METANÁLISE
META- ANÁLISIS
META-ANALYSIS

3

METANÁLISE
META- ANÁLISIS
META-ANALYSIS

Prof. Edson Zangiacomi Martinez

Faculdade de Medicina de Ribeirão Preto
Universidade de São Paulo (USP)
Ribeirão Preto, Brasil
edson@fmrp.usp.br

ESTIMANDO A MÉDIA E A VARIÂNCIA
ESTIMANDO EL PROMEDIO Y VARIANZA
ESTIMATING THE MEAN AND VARIANCE

BMC Medical Research Methodology

BioMed Central

Research article

Open Access

Estimating the mean and variance from the median, range, and the size of a sample

Stela Pudar Hozo[1], Benjamin Djulbegovic[2] and Iztok Hozo*[1]

Address: [1]Indiana University Northwest, Department of Mathematics, Gary, IN, 46408 USA and [2]Interdisciplinary Oncology Program, H. Lee Moffitt Cancer Center and Research Institute at the University of South Florida, Tampa, FL, USA

Email: Stela Pudar Hozo - spudar@iun.edu; Benjamin Djulbegovic - DjulbeBM@moffitt.usf.edu; Iztok Hozo* - ihozo@iun.edu

* Corresponding author

Published: 20 April 2005

BMC Medical Research Methodology 2005, **5**:13 doi:10.1186/1471-2288-5-13

Received: 26 September 2004
Accepted: 20 April 2005

This article is available from: http://www.biomedcentral.com/1471-2288/5/13

ESTIMANDO A MÉDIA E A VARIÂNCIA
ESTIMANDO EL PROMEDIO Y VARIANZA
ESTIMATING THE MEAN AND VARIANCE

m = median
a = the smallest value (minimum)
b = the largest value (maximum)
n = the size of the sample.

Mean: $$\bar{x} \approx \frac{a + 2m + b}{4} + \frac{a - 2m + b}{4n}$$

Standard deviation: $$s \approx \frac{b - a}{4}$$

ESTIMANDO A MÉDIA E A VARIÂNCIA
ESTIMANDO EL PROMEDIO Y VARIANZA
ESTIMATING THE MEAN AND VARIANCE

m = median = 49.7
a = minimum = 47.2
b = maximum = 52.0
n = the size of the sample = 20

50.1	49.6	49.1	48.1	51.7
49.8	50.6	51.8	48.8	49.9
52.0	48.7	47.2	50.4	51.0
50.7	49.3	49.5	49.5	48.5

Mean: $\bar{x} \approx \frac{a+2m+b}{4} + \frac{a-2m+b}{4n} = 49.6475$ (true mean = 49.815)

Desvio padrão: $s \approx \frac{b-a}{4} = 1.2$ (true sd = 1.263)

ESTIMANDO A MÉDIA E A VARIÂNCIA
ESTIMANDO EL PROMEDIO Y VARIANZA
ESTIMATING THE MEAN AND VARIANCE

Wan *et al. BMC Medical Research Methodology* 2014, **14**:135
http://www.biomedcentral.com/1471-2288/14/135

BMC
Medical Research Methodology

RESEARCH ARTICLE Open Access

Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range

Xiang Wan[1†], Wenqian Wang[2†], Jiming Liu[1] and Tiejun Tong[3*]

Abstract

Background: In systematic reviews and meta-analysis, researchers often pool the results of the sample mean and standard deviation from a set of similar clinical trials. A number of the trials, however, reported the study using the median, the minimum and maximum values, and/or the first and third quartiles. Hence, in order to combine results, one may have to estimate the sample mean and standard deviation for such trials.

Methods: In this paper, we propose to improve the existing literature in several directions. First, we show that the sample standard deviation estimation in Hozo et al.'s method (BMC Med Res Methodol 5:13, 2005) has some serious limitations and is always less satisfactory in practice. Inspired by this, we propose a new estimation method by

ESTIMANDO A MÉDIA E A VARIÂNCIA
ESTIMANDO EL PROMEDIO Y VARIANZA
ESTIMATING THE MEAN AND VARIANCE

m = median
a = the smallest value (minimum)
b = the largest value (maximum)
n = the size of the sample.

