General Linear Model

Quantitative Methods II

Merle Canfield, Ph.D.

**©**

There are two analyses presented in this chapter -- the formulae and computations are the same as the analysis in chapter 2. More data has been added to make it a more realistic problem. The second analysis is designed to show the similarity between analysis of regression and analysis of variance.

This chapter builds on chapter 2 and if there are points that you don't understand because of complexity of numbers it might be useful to refer to the more simplified set in chapter 2. A correlation between continuous variables is presented as the first example then a correlation between a continuous variable and a dichotomous variable will be presented. The similarities between this correlation and an analysis of variance will be shown.

The sample data was selected from a larger set that was administered to 5 different groups including psychiatric inpatients and professional staff that worked with psychiatric inpatients. The questionnaire follows on the next page.

Psychosocial Assessment Scale

Name:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ ID #\_\_\_\_\_\_\_\_\_ Date\_\_\_\_\_\_\_\_\_

Circle the number below that is the closest to how you have felt recently.

none all

IN THE PAST WEEK HOW OFTEN HAVE YOU ... of the several of the

time times time

1. ...enjoyed your leisure hours (evenings, weekends, etc.)? 0 1 2 3 4 5 6 7 8

2. ...felt fearful or afraid? 0 1 2 3 4 5 6 7 8

3. ...felt sad or depressed? 0 1 2 3 4 5 6 7 8

4. ...felt good about yourself or things you have done? 0 1 2 3 4 5 6 7 8

5. ...felt angry? 0 1 2 3 4 5 6 7 8

6. ...felt mixed up or confused? 0 1 2 3 4 5 6 7 8

7. ...felt like you've spent a worthwhile day? 0 1 2 3 4 5 6 7 8

8. ...felt tense? 0 1 2 3 4 5 6 7 8

9. ...felt useless? 0 1 2 3 4 5 6 7 8

very very

IN THE PAST WEEK ... dissatisfied satisfied

10. ...how satisfied have you been in general (with relationships, with finances, 0 1 2 3 4 5 6 7 8

and family?)

never often

11. ...how often did you get together with people outside of your home? 0 1 2 3 4 5 6 7 8

all of

never the time

12. ...did you handle the basic necessities such as paying bills, shopping and 0 1 2 3 4 5 6 7 8

and taking care of your room (home; apt.)?

none some much

13. ...how much time did you spend with friends & family talking or doing things 0 1 2 3 4 5 6 7 8

together?

a great

none some deal

14. ...how much conflict was there with the person(s) you live with? 0 1 2 3 4 5 6 7 8

15. ...have you used alcohol? 0 1 2 3 4 5 6 7 8

16. ...have you used drugs? 0 1 2 3 4 5 6 7 8

‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑

17. How much would friends or community members support you if you were sick 0 1 2 3 4 5 6 7 8

or having problems?

not part full

employed time time

18. Are you employed (includes self-employed, housewife, student, employee)? 0 1 2 3 4 5 6 7 8

not at all completely

19. Do you feel that you do a good job (whether self‑employed, housewife, 0 1 2 3 4 5 6 7 8

student, employee)?

