Poisson regression with equal observation periods

A medical experiment and (fabricated) data with equal observation periods

We begin by considering the first case above, i.e. where the time interval for all observations is equal. Later, we relax this assumption. Consider an investigation into how to reduce the number of epileptic fits. Participants who had been diagnosed with epilepsy for at least one year were randomly allocated to one of three treatments; high drug dose (treatment 1), low drug dose (treatment 2) and placebo (treatment 3). Details were also taken of two potentially confounding variables; ESTEEM, Rosenberg's Self-Esteem Scale (ranging from 0 to 30 with scores below 15 indicating low self-esteem) and ALCOHOL, a yes/no variable indicating whether the participant drinks more than the recommended upper limit of alcohol. The DV is the number of epileptic seizures (EVENTS) during the year following the end of the treatment. For the remainder of the chapter these are referred to as events. Data were collected for all participants for the whole year. Table 13.1 shows the first four cases within each treatment group.

Table 13.1

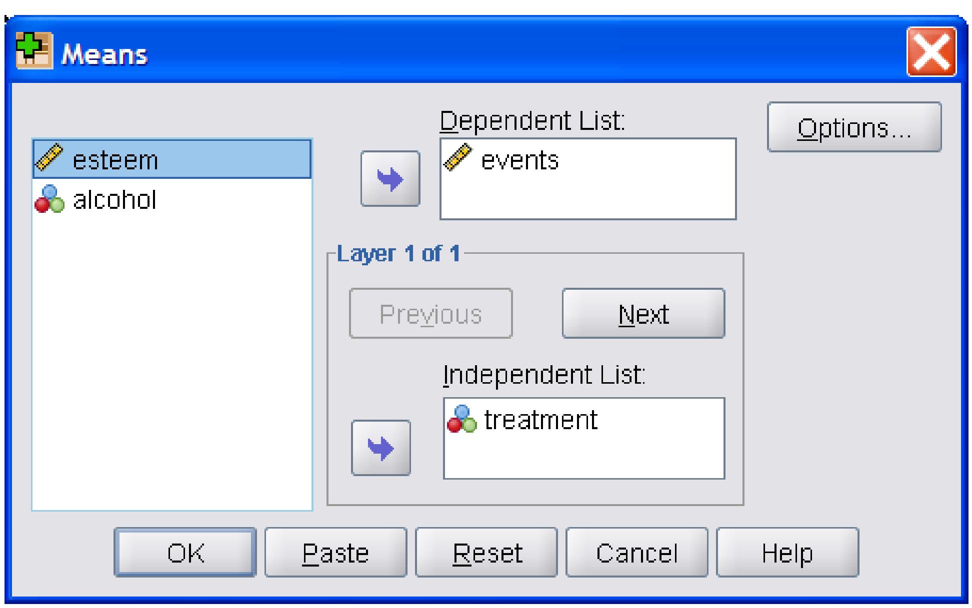
The first four cases from each treatment group in the epilepsy trial (the full dataset can be found as med.poissonregression.equaltimes.sav on the website)

Esteem	Alcohol[1]	Treatment[2]	Events
13	0	1	6
15	0	1	5
16	0	1	4
15	0	1	4
16	0	2	9
19	0	2	8
23	0	2	7
13	0	2	9
12	0	3	16
22	0	3	11
11	1	3	14
13	0	3	12

[1] Alcohol 0 = no, 1 = yes; [2]Treatment 1 = high dose treatment 2 = low dose, treatment 3 = placebo.

