PSYCGR01: Statistics

Moderation and mediation worksheet

Maarten Speekenbrink *UCL Experimental Psychology*
November, 2016

1 General Exercises

1. Wuensch *et al.* (2002) conducted a study on the relation between misanthropy, idealism, and attitudes towards animal rights. Their hypothesis was that when evaluating the ethical status of an action that harms a nonhuman animal, one might weight the benefit to humankind against the cost of the harm done to the animal. To the extent that one does not like humans (i.e., the extent of misanthropy), one will not be likely to think that benefits to humans can justify harm to animals. However, they also thought that this relation would be moderated by people's level of idealism: misanthropy would be less strongly related to support for animal rights among idealists (who tend not to do cost-benefit analysis) than among non-idealists. Using (centered) measures of support for animal rights (`AR`), misanthropy (`MT`) and idealism (`Id`), they estimated the following model

$$\texttt{AR}_i = .30 \times \texttt{MT}_i + .07 \times \texttt{Id}_i - .15 \times (\texttt{Id} \times \texttt{MT})_i + e_i$$

Assume that all parameters were significant.

(a) What is the predicted level of support for animal rights for someone with a score of 0 on misanthropy and idealism? What does this prediction reflect?

Answer: The predicted level is 0. As all variables are centered, the prediction is an average level of support when both idealism and misanthropy are at an average level.

(b) What is the predicted relation between support for animal rights and misanthropy for someone with a score of -1 on idealism?

Answer:

$$\begin{aligned}\texttt{AR}_i &= .30 \times \texttt{MT}_i + .07 \times (-1) - .15 \times ((-1) \times \texttt{MT})_i + e_i \\ &= -0.07 + .45 \times \texttt{MT}_i + e_i\end{aligned}$$

(c) What is the predicted relation between support for animal rights and misanthropy for someone with a score of 1 on idealism?

Answer:

$$\texttt{AR}_i = 0.07 + .15 \times \texttt{MT}_i + e_i$$

(d) What is the predicted relation between support for animal rights and idealism for someone with a score of -1 on misanthropy?

Answer:

$$\texttt{AR}_i = -.30 + .22 \times \texttt{Id}_i + e_i$$

(e) What is the predicted relation between support for animal rights and idealism for someone with a score of 1 on misanthropy?

Answer:

$$\texttt{AR}_i = .30 - .08 \times \texttt{Id}_i + e_i$$

(f) From these predictions, do you think the researchers' hypothesis is supported?

Answer: Yes. The interaction is significant and you can see that the relation between misanthropy and support for animal rights changes as a function of idealism. Note that there is no indication of causality, e.g., from the results, one might just as well say that the relation between support for animal rights and idealism is moderated by misanthropy.

2 SPSS Exercises

1. Open the dataset `Geller.sav`. This is the same dataset you analysed in the "Regression models" worksheet. For a description of the variables, you can look there. Recall that the hypothesis of Geller *et al.* was that the extent to which perceptions of weight and shape determine self-esteem affects the development of eating disorders. With that in mind, it seems plausible that for people who have a high level of self-esteem, the `SAWBS` measure might be largely irrelevant when predicting eating disorders. People with low self-esteem on the other hand, might develop eating disorders when perceptions of weight and shape play a defining role in their (low) self-esteem. This prediction amounts to an interaction between `SAWBS` and self-esteem (measured by the `RSES`), or in other words, that the relation between `SAWBS` and level of eating disorders is moderated by general self-esteem. Let's assess whether this prediction holds, using the `HIQ` measure of the level of eating disorders.

