Robust mediation analysis using the R package robmed

Andreas Alfons\textsuperscript{1}  Nüfer Y. Ateş\textsuperscript{2,3}  Patrick J.F. Groenen\textsuperscript{1}

\textsuperscript{1}Erasmus School of Economics, Erasmus University Rotterdam
\textsuperscript{2}Tilburg School of Economics and Management, Tilburg University
\textsuperscript{3}Faculty of Business Administration, Bilkent University

useR!, July 10, 2019
Effect of stimuli on behavior are mediated by various transformation processes

Simple mediation model:

\[ X \] independent variable
\[ Y \] dependent variable
\[ M \] mediator

Dependent variable is influenced by the independent variable through the mediator
Simple mediation model

→ $X$ affects $Y$ indirectly through $M$

Example: Task conflict ($M$) mediates the relationship between value diversity ($X$) and team commitment ($Y$)
Simple mediation model

Consider the following three regression models:

\[ M = i_1 + aX + e_1 \]
\[ Y = i_2 + c'X + e_2 \]
\[ Y = i_3 + bM + cX + e_3 \]

\[ \rightarrow \] Indirect effect \( ab \)
\[ \rightarrow \] Direct effect \( c \)
\[ \rightarrow \] Total effect \( c' = ab + c \)
Simple mediation model

Consider the following three regression models:

\[ M = i_1 + aX + e_1 \]
\[ Y = i_2 + c'X + e_2 \]
\[ Y = i_3 + bM + cX + e_3 \]

→ Indirect effect \( ab \)
→ Direct effect \( c \)
→ Total effect \( c' = ab + c \)
Simple mediation model

Estimate the following two regression models:

\[ M = i_M + aX + e_M \]
\[ Y = i_Y + bM + cX + e_Y \]

→ Indirect effect \( ab \)
→ Direct effect \( c \)
→ Total effect \( c' = ab + c \)
Estimation of the mediation model

Typically, a series of ordinary least squares (OLS) regressions is used to estimate the mediation model.

Sampling distribution of the estimator \( \hat{ab} \) of the indirect effect is asymmetric.

Bootstrap is typically used to construct confidence intervals (Preacher and Hayes, 2004, 2008).
Illustration: Estimation of the mediation model

\[ X \]

\[ \text{Equation}\]

\[ \begin{align*}
&\hat{Y}^\prime = i_3 + \hat{b} \cdot M \\
&\hat{Y}^\prime = i_2 + \hat{c}^\prime \\
&\hat{Y}^\prime = i_1 + \hat{a}
\end{align*} \]

Without outlier

Including outlier

<table>
<thead>
<tr>
<th>Without outlier</th>
<th>Including outlier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outlier</td>
<td>Outlier</td>
</tr>
<tr>
<td>Standard method</td>
<td>ROBMED</td>
</tr>
</tbody>
</table>

\[ \begin{align*}
X & 0 \quad \hat{M} = i_1 \\
1 & \hat{M} = i_1 + \hat{a}
\end{align*} \]

\[ \begin{align*}
Y^\prime & \Rightarrow \hat{a} \\
Y^\prime & \Rightarrow \hat{b} \cdot M + \hat{c}
\end{align*} \]
Standard bootstrap test

→ OLS and the bootstrap are easily distorted by deviations from the usual normality assumptions in regression
  ■ Heavily tailed errors
  ■ Outliers
  ■ . . .

→ Robust alternatives are needed to ensure reliable results in empirical research
MM-estimator of regression

→ Replace least squares loss with a more robust loss function

→ Can be seen as weighted least squares estimator (WLS) with outlyingness weights derived from data

→ See Yohai (1987) and Salibian-Barrera and Yohai (2006)
Linear regression: Loss function and weights
Fast and robust bootstrap

→ Not necessary to do exhaustive search for optimal weights on each bootstrap sample

→ On each bootstrap sample:
  1. Compute WLS estimate with outlyingness weights obtained from original sample
  2. Apply linear correction of estimates to account for additional uncertainty from obtaining the weights

Robust mediation analysis

1. Estimate the mediation model via series of MM-regressions
2. Compute asymmetric confidence intervals for the indirect effect via the fast and robust bootstrap