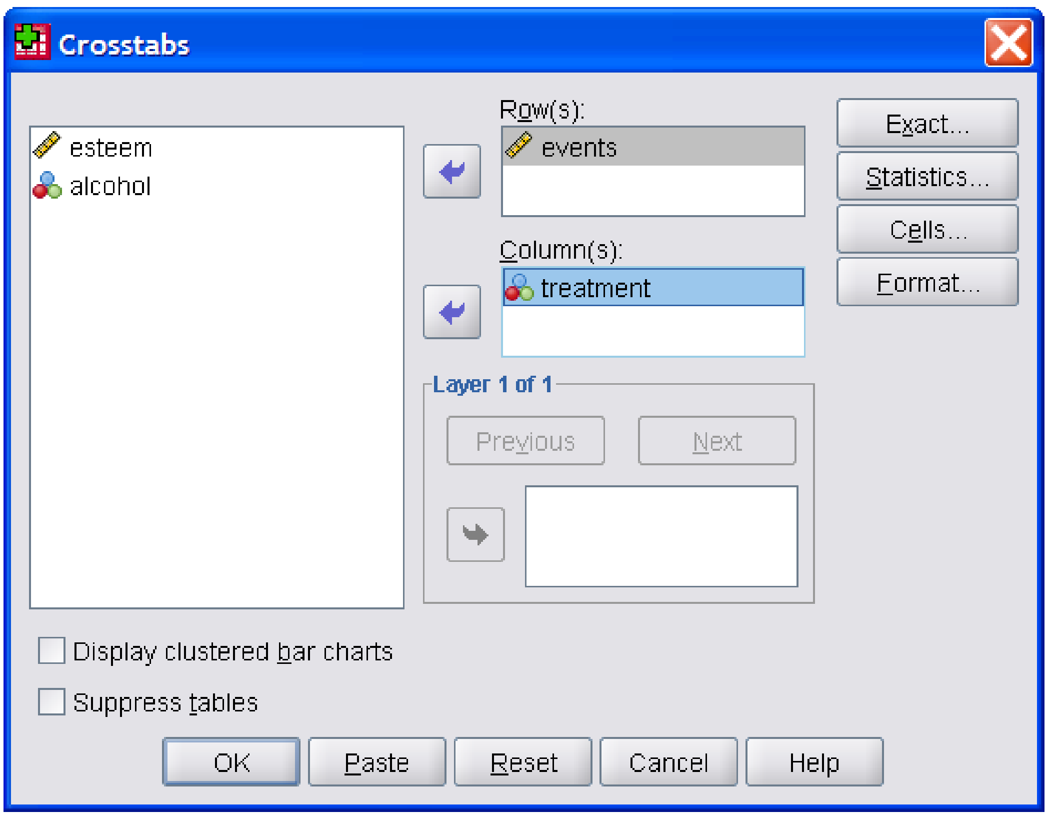
A first look at the data

Before starting our Poisson regression analysis we might begin by looking at the mean numbers of events for the three TREATMENT groups, and we find these are 2.95, 6.07 and 9.36 respectively. You can easily obtain these using **Analyze** then **Compare means** and then **Means**. Put EVENTS in the **Dependent List** box and TREATMENT in the **Independent List** box as shown in SPSS Dialog Box 13.1, and click **OK**.



SPSS Dialog Box 13.1. Obtaining the mean number of events for each treatment group

It is also interesting to consider the contingency table showing the numbers of events by treatment group. This can be obtained using **Analyze** then **Descriptive Statistics** and then **Crosstabs** to obtain SPSS Dialog Box 13.2. Put EVENTS in the **Row(s)** box and TREATMENT in the **Column(s)** box. Click **OK** to get SPSS Output 13.1.



SPSS Dialog Box 13.2. Producing a contingency table of the number of events for each treatment group.

events * treatment Crosstabulation

Count

		treatment			
		1	2	3	Total
events	1	3	0	0	3
	2	3	1	0	4
	3	9	3	0	12
	4	3	3	0	6
	5	1	5	0	6
	6	1	3	2	6
	7	0	8	4	12
	8	0	3	4	7
	9	0	4	5	9
	10	0	0	3	3
	11	0	0	3	3
	12	0	0	2	2
	14	0	0	1	1
	16	0	0	1	1
	Total	20	30	25	75

SPSS Output 13.1. The number of events (epileptic seizures) for three treatment arms. Treatment 1 = high dose, treatment 2 = low dose, treatment 3 = placebo.

The contingency table and means suggest that there are differences amongst the three treatments. The high dose group (treatment 1) has the lowest mean and 6 is the highest number of events for any participant in that group. The placebo (treatment 3) has the highest mean and 16 is the highest number of events for any participant in that group.

