a_measure:** tipo de medida de adherencia (autoinforme u otro)
- **c_measure:** tipo de medida de diligencia (NEO u otra)
- **meanage:** edad promedio de la muestra
- **quality:** calidad metodológica

1.2. Transformación de r de Pearson a z de Fisher

Por esto, vamos a transformar los coeficientes r de Pearson a z de Fisher, que no tienen este problema. Para esto, vamos a usar la función `escalc` del paquete `metafor`.

```
dat <- escalc(measure = "ZCOR",
             ri = ri,
             ni = ni,
             data= dat,
             slab = paste(authors, year, sep = ", "))
```

Esto ha creado dos nuevas variables en nuestra tabla: **yi**, que es el tamaño de efecto, y **vi** que es la varianza.

Tabla 2

Estructura de la base de datos, con transformación de los r de Pearson a z de Fisher

study_id	authors	year	ni	ri	controls	design	a_measure	c_measure	meanage	quality	yi	vi
1	Axelsson et al.	2009	109	0.187	none	cross-sectional	self-report	other	22.00	1	0.1892266	0.0094340
2	Axelsson et al.	2011	749	0.162	none	cross-sectional	self-report	NEO	53.59	1	0.1634399	0.0013405
3	Bruce et al.	2010	55	0.340	none	prospective	other	NEO	43.36	2	0.3540925	0.0192308
4	Christensen et al.	1999	107	0.320	none	cross-sectional	self-report	other	41.70	1	0.3316471	0.0096154
5	Christensen & Smith	1995	72	0.270	none	prospective	other	NEO	46.39	2	0.2768638	0.0144928
6	Cohen et al.	2004	65	0.000	none	prospective	other	NEO	41.20	2	0.0000000	0.0161290
7	Dobbels et al.	2005	174	0.175	none	cross-sectional	self-report	NEO	52.30	1	0.1768200	0.0058480
8	Ediger et al.	2007	326	0.050	multiple	prospective	self-report	NEO	41.00	3	0.0500417	0.0030960
9	Insel et al.	2006	58	0.260	none	prospective	other	other	77.00	2	0.2661084	0.0181818
10	Jerant et al.	2011	771	0.010	multiple	prospective	other	NEO	78.60	3	0.0100003	0.0013021
11	Moran et al.	1997	56	-0.090	multiple	prospective	other	NEO	57.20	2	-0.0902442	0.0188679
12	O'Cleirigh et al.	2007	91	0.370	none	prospective	self-report	NEO	37.90	2	0.3884231	0.0113636
13	Penedo et al.	2003	116	0.000	none	cross-sectional	self-report	NEO	39.20	1	0.0000000	0.0088496
14	Quine et al.	2012	537	0.150	none	prospective	self-report	other	69.00	2	0.1511404	0.0018727
15	Stilley et al.	2004	158	0.240	none	prospective	other	NEO	46.20	3	0.2447741	0.0064516
16	Wiebe & Christensen	1997	65	0.040	none	prospective	other	NEO	56.00	1	0.0400214	0.0161290

Nota: Datos tomados de Molloy et al., (2013).

1.3. Hacer el meta-análisis

Para hacer el meta-análisis, usaremos la función `rma` del paquete `metafor`, para el que tenemos que especificar los tamaños de efecto (`yi`) y varianzas (`vi`) de los estudios a meta-analizar. En este caso, las columnas donde tenemos estos valores, tienen los mismos nombres (`yi`, `vi`). Asignaré los resultados del meta-análisis a un objeto llamado `res`.

```
res <- rma(yi = yi, vi = vi, data = dat)
```

