|  |
| --- |
| **Suppression in Factorial ANOVA** |

The outcome variable is weight of student. The predictor variables are school and sex.

The GLM Procedure

| **Class Level Information** | | |
| --- | --- | --- |
| **Class** | **Levels** | **Values** |
| **School** | 2 | 0 1 |
| **Sex** | 2 | Female Male |

|  |  |
| --- | --- |
| **Number of Observations Read** | 200 |
| **Number of Observations Used** | 200 |

The GLM Procedure

Dependent Variable: Weight

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **Model** | 3 | 57186.3328 | 19062.1109 | 27.27 | <.0001 |
| **Error** | 196 | 137025.0222 | 699.1073 |  |  |
| **Corrected Total** | 199 | 194211.3550 |  |  |  |

| **R-Square** | **Coeff Var** | **Root MSE** | **Weight Mean** |
| --- | --- | --- | --- |
| 0.294454 | 20.84155 | 26.44064 | 126.8650 |

| **Source** | **DF** | **Type I SS** | **Mean Square** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **School** | 1 | 8385.12500 | 8385.12500 | 11.99 | 0.0007 |
| **Sex** | 1 | 48775.74818 | 48775.74818 | 69.77 | <.0001 |
| **School\*Sex** | 1 | 25.45960 | 25.45960 | 0.04 | 0.8489 |

| **Source** | **DF** | **Type III SS** | **Mean Square** | **F Value** | **Pr > F** |
| --- | --- | --- | --- | --- | --- |
| **School** | 1 | 866.47293 | 866.47293 | 1.24 | 0.2670 |
| **Sex** | 1 | 37801.01838 | 37801.01838 | 54.07 | <.0001 |
| **School\*Sex** | 1 | 25.45960 | 25.45960 | 0.04 | 0.8489 |

The observed means show that, on average, the students are heavier at School 0.

| **Level of School** | **N** | **Weight** | |
| --- | --- | --- | --- |
| **Mean** | **Std Dev** |
| **0** | **100** | 133.340000 | 31.6493477 |
| **1** | **100** | 120.390000 | 29.5863383 |

| **Level of Sex** | **N** | **Weight** | |
| --- | --- | --- | --- |
| **Mean** | **Std Dev** |
| **Female** | **130** | 114.615385 | 27.5186715 |
| **Male** | **70** | 149.614286 | 24.3133487 |

| **Level of School** | **Level of Sex** | **N** | **Weight** | |
| --- | --- | --- | --- | --- |
| **Mean** | **Std Dev** |
| **0** | **Female** | **40** | 109.950000 | 27.5177798 |
| **1** | **Female** | **90** | 116.688889 | 27.4167358 |
| **0** | **Male** | **60** | 148.933333 | 23.7078737 |
| **1** | **Male** | **10** | 153.700000 | 28.7365427 |

For both sexes, the students are lighter at School 0, but in the aggregate data the students are heavier at School 0. In other words, we have here a Reversal Paradox, aka Simpson’s Paradox.

|  |
| --- |
| Results for data given to Simpson |

The GLM Procedure

Least Squares Means

| **School** | **Weight LSMEAN** |
| --- | --- |
| **0** | 129.441667 |
| **1** | 135.194444 |

For the adjusted means, the students are lighter at School 0. These adjusted means address the question “what would the weights be if there were no correlation between Sex and School?”

Here are the means obtained with Proc Means.

The MEANS Procedure

School=0

| **Analysis Variable : Weight** |
| --- |
| **Mean** |
| 133.3400000 |

School=1

| **Analysis Variable : Weight** |
| --- |
| **Mean** |
| 120.3900000 |

The FREQ Procedure

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  | | --- | --- | | |  | | --- | | **Frequency** | | | | **Table of School by Sex** | | | | | --- | --- | --- | --- | | **School** | **Sex** | | | | **Female** | **Male** | **Total** | | **0** | |  | | --- | | 40 | | |  | | --- | | 60 | | |  | | --- | | 100 | | | **1** | |  | | --- | | 90 | | |  | | --- | | 10 | | |  | | --- | | 100 | | | **Total** | |  | | --- | | 130 | | |  | | --- | | 70 | | |  | | --- | | 200 | | |

The table above has the cell sizes. Notice that they are not proportional. At School 0 60% of the students are male, but at School 1 only 10% are male. This results in a large correlation between school and sex.

|  |
| --- |
| **Statistics for Table of School by Sex** |

| **Statistic** | **DF** | **Value** | **Prob** |
| --- | --- | --- | --- |
| **Chi-Square** | 1 | 54.9451 | <.0001 |
| **Phi Coefficient** |  | -0.5241 |  |