Now we will fit a Poisson regression with EVENTS as the DV, TREATMENT and ALCOHOL as factors and ESTEEM as a covariate. Because TREATMENT is a categorical variable with three levels, we will need two dummy variables to define it. The process is as described in the section on Creating dummy variables in Chapter 4, and illustrated in Table 4.2. The first of the dummy variables will take the value 1 for cases in treatment group 1 (high dose) and zero otherwise, the second will take the value 1 for cases in treatment group 2 (low dose) and zero otherwise. So the two dummy variables will have values (1, 0) for treatment group 1, (0, 1) for treatment group 2 and (0, 0) for treatment group 3. This arrangement, with treatment group 3 being (0, 0) is the SPSS default, and is referred to as making the last category the reference category.

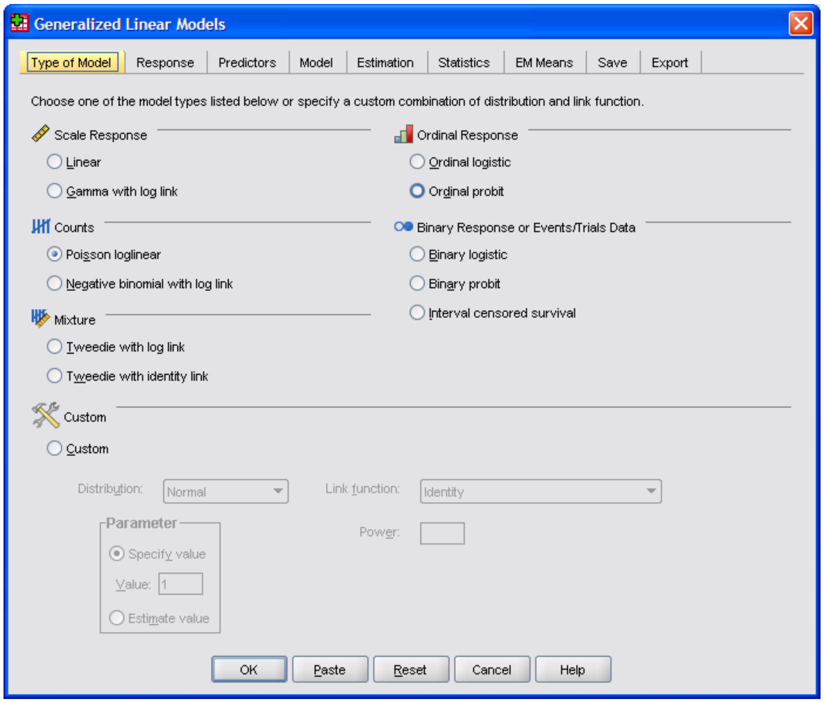
Since ALCOHOL is a categorical variable with only 2 categories, we can just code No as zero and Yes as 1, making just one dummy variable. Esteem is a numerical variable and can be treated as a covariate in the usual way.

The Poisson regression model uses a linear combination of factors and covariates, just like ANCOVA, but instead of using this to predict the DV as ANCOVA does, it is used to predict the log of the DV. This is analogous to the loglinear model where we used a linear combination of factors to predict the log of the cell probabilities in a contingency table. In the loglinear case, we used the logs so that the *multiplicative* law of probability could be modelled by a *sum* of the effects of factors and their interactions. In the Poisson regression case, if the count of events has a Poisson rather than a Normal distribution, it can be shown that we can predict the log of the count rather than the count itself, using a sum of the effects of factors and covariates. We will see later how this affects how we obtain predicted values from parameter estimates.

Requesting a Poisson regression with equal times in SPSS

With the data arranged as in Table 13.1 we can proceed to our Poisson regression model using TREATMENT, ALCOHOL and ESTEEM as our predictor variables. Choose **Analyze** then **Generalized Linear Models**, then **Generalized Linear Models**. Click the **Type of Model** tab if it is not already selected, so that you see SPSS Dialog Box

13.3. Our DV, EVENTS, is a count, so look at the **Counts** section, second down on the left. Click the **Poisson loglinear** radio button since we shall be using a linear model for the log of the number of events, as explained at the end of the Introduction.



