MANOVA between subjects

An alternative medical example and (fabricated) data for the same MANOVA between-subjects design

Neuropsychological investigations have found that early dementia often includes decline of recent memory, such that everyday tasks are forgotten, even though the individual remains capable of performing them. A study is conducted to compare the effectiveness of two electronic devices designed to help people with Alzheimer's Dementia to remember to carry out everyday tasks. Thirty people are recruited who are suffering dementia that includes the symptom of recent memory decline sufficient to necessitate prompts for everyday tasks to get done. Each recruit has a partner who is keen to help by acting as the recorder for the trial. There is a list of daily tasks such as washing, dressing, getting breakfast, making tea, checking cooker/kettle is switched off etc. Factor 1, with two levels, is the electronic METHOD used to prompt recall (time sensitive prompt; time + location sensitive prompt using geophysical location technology). In each case, the prompt is delivered in the form of a pre-recorded spoken message via a device worn on the wrist. Factor 2, with three levels, is the type of training given to the patient and partner before the trial starts. Level 1 gives a full day of practice for the patient with the partner there throughout and the Occupational Therapist supervising and instructing at frequent intervals. Level 2 also has a day of practice but the OT, after instructing the partner, only joins them to supervise for a few minutes four times during the day. At level 3, (the cheapest option) the OT instructs the partner and then leaves them unsupervised for the practice day, only checking on them at the end. During each day of the trial week the partner records each task on the list that is successfully completed following a prompt and the total is the DV COMPLETE. The partner is also trained to record the time in seconds between the prompt and commencement of each task. The mean of these processing times for completed tasks is the other DV, PROCESS.

The data appear in Table 3.1 (ignore for the moment the fifth column, labelled SPEED). Note that it is necessary to have more cases than DVs in every combination of factor levels (each combination of factor levels is often called a *cell*). This is a minimum requirement, and as in ANOVA, if this were a real experiment, the sample size should have been based on a power analysis. One way to tackle this is to consider the DV for which you want to be able to detect the smallest effect and base your sample size on the number needed for acceptable power for that one. Sample sizes for the other experimental designs in this chapter should also be decided using a power analysis.

Table 3.1
Data from a between groups experiment to compare reminder devices (med.manova.between.sav)

method	training	process	complete	speed
1	1	7	19	.1429
1	1	7	17	.1429
1	1	8	19	.1250
1	1	9	18	.1111
1	1	9	18	.1111
1	2	11	16	.0909
1	2	10	17	.1000
1	2	12	18	.0833
1	2	11	15	.0909
1	2	8	16	.1250
1	3	15	13	.0667
1	3	9	12	.1111
1	3	12	13	.0833
1	3	15	11	.0667
1	3	10	13	.1000
2	1	10	15	.1000
2	1	9	15	.1111
2	1	11	15	.0909
2	1	13	13	.0769
2	1	9	17	.1111
2	2	15	14	.0667
2	2	22	14	.0455
2	2	19	13	.0526
2	2	13	13	.0769
2	2	28	15	.0357
2	3	21	10	.0476
2	3	21	10	.0476
2	3	16	12	.0625
2	3	32	11	.0313
2	3	43	11	.0233

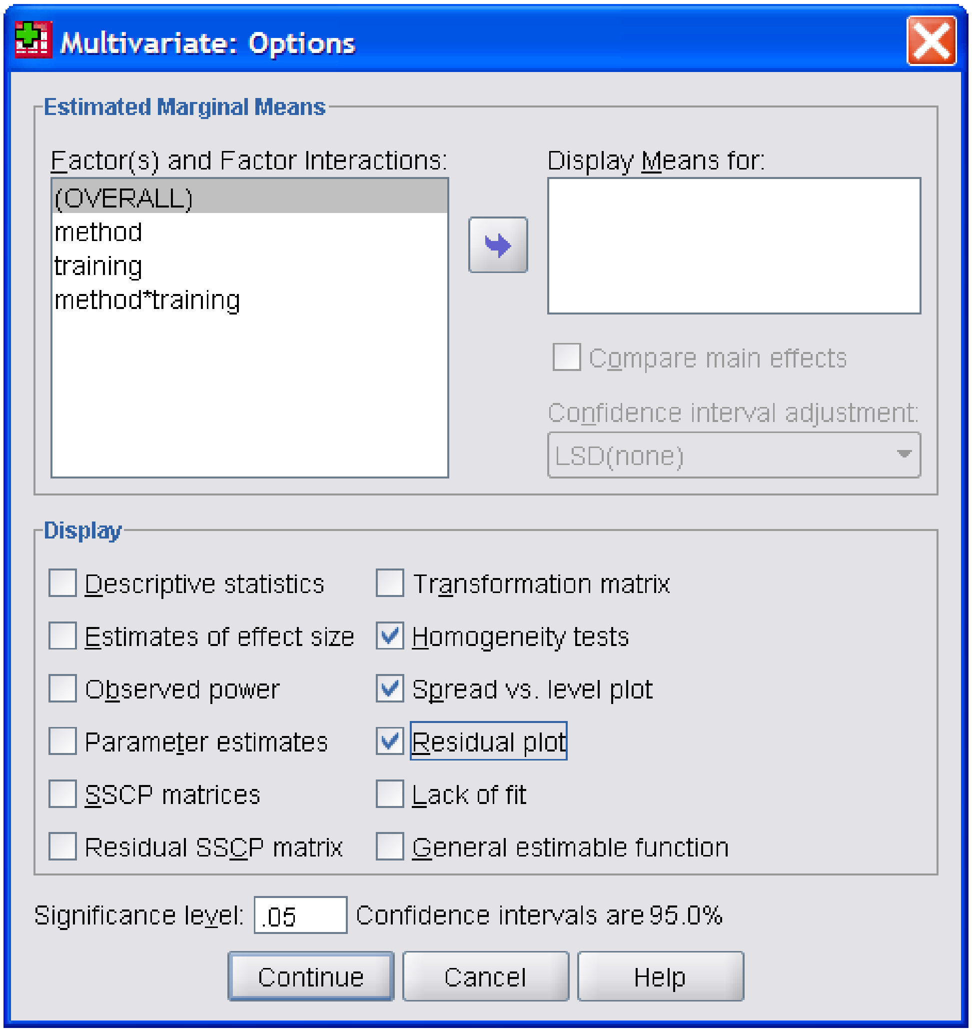
A between-subjects design with 2 DVs: setting it up in SPSS