20. How much do you like your work (or studies)? 0 1 2 3 4 5 6 7 8

21. Do emotional problems interfere with your work (or studies)? 0 1 2 3 4 5 6 7 8

no enough

22. In the past month, was the amount of money you had, enough to pay the bills? 0 1 2 3 4 5 6 7 8

poor good

23. Do you feel that you are in good physical condition? 0 1 2 3 4 5 6 7 8

5-6-93

The problems throughout this chapter use sample data from the preceding questionnaire. It should be recognized that is selected data -- that is incomplete and selected for the purpose of this example. At the same time is does represent results from a larger study. It is somewhat exaggerated here in that the two samples are: (1) patients at the time of admission to an inpatient hospital and (2) professional staff members. The data is randomly selected from those groups excluding subjects who had missing data. Only 20 cases were selected (10 from each group) so that the mathematical calculations can be followed.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| group | leisur | fear | depres | feelg | angry | confus | worth | tense | useles | satisf | outsid | bills | talkto | conflt | alcdrg | supprt | employ | goodj | likew | inway | money | health | distres | qualit | relate | job | abuse | slcar |
| 1 | 6 | 1 | 0 | 6 | 1 | 0 | 6 | 1 | 1 | 7 | 7 | 5 | 7 | 2 | 6 | 6 | 8 | 7 | 6 | 0 | 6 | 8 | 7.33 | 6.25 | 6.5 | 7.25 | 2 | 5 |
| 1 | 6 | 2 | 4 | 4 | 2 | 1 | 4 | 4 | 0 | 7 | 1 | 7 | 7 | 1 | 0 | 4 | 8 | 6 | 7 | 2 | 8 | 5 | 5.83 | 5.25 | 4.75 | 6.75 | 8 | 7 |
| 1 | 7 | 1 | 1 | 0 | 0 | 2 | 2 | 2 | 0 | 7 | 4 | 6 | 6 | 0 | 2 | 7 | 8 | 8 | 8 | 0 | 7 | 7 | 7 | 4 | 6.25 | 8 | 6 | 6 |
| 1 | 8 | 0 | 1 | 7 | 1 | 1 | 8 | 0 | 0 | 8 | 4 | 8 | 8 | 0 | 2 | 8 | 8 | 8 | 8 | 0 | 8 | 8 | 7.5 | 7.75 | 7 | 8 | 6 | 8 |
| 1 | 6 | 0 | 0 | 5 | 0 | 0 | 5 | 1 | 3 | 7 | 8 | 8 | 5 | 0 | 1 | 8 | 8 | 7 | 7 | 1 | 8 | 7 | 7.33 | 5.75 | 7.25 | 7.25 | 7 | 8 |
| 1 | 5 | 0 | 0 | 6 | 2 | 1 | 6 | 2 | 1 | 5 | 4 | 8 | 4 | 5 | 4 | 3 | 8 | 7 | 0 | 0 | 8 | 6 | 7 | 5.5 | 3.5 | 5.75 | 4 | 8 |
| 1 | 4 | 1 | 4 | 4 | 4 | 0 | 4 | 4 | 2 | 6 | 2 | 8 | 4 | 0 | 1 | 6 | 8 | 8 | 7 | 1 | 8 | 4 | 5.5 | 4.5 | 5 | 7.5 | 7 | 8 |
| 1 | 8 | 1 | 1 | 6 | 4 | 0 | 8 | 1 | 1 | 7 | 3 | 8 | 4 | 1 | 2 | 7 | 8 | 7 | 7 | 0 | 8 | 8 | 6.67 | 7.25 | 5.25 | 7.5 | 6 | 8 |
| 1 | 6 | 2 | 2 | 7 | 2 | 0 | 7 | 2 | 0 | 7 | 5 | 8 | 6 | 3 | 2 | 7 | 8 | 7 | 7 | 0 | 7 | 8 | 6.67 | 6.75 | 5.75 | 7.5 | 6 | 8 |
| 1 | 6 | 2 | 4 | 4 | 2 | 2 | 5 | 3 | 0 | 3 | 8 | 1 | 4 | 1 | 0 | 4 | 8 | 5 | 6 | 3 | 8 | 2 | 5.83 | 4.5 | 5.75 | 6 | 8 | 1 |
| 2 | 4 | 6 | 6 | 4 | 5 | 7 | 2 | 6 | 4 | 5 | 2 | 5 | 4 | 4 | 0 | 1 | 1 | 3 | 3 | 8 | 4 | 6 | 2.33 | 3.75 | 2.75 | 1.75 | 8 | 5 |
| 2 | 1 | 8 | 8 | 1 | 8 | 6 | 2 | 8 | 8 | 2 | 1 | 0 | 1 | 5 | 0 | 1 | 0 | 0 | 1 | 4 | 0 | 1 | 0.33 | 1.5 | 1.5 | 1.25 | 8 | 0 |
| 2 | 5 | 1 | 4 | 7 | 6 | 8 | 6 | 2 | 4 | 2 | 0 | 8 | 3 | 0 | 3 | 4 | 0 | 4 | 5 | 6 | 0 | 3 | 3.83 | 5 | 3.75 | 2.75 | 5 | 8 |
| 2 | 0 | 8 | 8 | 0 | 8 | 8 | 0 | 8 | 6 | 0 | 1 | 4 | 2 | 8 | 0 | 2 | 8 | 8 | 4 | 2 | 8 | 4 | 0.33 | 0 | 1.25 | 6.5 | 8 | 4 |
| 2 | 1 | 7 | 7 | 2 | 5 | 6 | 0 | 5 | 5 | 2 | 2 | 2 | 2 | 0 | 8 | 3 | 0 | 2 | 4 | 3 | 4 | 4 | 2.17 | 1.25 | 3.75 | 2.75 | 0 | 2 |
| 2 | 5 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 7.67 | 2 | 2.75 | 2.5 | 6 | 1 |
| 2 | 3 | 3 | 3 | 3 | 2 | 3 | 2 | 3 | 3 | 4 | 3 | 2 | 5 | 2 | 7 | 6 | 0 | 4 | 3 | 3 | 6 | 4 | 5.17 | 3 | 5 | 3 | 1 | 2 |
| 2 | 3 | 6 | 6 | 2 | 2 | 4 | 3 | 3 | 4 | 5 | 5 | 8 | 6 | 1 | 0 | 7 | 5 | 7 | 5 | 1 | 1 | 4 | 3.83 | 3.25 | 6.25 | 6 | 8 | 8 |
| 2 | 4 | 5 | 6 | 4 | 6 | 6 | 3 | 5 | 5 | 5 | 5 | 3 | 5 | 1 | 2 | 4 | 2 | 5 | 6 | 7 | 8 | 6 | 2.5 | 4 | 5.25 | 3.5 | 6 | 3 |
| 2 | 1 | 4 | 4 | 0 | 4 | 8 | 0 | 8 | 8 | 0 | 8 | 0 | 0 | 8 | 8 | 4 | 0 | 0 | 0 | 8 | 0 | 0 | 2 | 0 | 3 | 0 | 0 | 0 |

This example deals with only items # 3 and # 8 of the questionnaire. Those questions were "In the past week how often have you felt sad or depressed?" and "In the past week how often have you felt tense?" and are labeled as DEPRES and TENSE respectively. A discussion of the correlation between responses to these two items (items 3 and 8 on the questionnaire) follows.

Participant DEPRES # 3 TENSE # 8

1 0 1

2 4 4

3 1 2

4 1 0

5 0 1

6 0 2

7 4 4

8 1 1

9 2 2

10 4 3

11 6 6

12 8 8

13 4 2

14 8 8

15 7 5

16 0 0

17 3 3

18 6 3

19 6 5

20 4 8

We will now present this data in the same way as the more limited data was presented in chapter 2. So that all of the formulae are the same as those presented in chapter 2 you're not learning a new set. This is a more alive example and goes through the same process as in the previous chapter. The variable TENSE is labeled as the X variable (predictor or independent variable) and DEPRES as the Y variable (criterion or dependent variable). First the data will be presented and the SPSS jobstreams to compute it will be given.