   (a) Estimate the parameters of MODEL C:

   $$\texttt{HIQ}_i = \beta_0 + \beta_1 \texttt{RSES}_i + \beta_2 \texttt{SAWBS}_i + \epsilon_i$$

   and compute the SSE.

```
> library(foreign)
> geller <- as.data.frame(read.spss("Geller.sav"))
> MODEL_C <- lm(hiq ~ rses + sawbs,data=geller)
> summary(MODEL_C)
```

```
Call:
lm(formula = hiq ~ rses + sawbs, data = geller)

Residuals:
     Min       1Q   Median       3Q      Max
-14.3039  -4.6371   0.2091   4.4005  15.2705

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 29.29943    3.74596   7.822 1.69e-11 ***
rses        -0.48769    0.08999  -5.419 6.01e-07 ***
sawbs        0.08780    0.01587   5.533 3.76e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.223 on 81 degrees of freedom
Multiple R-squared:  0.5436,Adjusted R-squared:  0.5323
F-statistic: 48.24 on 2 and 81 DF,  p-value: 1.599e-14

> SSE_C <- sum(residuals(MODEL_C)^2)
> SSE_C

[1] 3137.012
```

Answer:

$$\texttt{HIQ}_i = 29.299 - 0.488 \times \texttt{RSES}_i + 0.088 \times \texttt{SAWBS}_i + \epsilon_i$$

and SSE(C) = 3137.012

(b) Estimate the parameters of MODEL A1:

$$\texttt{HIQ}_i = \beta_0 + \beta_1\texttt{RSES}_i + \beta_2\texttt{SAWBS}_i + \beta_3(\texttt{RSES} \times \texttt{SAWBS})_i + \epsilon_i$$

and compute the SSE.

```
> MODEL_A1 <- lm(hiq ~ rses + sawbs + rses:sawbs,data=geller)
> summary(MODEL_A1)


Call:
lm(formula = hiq ~ rses + sawbs + rses:sawbs, data = geller)

Residuals:
     Min       1Q   Median       3Q      Max
-14.2729  -4.4008   0.1306   4.4596  15.1809

Coefficients:
```

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 31.110904   5.451364   5.707 1.87e-07 ***
rses        -0.536468   0.139465  -3.847 0.000239 ***
sawbs        0.062231   0.057902   1.075 0.285711
rses:sawbs   0.000734   0.001598   0.459 0.647181
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.254 on 80 degrees of freedom
Multiple R-squared:  0.5448,Adjusted R-squared:  0.5277
F-statistic: 31.91 on 3 and 80 DF,  p-value: 1.144e-13

> SSE_A1 <- sum(residuals(MODEL_A1)^2)
> SSE_A1

[1] 3128.757
```

Answer:

$$\texttt{HIQ}_i = 31.111 - 0.536 \times \texttt{RSES}_i + 0.062 \times \texttt{SAWBS}_i + 0.001(\texttt{RSES} \times \texttt{SAWBS})_i + e_i$$

and SSE(A1) $= 3128.757$.

(c) What is the null-hypothesis you test when comparing MODEL A1 to MODEL C? Test this null-hypothesis.

```
> anova(MODEL_C,MODEL_A1)

Analysis of Variance Table

Model 1: hiq ~ rses + sawbs
Model 2: hiq ~ rses + sawbs + rses:sawbs
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     81 3137.0
2     80 3128.8  1    8.2546 0.2111 0.6472
```

Answer: $H_0 : \beta_3 = 0$, $F = 0.211$, $P(F_{1,80} \geq 0.221) = .647$, PRE $= 0.00263$, so there is no reason to reject the null hypothesis that there is no interaction.

(d) Compute the correlations between the three predictors in MODEL A1. Assess the level of multicollinearity.

```
> geller$rses_sawbs <- geller$rses*geller$sawbs
> cor(geller[,c("rses","sawbs","rses_sawbs")])

                  rses      sawbs  rses_sawbs
rses        1.00000000 -0.3782735 -0.04348224
sawbs      -0.37827350  1.0000000  0.90549922
rses_sawbs -0.04348224  0.9054992  1.00000000
```

```
> ## an alternative way is to get the predictors of
> ## MODEL_A1 via the model.matrix function:
> preds <- model.matrix(MODEL_A1)
> head(preds)