SPSS Dialog Box 13.3. The main dialog box for Poisson regression

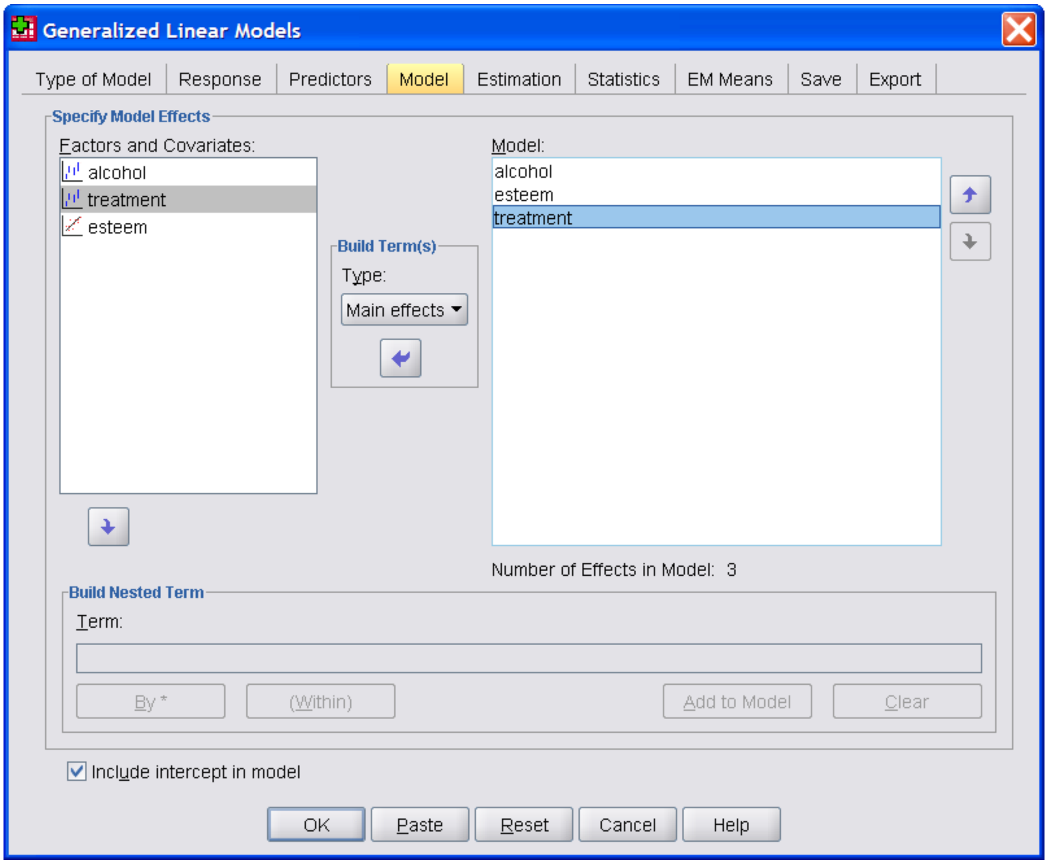
We now work our through the other tabs. Select the **Response** tab and put EVENTS in the **Dependent Variable** box.

Select the **Predictors** tab and put TREATMENT and ALCOHOL in the **Factors** box, and ESTEEM in the **Covariates** box.

Select the **Model** tab and make sure that **Main effects** appears in the **Build Terms** box (use the drop down list if it doesn't). Use the arrow to put ALCOHOL, ESTEEM and TREATMENT in the **Model** box, so your dialog box now looks like SPSS Dialog Box

13.4. (The order in which you put the predictor variables in the model box only affects the *order* in which parameter values are listed in the output, not their values.) We do not consider possible interactions among our factors and covariates in this introductory account of Poisson regression.