Los resultados, son los siguientes:

```
res

##
## Random-Effects Model (k = 16; tau^2 estimator: REML)
##
## tau^2 (estimated amount of total heterogeneity): 0.0081 (SE = 0.0055)
## tau (square root of estimated tau^2 value):      0.0901
## I^2 (total heterogeneity / total variability):    61.73%
## H^2 (total variability / sampling variability):   2.61
##
## Test for Heterogeneity:
## Q(df = 15) = 38.1595, p-val = 0.0009
##
## Model Results:
##
## estimate      se      zval      pval      ci.lb      ci.ub
## 0.1499 0.0316 4.7501 <.0001 0.0881 0.2118 ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Primero, nos confirma que ajustamos un modelo con efectos aleatorios (`Random-Effects Model`), a partir de 16 estudios (`k = 16`), y que para estimar τ^2 (tau cuadrado¹) usamos el método de **máxima verosimilitud restringida**² (`tau^2 estimator: REML`), que se designa como *REML* por sus siglas en inglés.

Posteriormente, nos provee los valores de una serie de estimadores de heterogeneidad o varianza:

¹ τ^2 es una estimación de la varianza de los tamaños de los efectos reales entre los estudios meta-analizados. Se usa, principalmente, para asignar pesos a cada estudio. Para más información, ver (Borenstein et al., 2009).

²Hay varios métodos disponibles como estimador, además de **máxima verosimilitud restringida** (REML). Sin embargo, si no estás seguro, REML es una buena opción. Cada método tiene ventajas y desventajas que, si tienes interés en mirar, están descritas en la [documentación](#) de la función `rma`.

- τ^2 : tau^2 (estimated amount of total heterogeneity): 0.0081 (SE = 0.0055)
- τ : tau (square root of estimated tau^2 value): 0.0901
- I^2 : I^2 (total heterogeneity / total variability): 61.73%, y
- H^2 : H^2 (total variability / sampling variability): 2.61

La tercera parte, reporta una prueba de heterogeneidad, usando el estadístico Q :

- Test for Heterogeneity:
Q(df = 15) = 38.1595, p-val = 0.0009

De todos estos, los más comúnmente reportados son τ^2 , τ , I^2 y Q . Cada una de estas medidas tiene ventajas y desventajas, por lo cual tiene sentido reportarlas todas.

I^2 , por ejemplo, tiene la ventaja de ser sencillo de interpretar, pues hay criterios generales para heterogeneidad baja, moderada y alta (típicamente 25 %, 50 %, and 75 %, respectivamente). Sin embargo, es muy sensible a los tamaños de muestra de los estudios meta-analizados (por ejemplo, si en tu meta-análisis hay estudios con tamaños de muestra muy grandes, esto va a sesgar tu I^2).

Q , aunque no es sensible al tamaño de muestra, es sensible al número de estudios meta-analizados. Tiene la ventaja de ser un test de hipótesis, y como tal, puede ser interpretado a partir de su valor p .

τ^2 no tiene estos problemas, pero es más difícil de interpretar.

En nuestro caso, el estadístico Q sugiere que hay una heterogeneidad significativa en los estudios meta-analizados ($p = 0.0009$). I^2 , sugiere una heterogeneidad moderada, lo que quiere decir que más de la mitad (61.73 %) de la varianza se estima que se deriva de diferencias en los tamaños de efecto.

Por último, tenemos los resultados del modelo de meta-análisis (**Model results**). Nos provee un estimado de la asociación positiva entre diligencia y adherencia a la medicación (0.1499 ± 0.0316), lo que equivale a un valor z de 4.7501, y sugiere que esa asociación es significativa ($p < .0001$). Así mismo, nos provee los límites inferior (0.0881) y superior (0.2118) de los intervalos de confianza.

1.3.1. Más información sobre heterogeneidad

Además de reportar los estadísticos τ^2 , τ , I^2 y Q , podemos fácilmente calcular los intervalos de confianza para τ^2 , τ , e I^2 con la función `confint`, que también pueden ser reportado junto a estos estadísticos.

```
confint(res)

##
##      estimate    ci.lb    ci.ub
## tau^2      0.0081  0.0017  0.0378
## tau        0.0901  0.0412  0.1944
## I^2(%)     61.7324 25.2799 88.2545
## H^2        2.6132  1.3383  8.5139
```

Para el τ^2 , el hecho de que los intervalos de confianza no crucen el 0 (en nuestro caso 0.0017 — 0.0378), sugiere que de hecho también que hay heterogeneidad entre los estudios que meta-analizamos.

1.3.2. Diagnóstico de influencia

Otro aspecto importante de un meta-análisis, es determinar si alguno(s) de los estudios meta-analizados es(son) particularmente influyente(s) en nuestro resultado³. Para esto, podemos usar la función `influence`, cuyo resultado en este caso asignaré a un objeto llamado `inf`.