Standard deviation: $$s \approx \frac{b-a}{2\Phi^{-1}\left(\frac{n-0.375}{n+0.25}\right)}$$

ESTIMANDO A MÉDIA E A VARIÂNCIA
ESTIMANDO EL PROMEDIO Y VARIANZA
ESTIMATING THE MEAN AND VARIANCE

m = median = 49.7
a = minimum = 47.2
b = maximum = 52.0
n = the size of the sample = 20

50.1	49.6	49.1	48.1	51.7
49.8	50.6	51.8	48.8	49.9
52.0	48.7	47.2	50.4	51.0
50.7	49.3	49.5	49.5	48.5

Standard deviation: $$s \approx \frac{b-a}{2\Phi^{-1}\left(\frac{n-0.375}{n+0.25}\right)} = 1.2846$$ (true sd = 1.263)

```
> (b-a)/(2*qnorm((n-0.375)/(n+0.25)))
[1] 1.28463
```

ESTIMANDO A MÉDIA E A VARIÂNCIA
ESTIMANDO EL PROMEDIO Y VARIANZA
ESTIMATING THE MEAN AND VARIANCE

m = median
q_1 = the first quartile
q_3 = the third quartile
n = the size of the sample.

Mean: $$\bar{x} \approx \frac{q_1 + m + q_3}{3}$$

Standard deviation: $$s \approx \frac{q_3 - q_1}{2\Phi^{-1}\left(\frac{0.75n - 0.125}{n + 0.25}\right)}$$

ESTIMANDO A MÉDIA E A VARIÂNCIA
ESTIMANDO EL PROMEDIO Y VARIANZA
ESTIMATING THE MEAN AND VARIANCE

m = median = 49.7
q_1 = the first quartile = 49.025
q_3 = the third quartile = 50.625
n = the size of the sample = 20

50.1	49.6	49.1	48.1	51.7
49.8	50.6	51.8	48.8	49.9
52.0	48.7	47.2	50.4	51.0
50.7	49.3	49.5	49.5	48.5

Mean: $\bar{x} \approx \frac{q_1 + m + q_3}{3} = 49.783$ (true mean = 49.815)

Standard deviation: $s \approx \frac{q_3 - q_1}{2\Phi^{-1}\left(\frac{0.75n - 0.125}{n + 0.25}\right)} = 1.2765$ (true sd = 1.263)

ESTIMANDO A MÉDIA E A VARIÂNCIA
ESTIMANDO EL PROMEDIO Y VARIANZA
ESTIMATING THE MEAN AND VARIANCE

- Hozo, S. P., Djulbegovic, B., & Hozo, I. (2005). Estimating the mean and variance from the median, range, and the size of a sample. *BMC Medical Research Methodology*, *5*(1), 13.
- Luo, D., Wan, X., Liu, J., & Tong, T. (2018). Optimally estimating the sample mean from the sample size, median, mid-range, and/or mid-quartile range. *Statistical Methods in Medical Research*, *27*(6), 1785-1805.
  - Excel spread sheet is freely available at www.math.hkbu.edu.hk/~tongt/papers/optimalmean.xlsx
- Wan, X., Wang, W., Liu, J., & Tong, T. (2014). Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. *BMC Medical Research Methodology*, *14*(1), 135.