Select the **Save** tab and tick **Standardized deviance residuals**. You may also want to save the values predicted by the model. If you tick **Predicted value of linear predictor** from the list you will get the predicted value of the log of the EVENT count for each case. If you tick **Predicted value of mean of response** you will get the predicted count. SPSS displays this to the nearest whole number as a default but you can change this. Click on Variable view in the data set then on the cell in the decimals column and the Mean Predicted row. Use the up arrow to select the required number of decimal places. Click **OK** to get the analysis.



SPSS Dialog Box 13.4. Specifying our Poisson regression model

Understanding the output: checking the fit of the Poisson regression model with equal times

The first four tables in the output show us the model we have chosen, the number of cases excluded because of missing data (none in our example), the numbers at each level of TREATMENT (20, 30 and 25 in our example) and ALCOHOL (52 and 23), and what continuous variables have been used. This last table lists variables that are not factors, i.e. are not categorical variables, and we see here our covariate ESTEEM and our DV EVENTS. If you are of a critical turn of mind you may notice that these variables are discrete not continuous. They are numerical variables on a ratio scale, but 'continuous' is not a correct description. However, just take it that variables listed here are not categorical.

Next come Goodness of fit tests, shown in SPSS Output 13.2. The Deviance and Pearson residuals are used to assess whether the model assumptions have been violated. Each should approximately equal its degrees of freedom and so Value/df (value divided by degrees of freedom) should be close to one. You can see that we have 0.560 and 0.549 in the Value/df column, both less than 1. This means that our values are probably more bunched up around the mean than would be expected in a Poisson distribution, but not seriously so. This is a case where we have mild *underdispersion*. A more serious departure from the Poisson assumption is *overdispersion*, which would result in our values being less tightly clustered around the mean. This gives values exceeding 1 in the Value/df column. Here we can be satisfied that there is no evidence that the Poisson regression model is a poor fit.

Goodness of Fit[b]

	Value	df	Value/df
Deviance	39.214	70	.560
Scaled Deviance	39.214	70	
Pearson Chi-Square	38.441	70	.549
Scaled Pearson Chi-Square	38.441	70	
Log Likelihood[a]	-153.371		
Akaike's Information Criterion (AIC)	316.742		
Finite Sample Corrected AIC (AICC)	317.611		
Bayesian Information Criterion (BIC)	328.329		
Consistent AIC (CAIC)	333.329		

Dependent Variable: events
Model: (Intercept), alcohol, esteem, treatment

a. The full log likelihood function is displayed and used in computing information criteria.

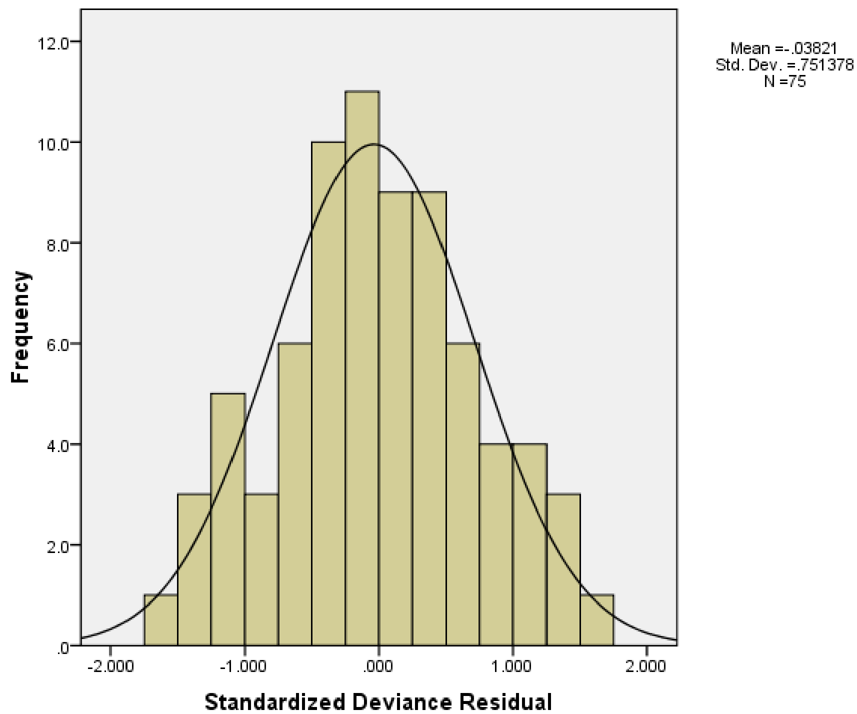
b. Information criteria are in small-is-better form.