  (Intercept)     rses     sawbs rses:sawbs
1           1 32.22844  56.93904  1835.0564
2           1 32.39728 117.16003  3795.6668
3           1 44.93232  48.46506  2177.6473
4           1 38.63428  66.88492  2584.0507
5           1 35.48696  88.12559  3127.3093
6           1 29.77724 -22.88933  -681.5811

> ## Note that the model.matrix has an (intercept) predictor
> ## which always equals 1. This is an artifical predictor
> ## such that the slope of this predictor equals the intercept
> ## (a slope for a predictor which always equals 1 is exactly
> ## like an intercept!)
> cor(preds[,2:4])

                  rses      sawbs  rses:sawbs
rses        1.00000000 -0.3782735 -0.04348224
sawbs      -0.37827350  1.0000000  0.90549922
rses:sawbs -0.04348224  0.9054992  1.00000000

> ## a better way to assess multicollinearity
> library(car)
> vif(MODEL_A1)

      rses      sawbs rses:sawbs
  2.775567  15.384581  13.208161

> # or tolerance
> 1/vif(MODEL_A1)

      rses      sawbs rses:sawbs
0.36028671 0.06500014 0.07571076
```

Answer: `SAWBS` and `RSES` have a moderate (but significant) negative correlation of $r = -.378$. The main problem is the correlation between `SAWBS` and the interaction term, which is $r = .905$.

(e) There has been a long-standing debate on the merits of centering predictors in interactive regression models. Let's see what happens when we center the predictors first. Compute two new predictors as

$\texttt{RSES0}_i = \texttt{RSES}_i - \overline{\texttt{RSES}}$

$\texttt{SAWBS0}_i = \texttt{SAWBS}_i - \overline{\texttt{SAWBS}}$

and estimate the MODEL A2:

$$\texttt{HIQ}_i = \beta_0 + \beta_1\texttt{RSES0}_i + \beta_2\texttt{SAWBS0}_i + \beta_3(\texttt{RSES0} \times \texttt{SAWBS0})_i + \epsilon_i$$

```
> geller$rses0 <- scale(geller$rses,scale=FALSE)
> geller$sawbs0 <- scale(geller$sawbs,scale=FALSE)
> MODEL_A2 <- lm(hiq ~ rses0*sawbs0,data=geller)
> summary(MODEL_A2)


Call:
lm(formula = hiq ~ rses0 * sawbs0, data = geller)

Residuals:
     Min       1Q   Median       3Q      Max
-14.2729  -4.4008   0.1306   4.4596  15.1809

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  16.904612   0.719333  23.500  < 2e-16 ***
rses0        -0.494188   0.091531  -5.399 6.68e-07 ***
sawbs0        0.088656   0.016055   5.522 4.03e-07 ***
rses0:sawbs0  0.000734   0.001598   0.459    0.647
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.254 on 80 degrees of freedom
Multiple R-squared:  0.5448,Adjusted R-squared:  0.5277
F-statistic: 31.91 on 3 and 80 DF,  p-value: 1.144e-13
```

Answer:

$$\texttt{HIQ}_i = 16.905 - 0.494 \times \texttt{RSES0}_i + 0.089 \times \texttt{SAWBS0}_i + 0.001 \times (\texttt{RSES0} \times \texttt{SAWBS0})_i + \epsilon_i$$

Note that the parameter value for the interaction stays the same, but those of the individual predictors change. By centering, the predictors will have a value of 0 when they are at their mean values, and the slopes of the predictors in an interaction model are the slopes when the other predictors are 0. For example, the slope of the centered predictor $\texttt{RSES0}_i$ is the slope when $\texttt{SAWBS0}_0 = 0$, so at $\overline{\texttt{SAWBS}}$.

(f) Compare the results of MODEL A2 (significance of individual predictors and overall model fit) to those obtained for MODEL A1. Are there differences? If so, can you explain these?

Answer: There is no difference for the interaction effect, but now the individual effect of `SAWBS` is significant. By centering, we are now estimating the "simple" effect of `SAWBS` when `RSES` is at its mean. This is different from the effect of `SAWBS` when `RSES` $= 0$. Note that the R^2 of this model is the same as for the uncentered model. So this model will make the same predictions (and is "in a deep sense" the same) as the uncentered model.

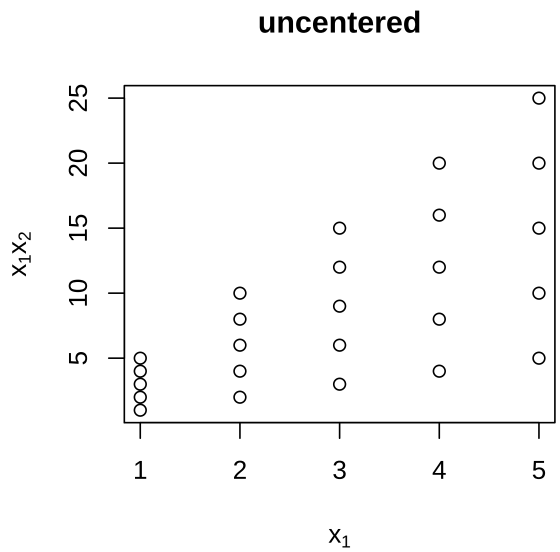
(g) Compute the correlations between the three predictors in MODEL A2 and assess the level of multicollinearity. Are there differences with model A1 in this respect?

