Reliability

Four types of reliability will be discussed in terms of classical test theory. The four types discussed are: (1) split-half, (2), parallel forms, (3) test-retest, and (4) internal consistency. These four methods can be reduced to basically two methods: (1) the reliability coefficient (split-half, parallel forms, and test-retest) and the standard error of measurement (internal consistency). The reliability coefficient assesses the degree that one test or part of test can predict another and uses some correlational method. The standard error of measurement assesses the degree that an individuals scores varies over parallel tests and uses ANOVA methods (although it has been shown in chapters 2 and 3 that these methods are different parts of the general linear model).

A notion of parallel tests is needed understand this section. A test is made up of items that are designed to measure psychological attributes. A test designed to measure a single attribute is called a univariate test --all items on the test are intended to measure the same thing. The Beck Depression Scale is such a test -- it is designed to measure the attribute of depression. A test that is designed to measure more than one attribute is called a multivariate test -- the MMPI is such a test with its many subscales. The idea of parallel tests is that two tests measure the same attribute. Since they measure the same thing they are identical or parallel. You should note that the split-half and parallel tests are similar. When a tests is split into two part (split-half) the two halves become parallel tests (in this chapter they can be thought of as the same concept). It is this notion of parallel items (or interchangable items) that we are testing when we assess reliability.

Reliability is a problem because psychological characteristics can't be measured perfectly. When considering psychological attributes, there is considerable unreliability. And the error in measurement is a problem that you must deal with in some way.

The Psychosocial Assessment Scale is a multivariate test since there are six subscales. But within those subscale, those items are univariate within those factors or subtests. There are six subscales on the PAS, however, 2 of the subtests have only 1 item. It may be debatable whether a subscale with only 1 item is really a subscale. The split-half and coefficient alpha cannot be used when there is only 1 item on a subscale. Each subscale is considered as a scale itself when assessing reliability.

In this example we will again use the data from the Psychosocial Assessment Scale (PAS).

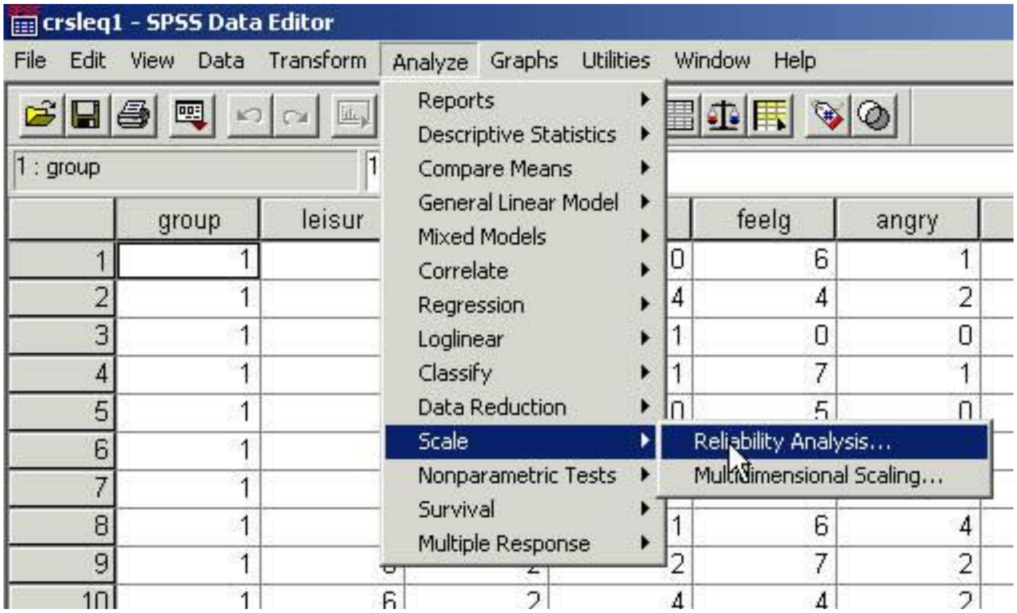
Click here to see the PAS. Click "Back" to return here.
Another view

Definition of reliability

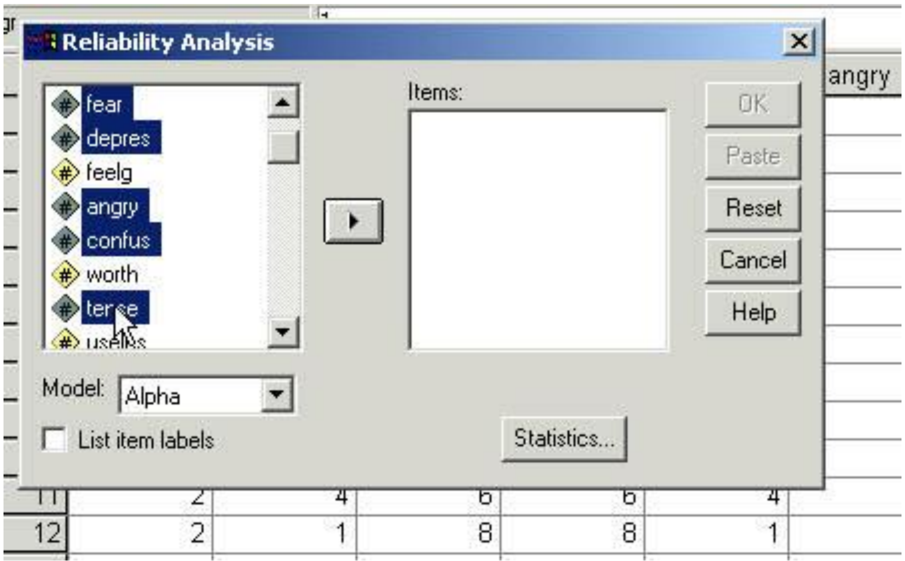
Split-half Method

Using this method you take half the items, and correlate them with the other half and that correlation is the index of reliability. The assumption is that all the items are measuring a single variable. Because the items should be comparable they should be considered interchangeable. It should be noted, however, that is this interchanability that we are testing when we test reliability.

Only one subscales of the questionnaire is assessed in this problem (it takes up too much paper to do them all -- in the coefficient alpha below all subscales are assessed). The following "click" procedure with produce a syntax file that we will change slightly for our purposes. The "click" method will not give us exactly what we want.

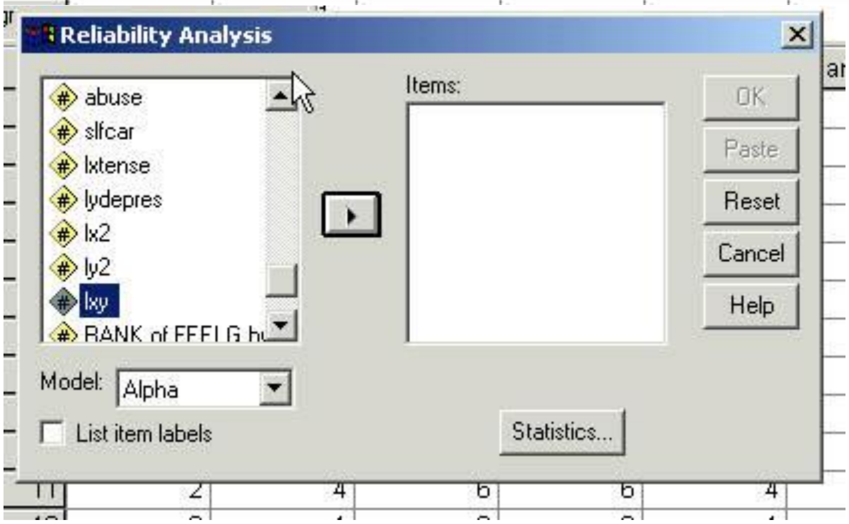


As you can see in the upper left hand corner we are using a file called crsleq1. It is an .sav file.
Click Analyze.
Select Scale
Click Reliability Analysis.
The following window appears:



Model: Alpha

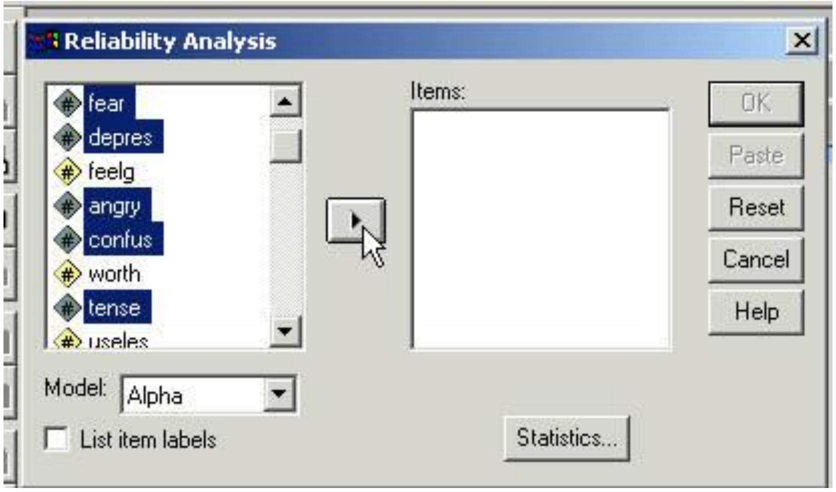
□ List item labels



Model: Alpha

□ List item labels

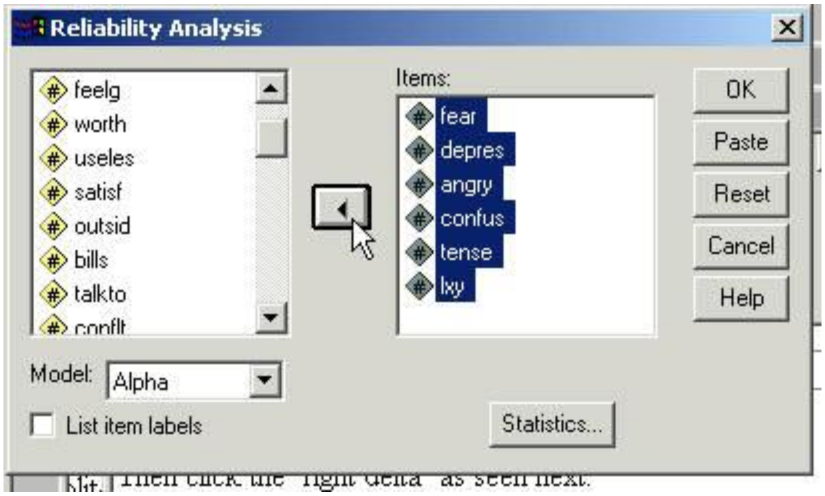
Select fear, depres, angry, confus, and tense by holding down the Ctrl key and clicking those variables and seen in the next screen.



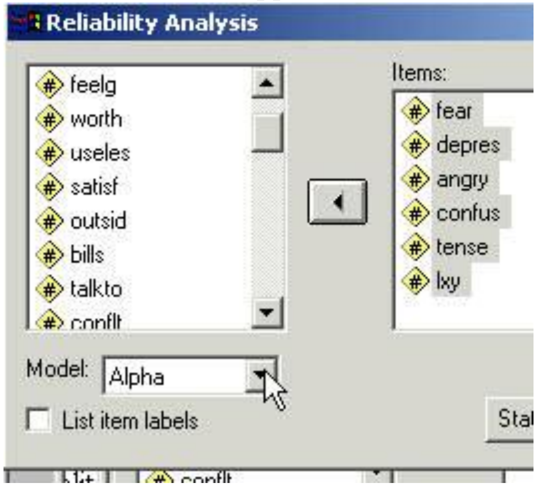
Model: Alpha

□ List item labels

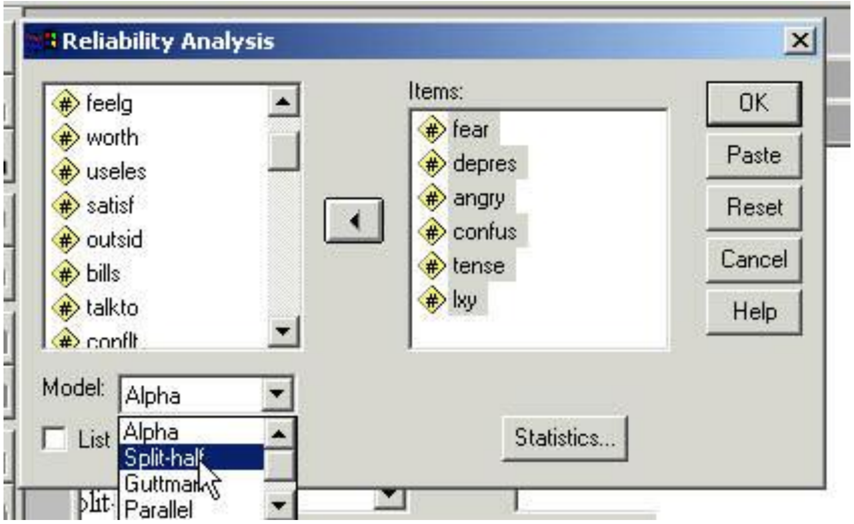
Then click the "right delta" as seen next.



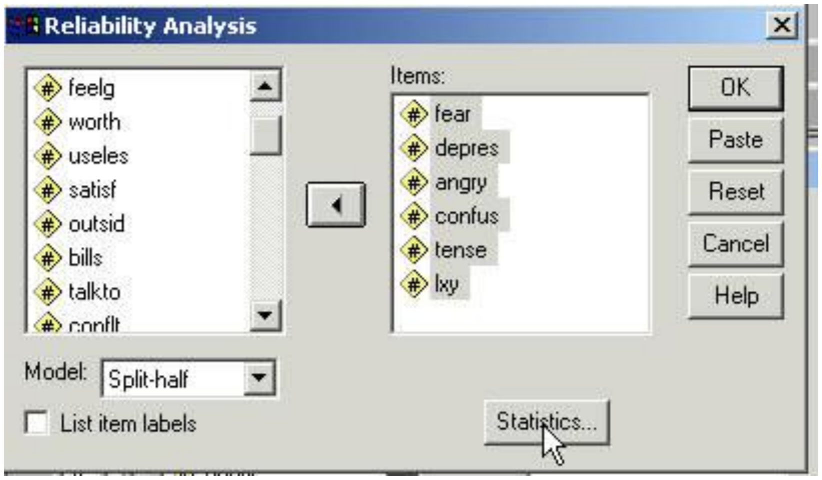
The variables will appear in the "Items:" window -- next.



Click the "Pull Down" box that has Alpha in it.



Click Split-Half

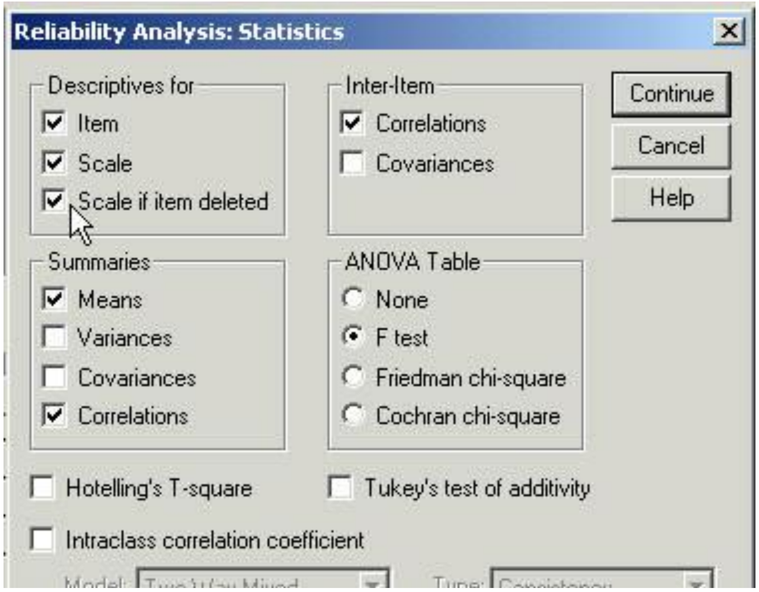


Items:
fear
depres
angry
confus
tense
lxy

Model: Split-half

□ List item labels

Click Statistics



Descriptives for
☑ Item
☑ Scale
☑ Scale if item deleted

Inter-Item
☑ Correlations
□ Covariances

Summaries
☑ Means
□ Variances
□ Covariances
☑ Correlations

ANOVA Table
○ None
● F test
○ Friedman chi-square
○ Cochran chi-square

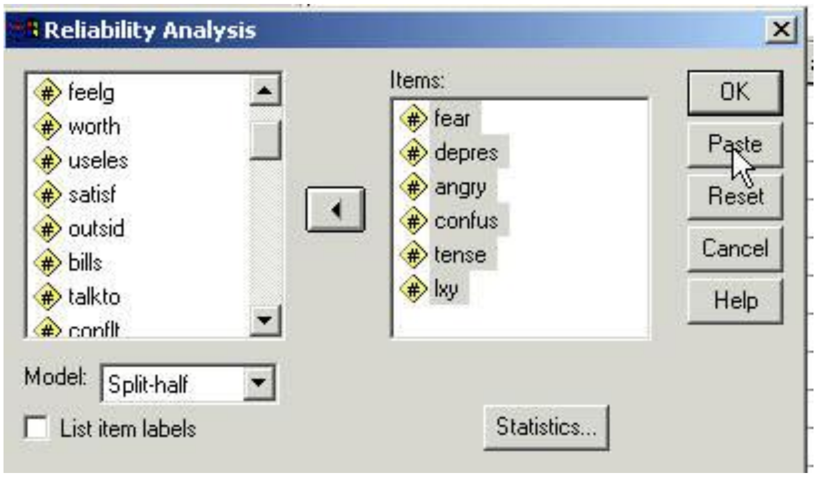
□ Hotelling's T-square

□ Tukey's test of additivity

□ Intraclass correlation coefficient

Model: Two-Way Mixed Type: Consistency

Check Item, Scale, Scale if item deleted, Correlations, Means, and a second Correlations. See next screen. Then click Continue.

