Poisson regression with unequal observation times

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

We now adapt our example using epileptic seizures to model the average number of events over the reporting period (seizures per year). We then estimate the expected count over the period by taking into account the reporting period. For example, if we have observed an individual for three years and their predicted Poisson rate was seven seizures per year, then we would predict that they would have had a total of 21 seizures during the three years they had been observed.

After reviewing the results of the experiment above where each participant was observed for one year, the research team wondered if the relative success of treatment 1 (high dose) was because participants in that group were less prone to seizures before the experiment began. Since records of seizures were retained throughout the time a patient was registered at the treatment unit, the team decided to count the number of seizures for each participant in the study from the time they were referred until the start of the intervention. The team would then test for differences amongst the treatment groups, allowing for ESTEEM and ALCOHOL, for the period preceding the original study. If the TREATMENT group effect is non-significant, then the research team will have more confidence that their results were due to their intervention.

Since the participants in the experiment have been with the treatment unit for different lengths of time, we need to modify the Poisson model. Instead of modelling the event count over a fixed interval of time we model the average number of events per year over the observed time. For each participant we shall need the number of seizures from when the participant was referred until the start of the experiment (which we call pre-treatment events, PRE_TRT_EVENTS), and also the number of years that the participant has been at the treatment unit before the experiment (YEARS). We shall also need the natural log of years which for a reason which will become apparent later is called OFFSET. Table 13.2 shows the first four cases from each treatment group. You can see that the first four columns are the same as Table 13.1, but we now have three extra columns containing the new variables. Our analysis will not include EVENTS this time, since we are now focussing on the period before the start of the experiment. YEARS will appear only through its log, OFFSET.

Table 13.2
The first four cases from each treatment group in the extended prison trial (the full dataset can be found as med.poissonregression.unequaltimes.sav on the website)

Esteem	**Alcohol**[1]	**Treatment**[2]	**Events**	**Years**	**Offset**	**Pre_trt_events**
13	0	1	6	1.95	0.667829	15
15	0	1	5	3.36	1.211941	24
16	0	1	4	3.30	1.193922	28
15	0	1	4	4.14	1.420696	27
16	0	2	9	3.07	1.121678	28
19	0	2	8	5.18	1.644805	28
23	0	2	7	4.44	1.490654	18
13	0	2	9	2.34	0.850151	30
12	0	3	16	3.47	1.244155	34
22	0	3	11	3.16	1.150572	28
11	1	3	14	4.77	1.562346	35
13	0	3	12	2.04	0.71295	21

[1] Alcohol 0 = no, 1 = yes; [2]Treatment 1 = high dose, treatment 2 = low dose, treatment 3 = placebo. Offset = log_e(years).

A first look at the unequal times data

Since participants have been registered at the treatment unit for varying times, it would not be particularly useful to compare the mean number of seizures to the start of the experiment for each treatment group. Instead, we find the mean number of pre-treatment events *per year* for each treatment group. To obtain this, we first calculate the number of events per year for each case by dividing the number of pre-treatment events (PRE_TRT_EVENTS) by the time the participant has been registered at the treatment unit before the experiment (YEARS). From the menu bar select **Transform** then **Compute Variable**. (You can see an example of the dialog box in the chapter on MANOVA, SPSS Dialog Box 3.2.) In the **Target Variable** box type a name for the new variable (we called it PRE_TRT_EVENTS_PER_YEAR). Use the arrow to put PRE_TRT_EVENTS in the **Numeric Expression** box, then click on the / for divide, then use the arrow again to put in YEARS. The numeric expression should now read PRE_TRT_EVENTS/ YEARS. Click **OK** and check that the datasheet now contains your new variable PRE_TRT_EVENTS_PER_YEAR.

Now we want a table giving the mean of PRE_TRT_EVENTS_PER_YEAR for each treatment group. You can obtain this by using **Analyze** then **Compare means** and then **Means**, as we did above (see SPSS Dialog Box 13.1). Put PRE_TRT_EVENTS_PER_YEAR in the **Dependent List** box and TREATMENT in the **Independent List** box and click **OK**.