This is a between subjects design, since each participant experiences just one of the six conditions, so the first four columns of Table 3.1 are arranged exactly as required for the SPSS datasheet, with each subject occupying one row, and a column for each of the variables (IVs in the first two columns, DVs in columns 3 and 4).

A between-subjects design with 2 DVs: requesting the analysis in SPSS

From the menus choose **Analyze**, then **General Linear Model**, then **Multivariate**, to get a dialog box just like SPSS Dialog Box 2.5, except that the box at the top is labelled **Dependent Variables** instead of **Dependent Variable**. Using the arrows, put METHOD and TRAINING into the **Fixed Factor(s)** box and PROCESS and COMPLETE into the **Dependent Variables** box. We could just click **OK** to get the analysis, but we will use the buttons on the right to get some extra information. If we click the **Model** button we get a dialog box just like SPSS Dialog Box 2.6, and just as with univariate ANOVA, we can accept the default full factorial model, or we can build a **Custom** model with the main effect of each of our factors and their interaction. The result will be the same so we will just accept the default. As before, the default **Type III Sums of Squares** is also appropriate since we have a balanced design (the same number of participants in each condition).

Clicking the **Options** button gives us SPSS Dialog Box 3.1. This time we will request displays of **Homogeneity test**s, **Spread vs level plots** and **Residual plots**. Click **Continue** to return to the main dialog box, then ignore the other buttons for now and click **OK**. These displays will allow us to check on the model assumptions, and since we shall find that our data do not conform to the assumptions, we will turn now to consider the output, leaving consideration of the other buttons till later.



SPSS Dialog Box 3.1. Displaying some diagnostics

A between-subjects design with 2 DVs: understanding the diagnostics output

First in the output is a table (not shown here) that summarises the data, telling us how many cases were at each level of each factor. Next comes the first table in SPSS Output 3.1, Box's test of equality of covariance matrices. For univariate ANOVA we need to assume that the DV has the same variance for all combinations of factor levels (the homogeneity of variance assumption). The analogous assumption for MANOVA is that the DVs have the same covariance matrices for all combinations of factor levels. We can see that our data fail this test, since the *F* calculated from Box's *M* is significant at the 1% level (probability = 0.005 from the Sig row in the table). We reject the hypothesis that the covariance matrices are equal.

Box's Test of Equality of Covariance Matrices[a]

Box's M	42.317
F	2.213
df1	15.000
df2	3150.554
Sig.	.005

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + method + training + method * training

Levene's Test of Equality of Error Variances[a]

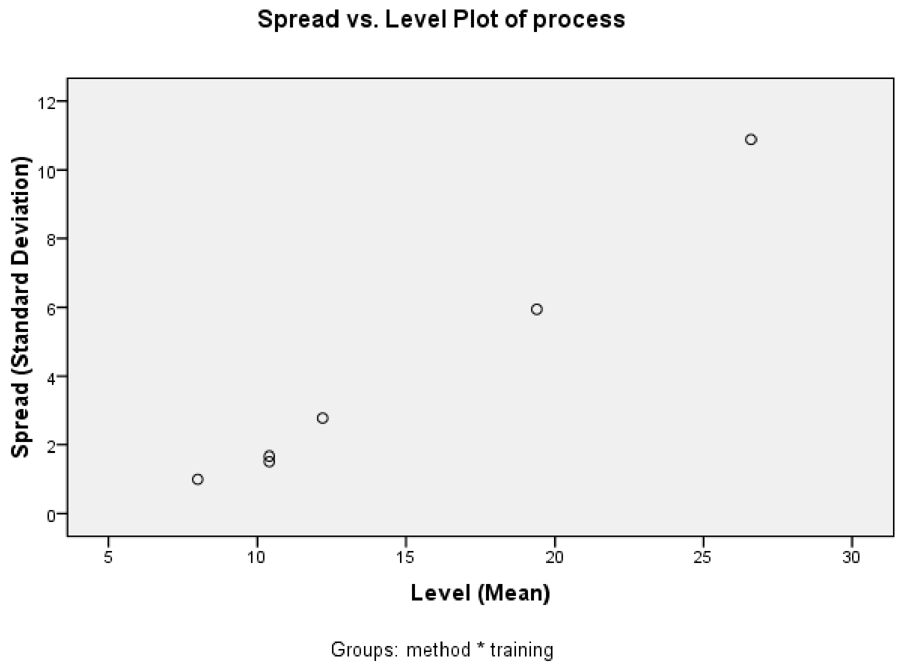
	F	df1	df2	Sig.
process	7.669	5	24	.000
complete	.137	5	24	.982