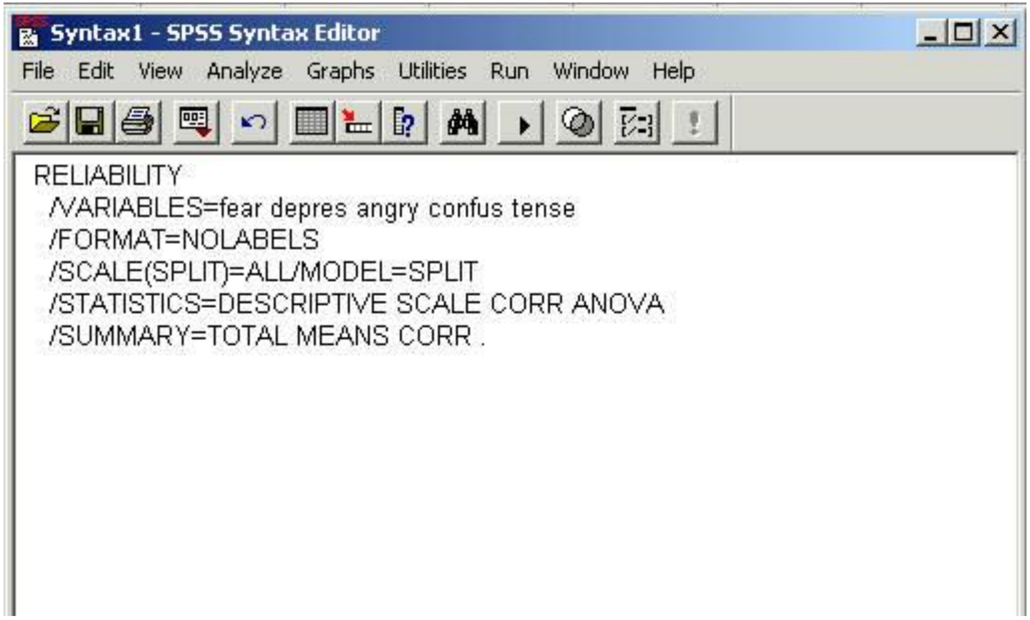


Items:
fear
depres
angry
confus
tense
lxy

Model: Split-half

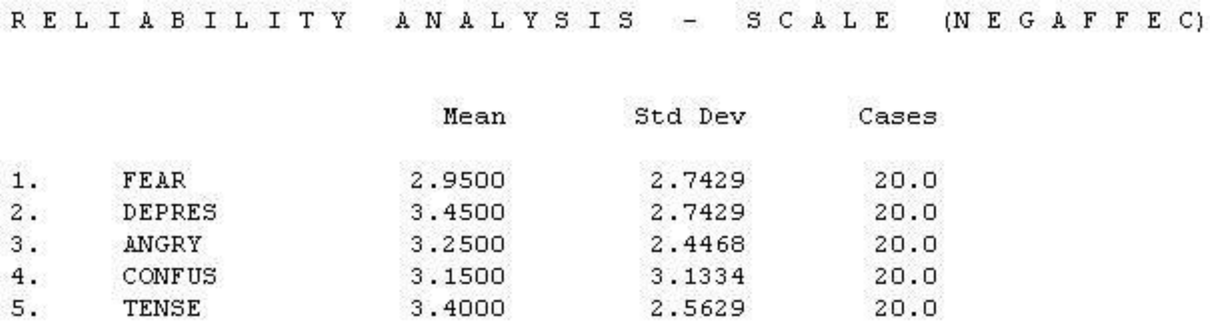
□ List item labels

Click Paste
The following Syntax File opens.
The following changes need to be made to that file:
Where it says "/SCALE(SPLIT)=ALL/MODEL=SPLIT" you need to change it so it reads as follows:
"/SCALE(NegAffect)=ALL/MODEL=SPLIT".
Now in the printout the scale of fear, depres, angry, confuse, and tense will be lableled NegAffect.



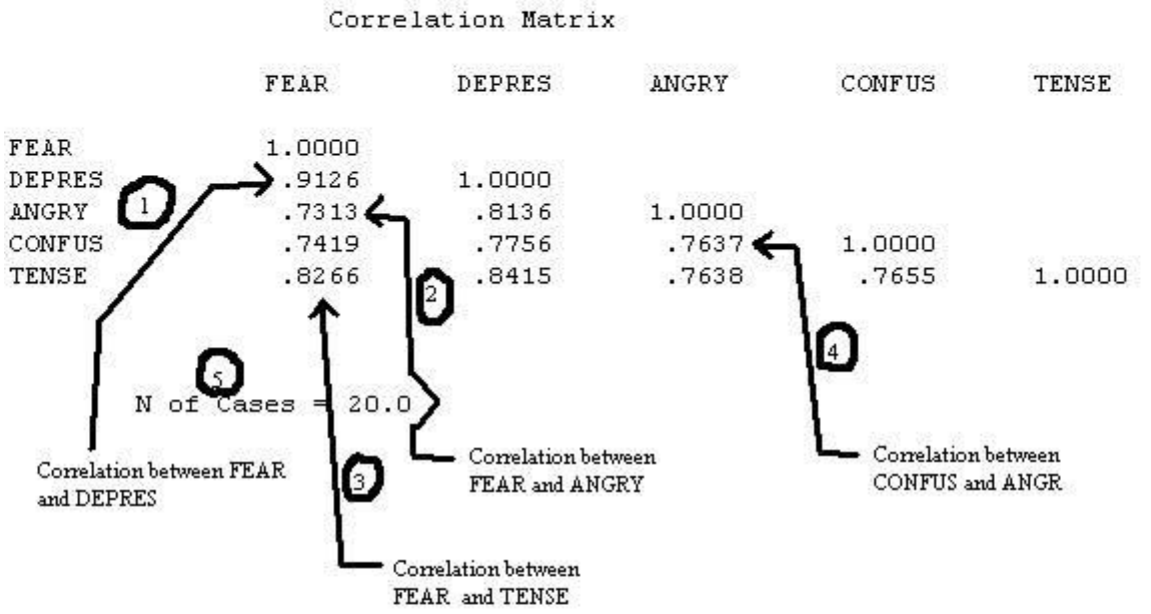
```
RELIABILITY
 /VARIABLES=fear depres angry confus tense
 /FORMAT=NOLABELS
 /SCALE(SPLIT)=ALL/MODEL=SPLIT
 /STATISTICS=DESCRIPTIVE SCALE CORR ANOVA
 /SUMMARY=TOTAL MEANS CORR .
```

Save and Run the Syntax File
Saving a Syntax File
Submitting a Syntax File -- Running an SPSS program



```
R E L I A B I L I T Y   A N A L Y S I S   -   S C A L E   (N E G A F F E C)


                              Mean        Std Dev       Cases

 1.     FEAR                2.9500         2.7429        20.0
 2.     DEPRES              3.4500         2.7429        20.0
 3.     ANGRY               3.2500         2.4468        20.0
 4.     CONFUS              3.1500         3.1334        20.0
 5.     TENSE               3.4000         2.5629        20.0
```

This first part of the output seems pretty much self descriptive. The name of the subscale is NEGAFFEC. It cut off the t of affect. It was supposed to be NegAffect for negative affect. The means, standard deviations and number of cases for each variable seem clear.



```
                    Correlation Matrix

                FEAR         DEPRES       ANGRY        CONFUS       TENSE

FEAR            1.0000
DEPRES           .9126       1.0000
ANGRY            .7313        .8136       1.0000
CONFUS           .7419        .7756        .7637       1.0000
TENSE            .8266        .8415        .7638        .7655       1.0000

        N of Cases =    20.0
```

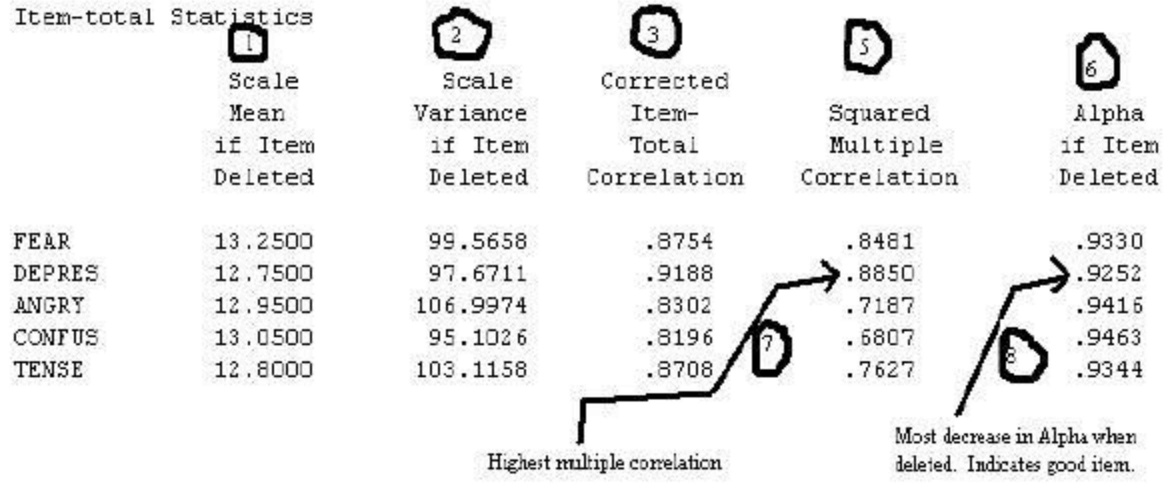
1
2
3
4
5
Correlation between FEAR and DEPRES
Correlation between FEAR and ANGRY
Correlation between FEAR and TENSE
Correlation between CONFUS and ANGR

In the next section of the output there is the correlation matrix and the number of cases used in the computation. The number is the correlation between FEAR and DEPRES.