Software

→ R package robmed available on CRAN:
   https://CRAN.R-project.org/package=robmed

→ R extension for SPSS under development:
   https://github.com/aalfons/ROBMED-SPSS
Illustrative hypotheses and data

Replicate known theory from organizational research on new data from a business strategy game played by students

Illustrative hypothesis Task conflict ($M$) mediates the relationship between value diversity ($X$) and team commitment ($Y$)

R> library("robmed")
R> data("BSG2014")
Empirical example
Empirical example

```r
R> # seed of random number generator
R> seed <- 20150601
R>
R> # perform standard method and proposed robust method
R> set.seed(seed, sample.kind = "Rounding") # mimic R 3.5.2
R> standard <- test_mediation(BSG2014,
+             x = "ValueDiversity",
+             y = "TeamCommitment",
+             m = "TaskConflict",
+             robust = FALSE)
R> set.seed(seed, sample.kind = "Rounding") # mimic R 3.5.2
R> robust <- test_mediation(BSG2014,
+             x = "ValueDiversity",
+             y = "TeamCommitment",
+             m = "TaskConflict")
```
Empirical example: Standard method (I)

R> summary(standard)

Bootstrap test for indirect effect via regression

x = ValueDiversity
y = TeamCommitment
m = TaskConflict

Sample size: 89
---
Outcome variable: TaskConflict

Coefficients:

|            | Data | Boot | Std. Error | z value | Pr(>|z|) |
|------------|------|------|------------|---------|---------|
| (Intercept)| 1.5007 | 1.4940 | 0.2265 | 6.596 | 4.23e-11 *** |
| ValueDiversity | 0.1552  | 0.1589 | 0.1266 | 1.255 | 0.209 |

Residual standard error: 0.3908 on 87 degrees of freedom
Multiple R-squared: 0.01857, Adjusted R-squared: 0.007289
F-statistic: 1.646 on 1 and 87 DF, p-value: 0.2029
---
Empirical example: Standard method (II)

Outcome variable: TeamCommitment

Coefficients:

|                          | Data  | Boot  | Std. Error | z value | Pr(>|z|) |
|--------------------------|-------|-------|------------|---------|----------|
| (Intercept)              | 4.49846 | 4.50162 | 0.32963 | 13.657  | <2e-16 *** |
| TaskConflict             | -0.36412 | -0.37036 | 0.16021 | -2.312  | 0.0208 *  |
| ValueDiversity           | -0.02088 | -0.01636 | 0.14524 | -0.113  | 0.9103 |

Residual standard error: 0.4296 on 86 degrees of freedom  
Multiple R-squared:  0.1031, Adjusted R-squared:  0.08227  
F-statistic: 4.944 on 2 and 86 DF,  p-value: 0.009279

---

Total effect of x on y:

|                          | Data       | Boot       | Std. Error | z value | Pr(>|z|) |
|--------------------------|------------|------------|------------|---------|----------|
| ValueDiversity           | -0.07738   | -0.07609   | 0.15855    | -0.48   | 0.631    |

Direct effect of x on y:

|                          | Data       | Boot       | Std. Error | z value | Pr(>|z|) |
|--------------------------|------------|------------|------------|---------|----------|
| ValueDiversity           | -0.02088   | -0.01636   | 0.14524    | -0.113  | 0.91     |
Empirical example: Standard method (III)

Indirect effect of x on y:

<table>
<thead>
<tr>
<th>TaskConflict</th>
<th>Data</th>
<th>Boot</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.0565</td>
<td>-0.05973</td>
<td>-0.2083</td>
<td>0.0251</td>
</tr>
</tbody>
</table>

---

Level of confidence: 95%

Number of bootstrap replicates: 5000

---

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

Function p_value() allows to extract the smallest $\alpha$ for which the $(1 - \alpha) \cdot 100\%$ confidence interval does not contain 0:

R> p_value(standard)
[1] 0.1584
Empirical example: ROBMED (I)