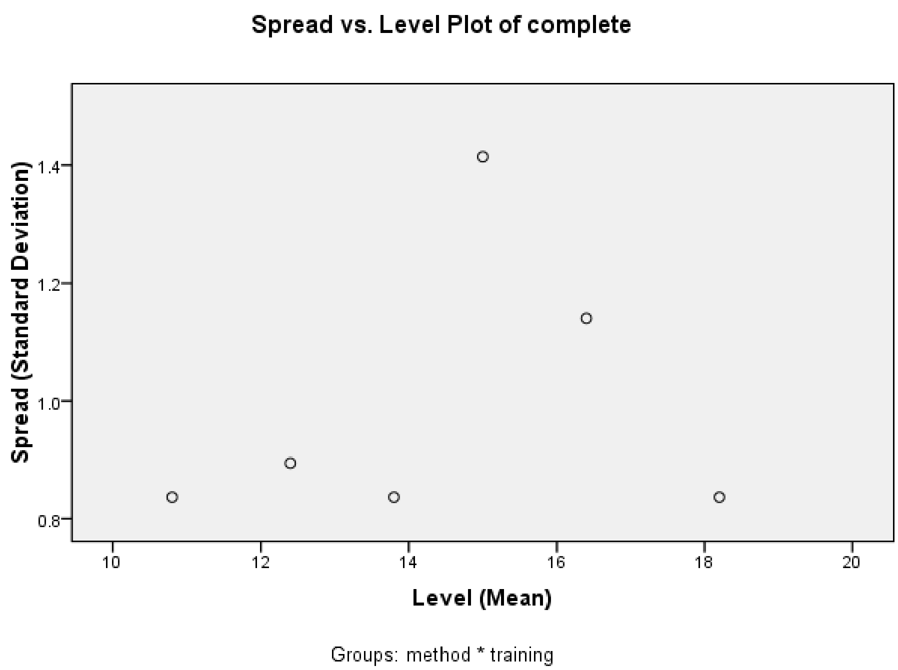
Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + method + training + method * training

SPSS Output 3.1. Tests of homogeneity of variances

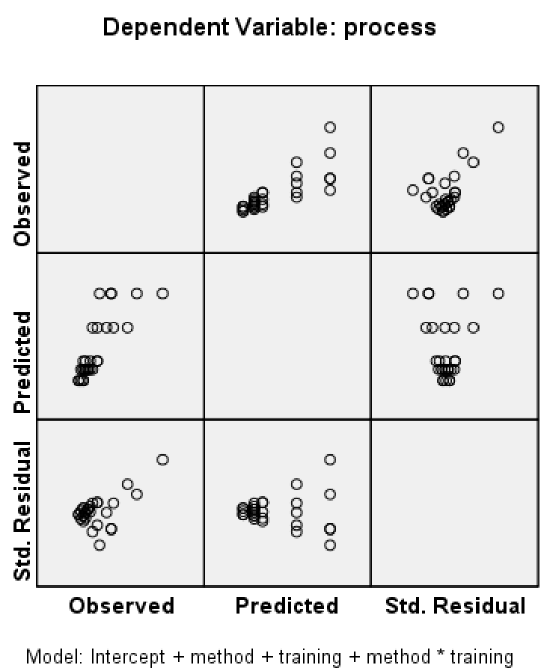
Ignoring the next table (Multivariate Tests) for the moment, we find below it the second table in SPSS Output 3.1. Here we have the homogeneity of variance test applied separately to each DV. For COMPLETE we would accept the hypothesis of equal variances (probability in the Sig column is 0.982), but for PROCESS we would reject it. Most likely it is only our DV PROCESS that is causing the problem. Since our data do not conform to the MANOVA homogeneity of variance assumption we will ignore the MANOVA tables and look at the plots we requested: perhaps these will suggest a possible remedy. SPSS Output 3.2 shows the first two Spread vs Level plots. For the first plot, the standard deviation (*sd*) of PROCESS has been calculated for the five observations in each of the six combinations of factor levels. Each *sd* has been plotted against the mean value of PROCESS for that combination of factor levels. We can see that there is a clear relationship: the bigger the mean, the bigger the *sd*.