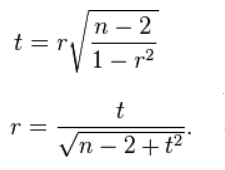
|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | C1 | C2 | C3 | C4 | C5 | C6 | C7 | C8 | C9 | C10 | C11 | C12 |
| A |  |  |  |  |  |  |  |  |  |  |  |  |
| *B* |  | *X* |  |  |  |  |  |  |  |  |  |  |
| C | actual score | deviation score | deviation squared | actual score | deviation score | deviation squared | cross product | predicted score | error:  actual minus predicted | error squared  or  within  squared | Y primed minus the mean of Y | Y primed minus the mean of Y squared |
| D | observed score | little x | little x squared | observed score | little y | little y squared |  | Y primed | residual score |  |  |  |
| E | raw score  TENSE | X minus the mean | little x times little x | raw score  DEPRES | Y minus the mean | little y times little y | little x times little y | Y predicted from X | error in prediction |  |  |  |
| predictor variable (X)  independent varialbe (IV) | | | | criterion variable (Y)  dependent variable (DV) | | |  |  |  |  |  |  |
| F | 1  4  2  0  1  2  4  1  2  3  6  8  2  8  5  0  3  3  5  8 | -2.40  .60  -1.40  -3.40  -2.40  -1.40  .60  -2.40  -1.40  ‑.40  2.60  4.60  -1.40  4.60  1.60  -3.40  ‑.40  ‑.40  1.60  4.60 | 5.7600  .3600  1.9600  11.5600  5.7600  1.9600  .3600  5.7600  1.9600  .1600  6.7600  21.1600  1.9600  21.1600  2.5600  11.5600  .1600  .1600  2.5600  21.1600 | 0  4  1  1  0  0  4  1  2  4  6  8  4  8  7  0  3  6  6  4 | -3.45  .55  -2.45  -2.45  -3.45  -3.45  .55  -2.45  -1.45  .55  2.55  4.55  .55  4.55  3.55  -3.45  ‑.45  2.55  2.55  .55 | 11.9025  .3025  6.0025  6.0025  11.9025  11.9025  .3025  6.0025  2.1025  .3025  6.5025  20.7025  .3025  20.7025  12.6025  11.9025  .2025  6.5025  6.5025  .3025 | 8.28  .33  3.43  8.33  8.28  4.83  .33  5.88  2.03  ‑.22  6.63  20.93  ‑.77  20.93  5.68  11.73  .18  ‑1.02  4.08  2.53 | 1.2884  3.9904  2.1891  .3878  1.2884  2.1891  3.9904  1.2884  2.1891  3.0897  5.7916  7.5929  2.1891  7.5929  4.8910  .3878  3.0897  3.0897  4.8910  7.5929 | ‑1.2884  .0096  ‑1.1891  .6122  ‑1.2884  ‑2.1891  .0096  ‑.2884  ‑.1891  .9103  .2084  .4071  1.8109  .4071  2.1090  ‑.3878  ‑.0897  2.9103  1.1090  ‑3.5929 | 1.6601  .0001  1.4139  .3748  1.6601  4.7921  .0001  .0832  .0358  .8286  .0434  .1657  3.2794  .1657  4.4479  .1504  .0080  8.4697  1.2299  12.9091 | ‑2.1616  .5404  ‑1.2609  ‑3.0622  ‑2.1616  ‑1.2609  .5404  ‑2.1616  ‑1.2609  ‑.3603  2.3416  4.1429  ‑1.2609  4.1429  1.4410  ‑3.0622  ‑.3603  ‑.3603  1.4410  4.1429 | 4.6723  .2920  1.5899  9.3771  4.6723  1.5899  .2920  4.6723  1.5899  .1298  5.4833  17.1638  1.5899  17.1638  2.0765  9.3771  .1298  .1298  2.0765  17.1638 |
| G | 68 | 0 | 124.80 | 69 | 0 | 142.95 | 112.40 | 68.999 | .0005 | 41.718 | 0 | 101.23 |
| H | 3.40 |  | Sum of x Squared | 3.45 |  | sum of y squared | sum of cross products |  |  | error sum of squares |  | between or regression sum of squares |
| I | mean of X |  | mean of Y |  |  |  |  |
| J |  |  |  |  |  | SST |  |  |  | SSE |  | SSB  SSR |

Table 2‑3. Rows A through F are either mathematical notation or verbal description of mathematical calculations of the numbers in the column. Rows 1 through 20 are associated numbers involved the calculation. Row G is the sum of the numbers in the column while row H is the mean for the column. Row I is the usual verbal description of the sum in the column and row J is an abbreviation of that description.