```
inf <- influence(res)
```

Ya que lo asigné a un objeto (`inf`), para ver el resultado, tengo que correrlo para ver su resultado.

³Por ejemplo, si estuviésemos meta-analizando 20 estudios, de los cuales 19 tienen un n de 100, pero el otro tiene un n de 10.000, éste último tendrá una influencia enorme en nuestro resultado. Sería preocupante que tu meta-análisis sea dependiente de un único estudio.

```
inf
```

```
##
##               rstudent  dffits  cook.d  cov.r  tau2.del  QE.del
## Axelsson et al., 2009    0.2918  0.0485  0.0025  1.1331    0.0091  37.7109
## Axelsson et al., 2011    0.1196 -0.0031  0.0000  1.2595    0.0100  36.7672
## Bruce et al., 2010       1.2740  0.2595  0.0660  0.9942    0.0075  35.3930
## Christensen et al., 1999 1.4711  0.3946  0.1439  0.9544    0.0068  33.5886
## Christensen & Smith, 1995 0.8622  0.1838  0.0339  1.0505    0.0082  36.5396
## Cohen et al., 2004      -0.9795 -0.2121  0.0455  1.0639    0.0084  37.1703
## Dobbels et al., 2005     0.2177  0.0296  0.0010  1.1740    0.0094  37.6797
## Ediger et al., 2007     -0.9774 -0.3120  0.1001  1.1215    0.0084  36.1484
## Insel et al., 2006       0.7264  0.1392  0.0195  1.0561    0.0083  37.0495
## Jerant et al., 2011     -1.8667 -0.5861  0.2198  0.8502    0.0047  25.0661
## Moran et al., 1997      -1.4985 -0.2771  0.0756  1.0073    0.0077  35.6617
## O'Cleirigh et al., 2007  1.8776  0.4918  0.2148  0.8819    0.0059  31.9021
## Penedo et al., 2003     -1.1892 -0.2939  0.0859  1.0550    0.0080  36.3291
## Quine et al., 2012      -0.0020 -0.0423  0.0021  1.2524    0.0100  37.7339
## Stilley et al., 2004     0.8066  0.2126  0.0459  1.0907    0.0083  35.8385
## Wiebe & Christensen, 1997 -0.7160 -0.1656  0.0280  1.0853    0.0087  37.7017
##
##               hat  weight  dfbs  inf
## Axelsson et al., 2009    0.0568  5.6776  0.0481
## Axelsson et al., 2011    0.1054 10.5396 -0.0032
## Bruce et al., 2010       0.0364  3.6432  0.2623
## Christensen et al., 1999 0.0562  5.6195  0.3994
## Christensen & Smith, 1995 0.0441  4.4069  0.1837
## Cohen et al., 2004       0.0411  4.1094 -0.2112
## Dobbels et al., 2005     0.0714  7.1362  0.0296
## Ediger et al., 2007     0.0889  8.8886 -0.3128
## Insel et al., 2006       0.0379  3.7886  0.1387
## Jerant et al., 2011     0.1058 10.5826 -0.5430
## Moran et al., 1997      0.0369  3.6922 -0.2791
## O'Cleirigh et al., 2007  0.0511  5.1150  0.5059
## Penedo et al., 2003     0.0587  5.8732 -0.2941
## Quine et al., 2012      0.0998  9.9778 -0.0434
## Stilley et al., 2004     0.0684  6.8403  0.2125
## Wiebe & Christensen, 1997 0.0411  4.1094 -0.1642
```

Esto me muestra gran cantidad de información de cada estudio (en este caso, una tabla sin formato, que es muy ancha para poder imprimirse en esta página, por lo cual está reportada en dos partes). Sin embargo, lo más importante ahora es mirar la última columna, también llamada `inf`. Si ahí aparecieran asteriscos (que no es nuestro caso), sugeriría que ese estudio es particularmente influyente.

Por último, podemos también ver esta información que tenemos guardada en el objeto `inf`, de manera gráfica, usando la función `plot`.