```
> cor(model.matrix(MODEL_A2)[,2:4])

                  rses0     sawbs0 rses0:sawbs0
rses0         1.0000000 -0.3782735    0.2102434
sawbs0       -0.3782735  1.0000000   -0.1841580
rses0:sawbs0  0.2102434 -0.1841580    1.0000000

> # tolerance
> 1/vif(MODEL_A2)

       rses0       sawbs0 rses0:sawbs0
   0.8364523    0.8454558    0.9430226
```

Answer: The correlation between `SAWBS` and `RSES` has not changed, but the correlation between `SAWBS` and the interaction term is much smaller. To see why this is, let's take a simple example of two predictors, X_1 and X_2, each with values from 1 to 5. We can compute the interaction $X_1 \times X_2$ as in the table below

	X_1				
X_2	1	2	3	4	5
1	1	2	3	4	5
2	2	4	6	8	10
3	3	6	9	12	15
4	4	8	12	16	20
5	5	10	15	20	25

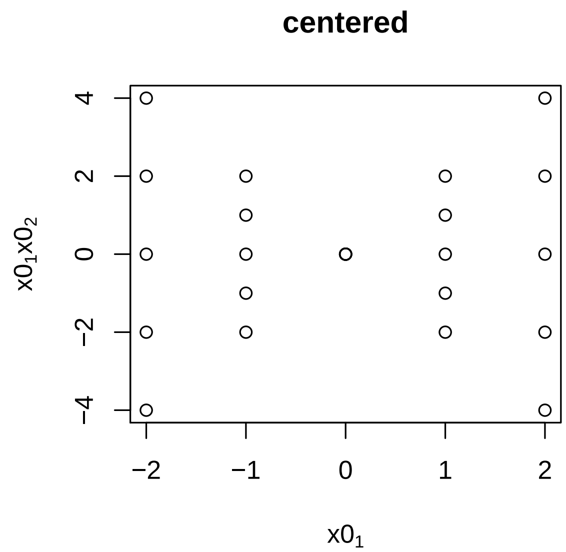
If we plot the values of X_1 against the interaction $X_1 X_2$ we get the following



Which shows that there is a positive relation between X_1 and the interaction X_1X_2. If we first center X_1 and X_2 (let's call these centered variables $X0_1$ and $X0_2$ respectively) and then compute the interaction, we get the values in the table below:

	$X0_1$				
$X0_2$	-2	-1	0	1	2
-2	4	2	0	-2	-4
-1	2	1	0	-1	-2
0	0	0	0	0	0
1	-2	-1	0	1	2
2	-4	-2	0	2	4

If we plot the values of $X0_1$ against the interaction $X0_1 \times X0_2$ we get the following



in which there is no evidence of a positive (or negative) correlation. The main reason why centering can reduce multicollinearity is that centering makes sure that for each predictor there are both positive and negative values. When we then compute a product of these variables, two negative values become positive in the product, while one negative value and one positive value become negative in the product. So, if a centered predictor has a negative value, the product (interaction) predictor can be both positive and negative (depending on the value of the other predictor).

2. This exercise is also based on the `Geller.sav` dataset. Let's attempt to investigate the researchers' hypothesis more directly, which we can view as one of mediation. Namely, it is assumed that perceptions of weight and shape influence self-esteem, and the extent to which they do determines the level of eating disorders. In other words, the effect of weight and shape perceptions on eating disorders is assumed to be mediated by self-esteem.

   (a) First, calculate an aggregated measure of weight and shape perceptions as the mean of these two values (for each participant) and call this measure `WSPercep`.

   ```
   > geller$wspercep <- (geller$wtpercep + geller$shpercep)/2
   > # alternatively, you can use
   > # geller£wspercep <- rowMeans(geller[,c("wtpercep","shpercep")])
   ```

   (b) Estimate the parameters of MODEL 1:

$$\texttt{HIQ}_i = \beta_0 + \beta_1 \texttt{WSPercep}_i + \epsilon_i$$