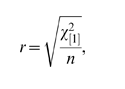
In the example below when there are scores for all 20 cases are individually computed only the first 4 will be given (this occurs with observation, little x, Y' and SSE).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| # | Description | Label | Symbol | Notation | Full Notation or Formula | Individual Case Values |
| 1 | Observation of the varialbe specified by "i" case number and the "j" group or variable | observation |  |  |  | Case #1 = 1  Case #2 = 4  Case #3 = 2  Case #4 = 0 |
| 2 | Number of observations in each (jth) group (set) | number of cases |  |  |  | 20 |
| 3 | Sum of X for each (jth) groujp or set | Sum of X, Total, Sum |  |  |  | 68 |
| 4 | Mean for each (jth) group or set | Average, Mean |  |  |  | 3.40 |
| 5 | Deviation Score (X minus the mean) | Little x, deviation score |  |  |  | Case #1 = 1 - 3.4 = -2.40  Case #2 = 4 - 3.4 = .60  Case #3 = 2 - 3.4 = -1.40  Case # 4 = 0 - 3.4 = -3.40 |
| 9 | Sum of little x |  |  |  |  | 0 |
| 10 | Sum of little x squared | Sum of Squares | SS |  |  | = ((-2.40)\*(-2.40)) + ((.60) \* (.60))+  ((-1.40)\*(-1.40)) + ((-3.40)\*(-3.40)) etc  =5.76 + .36 + 1.96 + 11.56 + etc. = 124.80 |
| 11 | Population variance of each (jth) group or set | Population variance |  |  |  | 124.80  --------- = 6.24  20 |
| 12 | Sample variance of each (jth) group or set | Sample variance |  |  |  |  |
| 13 | Population standard deviation for each (jth) group or set | Popultaion standard deviation |  |  |  |  |
| 14 | Sample standard deviation for each (jth) group or set | Sample standard deviation | SD |  |  | ==2.563 |
| 15 | *z*-score for each (ijth) observation | *z-*score | *z* |  |  | 1 - 3.40  case #1 = --------- = -.961  2.498  4 - 3.40  case #2 = ---------- = .241  2.498  2 - 3.40  case #3 = ----------- = -.560  2.498  0 - 3.40  case #4 = ----------- = -1.361  2.498 |
| 16 | T-score |  |  |  | 50+(z\*10) | case #1 =50+(-.961\*10)=40.39  case #2 =50+(.241\*10) =52.41  case #3 =50+(-.560\*10)=44.40  case #4 =50+(-1.361\*10)=36.39 |
| 17 | IO-score |  |  |  | 100+(z\*15) | case #1 = 100+(-.961\*15)=85.585  case #2 = 100+(.241\*15) =103.615  case #3 = 100+(-.560\*15)=91.60  case #4 = 100+(-1.361\*15)=79.585 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| # | Description | Label | Symbol | Notation | Full Notation or Formula | Individual Case Values |
| 18 | Observation of the varialbe specified by "i" case number and the "j" group or variable | observation |  |  |  | case #1 = 0  case #2 = 4  case #3 = 1  case #4 = 1 |
| 19 | Number of observations in each (jth) group (set) | number of cases |  |  |  | 20 |
| 20 | Sum of Y for each (jth) groujp or set | Sum of Y, Total, Sum |  |  |  | 69 |
| 21 | Mean for each (jth) group or set | Average, Mean |  |  |  | 3.45 |
| 22 | Deviation Score (Y minus the mean) | Little y, deviation score |  |  |  | case #1 = 0-3.45 = -3.45  case #2 = 4-3.45 = .55  case #3 = 1-3.45 = -2.45  case #4 = 1-3.45 = -2.45 |
| 23 | Sum of little y |  |  |  |  | 0 |
| 24 | Sum of little y squared | Sum of Squares | SS |  |  | ((-3.45)\*(-3.45))+(.55\*.55)+  ((-2.45)\*(-2.45))+((-2.45)\*(-2.45))+  etc  =11.90 + .3025 + 6.0025+6.0025 +etc etfetc |
| 25 | Population variance of each (jth) group or set | Population variance |  |  |  | 142.95  -------- = 7.148  20 |
| 26 | Sample variance of each (jth) group or set | Sample variance |  |  |  |  |
| 27 | Population standard deviation for each (jth) group or set | Popultaion standard deviation |  |  |  |  |
| 28 | Sample standard deviation for each (jth) group or set | Sample standard deviation | SD |  |  |  |
| 29 | *z*-score for each (ijth) observation | *z-*score | *z* |  |  | 0 - 3.45  case #1 = ---------- = -1.291  2.673  4 - 3.45  case #2 = ---------- = .206  2.673  1 - 3.45  case #3 = ---------- = -.917  2.673  1 - 3.45  case #4 = ---------- = -.917  2.673 |
| 30 |  |  |  |  |  |  |
| 31 | Cross products for two groups (sets) of scores | Cross Products |  |  |  | ((-2.40)\*(-3.45))+(.60\*.55)+  ((-1.40)\*(-2.45))+((-3.40)\*(-2.45))+  etc = 112.40 |
| 32 | Covariance | Covariance |  |  |  | 112.40  -------- = 5.62  20 |
| 33 | Slope of the regression line | slope, *b* | *b* |  |  | 112.40  --------- = .900641  124.80 |
| 34 | The value of Y at point where X is zero (0) | Constant, *a*, intercept | *a* |  |  | 3.45 - (.900641 \* 3.45) = .3878 |
| 35 | Y predicted from X | Y hat, Y primed, Y' |  |  |  | case #1=.3878+(.900641\*1)=1.2884  case #2=.3878+(.900641\*4)=3.9904  case #3=.3878+(.900641\*2)=2.1891  case #4=.3878+(.900641\*0)= .3873 |
| 36 | Regression sum of squares | sum of squares, between sum of squares | SSR, SSB |  | | 112.4 \* 112.4  ------------------ = 101.23  124.80 |
|  | Regression sum of Squares | sum of squares, between sum of squares | SSR, SSB |  | | case #1 ( 1.2884 - 3.45 )2 = 4.67251456  case #2 ( 3.9904 - 3.45 )2 = 0.29203216  case #3 ( 2.1891 - 3.45 )2 = 1.58986881  case #4 ( 0.3878 - 3.45 )2 = 9.37706884 |
| 37 | Residual sum of squres | Error, within sum of squares | SSE  SSW |  | | SST - SSR or SST - SSB  142.95 - 101.23 = 41.72 |
| case #1 = (0-1.2884)\*(0-3.9904)=1.66  case #2 = (4-3.9904)\*(4-3.9904)=.00  case #3 = (1-2.1891)\*(1-2.1891)=1.41  case #4 = (1- .3878)\*(1- .3878)= .37 |
| 38 | Total sum of squares | Total sum of squares | SST |  |  | 142.94 |
| 39 | Pearson correlation coefficient | correlation | *r* |  | *R*2 = *t*2 / (*t*2 + *DF*), |  |
| 40 | Analysis of Variance | ANOVA | F | SSB/SSW  (only applies when groups=2) | f=t2 |  |
| 41 | t-test | t-test | t |  | *t*2 = *DF\*R*2 / (1 - *R*2) |  |