The number shows the number of cases.

1

5



```
Item-total Statistics

              Scale        Scale       Corrected
              Mean        Variance       Item-       Squared       Alpha
             if Item      if Item        Total       Multiple     if Item
             Deleted      Deleted     Correlation   Correlation   Deleted

FEAR         13.2500      99.5658        .8754         .8481        .9330
DEPRES       12.7500      97.6711        .9188         .8850        .9252
ANGRY        12.9500     106.9974        .8302         .7187        .9416
CONFUS       13.0500      95.1026        .8196         .6807        .9463
TENSE        12.8000     103.1158        .8708         .7627        .9344
```

1
2
3
5
6
7
8
Highest multiple correlation
Most decrease in Alpha when deleted. Indicates good item.

In this next section of output item characteristics are identified by what happens when they are dropped. Good items are identified by what happens when they are good. Things go downhill when the best players leave a team.

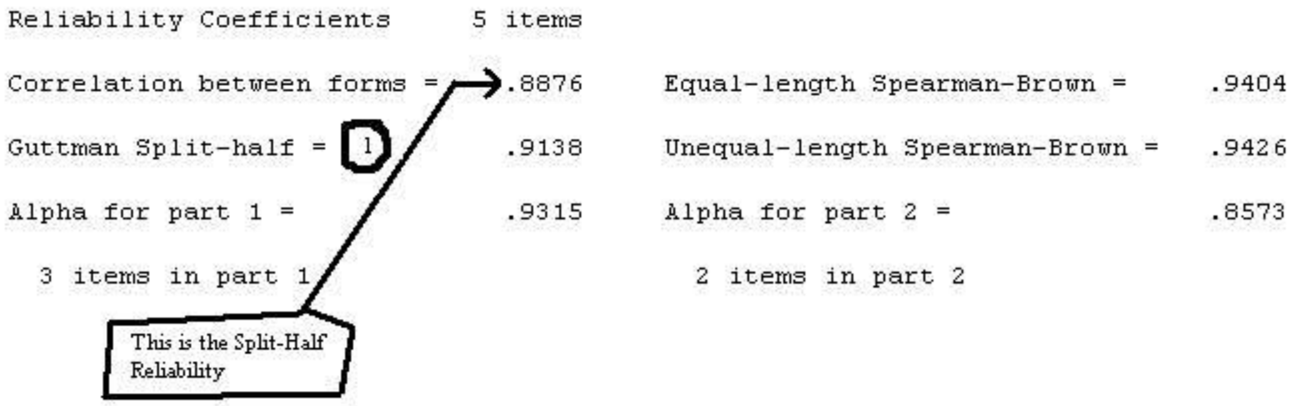
The most useful might be column What happens to the Alpha value when the item is dropped from the subtest?

6

We see that the Alpha goes down the most when DEPRES is dropped from the subtest. On this characteristic it would be considered the best it.
DEPRES is also seen to have the highest Squared Multiple Correlation.

5

Another characteristic of a good item.



```
Reliability Coefficients     5 items

Correlation between forms =   .8876     Equal-length Spearman-Brown =     .9404

Guttman Split-half =          .9138     Unequal-length Spearman-Brown =   .9426

Alpha for part 1 =            .9315     Alpha for part 2 =                .8573

  3 items in part 1                       2 items in part 2
```

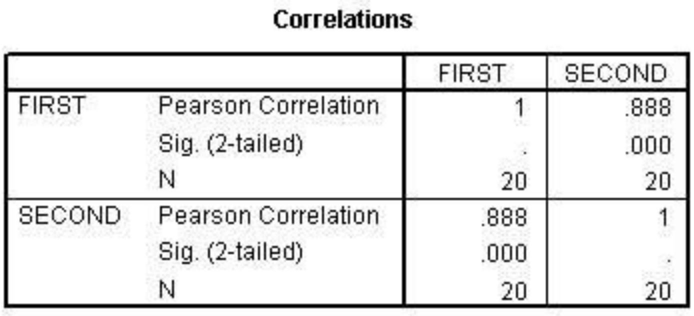
1

This is the Split-Half Reliability

Below number is where we are trying to get to. It is the Split-Half Reliability.

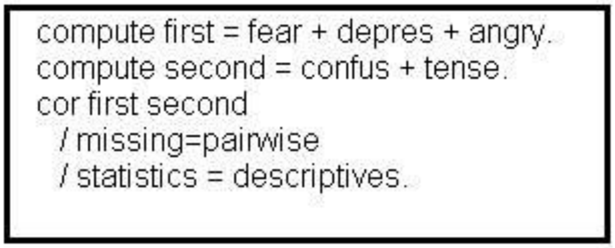
1

The Split-Half Reliability coefficient is the same as summing the items of the first half and summing the second half and then correlating the two results. This is demonstrated in the next example.



Correlations

		FIRST	SECOND
FIRST	Pearson Correlation	1	.888
	Sig. (2-tailed)	.	.000
	N	20	20
SECOND	Pearson Correlation	.888	1
	Sig. (2-tailed)	.000	.
	N	20	20



```
compute first = fear + depres + angry.
compute second = confus + tense.
cor first second
  / missing=pairwise
  / statistics = descriptives.
```

The above syntax produces the following output. Notice that the correlation is the same (after rounding) as the Split-Halt Reliability above. Consequently, it is the split-halt reliability.

Coefficient Alpha Method

In this example we will again use the data from the Psychosocial Assessment Scale.

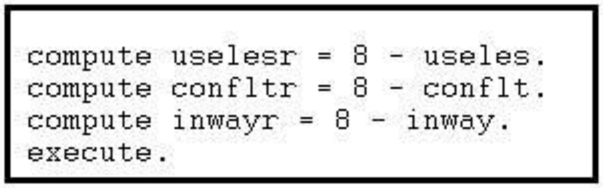
Click here to see the scale and data

The subscales in this example are:

1. Negative Emotion made up of items FEAR, DEPRES, ANGRY, CONFUS and TENSE.
2. Quality of Life made up of LEISUR, FEELG, WORTH, SATISF and USELES. [USELES would be reversed.]
3. Human Contact made up of OUTSID, TALKTO, CONFLT, and SUPPRT. [CONFLT would be reversed.]
4. Job or Employment made up EMPLOY, GOODJ, LIKEW and INWAY. [INWAY would be reversed.]

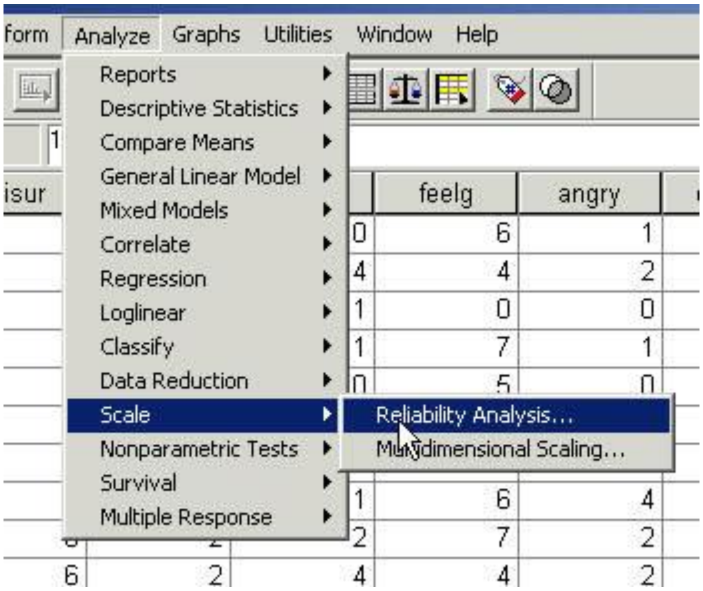
For a discussion of reversed items click here.

For our purposes here the items USELES, CONFLT and INWAY need to be reversed. The following syntax provides that reversal.