```
> MODEL_1 <- lm(hiq ~ wspercep,data=geller)
> summary(MDOEL_1)

Error in summary(MDOEL_1):  object 'MDOEL_1' not found
```

Answer:

$$\texttt{HIQ}_i = 41.69 - 6.73 \times \texttt{WSPercep}_i + e_i$$

(c) Estimate the parameters of MODEL 2:

$$\texttt{RSES}_i = \beta_0 + \beta_1 \texttt{WSPercep}_i + \epsilon_i$$

```
> MODEL_2 <- lm(rses ~ wspercep,data=geller)
> summary(MODEL_2)


Call:
lm(formula = rses ~ wspercep, data = geller)

Residuals:
     Min       1Q   Median       3Q      Max
-17.4672  -4.2555   0.2929   4.1053  19.2571

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  19.7249     3.6141   5.458 5.01e-07 ***
wspercep      4.3987     0.9524   4.618 1.42e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.349 on 82 degrees of freedom
Multiple R-squared:  0.2064,Adjusted R-squared:  0.1967
F-statistic: 21.33 on 1 and 82 DF,  p-value: 1.417e-05
```

Answer:

$$\texttt{RSES}_i = 19.72 + 4.40 \times \texttt{WSPercep}_i + e_i$$

(d) Estimate the parameters of MODEL 3:

$$\texttt{HIQ}_i = \beta_0 + \beta_1 \texttt{WSPercep}_i + \beta_2 \texttt{RSES}_i + \epsilon_i$$

```
> MODEL_3 <- lm(hiq ~ wspercep + rses,data=geller)
> summary(MODEL_3)


Call:
lm(formula = hiq ~ wspercep + rses, data = geller)

Residuals:
     Min       1Q   Median       3Q      Max 
-11.7633  -4.1639   0.4411   4.3885  16.3544 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  50.6493     3.6454  13.894  < 2e-16 ***
wspercep     -4.7309     0.9236  -5.122 2.01e-06 ***
rses         -0.4540     0.0954  -4.759 8.37e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.349 on 81 degrees of freedom
Multiple R-squared:  0.525,Adjusted R-squared:  0.5132 
F-statistic: 44.76 on 2 and 81 DF,  p-value: 8.085e-14
```

Answer:

$$\texttt{HIQ}_i = 50.65 - 4.73 \times \texttt{WSPercep}_i - 0.45 \times \texttt{RSES}_i + e_i$$

(e) Assess the mediation hypothesis by assessing the significance of the parameters in the three models estimated above (i.e. following the causal steps procedure).

Answer: The regression coefficient (slope) for `WSPercep` in the first model (with `HIQ` as dependent variable) is significant, as is the coefficient for `WSPercep` in the second model (with `RSES` as dependent variable). In the third model, both regression coefficients are significant. However, the significance of the `WSPercep` effect is reduced, from $F = 52.90$ in the first model to $F = 26.24$ in the third model. So, while the effect of weight and shape perceptions is not completely mediated by self-esteem, there is evidence of *partial* mediation.

(f) Use the Sobel-Aroian test to test the significance of the indirect effect of `WSPercep` on `HIQ`.