SPSS Output 13.2. Goodness of fit tests for the Poisson regression with equal times

An additional check on the model fit uses the standardized deviance residuals. They should approximately follow a standard normal distribution. Since we saved them (using the **Save** tab in SPSS Dialog Box 13.3), they now occupy a new column in the datasheet and we can obtain a histogram. Select **Graph** from the menu bar, then

Chart builder. Drag the icon showing a simple histogram into the chart area, and drag Standardized deviance residual from the variable list into the x-axis box. In the Element Properties dialog box to the right of the Chart builder dialog box, tick **Display normal curve** and click **Apply**. Then click **OK** in the Chart builder dialog box to get SPSS Output 13.3. This shows that the standardized deviance residuals are approximately normally distributed and that all values are between -2 and +2.

These two checks give us some confidence that the Poisson regression model assumptions are not being violated.



SPSS Output 13.3. Histogram of standardized deviance residuals for the Poisson regression with equal times

Understanding the output: testing hypotheses for the Poisson regression model with equal times

We shall ignore the next two tables since we can get all we want from the table of parameter estimates, shown as SPSS Output 13.4. Look first at the two rows of the table for ALCOHOL. This is a dummy variable coded zero for No (does not exceed recommended upper limit) and 1 for Yes. The default for SPSS is to use the last category as reference, so the parameter value for ALCOHOL = 1 is zero (look in the B column). The parameter value for ALCOHOL = 0 is -0.004. Because it is negative, the effect of ALCOHOL = 0 is to reduce the expected count of EVENTS compared with the reference category, ALCOHOL = 1 (assuming TREATMENT and ESTEEM are fixed). However, as the B value is very close to zero this effect is very small. Now look at the Sig column on the right: the *p* value for ALCOHOL is close to 1 (0.967) so it is almost certain that this small effect is only random variation. We certainly can't reject the null hypothesis that ALCOHOL has no effect on EVENTS, consistent with its very low value of B. We conclude that abusing alcohol or drugs does not affect the count of EVENTS significantly.

The *p* value for ESTEEM is 0.015 and hence this covariate is significant. The negative value of B for ESTEEM shows that the predicted value of the DV decreases as ESTEEM increases.

Because TREATMENT is a categorical variable with 3 levels, and once again the last category is used as the reference, we have two parameter estimates and corresponding *p* values, for TREATMENT = 1 (high dose) and TREATMENT = 2 (low dose). Both parameter values (the B column) are negative, so both high and low doses have lower predicted events counts than the placebo (assuming ALCOHOL and ESTEEM are fixed). The *p* value for TREATMENT = 1 is less than 0.001, so we can certainly reject the null hypothesis that, compared with the placebo, the high dose has no effect on EVENTS. The *p* value for TREATMENT = 2 is also less than 0.001, so we can certainly also reject the null hypothesis that, compared with the placebo, the low dose has no effect on

EVENTS. Notice that, so far we have no way to see whether the high and low doses differ significantly from each other: all we know so far is that each differs significantly from the placebo, after allowing for ALCOHOL and ESTEEM.

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval Lower	Upper	Hypothesis Test Wald Chi-Square	df	Sig.
(Intercept)	2.753	.2084	2.345	3.162	174.559	1	.000
[alcohol=0]	-.004	.1032	-.206	.198	.002	1	.967
[alcohol=1]	0[a]	.	.	.	.	.	.
esteem	-.034	.0139	-.061	-.007	5.958	1	.015
[treatment=1]	-1.153	.1457	-1.439	-.868	62.637	1	.000
[treatment=2]	-.457	.0994	-.652	-.263	21.178	1	.000
[treatment=3]	0[a]	.	.	.	.	.	.
(Scale)	1[b]						

Dependent Variable: events
Model: (Intercept), alcohol, esteem, treatment

a. Set to zero because this parameter is redundant.

b. Fixed at the displayed value.