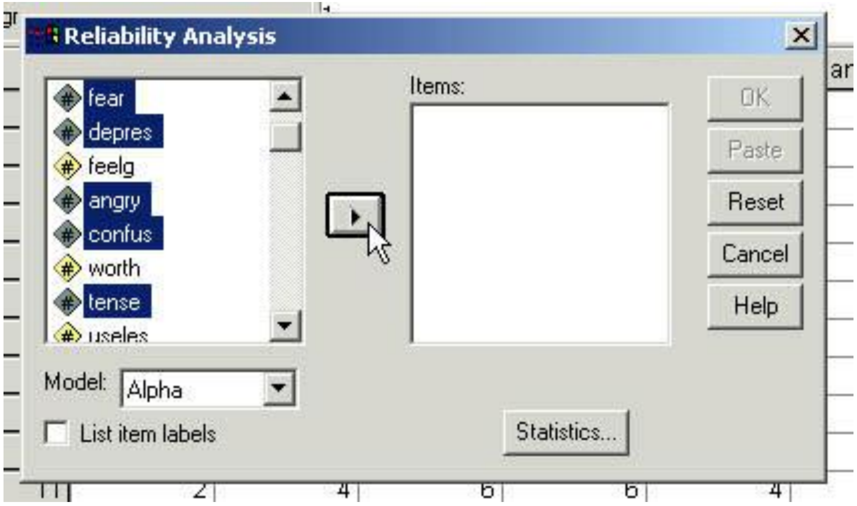


```
compute uselesr = 8 - useles.
compute confltr = 8 - conflt.
compute inwayr = 8 - inway.
execute.
```

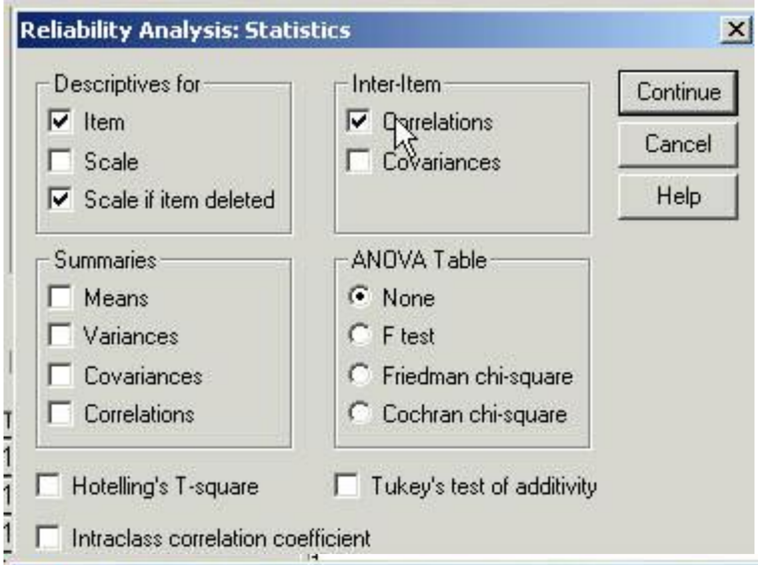
Now instead of use the variables USELES, CONFLT AND INWAY one should the variables USELESR, CONFLTR and INWAYR if the subscales are positive.
Chronbach's Alpha is run in the following way:



Click Analyze
Select Scale
Click Reliability Analysis...



Select the variables for the subscale and click the "right delta"



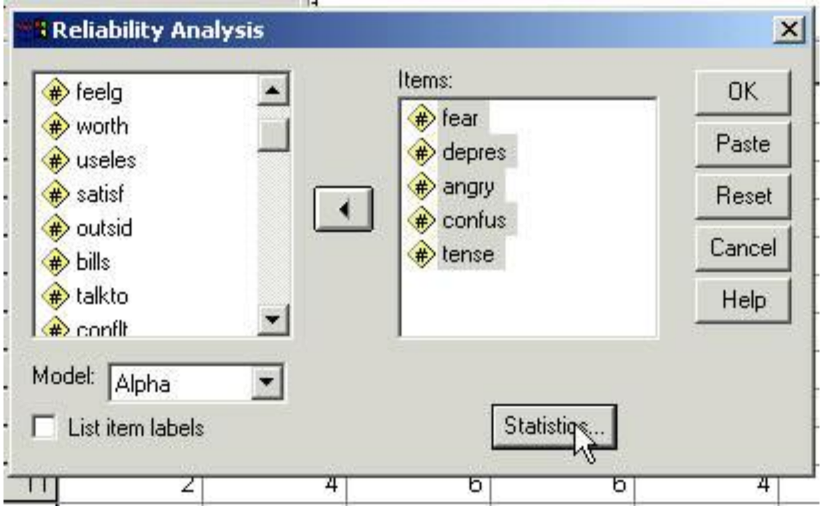
Descriptives for
☑ Item
☐ Scale
☑ Scale if item deleted

Inter-Item
☑ Correlations
☐ Covariances

Summaries
☐ Means
☐ Variances
☐ Covariances
☐ Correlations

ANOVA Table
● None
○ F test
○ Friedman chi-square
○ Cochran chi-square

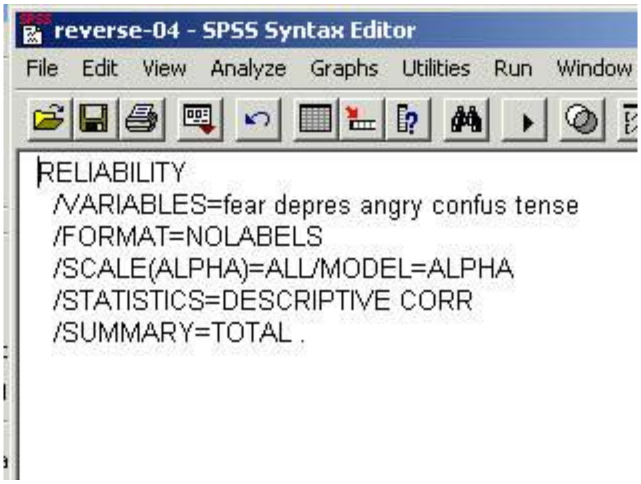
☐ Hotelling's T-square
☐ Tukey's test of additivity
☐ Intraclass correlation coefficient



Items:
fear
depres
angry
confus
tense

Model: Alpha
☐ List item labels

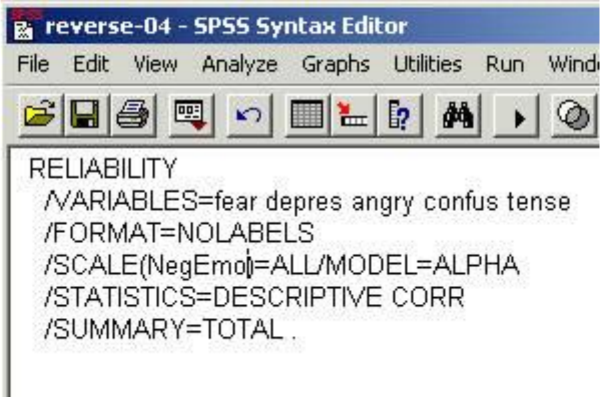
Click Statistics
Check Item, Scale if item deleted and Correlations
Then Click Continue
Click Paste



```
RELIABILITY
  /VARIABLES=fear depres angry confus tense
  /FORMAT=NOLABELS
  /SCALE(ALPHA)=ALL/MODEL=ALPHA
  /STATISTICS=DESCRIPTIVE CORR
  /SUMMARY=TOTAL .
```

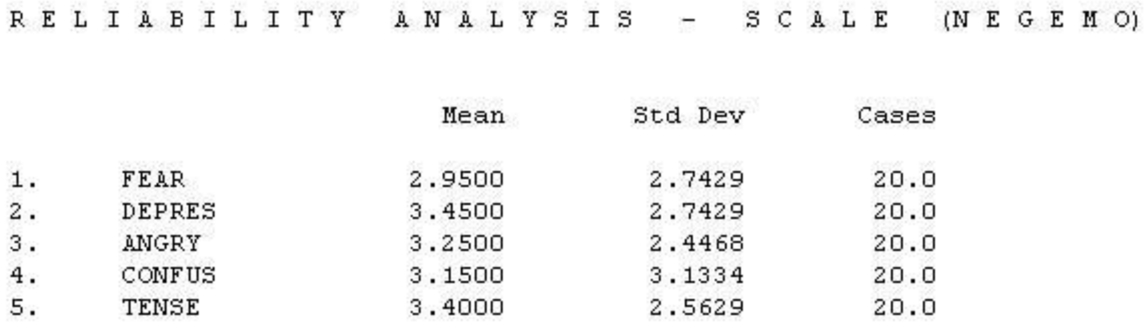
The Paste produces the following syntax file.