```
plot(inf)
```

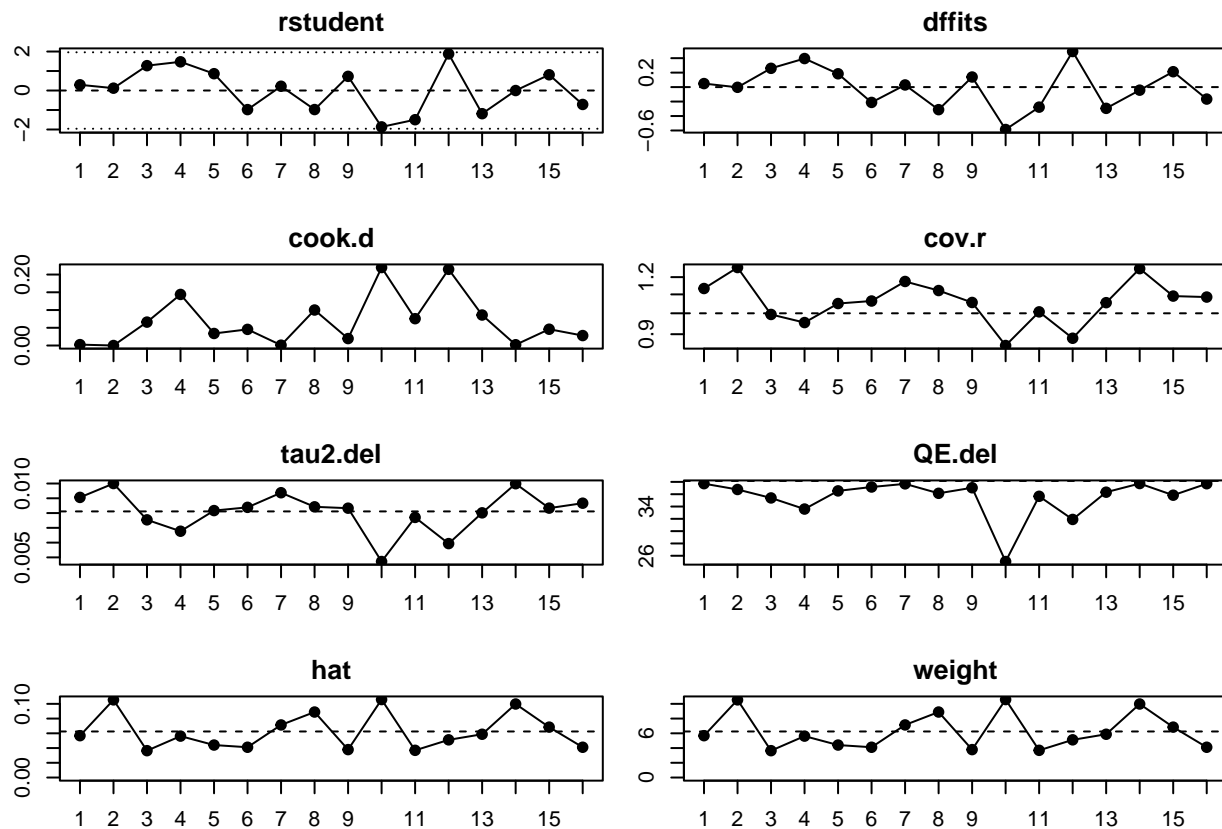


Figura 1. Diagnóstico de influencia. Estudios particularmente influyentes serían representados con un punto rojo. En este caso, no hay ningún estudio que se considere demasiado influyente, por lo que podemos estar tranquilos con nuestro meta-análisis.

1.3.3. *Forest plot* (diagrama de bosque)

Para hacer un diagrama de bosque (*forest plot*) resumiendo nuestro meta-análisis, solo tenemos que usar la función `forest`, usando como argumento el objeto al que asignamos los resultados de nuestro meta-análisis (`res`).