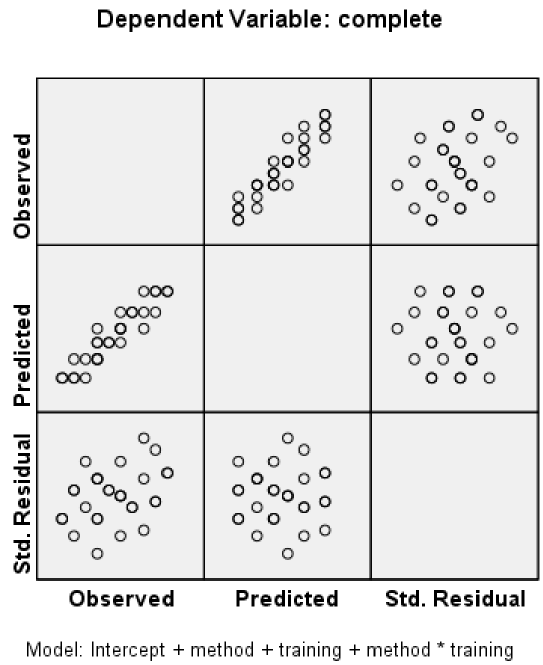




SPSS Output 3.2. Spread vs Level plots show sd of PROCESS *increases with mean*

The plot for COMPLETE does not show such a relationship, although we can see some variation in the *sd*s for complete. With only five observations used to calculate each *sd*, we expect some variability due to random sampling, even if the *sd*s would be the same if we could observe the populations of all possible participants in each of the six conditions. For COMPLETE, the largest *sd* is less than twice the smallest, but for PROCESS, the largest is more than ten times the smallest. Two more Spread vs Level plots (not shown here) plot the variance instead of the *sd* against the mean. Turn now to the residual plots shown in SPSS Output 3.3.



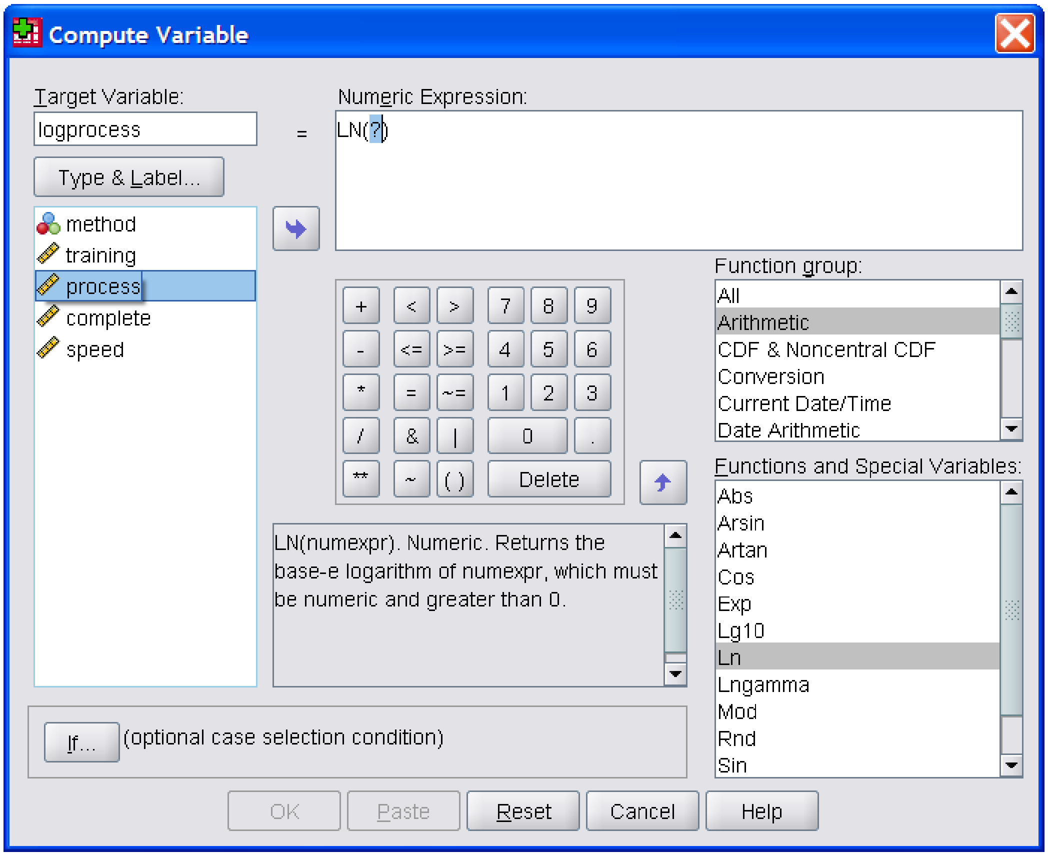


SPSS Output 3.3. Residual plots show variance of residuals increasing with predicted value for PROCESS*, but not for* COMPLETE

The centre plot on the bottom row of each matrix is the one we want. If the normality and homogeneity of variance assumptions are met, then the residuals will just be a random sample from a standard Normal distribution, and the plot of residuals against predicted values will show a shapeless cloud of points. For COMPLETE, this is just what we do see. However, for PROCESS we see the range of the residuals expanding as we move from left to right, from lower to higher predicted values. The variance is not the same for all values of the IVs, reinforcing what we already know from the Spread vs Level plots.