Change the /SCALE(ALPHA)=ALL
So that it looks like the following with the name of your subtest.


reverse-04 - SPSS Syntax Editor
File Edit View Analyze Graphs Utilities Run Wind

```
RELIABILITY
 /VARIABLES=fear depres angry confus tense
 /FORMAT=NOLABELS
 /SCALE(NegEmo)=ALL/MODEL=ALPHA
 /STATISTICS=DESCRIPTIVE CORR
 /SUMMARY=TOTAL .
```

I changed this one for negative emotions (NegEmo) see above.
Run the syntax file.
Saving a Syntax File
Submitting a Syntax File -- Running an SPSS program

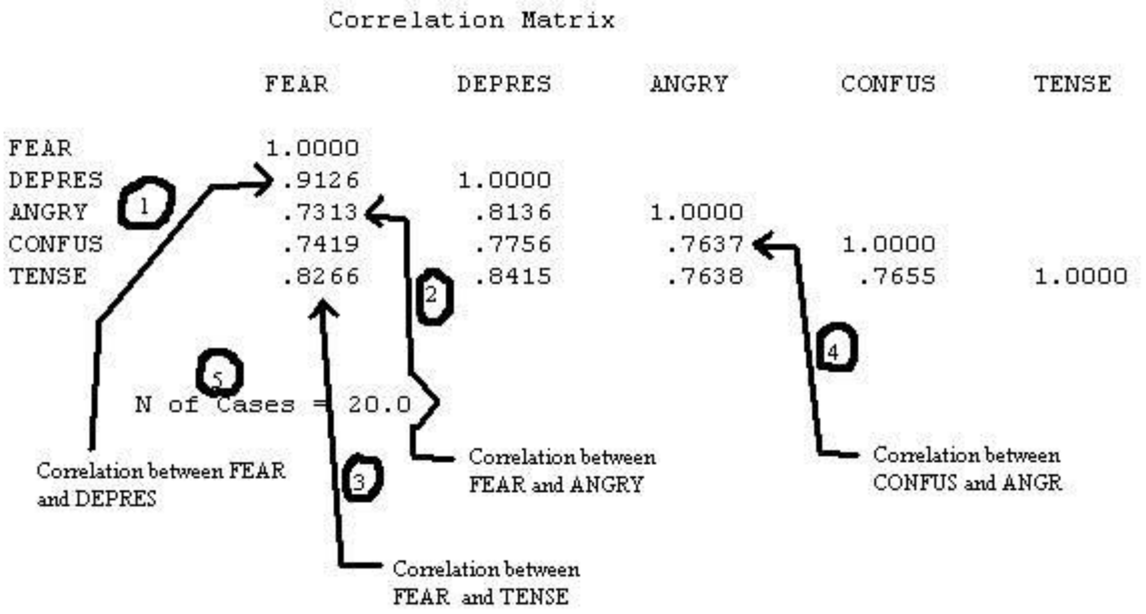


```
R E L I A B I L I T Y   A N A L Y S I S   -   S C A L E   (N E G E M O)


                            Mean        Std Dev       Cases

1.     FEAR               2.9500        2.7429        20.0
2.     DEPRES             3.4500        2.7429        20.0
3.     ANGRY              3.2500        2.4468        20.0
4.     CONFUS             3.1500        3.1334        20.0
5.     TENSE              3.4000        2.5629        20.0
```

The first part of the output is descriptive.

In the next section of the output there is the correlation matrix and the number of cases used in the computation. The number is the correlation between FEAR and DEPRES.

The number shows the number of cases.



```
                    Correlation Matrix

                FEAR        DEPRES      ANGRY       CONFUS      TENSE

FEAR           1.0000
DEPRES          .9126       1.0000
ANGRY           .7313        .8136      1.0000
CONFUS          .7419        .7756       .7637      1.0000
TENSE           .8266        .8415       .7638       .7655      1.0000


        N of Cases =    20.0
```

1
2
3
4
5
Correlation between FEAR and DEPRES
Correlation between FEAR and ANGRY
Correlation between FEAR and TENSE
Correlation between CONFUS and ANGR

1

5

In this next section of output item characteristics are identified by what happens when they are dropped. Good items are identified by what happens when they are good. Things go downhill when the best players leave a team.

The most useful might be column What happens to the Alpha value when the item is dropped from the subtest?

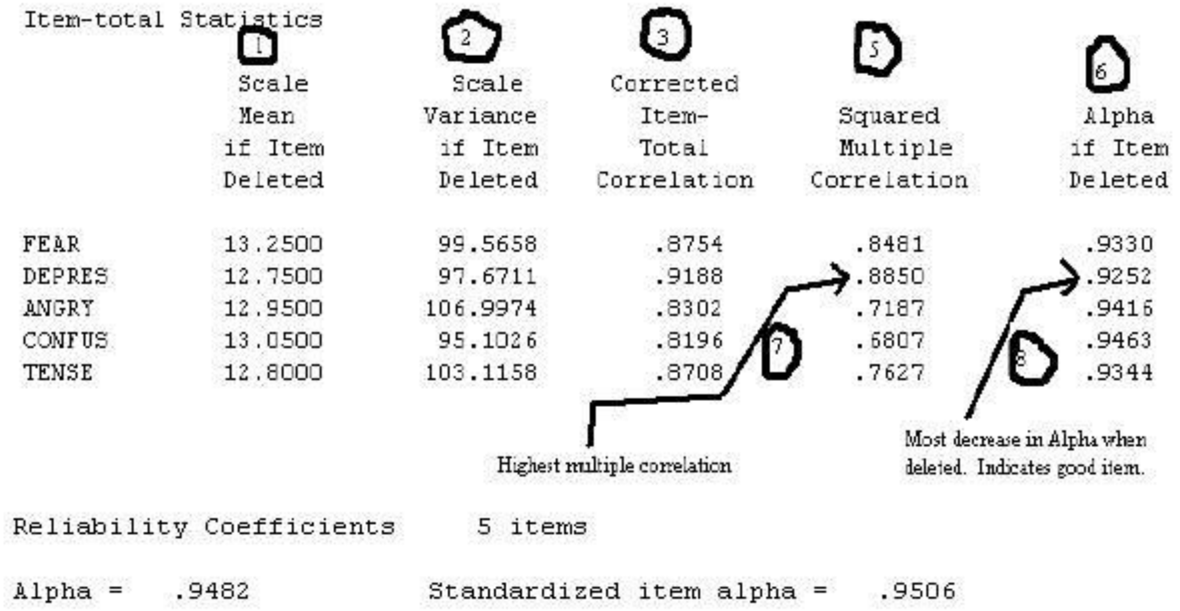
6

We see that the Alpha goes down the most when DEPRES is dropped from the subtest. On this characteristic it would be considered the best it.

DEPRES is also seen to have the highest Squared Multiple Correlation.

5

Another characteristic of a good item.



```
Item-total Statistics

               Scale          Scale      Corrected
               Mean          Variance      Item-         Squared         Alpha
              if Item        if Item       Total         Multiple       if Item
              Deleted        Deleted    Correlation    Correlation      Deleted

FEAR          13.2500        99.5658       .8754          .8481          .9330
DEPRES        12.7500        97.6711       .9188          .8850          .9252
ANGRY         12.9500       106.9974       .8302          .7187          .9416
CONFUS        13.0500        95.1026       .8196          .6807          .9463
TENSE         12.8000       103.1158       .8708          .7627          .9344

Reliability Coefficients     5 items

Alpha =   .9482            Standardized item alpha =   .9506
```

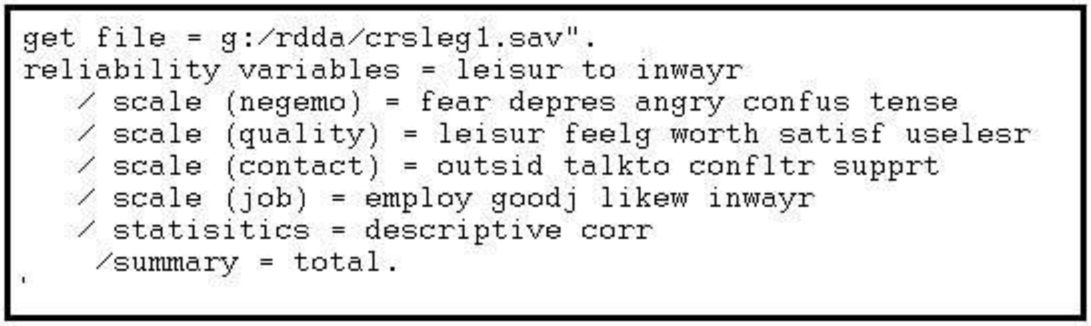
Highest multiple correlation

Most decrease in Alpha when deleted. Indicates good item.

Finally we are looking for the Reliability Coefficient of Alpha.

The Standardized Item Alpha first converts the scores of each respondent to a standard score before computing the coefficient alpha.

When there is more than one subtest it might be more efficient to use a syntax than the "clicking methdo." Below is a syntax file that will run reliabilities on four subtests. Notice that the reversed variables are used (uselesr, confltr, and inwayr).



```
get file = g:/rdda/crsleg1.sav".
reliability variables = leisur to inwayr
   / scale (negemo) = fear depres angry confus tense
   / scale (quality) = leisur feelg worth satisf uselesr
   / scale (contact) = outsid talkto confltr supprt
   / scale (job) = employ goodj likew inwayr
   / statisitics = descriptive corr
    /summary = total.
```

Some items need to be reversed when the item stem implies an opposing direrection. For example, in the "Quality" scale the items LEISUR, FEELG, WORTH, SATISF are in the positive direction (LEISUR--have you felt good about your leisure hours? FEELG -- have you felt good about things you have done?) while USELES was negative (USELES -- have you felt useless?). USELES is reversed and the reversed result is put in the variable USELESR.