DIFERENÇA ENTRE MÉDIAS
DIFERENCIAS ENTRE PROMEDIOS
DIFFERENCE BETWEEN MEANS

Consider the data reported in Normand (1999) on the length of the hospital stay of stroke patients under specialized care and under conventional/routine non-specialist care:

```
> library(metafor)
> dat.normand1999
  study              source n1i m1i sd1i n2i m2i sd2i
1     1           Edinburgh 155  55   47 156  75   64
2     2      Orpington-Mild  31  27    7  32  29    4
3     3 Orpington-Moderate  75  64   17  71 119   29
4     4    Orpington-Severe  18  66   20  18 137   48
5     5       Montreal-Home   8  14    8  13  18   11
6     6  Montreal-Transfer  57  19    7  52  18    4
7     7           Newcastle  34  52   45  33  41   34
8     8                Umea 110  21   16 183  31   27
9     9             Uppsala  60  30   27  52  23   20
>
```

n1i : sample size, group 1
m1i : mean length of stay (in days), group 1
sd1i : standard deviation, group2
n2i : sample size, group 2
m2i : mean length of stay (in days), group 2
sd2i : standard deviation, group2

DIFERENÇA ENTRE MÉDIAS
DIFERENCIAS ENTRE PROMEDIOS
DIFFERENCE BETWEEN MEANS

"MD": The raw mean difference is equal to m1i - m2i
"SMD": The standardized mean dierence is equal to (m1i - m2i)/spi, where spi is the pooled standard deviation of the two groups

```
> meta <- escalc(measure="MD", m1i=m1i, sd1i=sd1i, n1i=n1i,
                               m2i=m2i, sd2i=sd2i, n2i=n2i, data=dat.normand1999)
> meta
  study              source n1i m1i sd1i n2i m2i sd2i       yi       vi
1     1           Edinburgh 155  55   47 156  75   64 -20.0000  40.5080
2     2     Orpington-Mild  31  27    7  32  29    4  -2.0000   2.0806
3     3 Orpington-Moderate  75  64   17  71 119   29 -55.0000  15.6984
4     4   Orpington-Severe  18  66   20  18 137   48 -71.0000 150.2222
5     5      Montreal-Home   8  14    8  13  18   11  -4.0000  17.3077
6     6  Montreal-Transfer  57  19    7  52  18    4   1.0000   1.1673
7     7           Newcastle  34  52   45  33  41   34  11.0000  94.5891
8     8               Umea 110  21   16 183  31   27 -10.0000   6.3109
9     9             Uppsala  60  30   27  52  23   20   7.0000  19.8423
>
```