A between-subjects design with 2 DVs: trying out transformations

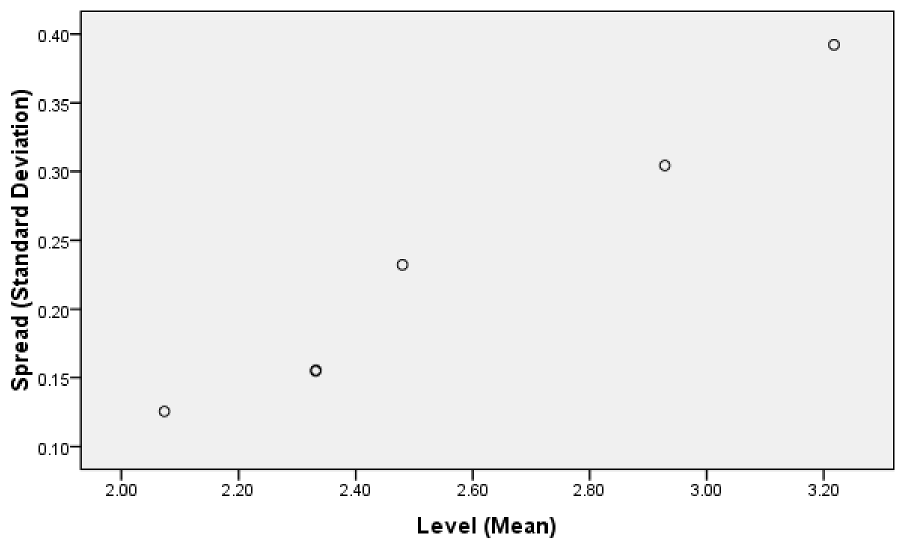
There are several transformations we could consider for a variable where the variance increases with the mean, as is the case for PROCESS. A log, a square-root or a reciprocal transformation would tend to reduce the variance at higher values of the mean. To apply a transformation, use **Transform**, then **Compute** from the menu bar, to get SPSS Dialog Box 3.2. In the **Target Variable** box, type a suitable name for the transformed variable: here we use LOGPROCESS for the transformed variable log(PROCESS). Then choose a function group (here we want the **Arithmetic** group) to open up a list of functions in the **Functions and Special Variables** box. From this list find the one you want and use the up arrow to put it in the **Numeric Expression** box. There are two log functions, Lg10 gives log to the base 10 and Ln gives natural logs: either will do but we chose Ln here. Now we use the right arrow to put PROCESS in the **Numeric Expression** box where the **?** is, and click OK. We now have a new variable in the datasheet called LOGPROCESS. For some transformations you need the arithmetic buttons instead of the function list. For instance, to get the reciprocal, click the 1, then / for divide, then use the arrow to enter PROCESS, so the **Numeric Expression** reads 1/PROCESS.



SPSS Dialog Box 3.2. Making a transformed variable

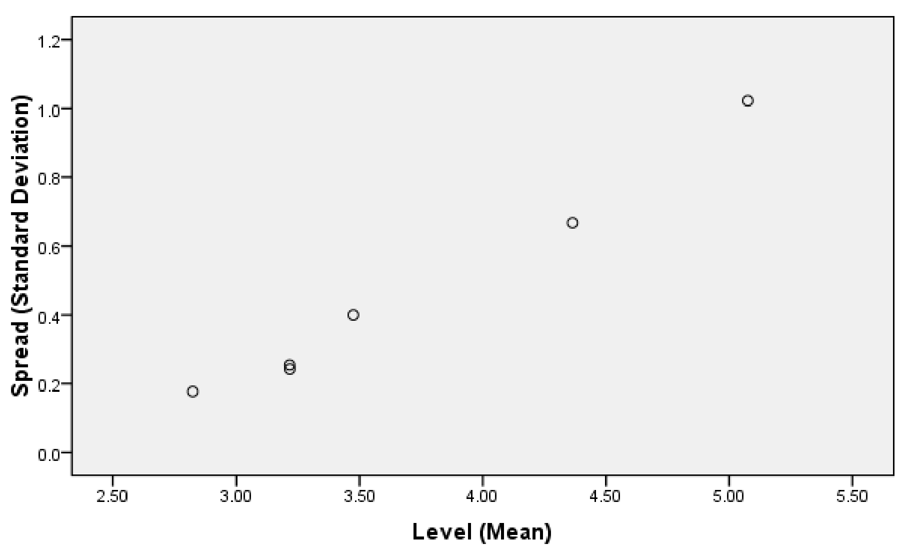
We used these methods to make two more transformations of PROCESS, the square root (called SQUROOT) and the reciprocal, called SPEED. Using these as the DVs in a MANOVA we obtained the Spread vs Level plots shown in SPSS Output 3.4. For LOGPROCESS and SQUROOT we see the same kind of dependence of the *sd* on the mean that we saw for PROCESS. However, if you compare these graphs with the one for PROCESS in SPSS Output 3.2 you will see from the vertical scale that the largest *sd* for PROCESS is about 11 times the smallest. For LOGPROCESS, the largest *sd* is about 4 times the smallest, and the same is true for SQUROOT. So both of these transformations have made the increase in *sd* with the mean less extreme. The reciprocal transformation, SPEED, has removed the dependence of the *sd* on the mean, though the *sd* still shows some variation.