`

= eta squared = SSB/SST



The correlation and regression can be shown graphically to develop understanding. The two variables are plotted by using the following SPSS jobstream.

|  |
| --- |
| File Name = crsplt8.sps |
| get file = '\rdda\crsleq1.sav'.  GRAPH  /SCATTERPLOT(BIVAR)=tense with depres  /MISSING=LISTWISE . |

That jobstream will produce the following plot:



This plot represents two variables DEPRES and TENSE. There was one person who answered both questions 0. If you look back at the raw data you will that was participant # 12. That person is represented by the # 1 in the lower left corner of the plot. There was one person who gave a 0 to TENSE and a 2 to DEPRES -- that person is represented by the number 1 on the left hand side of the plot near the number 2 on the DEPRES variable. There were two people that gave a 3 to TENSE and a 2 to DEPRES. Those two people are represented by the number 2 in the lower left quadrant of the plot. You might want to identify some more of the cases to convince yourself of the relationship of the data to the plot.

This scattergram represents all of the respondents on the items of TENSE and DEPRES. People who responded with smaller numbers to the item TENSE also responded with smaller numbers to DEPRES. At the same time people who responded with larger numbers to TENSE also responded with larger numbers to DEPRES. Two people answered 8 to both items that is what the 2 in the upper right hand corner indicates. One person answered 0 to both questions -- that is indicated by the 1 down in the left hand corner. Then there were those who responded in the middle range to both items. There are 2 people who answered 4 for depressed and 4 for tense. So now you have the people in the middle going together. They put down the same numbers for both questions, and that indicates a relationship.

The next three plots all have the same data as the previous but have modifications drawn to show characteristics of the correlation or regression. The next plot shows the sum of squares due to error or residual. It is the error in prediction of Y from X.



This next plot shows the sum of squares do to regression.



This nex plot shows the sum of squares total: it is the same as the sum of squares for little y.



The output below was created from SPSS with the following jobstream.

|  |
| --- |
| File Name = crsreg12.sps |
| get file = '\rdda\crsleq1.sav'.  REGRESSION  descriptives = mean stddev cov corr n  /MISSING LISTWISE  /STATISTICS COEFF OUTS R ANOVA  /CRITERIA=PIN(.05) POUT(.10)  /NOORIGIN  /DEPENDENT depres  /METHOD=ENTER tense . |

Table 3-1 Printout of regression of DEPRES and TENSE.

\* \* \* \* M U L T I P L E R E G R E S S I O N \* \* \* \*

Listwise Deletion of Missing Data

Mean Std Dev Label

DEPRES 3.450 2.743

TENSE 3.400 2.563

N of Cases = 20

Correlation, Covariance:

DEPRES TENSE

DEPRES 1.000 .842

7.524 5.916

TENSE .842 1.000

5.916 6.568

\* \* \* \* M U L T I P L E R E G R E S S I O N \* \* \* \*

Equation Number 1 Dependent Variable.. DEPRES

Descriptive Statistics are printed on Page 4

Block Number 1. Method: Enter TENSE

Variable(s) Entered on Step Number

1.. TENSE

Multiple R .84152 **A**

R Square .70816 **B**

Adjusted R Square .69195

Standard Error 1.52239

Analysis of Variance

DF Sum of Squares Mean Square

Regression 1 101.23205 101.23205 **C**

Residual 18 41.71795 2.31766 **D**

F = 43.67849 Signif F = .0000

‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑ Variables in the Equation ‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑

Variable B SE B Beta T Sig T

TENSE .900641 .136276 .841525 6.609 .0000

(Constant) .387821 .574947 .675 .5085

End Block Number 1 All requested variables entered.

Throughout this manual computer output will have the following characteristics:

(1) It will not contain "job stream" information.

(2) only those portions germane to the discussion will be presented.

(3) any added information will be in "bold face" print.

(4) the page breaks of SPSS will be removed.

(5) reference points of discussion will be identified by underlining and a letter inside a box toward the end of that line.

R squared is at "A" of the printout. Well since there are only two variables in our equation here, we have DEPRES and TENSE. It says multiple but its really unitary. Its only two variables simple at this point. Even though it says multiple the computer can change its mind on those things. So, but at any rate its .84.