R> summary(robust)
Robust bootstrap test for indirect effect via regression

x = ValueDiversity
y = TeamCommitment
m = TaskConflict

Sample size: 89
---

Outcome variable: TaskConflict

Coefficients:

| Data | Boot | Std. Error | z value | Pr(>|z|) |
|------|------|------------|---------|----------|
| (Intercept) | 1.1182 | 1.1162 | 0.1778 | 6.279 | 3.42e-10 *** |
| ValueDiversity | 0.3197 | 0.3211 | 0.1071 | 2.998 | 0.00272 ** |

Robust residual standard error: 0.3033 on 87 degrees of freedom
Robust R-squared: 0.1181, Adjusted robust R-squared: 0.108
Robust F-statistic: 9.113 on 1 and Inf DF, p-value: 0.002539
---
Empirical example: ROBMED (II)

Outcome variable: TeamCommitment

Coefficients:

|                  | Data     | Boot     | Std. Error | z value | Pr(>|z|) |
|------------------|----------|----------|------------|---------|---------|
| (Intercept)      | 4.33385  | 4.34430  | 0.34088    | 12.744  | <2e-16  *** |
| TaskConflict     | -0.33659 | -0.34353 | 0.17761    | -1.934  | 0.0531 . |
| ValueDiversity   | 0.06523  | 0.06507  | 0.18594    | 0.350   | 0.7264  |

Robust residual standard error: 0.3899 on 86 degrees of freedom
Robust R-squared: 0.08994, Adjusted robust R-squared: 0.06878
Robust F-statistic: 1.497 on 2 and Inf DF, p-value: 0.2239

---

Total effect of x on y:

|                  | Data     | Boot     | Std. Error | z value | Pr(>|z|) |
|------------------|----------|----------|------------|---------|---------|
| ValueDiversity   | -0.04239 | -0.04501 | 0.18671    | -0.241  | 0.81    |

Direct effect of x on y:

|                  | Data     | Boot     | Std. Error | z value | Pr(>|z|) |
|------------------|----------|----------|------------|---------|---------|
| ValueDiversity   | 0.06523  | 0.06507  | 0.18594    | 0.35    | 0.726   |
Empirical example: ROBMED (III)

Indirect effect of x on y:

<table>
<thead>
<tr>
<th>TaskConflict</th>
<th>Data</th>
<th>Boot</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.1076</td>
<td>-0.1101</td>
<td>-0.294</td>
<td>-0.01042</td>
</tr>
</tbody>
</table>

---

Level of confidence: 95%

Number of bootstrap replicates: 5000

---

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

Function `p_value()` allows to extract the smallest $\alpha$ for which the $(1 - \alpha) \cdot 100\%$ confidence interval does not contain 0:

R> p_value(robust)
[1] 0.0271
Empirical example

R> plot_mediation(list(Standard = standard, ROBMED = robust), + method = "density")
R package robmed: Further details

- The usual coef(), confint(), plot(), print() and summary() methods

- Other techniques: based on winsorization (Zu and Yuan, 2010) or median regression (Yuan and MacKinnon, 2014)

- Multiple mediators or additional covariates
Conclusions and discussion

Conclusions

→ Standard bootstrap test for mediation analysis is easily distorted by deviations from the usual normality assumptions
→ Fast and robust bootstrap allows for much more reliable empirical results than other methods
→ R package robmed available on CRAN

Future work

→ Binary/nominal/ordinal dependent variable or mediators
→ Mediated moderation, moderated mediation, ...


Empirical example: Further analysis
Technical details: MM-estimator

\[
\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} \rho \left( \frac{r_i(\beta)}{\hat{\sigma}} \right),
\]

where \( \hat{\sigma} \) is a robust scale estimate of the residuals from a highly robust initial regression estimator.

Loss function \( \rho \) can be tuned for high efficiency.

High robustness is inherited from initial residual scale \( \hat{\sigma} \).
Technical details: MM-estimator

The MM-estimate can be written as

\[
\hat{\beta} = \left( \sum_{i=1}^{n} w_i x_i x_i^T \right)^{-1} \sum_{i=1}^{n} w_i x_i y_i
\]

with

\[
w_i = \rho'(r_i(\beta)/\hat{\sigma}) \frac{r_i(\beta)/\hat{\sigma}}{r_i(\beta)/\hat{\sigma}}, \quad i = 1, \ldots, n
\]

→ Weighted least squares estimator (WLS) with outlyingness weights derived from data