Spread vs. Level Plot of logprocess

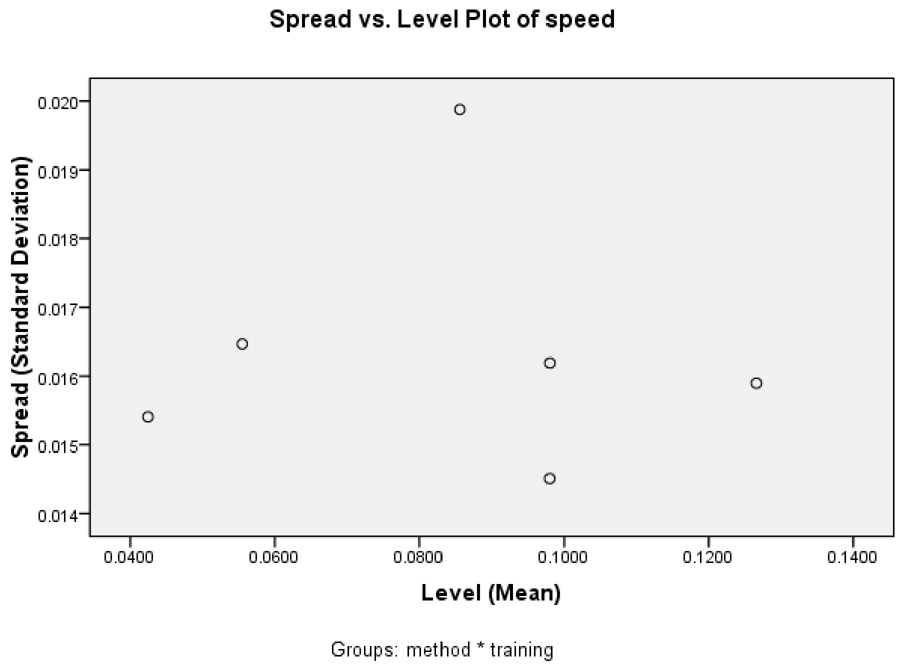


Groups: method * training

Spread vs. Level Plot of squroot



Groups: method * training



SPSS Output 3.4. Spread vs Level plots for three transformations of PROCESS

A between-subjects design with 2 DVs: meaningfulness of transformed data

If we transform a variable to make it conform better to our assumptions, we also need some sensible interpretation of the transformed variable. Of our three trial transformations, the one that does the best job on the homogeneity of variance assumption is the reciprocal (which we called SPEED). But does the reciprocal of PROCESS have any meaning? It does: the time between the reminder and starting the task we called the processing time, and the reciprocal of this is proportional to the SPEED of processing, just as good a measure of processing time as PROCESS was. The log and square-root transformations, on the other hand, have no obvious interpretation. Happily in this case we have a transformation that gives a new sensible variable that also conforms better with our assumptions than the original did. So using SPEED instead of PROCESS, we will repeat our MANOVA.

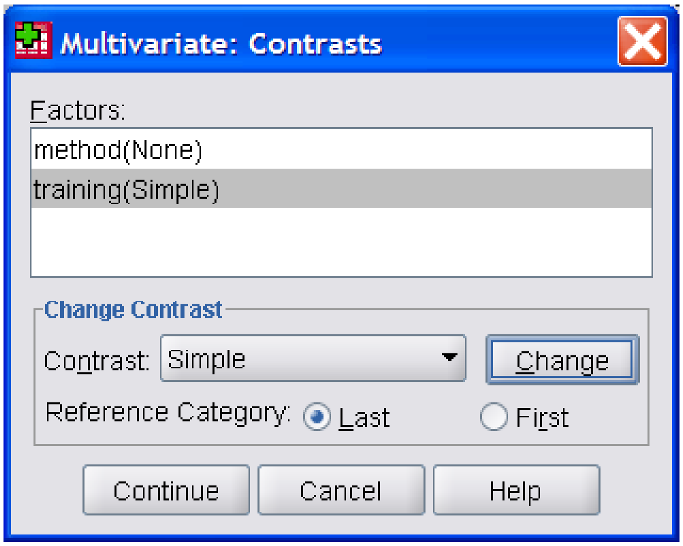
A between-subjects design with 2 DVs: requesting a re-analysis in SPSS

Now we include the extra column SPEED in our datasheet, all we have to do is choose **Analyze**, **General Linear Model**, **Multivariate** again, and now the variable list in the dialog box also includes SPEED. Put SPEED and COMPLETE in the **Dependent Variables** box instead of PROCESS and COMPLETE. In the **Options** Dialog Box (SPSS Dialog Box 3.1) click **Estimates of effect size** and **Observed power** as well as the items already selected.

Clicking the **Plots** button gives a dialog box just like SPSS Dialog Box 2.7. Put TRAINING into the **Horizontal Axis** box and METHOD into the **Separate Lines** box using the arrows, then click **Add**. Click **Continue** to return to the main dialog box.

The **Contrasts** button gives us SPSS Dialog Box 3.3, and we could use this to obtain contrasts between level 3 of TRAINING (OT instructs partner) and each of the two more expensive levels (1 and 2). This could be useful in helping to decide what sort of training is needed if the TRAINING factor is significant. Select TRAINING, change the **Contrast** to **Simple** using the arrow and the **Change** button, and check that the **Reference Category** is **Last**. Click **Continue** and **OK** to get the analysis. Using the simple option as here tests each of the other levels of the chosen factor against the specified reference level (here the last level). The Helmert options tests each level against the mean of the subsequent levels. The mirror image is difference, which tests each level against the mean of the previous levels. The repeated option tests each level against the next level.