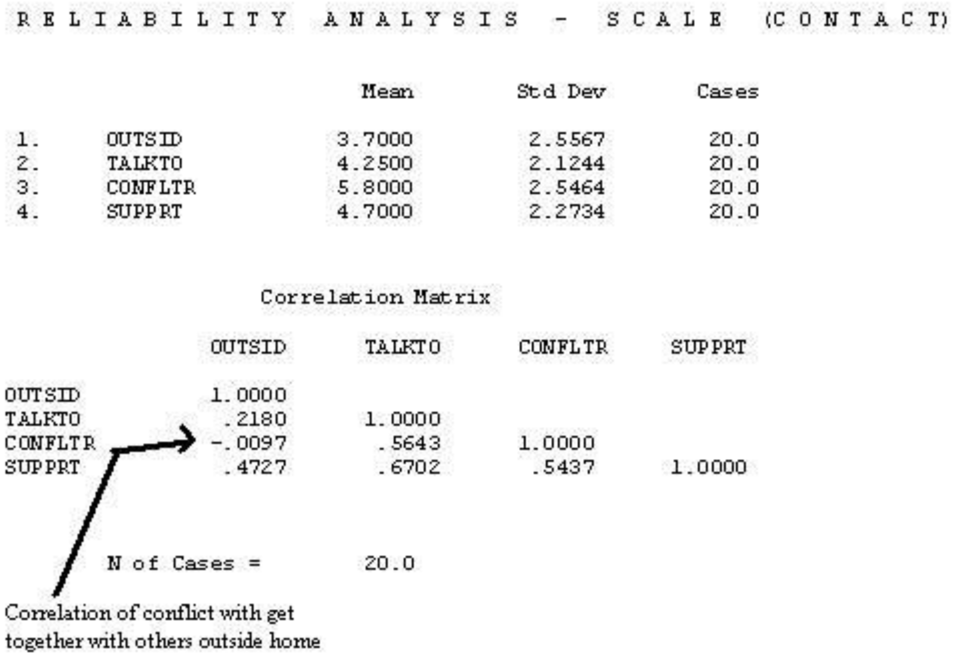
Click here to see the scale and data

For a discussion of reversed items click here.

The alpha coefficients for each of the subscales are:

1. Negative Emotions (negemo) was .93
2. Quality of Life (quality) was .92
3. Human Contact (contact) was .71
4. Employment (job) was 94.

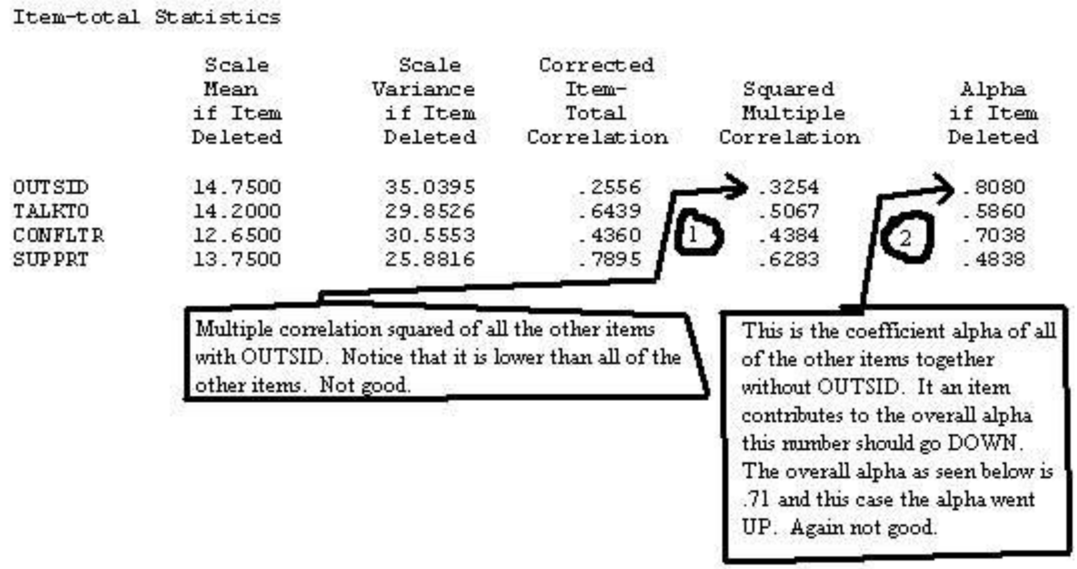
Only the CONTACT subscale will be discussed here to point out the variaous characteristics of the the output from SPSS. First is the means, standard deviations and number of cases.
Next the correlation matrix is printed showing the correlation of each item with every other item. The arrow points to a problematic correlation. Since it is assumed that each item of a subtest is measuring the same thing each item should correlate highly with every other item. The correlation of CONFLTR with OUTSID is essentially a zero correlation. Note also that OUTSID correlates poorly with all of the other variables.



```
  R E L I A B I L I T Y   A N A L Y S I S   -   S C A L E   (C O N T A C T)

                            Mean        Std Dev       Cases

  1.     OUTSID            3.7000        2.5567        20.0
  2.     TALKTO            4.2500        2.1244        20.0
  3.     CONFLTR           5.8000        2.5464        20.0
  4.     SUPPRT            4.7000        2.2734        20.0


                     Correlation Matrix

                  OUTSID      TALKTO      CONFLTR     SUPPRT

 OUTSID           1.0000
 TALKTO            .2180      1.0000
 CONFLTR         -.0097        .5643      1.0000
 SUPPRT            .4727       .6702       .5437      1.0000


        N of Cases =          20.0
```

Correlation of conflict with get together with others outside home

Next is the Item-total Statistics. Mostly we are interested in the Squared Multiple Correlation and Alpha if Item Deleted.

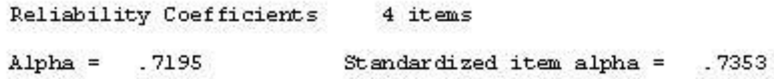


```
Item-total Statistics

                Scale          Scale        Corrected
                 Mean        Variance         Item-          Squared          Alpha
               if Item        if Item         Total         Multiple         if Item
               Deleted        Deleted      Correlation     Correlation       Deleted

OUTSID         14.7500        35.0395         .2556           .3254            .8080
TALKTO         14.2000        29.8526         .6439           .5067            .5860
CONFLTR        12.6500        30.5553         .4360           .4384            .7038
SUPPRT         13.7500        25.8816         .7895           .6283            .4838
```

Multiple correlation squared of all the other items with OUTSID. Notice that it is lower than all of the other items. Not good.

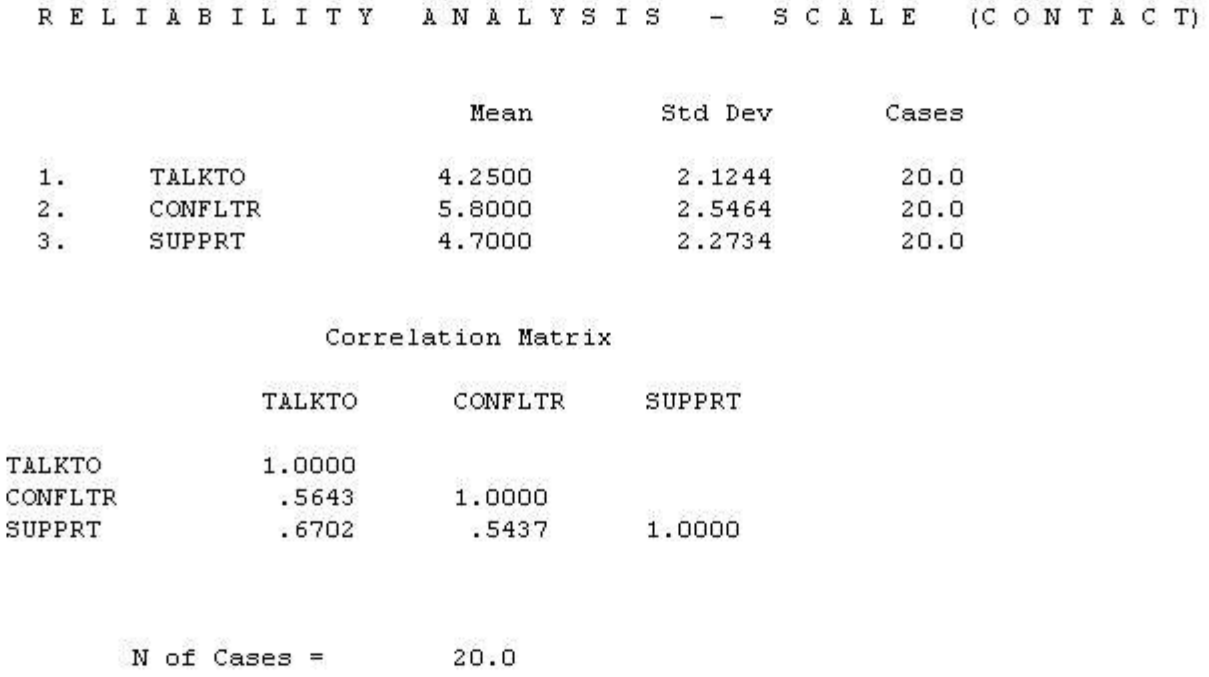
This is the coefficient alpha of all of the other items together without OUTSID. It an item contributes to the overall alpha this number should go DOWN. The overall alpha as seen below is .71 and this case the alpha went UP. Again not good.