The CORR Procedure

|  |  |
| --- | --- |
| **3 Variables:** | Weight School Sex |

| **Simple Statistics** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **N** | **Mean** | **Std Dev** | **Sum** | **Minimum** | **Maximum** |
| **Weight** | 200 | 126.86500 | 31.23998 | 25373 | 39.00000 | 218.00000 |
| **School** | 200 | 0.50000 | 0.50125 | 100.00000 | 0 | 1.00000 |
| **Sex** | 200 | 0.35000 | 0.47817 | 70.00000 | 0 | 1.00000 |

| **Pearson Correlation Coefficients, N = 200  Prob > |r| under H0: Rho=0** | | | |
| --- | --- | --- | --- |
|  | **Weight** | **School** | **Sex** |
| **Weight** | |  | | --- | | 1.00000 | |  | | |  | | --- | | -0.20779 | | 0.0032 | | |  | | --- | | 0.53570 | | <.0001 | |
| **School** | |  | | --- | | -0.20779 | | 0.0032 | | |  | | --- | | 1.00000 | |  | | |  | | --- | | -0.52414 | | <.0001 | |
| **Sex** | |  | | --- | | 0.53570 | | <.0001 | | |  | | --- | | -0.52414 | | <.0001 | | |  | | --- | | 1.00000 | |  | |

The REG Procedure

Model: MODEL1

Dependent Variable: Weight

|  |  |
| --- | --- |
| **Number of Observations Read** | 200 |
| **Number of Observations Used** | 200 |

| **Analysis of Variance** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 2 | 57161 | 28580 | 41.08 | <.0001 |
| **Error** | 197 | 137050 | 695.68772 |  |  |
| **Corrected Total** | 199 | 194211 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 26.37589 | **R-Square** | 0.2943 |
| **Dependent Mean** | 126.86500 | **Adj R-Sq** | 0.2872 |
| **Coeff Var** | 20.79052 |  |  |

| **Parameter Estimates** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** | **Standardized Estimate** |
| **Intercept** | **1** | 110.27273 | 3.81395 | 28.91 | <.0001 | 0 |
| **School** | **1** | 6.27273 | 4.37997 | 1.43 | 0.1537 | 0.10065 |
| **Sex** | **1** | 38.44545 | 4.59145 | 8.37 | <.0001 | 0.58846 |

The zero-order correlation between school and weight was -0.20779, but the beta weight for school is of the opposite sign. School is functioning as a net suppressor.

The zero-order correlation between sex and weight was 0.53570, but the beta weight is of greater magnitude.

The CORR Procedure

Sex=Female

| **Pearson Correlation Coefficients, N = 130  Prob > |r| under H0: Rho=0** | | |
| --- | --- | --- |
|  | **School** | **Weight** |
| **School** | |  | | --- | | 1.00000 | |  | | |  | | --- | | 0.11346 | | 0.1987 | |

The CORR Procedure

Sex=Male

|  |  |
| --- | --- |
| **2 Variables:** | School Weight |

| **Pearson Correlation Coefficients, N = 70  Prob > |r| under H0: Rho=0** | | |
| --- | --- | --- |
|  | **School** | **Weight** |
| **School** | |  | | --- | | 1.00000 | |  | | |  | | --- | | 0.06910 | | 0.5698 | |

In the aggregate data, the correlation between school and weight was negative, but within each sex it is positive.

The data for this lesson were simulated using this SAS code:

**PROC** **FORMAT**;VALUE sx **0**='Female' **1**='Male'; **run**;

**%MACRO** ***XDANOVA2***; DATA YS;KEEP School Sex Weight; FORMAT Sex sx. ;

School = **0**; Sex = **0**; DO S=**1** TO **40**;

Weight=ROUND(**110**+**25**\*NORMAL(**0**));

OUTPUT; FILE "D:\\_Stats\Simulators\%name.txt"; put Juror Defendant Sentence; END;

School = **0**; Sex = **1**; DO S=**1** TO **60**;

Weight=ROUND(**150**+**25**\*NORMAL(**0**));

OUTPUT; FILE "D:\\_Stats\Simulators\%name.txt"; put Juror Defendant Sentence; END;

School = **1**; Sex = **0**; DO S=**1** TO **90**;

Weight=ROUND(**120**+**25**\*NORMAL(**0**));

OUTPUT; FILE "D:\\_Stats\Simulators\%name.txt"; put Juror Defendant Sentence; END;

School = **1**; Sex = **1**; DO S=**1** TO **10**;

Weight=ROUND(**160**+**25**\*NORMAL(**0**));

if Sentence < **1** then Sentence=**1**; if Sentence > **15** then sentence = **15**;

OUTPUT; FILE "D:\\_Stats\Simulators\%name.txt"; put Juror Defendant Sentence; END;

TITLE1 "Results for data given to %ST"; run;

PROC GLM; CLASS School Sex; MODEL Weight=School|Sex;

MEANS School|Sex ; LSMEANS School; run; quit;

Proc Sort; By School; Proc Means mean; Var weight; By School; run;

Proc Freq; Tables School\*Sex / NOCOL NOROW NOPERCENT CHISQ; run;

Proc Corr; Var Weight School Sex; run;

Proc Reg; Model Weight = School Sex / stb; run; quit;

Proc Sort; By Sex; Proc Corr; Var School Weight; By Sex; run;

**%mend** XDANOVA2;

**%MACRO** ***OP***; %***XDANOVA2*** **%MEND** OP;

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* NAMES FILE FOLLOWS \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

**%MACRO** ***ST***; Simpson **%MEND** ST;

**%MACRO** ***NAME***; Simpson-**01** **%MEND** NAME; %***OP*** **run**;

[Karl L. Wuensch](http://core.ecu.edu/psyc/wuenschk/klw.htm), April, 2021.