Note that multivariate contrasts can be obtained but only by using the command language. The syntax is given by Tabachnik and Fidell (see Further reading at the end of the book).



SPSS Dialog Box 3.3. Requesting contrasts for the TRAINING *factor*

A between-subjects design with 2 DVs and transformed data: understanding the multivariate output

With our transformed DV SPEED replacing PROCESS, we find that Box's test of equality of covariance matrices is not significant (probability = 0.915 but we don't show the table this time). The Levene's tests for the equality of variance for the individual DVs are also not significant. The Spread vs Level plot for SPEED was shown in SPSS Output 3.4 and was satisfactory. The residual plot for SPEED is not shown but is similar to that for COMPLETE shown in SPSS Output 3.3. So, we are satisfied that our data conform to the assumptions needed for MANOVA and we will now look at the analysis.

Multivariate Tests[d]

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Intercept	Pillai's Trace	.996	3.136E3	2.000	23.000	.000	.996	6272.072	1.000
	Wilks' Lambda	.004	3.136E3	2.000	23.000	.000	.996	6272.072	1.000
	Hotelling's Trace	272.699	3.136E3	2.000	23.000	.000	.996	6272.072	1.000
	Roy's Largest Root	272.699	3.136E3	2.000	23.000	.000	.996	6272.072	1.000
method	Pillai's Trace	.766	37.550[a]	2.000	23.000	.000	.766	75.101	1.000
	Wilks' Lambda	.234	37.550[a]	2.000	23.000	.000	.766	75.101	1.000
	Hotelling's Trace	3.265	37.550[a]	2.000	23.000	.000	.766	75.101	1.000
	Roy's Largest Root	3.265	37.550[a]	2.000	23.000	.000	.766	75.101	1.000
training	Pillai's Trace	1.128	15.524	4.000	48.000	.000	.564	62.094	1.000
	Wilks' Lambda	.099	25.135[a]	4.000	46.000	.000	.686	100.542	1.000
	Hotelling's Trace	6.849	37.672	4.000	44.000	.000	.774	150.686	1.000
	Roy's Largest Root	6.495	77.945[c]	2.000	24.000	.000	.867	155.890	1.000
method * training	Pillai's Trace	.167	1.092	4.000	48.000	.371	.083	4.366	.317
	Wilks' Lambda	.835	1.083[a]	4.000	46.000	.376	.086	4.332	.314
	Hotelling's Trace	.195	1.071	4.000	44.000	.382	.089	4.286	.309
	Roy's Largest Root	.182	2.178[c]	2.000	24.000	.135	.154	4.356	.401

a. Exact statistic

b. Computed using alpha = .05

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

d. Design: Intercept + method + training + method * training

SPSS Output 3.5. The MANOVA table

SPSS Output 3.5 shows the MANOVA table. The intercept or grand mean is rarely of interest: the null hypothesis that it is zero is rejected but we should be very surprised if it were not. For each of the two main effects and the interaction we have four test statistics. Wilks' Lambda is the matrix analogue of the ratio of the residual sum of squares to the total sum of squares, so for METHOD it compares the within treatment variability with the total variability. So the Wilks statistic falls between zero and one. The total variability will always be greater, but the bigger the difference between the two METHOD means, the more of the total this accounts for, and the closer to 0 is Wilks' ratio. If the two METHOD means are almost the same, the ratio will be almost one. Our value for method is 0.234. The other three statistics use different combinations of residual and between group variances and each of the four statistics is converted to an approximate *F*. The degrees of freedom are shown and the probability of an *F* this large if the null hypothesis of equal group means is true is given. The four statistics may not all agree about whether or not a result is significant. Tabachnik and Fidell discuss the power and robustness of these statistics in various circumstances (see Further reading at the end of the book). SPSS also offers some advice: if Hotelling's Trace and Pillai's Trace are nearly equal then the effect is likely to be non-significant, and if Roy's largest root and Hotelling's Trace are equal or nearly equal then the effect is due mainly to just one of the DVs, or else there is a strong correlation between the DVs, or else the effect does not contribute much to the model.

For the main effect of METHOD we see in the Sig column that the null hypothesis is rejected by all the test statistics (probability < 0.001 in every case): there is a significant difference between the two reminder devices on our pair of DVs SPEED and COMPLETE. However, the fact that Roy's and Hotelling's statistics are equal tells us that probably this effect is mainly due to just one of the DVs, or else the DVs are rather highly correlated. We can easily check the correlation using **Analyze**, **Correlate**, **Bivariate** from the menu bar and entering SPEED and COMPLETE into the **Variables** box. We find that the correlation is 0.74, which is indeed rather high: our two DVs are measuring rather similar things. If we design further experiments to assess these reminder devices we could look for other aspects of performance to measure to obtain a more complete picture.

The results for TRAINING are similar: this main effect is also highly significant. The Roy and Hotelling statistics are similar but not equal, but we already know the correlation of the DVs is high. The interaction METHOD*TRAINING is not significant, the probabilities in the Sig column are all much greater than 0.05.