```
forest(res)
```

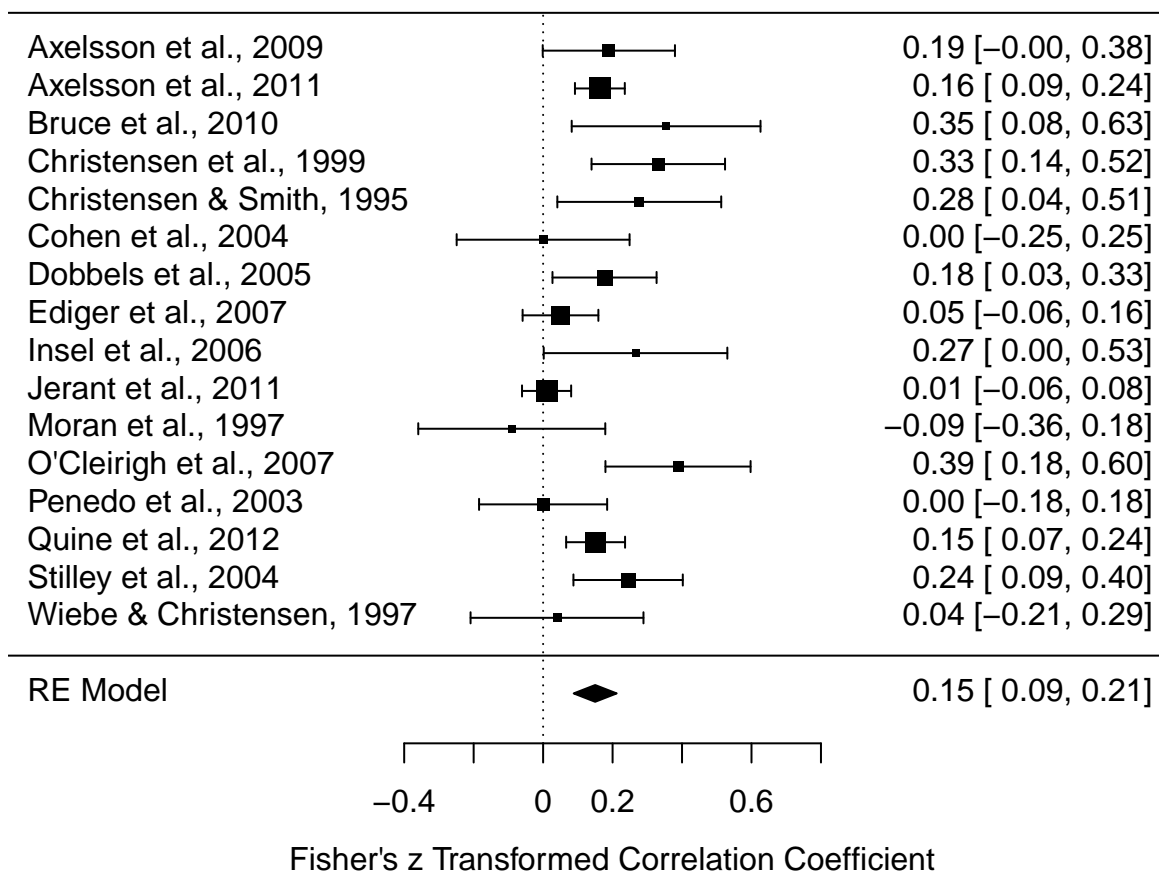


Figura 2. Forest plot básico de metafor. Para cada estudio meta-analizado, tenemos el efecto (correlación, en este caso en valores z de Fisher), así como sus intervalos de confianza entre paréntesis cuadrados. Esta misma información está representada gráficamente, con los cuadrados representando el efecto de cada estudio así como sus intervalos de confianza, y el tamaño de muestra (representado por el tamaño del cuadrado). Bajo estos resultados, tenemos nuestro meta-análisis, con el mismo formato en texto, pero representando el efecto y sus intervalos de confianza con un diamante.

Para una versión más completa y anotada, también usando el plot básico de `metafor`, pero representando coeficientes de correlación de Pearson (r) en vez de valores z , podemos agregar algunas opciones (explicadas [aquí](#)):