Report

pre_trt_events_per_year

tre...	Mean	N	Std. Deviation
1	7.6624	20	1.17110
2	8.0759	30	2.33067
3	8.4821	25	2.35803
Total	8.1010	75	2.09418

SPSS Output 13.6. Mean annual number of pre-treatment events by treatment group

The result is shown in SPSS Output 13.6. We can see that the mean annual number of seizures for each participant in the period before the experiment is approximately equal for each treatment group at around eight incidents per year. So the *rate* at which seizures were occurring before the experiment began was similar for all the treatment groups. Now we will do a Poisson regression analysis, taking account of ESTEEM and ALCOHOL, which we have ignored in this first look at the data.

Requesting a Poisson regression with unequal times in SPSS

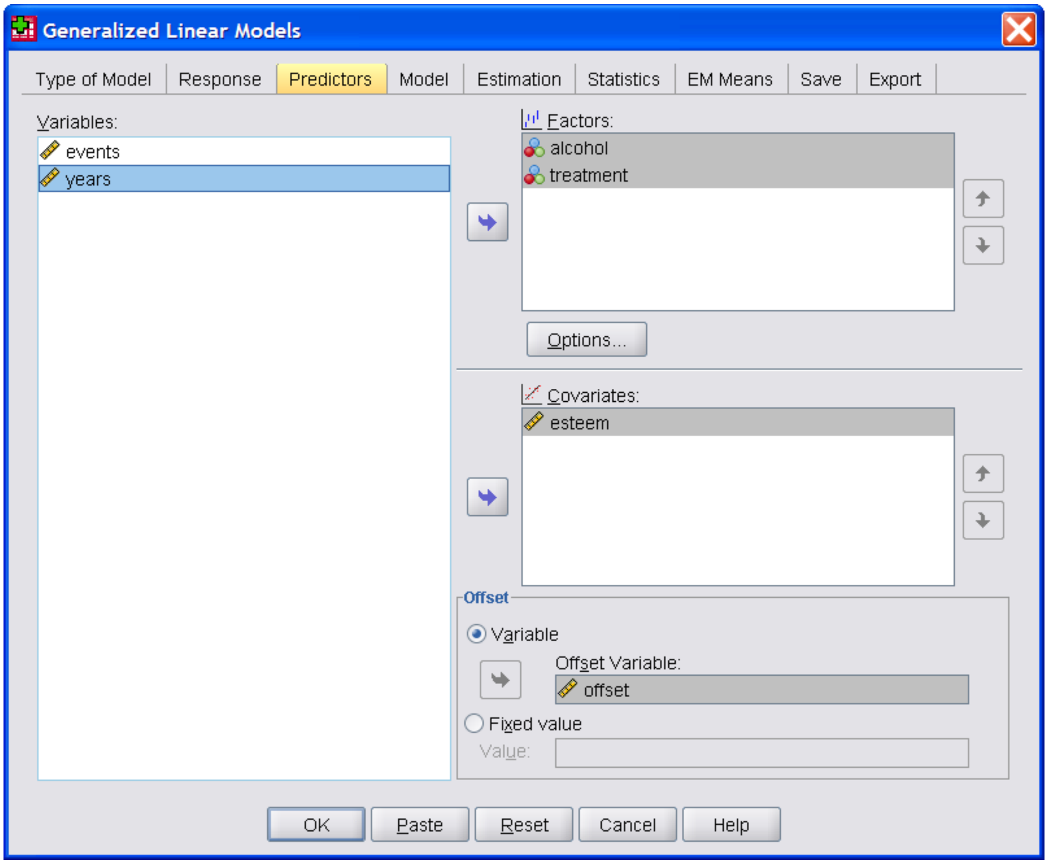
In this analysis we attempt to predict the rate of our PRE_TRT_EVENTS. Our DV is PRE_TRT_EVENTS, and as before we use ALCOHOL and ESTEEM as predictors. Also, because we want to know whether the treatment groups for the one year experiment differed in their levels of violence before the experiment started, we include TREATMENT to see whether there are significant differences in PRE_TRT_EVENTS among treatment groups. It remains to consider YEARS. Since the observation periods for PRE_TRT_EVENTS are not all the same, we have to allow for time during which PRE_TRT_EVENTS was counted for each participant. We do this by adding the log of YEARS to the linear combination of predictor variables, and this extra term is called the *offset* (hence our name for it in Table 13.2). If you are old enough to have used log tables before the arrival of calculators you can see how this works by remembering that to multiply two numbers you add their logs. If the linear combination of the predictor variables gives the log of the event count per year, then adding the log of the number of years will give the total event count over that number of years. It's called the offset because it shifts the prediction up or down according to whether the observation period was long or short.