We are now going to repeat the same procedure but change the X variable to GROUP. The important thing to notice is that GROUP is a dichotomous variable and TENSE was a continuous variable. It will also be shown how regression is like analysis in this case. First the data an computations are shown.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | C1 | C2 | C3 | C4 | C5 | C6 | C7 | C8 | C9 | C10 | C11 | C12 |
| A |  |  |  |  |  |  |  |  |  |  |  |  |
| *B* |  | *X* |  |  |  |  |  |  |  |  |  |  |
| C | actual score | deviation score | deviation squared | actual score | deviation score | deviation squared | cross product | predicted score | error:  actual minus predicted | error squared  or  within  squared | Y primed minus the mean of Y |  |
| D | observed score | little x | little x squared | observed score | little y | little y squared |  | Y primed | residual score |  |  |  |
| E | raw score:  group | X minus the mean | little x times little x | raw score:  depres | Y minus the mean | little y times little y | little x times little y | Y predicted from X | error in prediction |  |  |  |
| predictor variable (X)  independent varialbe (IV) | | | | criterion variable (Y)  dependent variable (DV) | | |  |  |  |  |  |  |
| F | 1  1  1  1  1  1  1  1  1  1  2  2  2  2  2  2  2  2  2  2 | ‑.50  ‑.50  ‑.50  ‑.50  ‑.50  ‑.50  ‑.50  ‑.50  ‑.50  ‑.50  .50  .50  .50  .50  .50  .50  .50  .50  .50  .50 | .25  .25  .25  .25  .25  .25  .25  .25  .25  .25  .25  .25  .25  .25  .25  .25  .25  .25  .25  .25 | 0  4  1  1  0  0  4  1  2  4  6  8  4  8  7  0  3  6  6  4 | -3.45  .55  -2.45  -2.45  -3.45  -3.45  .55  -2.45  -1.45  .55  2.55  4.55  .55  4.55  3.55  -3.45  ‑.45  2.55  2.55  .55 | 11.9025  .3025  6.0025  6.0025  11.9025  11.9025  .3025  6.0025  2.1025  .3025  6.5025  20.7025  .3025  20.7025  12.6025  11.9025  .2025  6.5025  6.5025  .3025 | 1.73  ‑.28  1.23  1.23  1.73  1.73  ‑.28  1.23  .73  ‑.28  1.28  2.28  .28  2.28  1.78  ‑1.73  ‑.23  1.28  1.28  .28 | 1.70  1.70  1.70  1.70  1.70  1.70  1.70  1.70  1.70  1.70  5.20  5.20  5.20  5.20  5.20  5.20  5.20  5.20  5.20  5.20 | ‑1.70  2.30  ‑.70  ‑.70  ‑1.70  ‑1.70  2.30  ‑.70  .30  2.30  .80  2.80  ‑1.20  2.80  1.80  ‑5.20  ‑2.20  .80  .80  ‑1.20 | 2.89  5.29  .49  .49  2.89  2.89  5.29  .49  .09  5.29  .64  7.84  1.44  7.84  3.24  27.04  4.84  .64  .64  1.44 | ‑1.75  ‑1.75  ‑1.75  ‑1.75  ‑1.75  ‑1.75  ‑1.75  ‑1.75  ‑1.75  ‑1.75  1.75  1.75  1.75  1.75  1.75  1.75  1.75  1.75  1.75  1.75 | 3.06  3.06  3.06  3.06  3.06  3.06  3.06  3.06  3.06  3.06  3.06  3.06  3.06  3.06  3.06  3.06  3.06  3.06  3.06  3.06 |
| G | 30 | 0 | 5 | 69 | 0 | 142.95 | 17.50 | 69.00 | 0 | 81.70 | 0 | 61.25 |
| H | 1.5 |  | Sum of x Squared | 3.45 |  | sum of y squared | sum of cross products |  |  | error sum of squares |  | between or regression sum of squares |
| I | mean of X |  | mean of Y |  |  |  |  |
| J |  |  |  |  |  | SST |  |  |  | SSE |  | SSB  SSR |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| # | Description | Label | Symbol | Notation | Full Notation or Formula | Individual Case Values |
| 1 | Observation of the varialbe specified by "i" case number and the "j" group or variable | observation |  |  |  | Case #1 = 1  Case #2 = 1  Case #3 = 1  Case #4 = 1 |
| 2 | Number of observations in each (jth) group (set) | number of cases |  |  |  | 20 |
| 3 | Sum of X for each (jth) groujp or set | Sum of X, Total, Sum |  |  |  | 30 |
| 4 | Mean for each (jth) group or set | Average, Mean |  |  |  | 1.5 |
| 5 | Deviation Score (X minus the mean) | Little x, deviation score |  |  |  | Case #1 = 1 - 1.5  Case #2 = 1 - 1.5  Case #3 = 1 - 1.5  Case # 4 =1 - 1.5 |
| 9 | Sum of little x |  |  |  |  | 0 |
