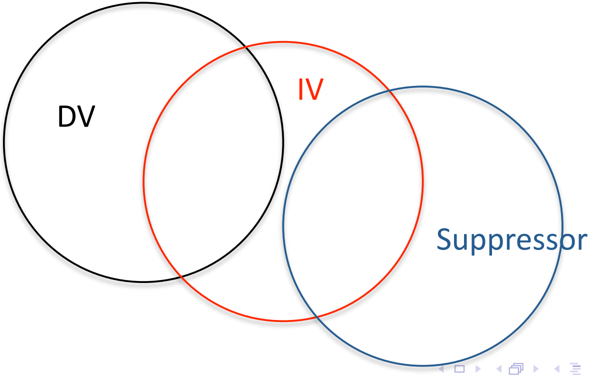
Suppressor Effects in Linear Models

Dr. J. Kyle Roberts

Southern Methodist University
Simmons School of Education and Human Development
Department of Teaching and Learning

Thinking Graphically About Suppressors

- A suppressor effect occurs when a variable has a non-zero β weight but a zero structure coefficient.
- The inclusion of a suppressor in a regression equation removes the unwanted variance from the predictor variable, thus enhancing the relationship between the other independent variable and the dependent variable.



Illustrating Suppressors in R

```
> library(MASS)
> correlation <- matrix(c(1, 0.5, 0, 0.5, 1, 0.5,
+     0, 0.5, 1), ncol = 3, nrow = 3, dimnames = list(c("dv",
+     "iv1", "iv2"), c("dv", "iv1", "iv2")))
> correlation

     dv iv1 iv2
dv  1.0 0.5 0.0
iv1 0.5 1.0 0.5
iv2 0.0 0.5 1.0

> set.seed(12346)
> suppressor.set <- data.frame(mvrnorm(n = 1000,
+     rep(10, 3), correlation))
```

Illustration cont.

```
> m1 <- lm(dv ~ iv1 + iv2, suppressor.set)
> summary(m1)

Call:
lm(formula = dv ~ iv1 + iv2, data = suppressor.set)

Residuals:
     Min       1Q   Median       3Q      Max
-2.75190 -0.54804  0.03108  0.53789  2.93339

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.00115    0.29836  20.114   <2e-16
iv1          0.67842    0.03044  22.285   <2e-16
iv2         -0.28112    0.03015  -9.323   <2e-16

Residual standard error: 0.8217 on 997 degrees of freedom
Multiple R-squared: 0.3356, Adjusted R-squared: 0.3343
F-statistic: 251.8 on 2 and 997 DF,  p-value: < 2.2e-16
```

Illustration cont.

```
> library(yhat)
> regr(m1)$Beta_Weights

            dv
iv1   0.673360
iv2  -0.281694

> regr(m1)$Structure_Coefficients

      iv1       iv2
0.9096263 0.1178144

> regr(m1)$Commonality_Data$CC

                        Coefficient     % Total
Unique to iv1                0.3310       98.61
Unique to iv2                0.0579       17.26
Common to iv1, and iv2      -0.0533      -15.87
Total                        0.3356      100.00
```

Types of Effect Size Corrections

$$\text{Wherry-1:} = 1 - \frac{N-1}{N-p-1}(1-R^2)$$

$$\text{Claudy3:} = 1 - \frac{(N-4)(1-R^2)}{N-p-1}\left[1 + \frac{2(1-R^2)}{N-p+1}\right]$$

$$\text{Smith-1:} = 1 - \frac{N}{N-p}(1-R^2)$$

$$\text{Wherry-2:} = 1 - \frac{N-1}{N-p}(1-R^2)$$

$$\text{Olkin\&Pratt:} = 1 - \frac{(N-3)(1-R^2)}{N-p-1}\left[1 + \frac{2(1-R^2)}{N-p+1}\right]$$

$$\text{Pratt:} = 1 - \frac{(N-3)(1-R^2)}{N-p-1}\left[1 + \frac{2(1-R^2)}{N-p-2.3}\right]$$

Yin and Fan, 2001

- Yin and Fan (2002) showed how the multiple R^2 is really an upward bound for the actual R^2 in the population.
- Perhaps most important, they noted that the Wherry formula (basis for the default adjusted R^2 in commonly used statistical software) has been found to be an ineffective analytical formula when sample sizes are small.
- What `regr` does is compute the multiple R^2 correction that has the least amount of bias given the simulations done by Yin and Fan (2002).

Homework

- Find a journal article in a journal in your field that has used multiple regression.
- Make sure that the one you identify has used both multiple regression *and* has the correlation matrix attached in a table.
- Reproduce their results by simulating their dataset.
- Try other competing models against their full model and see if you can find a different, better (or at least as good as) model.
- Use the `anova` function to compare the models.

```
> m0 <- lm(dv ~ iv1, suppressor.set)
> anova(m0, m1)
```

```
Analysis of Variance Table

Model 1: dv ~ iv1
Model 2: dv ~ iv1 + iv2
  Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1    998 731.86
2    997 673.17  1    58.685 86.915 < 2.2e-16
```