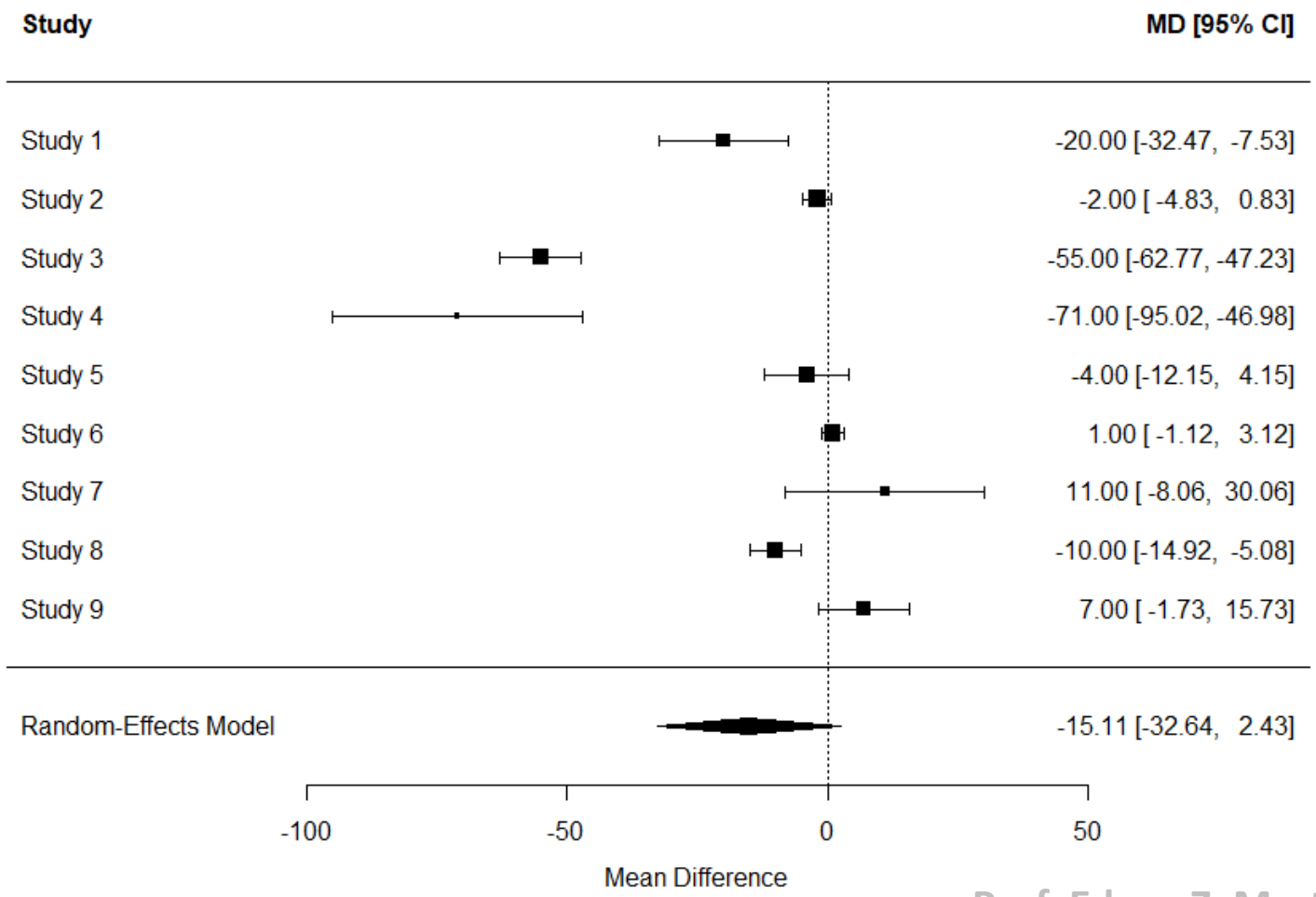


DIFERENÇA ENTRE MÉDIAS
DIFERENCIAS ENTRE PROMEDIOS
DIFFERENCE BETWEEN MEANS

```
> model1 <- rma(yi, vi, data=meta)
> model1

Random-Effects Model (k = 9; tau^2 estimator: REML)

tau^2 (estimated amount of total heterogeneity): 684.6462 (SE = 359.7541)
tau (square root of estimated tau^2 value):      26.1657
I^2 (total heterogeneity / total variability):   98.97%
H^2 (total variability / sampling variability):  97.21

Test for Heterogeneity:
Q(df = 8) = 238.9158, p-val < .0001

Model Results:

estimate      se     zval    pval     ci.lb   ci.ub
-15.1060  8.9466  -1.6885  0.0913  -32.6409  2.4289  .

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

>
```