| 10 | Sum of little x squared | Sum of Squares | SS |  |  | = ((-.50)\*(-.50)) + ((-.50) \* (-.50))+  ((-.50)\*(-.50)) + ((-.50)\*(-.50)) etc  = .25 + .25 + .25 + .25 + etc= 5.00 |
| 11 | Population variance of each (jth) group or set | Population variance |  |  |  | 5.00  --------- = .25  20 |
| 12 | Sample variance of each (jth) group or set | Sample variance |  |  |  |  |
| 13 | Population standard deviation for each (jth) group or set | Popultaion standard deviation |  |  |  |  |
| 14 | Sample standard deviation for each (jth) group or set | Sample standard deviation | SD |  |  |  |
| 15 | *z*-score for each (ijth) observation | *z-*score | *z* |  |  | 1 -1.5  case #1 = --------- = -1.00  .5  1 - 1.5  case #2 = ---------- = -1.00  .5  1 - 1.5  case #3 = ----------- = -1.00  .5  1 - 1.5  case #4 = ----------- = -1.00  .5 |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| # | Description | Label | Symbol | Notation | Full Notation or Formula | Individual Case Values |
| 1 | Observation of the varialbe specified by "i" case number and the "j" group or variable | observation |  |  |  | case #1 = 0  case #2 = 4  case #3 = 1  case #4 = 1 |
| 2 | Number of observations in each (jth) group (set) | number of cases |  |  |  | 20 |
| 3 | Sum of Y for each (jth) groujp or set | Sum of Y, Total, Sum |  |  |  | 69 |
| 4 | Mean for each (jth) group or set | Average, Mean |  |  |  | 3.45 |
| 5 | Deviation Score (Y minus the mean) | Little y, deviation score |  |  |  | case #1 = 0-3.45 = -3.45  case #2 = 4-3.45 = .55  case #3 = 1-3.45 = -2.45  case #4 = 1-3.45 = -2.45 |
| 9 | Sum of little y |  |  |  |  | 142.95 |
| 10 | Sum of little y squared | Sum of Squares | SS |  |  | ((-3.45)\*(-3.45))+(.55\*.55)+  ((-2.45)\*(-2.45))+((-2.45)\*(-2.45))+  etc  =11.90 + .3025 + 6.0025+6.0025 +etc |
| 11 | Population variance of each (jth) group or set | Population variance |  |  |  | 142.95  -------- = 7.148  20 |
| 12 | Sample variance of each (jth) group or set | Sample variance |  |  |  |  |
| 13 | Population standard deviation for each (jth) group or set | Popultaion standard deviation |  |  |  |  |
| 14 | Sample standard deviation for each (jth) group or set | Sample standard deviation | SD |  |  |  |
| 15 | *z*-score for each (ijth) observation | *z-*score | *z* |  |  | 0 - 3.45  case #1 = ---------- = -1.291  2.673  4 - 3.45  case #2 = ---------- = .206  2.673  1 - 3.45  case #3 = ---------- = -.917  2.673  1 - 3.45  case #4 = ---------- = -.917  2.673 |
| 31 | Cross products for two groups (sets) of scores | Cross Products |  |  |  | ((-.50)\*(-3.45))+(-.50\*.55)+  ((-.50)\*(-2.45))+((-.50)\*(-2.45))+  etc = 17.50 |
| 32 | Covariance | Covariance |  |  |  | 17.50  -------- = .875  20 |
| 33 | Slope of the regression line | slope,regression coefficient, *b* | *b* |  |  | 17.50  ------- = 3.50  5.00 |
| 33a | Standardized slope of the regression line (slope when X and Y are z-scores) | standardized regression coefficient, beta |  |  |  |  |
| 34 | The value of Y at point where X is zero (0) | Constant, *a*, intercept | *a* |  |  | 3.45 - (3.50 \* 1.50) = -1.80 |
| 35 | Y predicted from X | Y hat, Y primed, Y' |  |  |  | case #1=-1.80+(3.50\*1)=1.70  case #2=-1.80+(3.50\*1)=1.70  case #3=-1.80+(3.50\*1)=1.70  case #4=-1.80+(3.50\*1)=1.70 |
| 36 | Regression sum of squares | sum of squares, between sum of squares | SSR, SSB |  | | 17.50 \* 17.50  ------------------ = 61.25  5.00 |
| 37 | Residual sum of squres | Error, witin sum of squares | SSE  SSW |  | | SST - SSR or SST - SSB  142.95 - 61.25 = 81.70 |
| case #1 = (0-1.70)\*(0-1.70)=2.89  case #2 = (4-1.70)\*(4-1.70)=5.29  case #3 = (1-1.70)\*(1-1.70)=.49  case #4 = (1- .70)\*(1-1.70)= ..49 |
| 38 | Total sum of squares | Total sum of squares | SST |  |  | 142.94 |
| 39 | Pearson correlation coefficient | correlation | *r* |  |  |  |