```
> mySobelTest <- function(a,b,se_a,se_b) {
+   Z <- (a*b)/sqrt((a^2 * se_b^2 + b^2 * se_a^2 + se_a^2*se_b^2))
+   # two-sided p-value
+   p <- 2*(1-pnorm(abs(Z)))
+   cat("Sobel-Aroian Test \n")
+   cat(paste("Z = ",Z,", p = ",p,"\n",sep=""))
```

```
+ }
>
> a <- summary(MODEL_2)$coefficients["wspercep","Estimate"]
> se_a <- summary(MODEL_2)$coefficients["wspercep","Std. Error"]
> b <- summary(MODEL_3)$coefficients["rses","Estimate"]
> se_b <- summary(MODEL_3)$coefficients["rses","Std. Error"]
>
> mySobelTest(a,b,se_a,se_b)

Sobel-Aroian Test
Z = -3.2772581185632, p = 0.00104820497747915
```

Answer:

$$
\begin{aligned}
Z &= \frac{a \times b}{\sqrt{a^2 \times s_b^2 + b^2 \times s_a^2 + s_a^2 \times s_b^2}} \\
&= \frac{4.40 \times (-0.45)}{\sqrt{4.40^2 \times (0.095)^2 + (-0.45)^2 \times (0.952)^2 + (0.952)^2 \times (0.095)^2}} \\
&= -3.28
\end{aligned}
$$

Using the standard-normal distribution, the probability of obtaining this value or a smaller one is $P(Z \leq -3.28) = .0005$. To perform a two-sided test (which is what we'd normally want to do), we need $P(|Z| \geq 3.28) = P(Z \leq -3.28) + P(Z \geq 3.28) = 2 \times .0005 = .001$, which is smaller than $\alpha = .05$, so the mediated effect of `WSPercep` on `HIQ`, via `RSES`, is significant.

(g) Now, let's look at another possibly mediated effect. Previous research (e.g., Miotto *et al.*, 2002) has found a significant relation between social desirability and level of eating disorders. Let's assess whether this effect is mediated by depression. Use both the causal steps and Sobel-Aroian procedures to assess whether the effect of `SocDesir` on `HIQ` is mediated by `BDI`.

```
> # causal steps
> summary(lm(hiq ~ socdesir,data=geller))


Call:
lm(formula = hiq ~ socdesir, data = geller)

Residuals:
     Min       1Q   Median       3Q      Max
-14.7563  -6.1382  -0.7375   5.7025  21.7547

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  24.9082     2.5261   9.860 1.41e-15 ***
```

```
socdesir     -1.6216     0.4695  -3.454 0.000876 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.554 on 82 degrees of freedom
Multiple R-squared:  0.127,Adjusted R-squared:  0.1164
F-statistic: 11.93 on 1 and 82 DF,  p-value: 0.0008762

> summary(lm(bdi ~ socdesir,data=geller))


Call:
lm(formula = bdi ~ socdesir, data = geller)

Residuals:
     Min       1Q   Median       3Q      Max
-15.8848  -4.8417   0.3759   4.1713  17.4800

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  16.4399     2.0193   8.141 3.67e-12 ***
socdesir     -1.1880     0.3753  -3.166  0.00217 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.838 on 82 degrees of freedom
Multiple R-squared:  0.1089,Adjusted R-squared:  0.09803
F-statistic: 10.02 on 1 and 82 DF,  p-value: 0.002173

> summary(lm(hiq ~ socdesir + bdi,data=geller))