DIFERENÇA ENTRE MÉDIAS
DIFERENCIAS ENTRE PROMEDIOS
DIFFERENCE BETWEEN MEANS

```
> forest(model1)
```



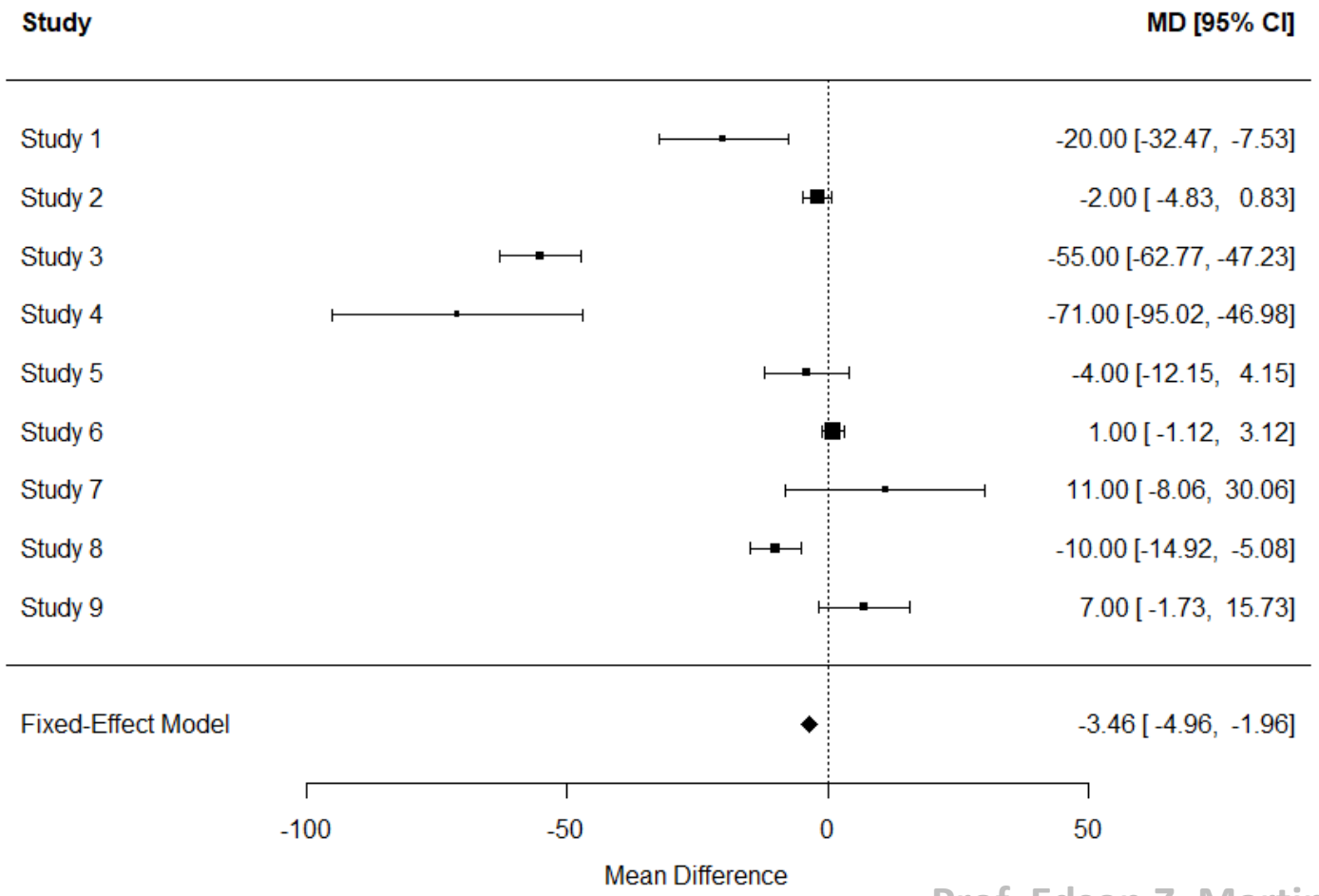


DIFERENÇA ENTRE MÉDIAS
DIFERENCIAS ENTRE PROMEDIOS
DIFFERENCE BETWEEN MEANS

```
> model2 <- rma(yi, vi, data=meta, method="FE")
> model2

Fixed-Effects Model (k = 9)

I^2 (total heterogeneity / total variability):   96.65%
H^2 (total variability / sampling variability):  29.86

Test for Heterogeneity:
Q(df = 8) = 238.9158, p-val < .0001

Model Results:

estimate      se     zval    pval    ci.lb    ci.ub
 -3.4636  0.7648  -4.5286  <.0001  -4.9626  -1.9646  ***

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

DIFERENÇA ENTRE MÉDIAS
DIFERENCIAS ENTRE PROMEDIOS
DIFFERENCE BETWEEN MEANS