Again we see that the item OUTSID is problematic. The analysis should be computed again without the variable OUTSID. Actually we know what the result will be. It will be .8080. However, we will get new diagnostic data. The alpha in this run follows:

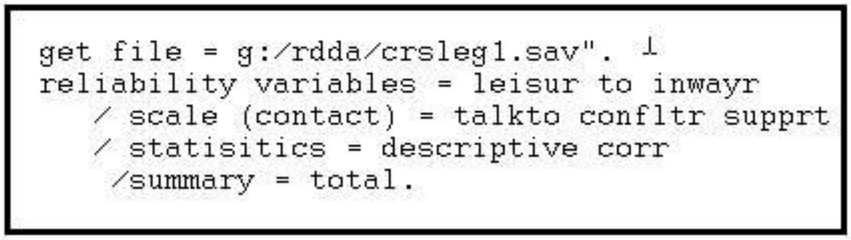


```
Reliability Coefficients      4 items

Alpha =    .7195            Standardized item alpha =    .7353
```

The criteria for the strength of alpha for including a variable in a test is not settled. Some say it can be as low as .70 while others say the lower cutoff should be .80. I believe that the item OUTSID is not a good measure of our subscale. Next will be a recomputation of the subscale without the item OUTSID. The syntax file is shown.

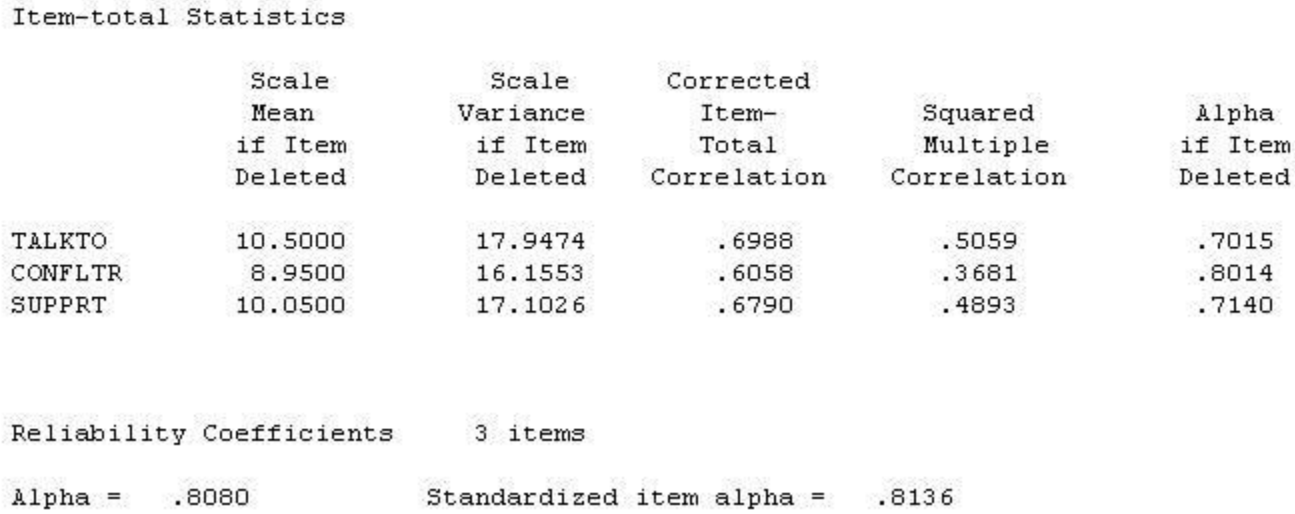


```
R E L I A B I L I T Y   A N A L Y S I S  -  S C A L E   (C O N T A C T)


                                  Mean         Std Dev       Cases

  1.     TALKTO                  4.2500         2.1244        20.0
  2.     CONFLTR                 5.8000         2.5464        20.0
  3.     SUPPRT                  4.7000         2.2734        20.0


                        Correlation Matrix

                TALKTO       CONFLTR      SUPPRT

TALKTO          1.0000
CONFLTR          .5643       1.0000
SUPPRT           .6702        .5437       1.0000



      N of Cases =        20.0
```



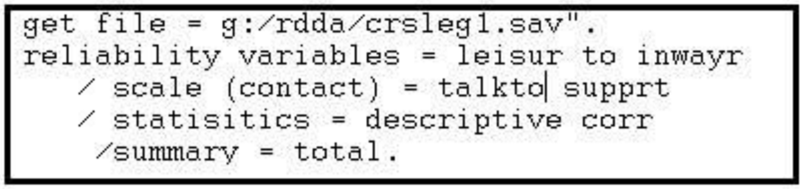
```
get file = g:/rdda/crsleg1.sav".
reliability variables = leisur to inwayr
   / scale (contact) = talkto confltr supprt
   / statisitics = descriptive corr
    /summary = total.
```

And the first part of the output.
This part did not change from the last run.

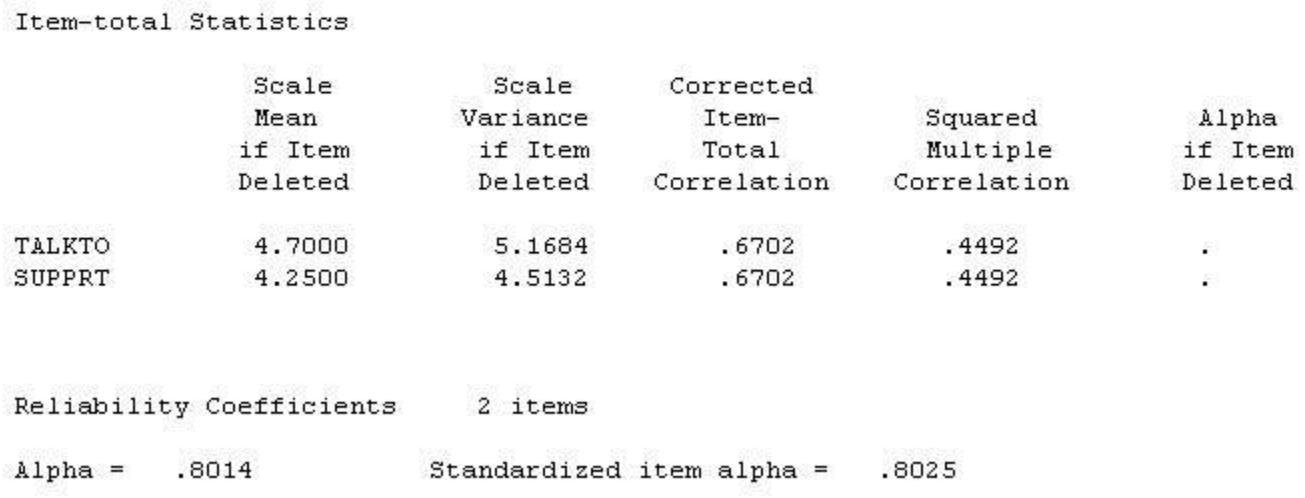


```
Item-total Statistics

              Scale        Scale      Corrected
              Mean        Variance      Item-       Squared         Alpha
             if Item      if Item       Total       Multiple       if Item
             Deleted      Deleted    Correlation   Correlation     Deleted

TALKTO       10.5000      17.9474       .6988         .5059          .7015
CONFLTR       8.9500      16.1553       .6058         .3681          .8014
SUPPRT       10.0500      17.1026       .6790         .4893          .7140



Reliability Coefficients     3 items

Alpha =   .8080           Standardized item alpha =   .8136
```

The remainder of the output follows:
The alpha is .80 but still the Squared Multiple Rs are not great and somewhat lower than the original run. Lets try one more.



```
get file = g:/rdda/crsleg1.sav".
reliability variables = leisur to inwayr
   / scale (contact) = talkto supprt
   / statisitics = descriptive corr
    /summary = total.
```



```
Item-total Statistics

               Scale          Scale      Corrected
               Mean         Variance       Item-         Squared          Alpha
              if Item        if Item       Total         Multiple        if Item
              Deleted        Deleted    Correlation    Correlation       Deleted

TALKTO         4.7000         5.1684        .6702          .4492             .
SUPPRT         4.2500         4.5132        .6702          .4492             .


Reliability Coefficients     2 items

Alpha =    .8014          Standardized item alpha =    .8025
```

Not much change. The multiple Rs have not improved. I don't think this is much of a test (in this instance subscale.)

NOTE: The Coefficient Alpha has at least three different names: (1) Internal Consistency (2)Coefficient Alpha, (3) Chronbach's Alphs, and (4) Alpha. They are used interchangably here.

Which items to reverse? The name of the scale might be one way to decide. For example, above the name of the first scale is Negative Emotion. One might expect that a high score on such a scale would indicate high negative emotion. So that none of the items of fear, depression, anger, confusion and tenseness would be reversed. However, on the scale Ouality of Life where you might expect a high score to represent a high quality of life the negative emotion of feeling useless would be reversed.