Partial eta squared and observed power were requested and their values are shown at the right of the MANOVA table. As always, the closer the value of partial eta squared is to 1, the more important the effect is in the model. It is a matrix analogue of the ratio of the variance accounted for by the effect, to the sum of this and the variance due to error. For the main effects, all of the partial eta squared values exceed 0.5, while for the interaction the highest is 0.154. For each of the two highly significant main effects the retrospective power is 1.00 (you would be unlikely to miss an effect of these sizes in a replication using the same sample size) and for the non-significant interaction the highest value for power (Hotelling's Trace) is 0.38.

A between-subjects design with 2 DVs and transformed data: understanding the univariate output

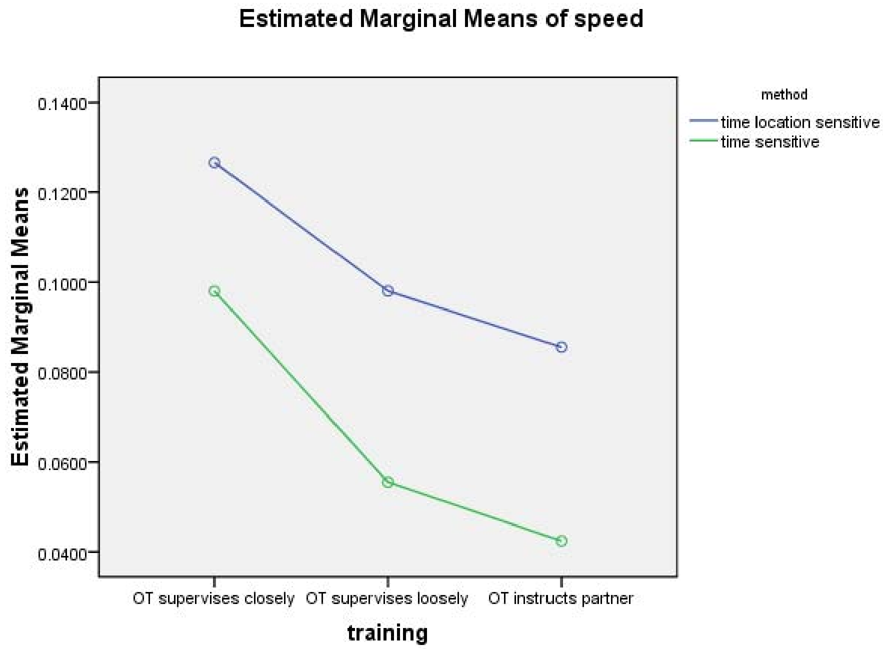
The table Tests of Between Subjects Effects (SPSS Output 3.6) shows the univariate ANOVA for each of the DVs. If you requested a univariate ANOVA of COMPLETE with METHOD and TRAINING as fixed factors, you would get exactly the results printed in the COMPLETE rows of this table, and similarly for SPEED. Generally speaking, however, the correct procedure is to carry out the MANOVA first and to only proceed to univariate ANOVAs if effects on the linear combination of DVs are significant. Looking at this table we can see that both main effects are significant for each of our DVs, and the interaction is significant for neither. The partial eta squared and retrospective (observed) power values are similar for both COMPLETE and SPEED to those found in the multivariate output.

SPSS Output 3.6. The table of univariate ANOVA results

If you find a significant MANOVA result as we did here you may wish to determine which variables are the most important in producing the significance. You might be tempted to use the univariate ANOVA above for this purpose, but this is not recommended for two reasons. Firstly, by performing a series of univarate ANOVAs you are ignoring the correlation between the DVs. Secondly, when you perform a number of tests each at the 5% significance level, you encounter the problem of multiple testing as described in the previous chapter (see A one-way design: post hoc tests). This effectively increases above 5% the chance of rejecting the null hypothesis when it is actually true. One way to avoid the latter problem is the step-down MANOVA or the Roy-Bargman F test. This ensures that the overall significance level remains at a specified level (e.g 5%) irrespective of how many tests are performed. You must first order the DVs in decreasing order of importance prior to any data collection. Test the first DV using a univariate ANOVA (as shown in Chapter 2). Then test the next most important DV in a model that also contains the first DV using analysis of covariance, which we consider in chapter 5. Continue in the same way, adding the DVs one at a time in order of importance, keeping the previous ones in the model. This allows you to gauge the relative importance of the DVs. Although this procedure ensures that you don't increase the chance of rejecting the null hypothesis when it is true, the results are heavily dependent on the ordering of the DVs. To implement this technique you need to use the command language. The syntax is given by Tabachnick and Fidell (see Further reading at the end of the book).

In SPSS Output 3.7 we show just one of the plots: for each METHOD the mean SPEED of processing reduces as the OT involvement in the training is reduced. The lines are nearly parallel because there is no significant interaction between METHOD and

TRAINING. At each level of TRAINING, method 1 (time + location sensitive prompt) has a higher processing speed. It appears that the most expensive device and the most expensive training are needed to optimise results for these patients! The plot for COMPLETE is similar.



SPSS Output 3.7. Plot showing the effect on SPEED *of* TRAINING *for each* METHOD