SPSS Output 13.4. Parameter estimates for the Poisson regression with equal times

Obtaining expected EVENTS counts from the parameter values

We told you how to save predicted values of the EVENT count and its log at the end of the section on Requesting a Poisson regression analysis with equal times in SPSS. However, you may find it helps your understanding of the Poisson regression model if you work out some of the expected values yourself from the parameter values. Just be aware that the table in SPSS Output 13.4 gives parameter values correct to three decimal places, but more precise values are used by SPSS in its own calculation of expected values. Hence in the calculations shown below and in any you do yourself, you may find small discrepancies in the last decimal place.

To see how to obtain expected values of events using the parameter values, first consider a case ALCOHOL = 1 and TREATMENT = 3. Because these are the reference categories for the two factors, the parameter values for TREATMENT and ALCOHOL for this case are both zero. You can see such a case in the penultimate row of Table 13.1.

This participant had an esteem score of 11, was abused as a child, was in treatment group 3 and had an events count of 14. In SPSS Output 13.4 you see that the intercept is 2.753 and the B value for ESTEEM is -0.034. So, with zeros for ABUSE and TREATMENT, and 11 for ESTEEM, our predictor for this case is

$$2.753 - 0.004*(0) + 0*(1) - 0.034*11 - 1.153*(0) - 0.457*(0) + 0*(1) = 2.379$$

If this were ordinary regression, this would be the predicted value of the DV. However, in Poisson regression this predictor is for the *log* of EVENTS, so to get the predicted value of EVENTS we need exp(2.379) = 10.8, or 11 to the nearest whole number. Remember though that you can display more accurate values in the data set file by specifying the number of decimal places using variable view. To get this from your calculator, look for e^x or the inverse of ln (the natural log). You may have to use the Shift or Inverse key. The observed value of EVENTS for this individual was 14.

Now consider a case with ALCOHOL = 0 and TREATMENT = 3. You can find such a case in the last row of Table 13.1. The ESTEEM score is 13 and the EVENTS count is 12. This time the B value for ALCOHOL is -0.004, that for TREATMENT is still zero. Our predictor of log(EVENTS) will be

$$2.753 - 0.034*13 - 0.004 = 2.307$$

and the predicted value for EVENTS is exp(2.307) = 10.0, again slightly below the observed value of 12.

Now consider a case with ALCOHOL = 0 and TREATMENT = 1. You can find such a case in the first row of Table 13.1. The ESTEEM score is 13 and the EVENTS count is 6. This time the B value for ALCOHOL is -0.004, and for TREATMENT is -1.153. Our predictor of log events will be

$$2.753 - 0.034*13 - 0.004 - 1.153 = 1.154$$

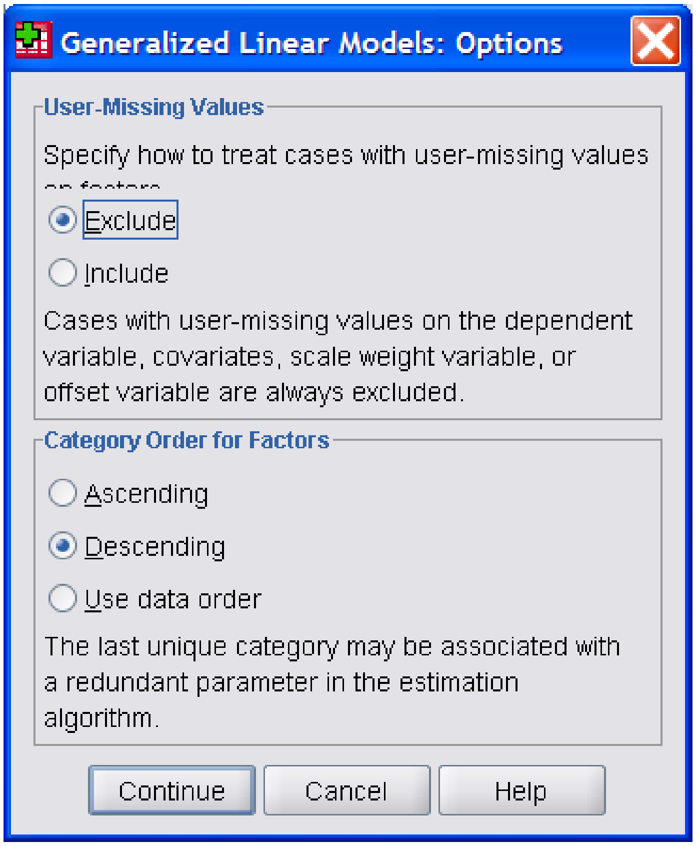
and the predicted value for EVENTS is exp(1.154) = 3.17, or 3 to the nearest whole number, compared with the observed value of 6. To find the predicted value for a person in treatment group 2, replace -1.153 by -0.457 and insert the correct value of esteem where we have 13 here.