```
# forest plot con anotaciones adicionales
forest(res, cex = 0.8, xlim = c(-1.6, 1.6),
       attransf = transf.ztor,
       at = transf.rtoz(c(-0.4, -0.2, 0, 0.2, 0.4, 0.6)))
op <- par(cex = 0.8, font=2)
text(-1.6, 18, "Autor(es), Año", pos = 4)
text(1.6, 18, "Correlación [95% CI]", pos = 2)
par(op)
# añadir texto con valor Q, dfs, valor p y estadístico I^2
text(-1.6, -1, pos = 4, cex = 0.8,
     bquote(paste("RE Model: Q(", .(res$k - res$p), ") = ",
                  .(formatC(res$QE, digits=2, format="f")),
                  ", p ", .(scales::pvalue(res$pval)), "; ", I^2, " = ",
                  .(formatC(res$I2, digits=1, format="f")), "%"))))
```

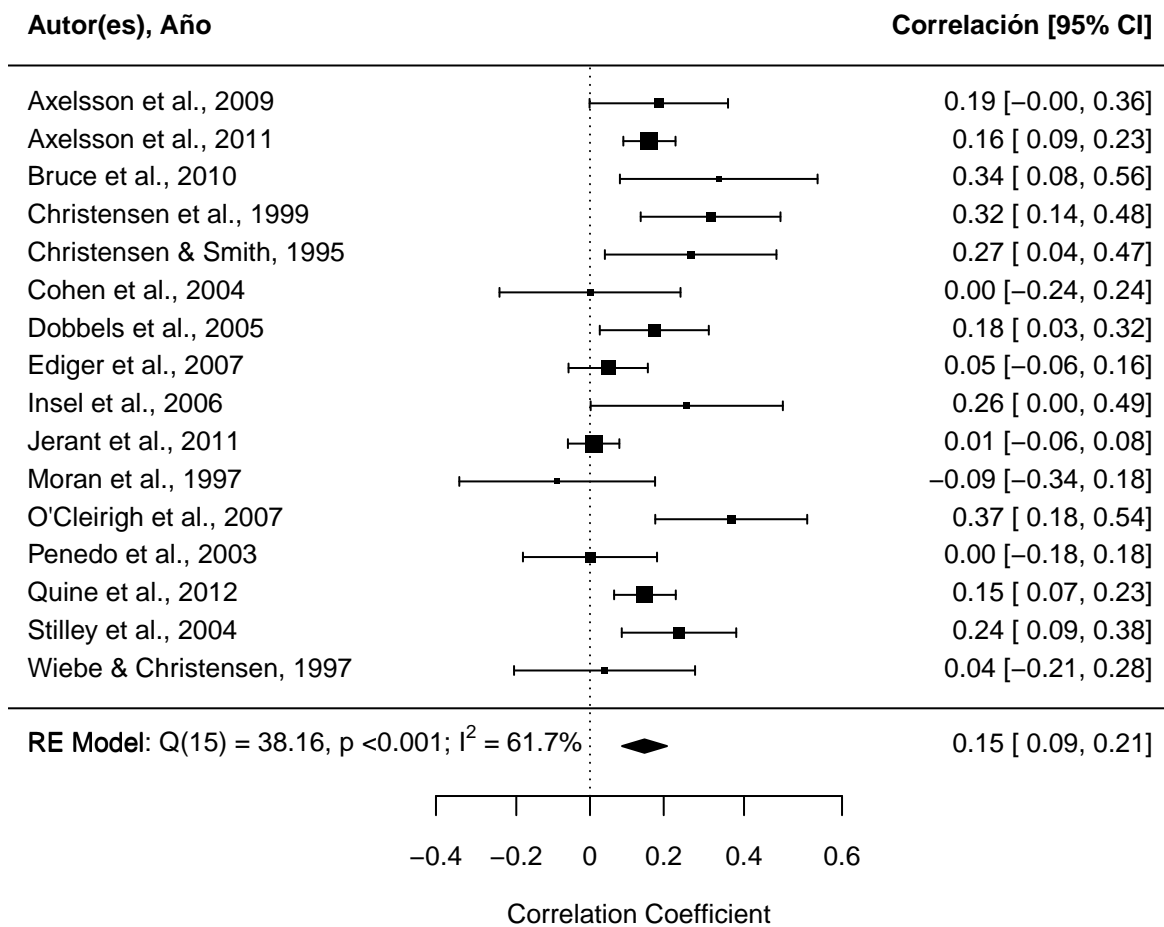


Figura 3. Forest plot anotado. En esta versión agregué algunos encabezados en español, así como estadísticos generales del modelo de meta-análisis.

O, para una incluso más sofisticada, se puede usar la función `viz_forest` del paquete `metaviz`, usando la variante `rain`.