Call:
lm(formula = hiq ~ socdesir + bdi, data = geller)

Residuals:
     Min       1Q   Median       3Q      Max
-13.4835  -6.3022   0.3493   5.0465  17.0102

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  12.3511     2.7068   4.563 1.77e-05 ***
socdesir     -0.7142     0.3963  -1.802   0.0752 .
bdi           0.7638     0.1101   6.939 8.85e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 6.816 on 81 degrees of freedom
Multiple R-squared:  0.4525,Adjusted R-squared:  0.439
F-statistic: 33.47 on 2 and 81 DF,  p-value: 2.543e-11

> # sobel-aroian test
> # As an example of "good" R use
> # let's write a simpler R function for this...
>
> simpleSobelTest <- function(source,mediator,dependent) {
+   # fit model 2
+   mod <- lm(mediator ~ source)
+   a <- summary(mod)$coefficients["source","Estimate"]
+   se_a <- summary(mod)$coefficients["source","Std. Error"]
+   # fit model 3
+   mod <- lm(dependent ~ source + mediator)
+   b <- summary(mod)$coefficients["mediator","Estimate"]
+   se_b <- summary(mod)$coefficients["mediator","Std. Error"]
+   Z <- (a*b)/sqrt((a^2 * se_b^2 + b^2 * se_a^2 + se_a^2*se_b^2))
+   # two-sided p-value
+   p <- 2*(1-pnorm(abs(Z)))
+   result <- list(
+     "Z" = Z,
+     "p-value" = p
+   )
+   cat("Sobel-Aroian Test \n\n")
+   cat("Path:",deparse(substitute(source)),"->",
+       deparse(substitute(mediator)),"->",
+       deparse(substitute(dependent)),"\n")
+   cat("a * b = ",a*b,"\n\n")
+   # note: the signif function "rounds" things by a specified
+   #        number of significant (non-zero) digits
+   cat(paste("Z = ",signif(Z,3),", p = ",signif(p,3),"\n",sep=""))
+   return(result)
+ }
>
> result <- simpleSobelTest(source=geller$socdesir,
+                              mediator=geller$bdi,
+                              dependent=geller$hiq)

Sobel-Aroian Test

Path: geller$socdesir -> geller$bdi -> geller$hiq
a * b =  -0.9073957

Z = -2.86, p = 0.0043
```

```
> result

$Z
[1] -2.855555

$`p-value`
[1] 0.004296164
```

Answer: In model

$$\texttt{HIQ}_i = 24.91 - 1.62 \times \texttt{SocDesir}_i + e_i$$

the effect of `SocDesir` is significant, $F(1, 82) = 11.93$, $p < .001$. In model

$$\texttt{BDI}_i = 16.44 - 1.19 \times \texttt{SocDesir}_i + e_i$$

the effect of `SocDesir` is significant, $F(1, 82) = 10.02$, $p = .002$. In model

$$\texttt{HIQ}_i = 12.35 + 0.76 \times \texttt{BDI}_i - 0.71 \times \texttt{SocDesir}_i + e_i$$

the effect of `BDI` is significant, $F(1, 81) = 48.15$, $p < .001$, but the effect of `SocDesir` is not, $F(1, 81) = 3.25$, $p = .075$. So, according to the causal steps procedure, there is evidence that the effect of social desirability on eating disorders is *completely* mediated by depression. Using the Sobel-Aroian test, we can assess whether this mediated effect itself is significant:

$$\begin{aligned} Z &= \frac{(-1.19) \times (0.76)}{\sqrt{(-1.19)^2 \times (0.11)^2 + (0.76)^2 \times (0.375)^2 + (0.375)^2 \times (0.11)^2}} \\ &= -2.85 \end{aligned}$$

As $P(Z \leq -2.85) = .0022$, a two sided tests gives $P(|Z| > 2.85) = 2 \times .0022 = .0044$. So the indirect effect of social desirability on eating disorders (mediated by depression) is significant.