|  |
| --- |
| File Name = crsplt9.sps |
| get file = '\rdda\crsleq1.sav'.  GRAPH  /SCATTERPLOT(BIVAR)=group WITH depres  /MISSING=LISTWISE . |



This plot represents two variables DEPRES and GROUP. There were three people in GROUP # 1 who answered 0 to the question of "sad or depressed." If you look back at the raw data you will that was participants 1, 5, and 6. There was one person in GROUP # 2 that answered the question as 0. In looking at the raw data you will see that it was person # 16. There were two people in GROUP # 2 that answered the question as 8. There were person number # 12 and person # 14. This scattergram represents all the people of both groups. Once again the scattergram represents a relationship. The smaller going with the small and the large with the large. People in GROUP # 1 gave responses which were smaller and people in GROUP # 2 (2 is larger than one) gave responses which were larger than those in GROUP # 1.







|  |
| --- |
| File Name = crsreg13.sps |
| get file = '\rdda\crsleq1.sav'.  REGRESSION  descriptives = mean stddev cov corr n  /MISSING LISTWISE  /STATISTICS COEFF OUTS R ANOVA  /CRITERIA=PIN(.05) POUT(.10)  /NOORIGIN  /DEPENDENT depres  /METHOD=ENTER group . |

\* \* \* \* M U L T I P L E R E G R E S S I O N \* \* \* \*

Listwise Deletion of Missing Data

Mean Std Dev Label

DEPRES 3.450 2.743

GROUP 1.500 .513

N of Cases = 20

Correlation, Covariance:

DEPRES GROUP

DEPRES 1.000 .655

7.524 .921

GROUP .655 1.000

.921 .263

\* \* \* \* M U L T I P L E R E G R E S S I O N \* \* \* \*

Equation Number 1 Dependent Variable.. DEPRES

Descriptive Statistics are printed on Page 2

Block Number 1. Method: Enter GROUP

Variable(s) Entered on Step Number

1.. GROUP

Multiple R .65458

R Square .42847

Adjusted R Square .39672

Standard Error 2.13047

Analysis of Variance

DF Sum of Squares Mean Square

Regression 1 61.25000 61.25000

Residual 18 81.70000 4.53889

F = 13.49449 Signif F = .0017

‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑ Variables in the Equation ‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑

Variable B SE B Beta T Sig T

GROUP 3.500000 .952774 .654577 3.673 .0017

(Constant) ‑1.800000 1.506468 ‑1.195 .2477

End Block Number 1 All requested variables entered.

We now have the sum of squares, between, sum to squares within, and sum of squares total. Now if you divide the sums of squares total into the sums of squares between, you get R squared. And so if you take the square root of that then you get the correlation.

|  |
| --- |
| File Name = crsano1.sps |
| get file = '\rdda\crsleq1.sav'.  ANOVA  VARIABLES=depres  BY group(1 2)  /MAXORDERS ALL  /STATISTICS MEAN  /METHOD EXPERIM  /FORMAT LABELS . |

\* \* \* C E L L M E A N S \* \* \*

DEPRES

by GROUP

Total Population

3.45

( 20)

GROUP

1 2

1.70 5.20

( 10) ( 10)

\* \* \* A N A L Y S I S O F V A R I A N C E \* \* \*

DEPRES

by GROUP

EXPERIMENTAL sums of squares

Covariates entered FIRST

Sum of Mean Sig

Source of Variation Squares DF Square F of F

Main Effects 61.250 1 61.250 13.494 .002

GROUP 61.250 1 61.250 13.494 .002

Explained 61.250 1 61.250 13.494 .002

Residual 81.700 18 4.539

Total 142.950 19 7.524

|  |
| --- |
| File Name = crsdsc3.sps |
| get file = '\rdda\crsleq1.sav'.  DISCRIMINANT  /GROUPS=group(1 2)  /VARIABLES=depres  /ANALYSIS ALL  /PRIORS EQUAL  /STATISTICS=MEAN STDDEV TABLE  /CLASSIFY=NONMISSING POOLED . |

Number of cases by group

Number of cases

GROUP Unweighted Weighted Label

1 10 10.0

2 10 10.0

Total 20 20.0

Group means

GROUP DEPRES

1 1.70000

2 5.20000

Total 3.45000

Group standard deviations

GROUP DEPRES

1 1.70294

2 2.48551

Total 2.74293

‑ ‑ ‑ ‑ ‑ ‑ ‑ ‑ D I S C R I M I N A N T A N A L Y S I S ‑ ‑ ‑ ‑ ‑ ‑ ‑ ‑

On groups defined by GROUP

Analysis number 1

Direct method: all variables passing the tolerance test are entered.

Minimum tolerance level.................. .00100

Canonical Discriminant Functions

Maximum number of functions.............. 1

Minimum cumulative percent of variance... 100.00

Maximum significance of Wilks' Lambda.... 1.0000

Prior probability for each group is .50000

Canonical Discriminant Functions

Pct of Cum Canonical After Wilks'

Fcn Eigenvalue Variance Pct Corr Fcn Lambda Chi‑square df Sig

: 0 .571529 9.790 1 .0018

1\* .7497 100.00 100.00 .6546 :

\* Marks the 1 canonical discriminant functions remaining in the analysis.

Standardized canonical discriminant function coefficients

Func 1

DEPRES 1.00000

Structure matrix:

Pooled within‑groups correlations between discriminating variables

and canonical discriminant functions

(Variables ordered by size of correlation within function)

Func 1

DEPRES 1.00000

Canonical discriminant functions evaluated at group means (group centroids)

Group Func 1

1 ‑.82142

2 .82142

Classification results ‑

No. of Predicted Group Membership

Actual Group Cases 1 2

‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑‑ ‑‑‑‑‑‑ ‑‑‑‑‑‑‑‑ ‑‑‑‑‑‑‑‑

Group 1 10 7 3

70.0% 30.0%

Group 2 10 2 8

20.0% 80.0%

Percent of "grouped" cases correctly classified: 75.00%-------------------------------------------

Y' becomes the general formula for all parametric statistics or what is refered to as the general linear model. It follows

*Y' = a + b1X1 + b2X2 + ... bnXn*

*In this section it is noted that it is the fundamental calculation of the correlation, the t-test, ANOVA, and discriminant function. Further, it can be used to express a relationship, a prediction, and with conversions a test of statistical significance. Since more dimensions or weights can be used the contribution of the added dimensions or variables can be assessed. This allows one to test or falsifiy ones theory.*