```
> forest(model2)
```



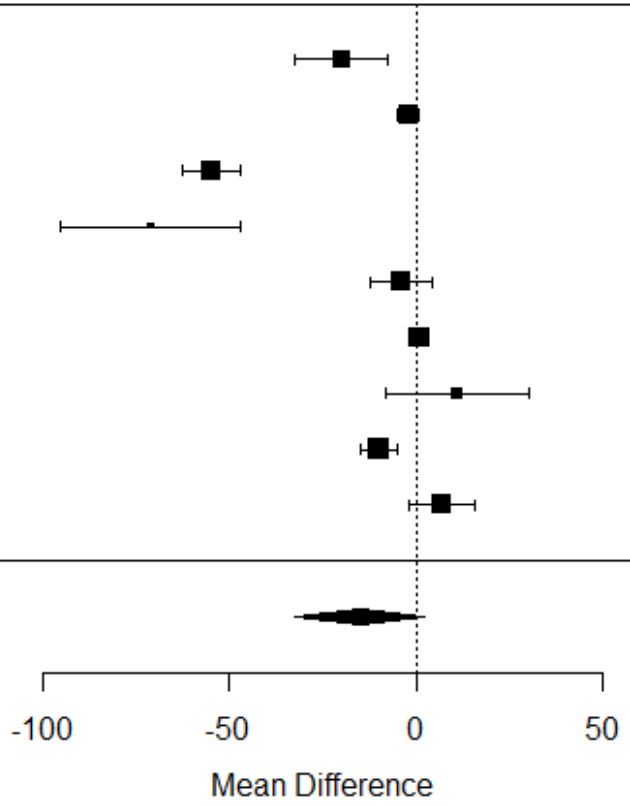


DIFERENÇA ENTRE MÉDIAS
DIFERENCIAS ENTRE PROMEDIOS
DIFFERENCE BETWEEN MEANS

```
model3 <- rma(yi, vi, data=meta, slab=source)
forest(model3,xlim=c(-350,160), mlab="", header=FALSE,
       ilab=cbind(dat.normand1999$m1i,
                  dat.normand1999$sd1i,
                  dat.normand1999$m2i,
                  dat.normand1999$sd2i),
       ilab.xpos=c(-220,-190,-150,-120))

text(c(-220,-190,-150,-120), 11,   c("Mean", "SD", "Mean", "SD"), font=2)
text(c(-205,-135),           11.8, c("Treated", "Control"), font=2)
text(-350,                   11,   "Author",  pos=4, font=2)
text(160,                    11,   "Diff. [95% CI]", pos=2, font=2)

text(-350, -1, pos=4, cex=0.8, bquote(paste("RE Model for all studies (Q = ",
     .(formatC(model3$QE, digits=2, format="f")), ", df = ", .(model3$k - model3$p),
     ", p = ", .(formatC(model2$QEp, digits=2, format="f")), "; ", I^2, " = ",
     .(formatC(model3$I2, digits=1, format="f")), "%)")))
```

DIFERENÇA ENTRE MÉDIAS
DIFERENCIAS ENTRE PROMEDIOS
DIFFERENCE BETWEEN MEANS

Author	Treated Mean	Treated SD	Control Mean	Control SD	Diff. [95% CI]
Edinburgh	55	47	75	64	-20.00 [-32.47, -7.53]
Orpington-Mild	27	7	29	4	-2.00 [-4.83, 0.83]
Orpington-Moderate	64	17	119	29	-55.00 [-62.77, -47.23]
Orpington-Severe	66	20	137	48	-71.00 [-95.02, -46.98]
Montreal-Home	14	8	18	11	-4.00 [-12.15, 4.15]
Montreal-Transfer	19	7	18	4	1.00 [-1.12, 3.12]
Newcastle	52	45	41	34	11.00 [-8.06, 30.06]
Umea	21	16	31	27	-10.00 [-14.92, -5.08]
Uppsala	30	27	23	20	7.00 [-1.73, 15.73]
RE Model for all studies (Q = 238.92, df = 8, p = 0.00; I^2 = 99.0%)					-15.11 [-32.64, 2.43]



METANÁLISE
META- ANÁLISIS
META-ANALYSIS

Prof. Edson Zangiacomi Martinez

Faculdade de Medicina de Ribeirão Preto
Universidade de São Paulo (USP)
Ribeirão Preto, Brasil
edson@fmrp.usp.br