```
library(metaviz)
viz_forest(res,
  study_labels = paste(dat$authors, dat$year, sep = ", "),
  xlab = "Correlación",
  variant = "rain",
  annotate_CI = TRUE,
  summary_label = "Resumen",
  text_size = 2.6)
```

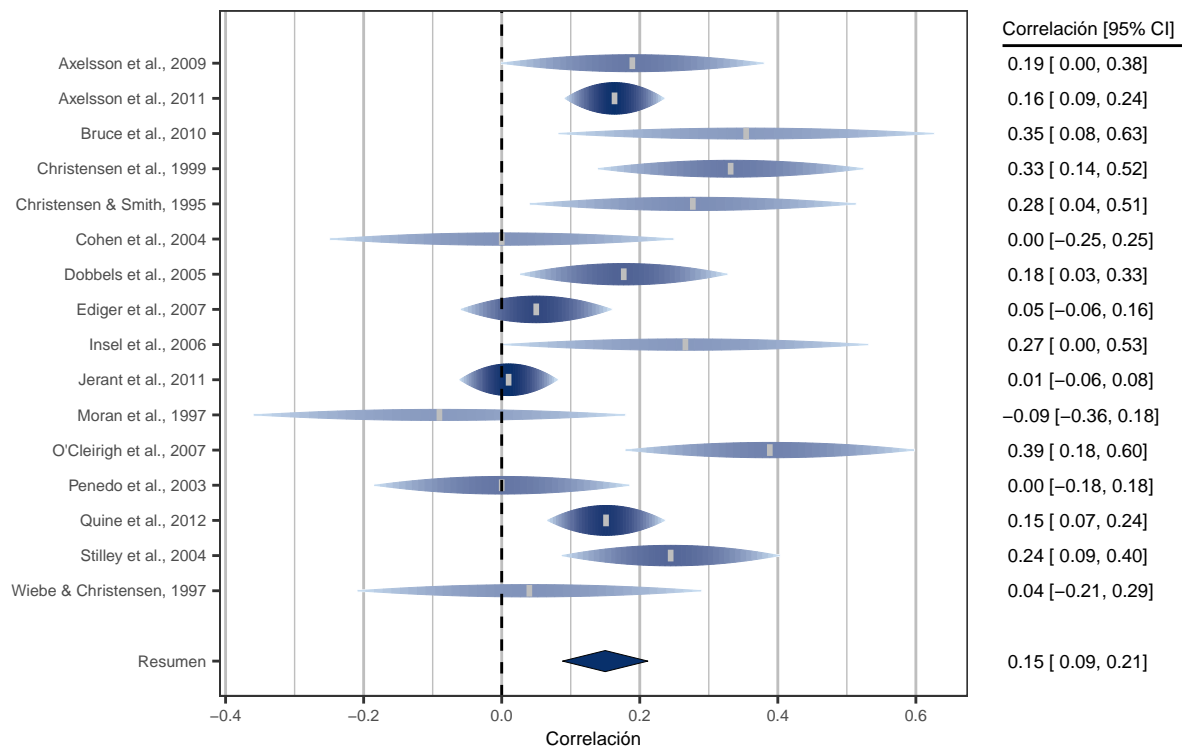



Figura 4. Forest plot creado con metaviz.

Referencias

- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). Identifying and Quantifying Heterogeneity. In *Introduction to Meta-Analysis* (pp. 107–125). Wiley. <https://doi.org/10.1002/9780470743386.ch16>
- Fisher, Z., & Tipton, E. (2015). Robumeta: An R-package for robust variance estimation in meta-analysis. *arXiv:1503.02220 [Stat]*. <http://arxiv.org/abs/1503.02220>
- Molloy, G. J., O'Carroll, R. E., & Ferguson, E. (2013). Conscientiousness and Medication Adherence: A Meta-analysis. *Annals of Behavioral Medicine*, 47(1), 92–101. <https://doi.org/10.1007/s12160-013-9524-4>
- Quintana, D. S. (2021). *How to perform a meta-analysis in R*. [Archivo de Vídeo]. YouTube. <https://youtu.be/IH4VZMTEZSc>.
- Viechtbauer, W. (2010). Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software*, 36(3). <https://doi.org/10.18637/jss.v036.i03>