Notice that other time units might be appropriate in other problems, and you can even use the same idea when counting in space rather than time. In the forestry example briefly mentioned above, you might count trees above 40 metres tall in several patches of forest, and the area of each patch of forest would take the place of years of observation. The offset would be log(area). In this case you would most likely use hectares as the unit of area.

Now we fit the model. 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 again. Our DV, PRE_TRT_EVENTS, is a count as before, so look at the **Counts** section, and again click the **Poisson loglinear** radio button.

Select the **Response** tab and put PRE_TRT_EVENTS in the **Dependent Variable** box.

Select the **Predictors** tab and put TREATMENT and ALCOHOL in the **Factors** box, ESTEEM in the **Covariates** box and OFFSET in the **Offset** box, as in SPSS Dialog Box 13.6. Select the **Model** tab and specify TREATMENT, ALCOHOL and ESTEEM as main effects, exactly as in SPSS Dialog Box 13.4. Select the **Save** tab and check **Standardized deviance residuals**. Click **OK**.



SPSS Dialog Box 13.6. The Predictors dialog box for Poisson regression with unequal times.

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

As with the first experiment, we consider goodness of fit in two ways. These help to determine whether there have been gross violations of the assumptions made by our model. Note that these results are identical whether we use the ascending or descending order of factor categories. First consider SPSS Output 13.7, the goodness of fit table which appears near the beginning of the output.

Goodness of Fit[b]

	Value	df	Value/df
Deviance	70.361	70	1.005
Scaled Deviance	70.361	70	
Pearson Chi-Square	70.888	70	1.013
Scaled Pearson Chi-Square	70.888	70	
Log Likelihood[a]	-222.601		
Akaike's Information Criterion (AIC)	455.201		
Finite Sample Corrected AIC (AICC)	456.071		
Bayesian Information Criterion (BIC)	466.789		
Consistent AIC (CAIC)	471.789		

Dependent Variable: pre_trt_events
Model: (Intercept), alcohol, esteem, treatment, offset = offset

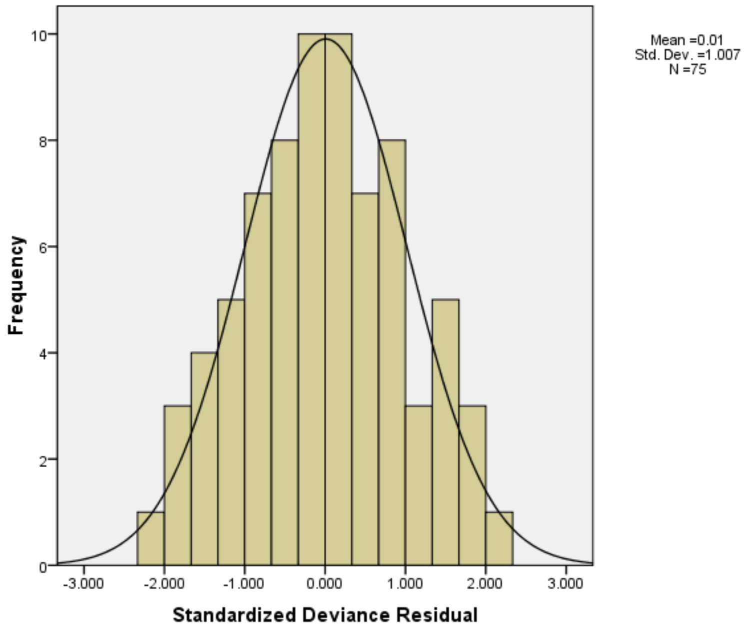
a. The full log likelihood function is displayed and used in computing information criteria.

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

SPSS Output 13.7. Goodness of fit tests for the Poisson regression with unequal time intervals.

As in the equal times model, 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. We see that our values are 1.005 and 1.013, both very close to 1.