Changing the reference category from last to first

If we want to test the null hypothesis that there is no difference between the effects of treatments 1 and 2 on EVENTS, we can do it by changing the reference category from the last to the first, so that treatments 2 and 3 will each be compared to treatment 1. Note that all factors will have the reference category changed to the first, so ALCOHOL = 1 will be compared to ALCOHOL = 0 instead of the other way round, but because ALCOHOL has only two categories this will just reverse the sign of the parameter.

Before you rerun the analysis you need to delete the columns containing the predicted values of the event count, the predicted value of the linear predictor and the standardized deviance residuals, if you created them. Select **Data View** within the data set, click on MeanPredicted ensuring that the column of data is highlighted and then press the delete key. Repeat for StdDeviance Residual and XBPredicted.

When you click the **Predictors** tab in SPSS Dialog Box 13.3, you see an **Options** button at the bottom of the **Factors** box. Click this after putting ALCOHOL and TREATMENT in the box. Click the radio button for **Descending**, as shown in SPSS Dialog Box 13.5. Repeat the analysis, and this time you see the table of parameters shown in SPSS Output 13.5.



SPSS Dialog Box 13.5. Changing the reference category for factors from last to first

As expected, you can see that in the B column we now have 0.004 for ALCOHOL = 0 instead of -0.004 for ALCOHOL = 1. The *p* value remains unchanged. The parameter for ESTEEM is also unchanged. The B value for TREATMENT = 3 (compared to TREATMENT = 1) is 1.153. Previously we had -1.153 for TREATMENT = 1 compared to TREATMENT = 3. But the comparison of interest is TREATMENT = 2 compared to TREATMENT = 1. Here we have B = 0.696 and a *p* value of less than 0.001. The B value is positive, so compared with TREATMENT = 1, TREATMENT = 2 increases the expected value of EVENTS (assuming ALCOHOL and ESTEEM are fixed). As the *p* value is so low we can certainly reject the null hypothesis that there is no difference between treatments 1 and 2. It appears that the low dose increases the number of EVENTS significantly, compared with the high dose.

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval Lower	95% Wald Confidence Interval Upper	Hypothesis Test Wald Chi-Square	Hypothesis Test df	Hypothesis Test Sig.
(Intercept)	1.596	.2552	1.096	2.096	39.102	1	.000
[alcohol=1]	.004	.1032	-.198	.206	.002	1	.967
[alcohol=0]	0[a]	.	.	.	.	.	.
esteem	-.034	.0139	-.061	-.007	5.958	1	.015
[treatment=3]	1.153	.1457	.868	1.439	62.637	1	.000
[treatment=2]	.696	.1504	.401	.991	21.404	1	.000
[treatment=1]	0[a]	.	.	.	.	.	.
(Scale)	1[b]						

Dependent Variable: events
Model: (Intercept), alcohol, esteem, treatment

a. Set to zero because this parameter is redundant.

b. Fixed at the displayed value.

SPSS Output 13.5. Parameter estimates with the reference category for factors changed from last to first