SPSS Output 13.8 shows the histogram of the standardized deviance residuals which we saved as we did for the equal times model (and we obtained the histogram in the same way as SPSS Output 13.3). As before, you can see that we have an approximately standard normal distribution. Note that two values fall just outside the interval -2 to +2. However, this is not particularly worrying as -2 to +2 is the 95% confidence interval for the standard normal distribution. Hence we might expect that a small number of the 75 observations to fall outside the range -2 to +2.



SPSS Output 13.8. Histogram of standardized deviance residuals for the unequal time interval experiment

Both our checks suggest that we have a well fitting model for the unequal times data.

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

As before, we can find most of what we need in the table of parameter estimates, shown in SPSS Output 13.9. First consider the *p* values (from the Sig. column) for the two covariates. The *p* values for ALCOHOL and ESTEEM are 0.246 and 0.000 (i.e. < 0.001) respectively. Consistent with the results of the experiment, ALCOHOL is found to be nonsignificant whilst ESTEEM is very highly significant.

After allowing for these covariates, the two TREATMENTs are found to be nonsignificant (with *p* values of 0.105 and 0.150). This shows that neither treatment 1 nor treatment 2 has a PRE_TRT_EVENTS rate that is significantly different from the PRE_TRT_EVENTS rate for treatment 3. However, it is still possible that there is a difference between treatments 1 and 2. We cannot use the results in SPSS Output 13.9 to ascertain whether this is the case since treatment 3 is the reference category. The values of B for treatments 1 and 2 are very similar (-0.095 and -0.078) and their 95% confidence intervals have a large overlap. However, to be sure that treatment groups 1 and 2 do not have a significantly different rate of PRE_TRT_EVENTS we change the reference category to treatment 1.

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	2.634	.1029	2.432	2.836	655.168	1	.000
[alcohol=0]	.060	.0514	-.041	.160	1.345	1	.246
[alcohol=1]	0[a]	.	.	.	.	.	.
esteem	-.037	.0070	-.050	-.023	27.604	1	.000
[treatment=1]	-.095	.0585	-.209	.020	2.622	1	.105
[treatment=2]	-.078	.0540	-.184	.028	2.077	1	.150
[treatment=3]	0[a]	.	.	.	.	.	.
(Scale)	1[b]						

Dependent Variable: pre_trt_events
Model: (Intercept), alcohol, esteem, treatment, offset = offset

a. Set to zero because this parameter is redundant.

b. Fixed at the displayed value.

SPSS Output 13.9. Poisson regression model for unequal times, including TREATMENT, ALCOHOL, ESTEEM and OFFSET.

First, delete the columns containing the predicted value of the linear predictor and the predicted event counts if you have saved them. We have already seen how to change the reference category: 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 13.10.

Parameter Estimates

Parameter	B	Std. Error	95% Wald Confidence Interval Lower	Upper	Hypothesis Test Wald Chi-Square	df	Sig.
(Intercept)	2.599	.1188	2.366	2.832	478.958	1	.000
[alcohol=1]	-.060	.0514	-.160	.041	1.345	1	.246
[alcohol=0]	0[a]	.	.	.	.	.	.
esteem	-.037	.0070	-.050	-.023	27.604	1	.000
[treatment=3]	.095	.0585	-.020	.209	2.622	1	.105
[treatment=2]	.017	.0561	-.093	.127	.090	1	.764
[treatment=1]	0[a]	.	.	.	.	.	.
(Scale)	1[b]						

Dependent Variable: pre_trt_events
Model: (Intercept), alcohol, esteem, treatment, offset = offset

a. Set to zero because this parameter is redundant.

b. Fixed at the displayed value.

SPSS Output 13.10. Poisson regression model including TREATMENT, ALCOHOL, ESTEEM *and* OFFSET*. Reference category levels are the lowest.*

Comparing this model with the version that used the highest category as the reference we note that the estimate associated with ESTEEM is unchanged. The estimate for ALCOHOL is -0.060 rather than +0.060 as we are measuring the difference from 1 to 0 (yes to no) rather than 0 to 1 (no to yes). Similarly the parameter estimate for treatment 3 compared to treatment 1 is just the negative of treatment 1 compared to treatment 3. Note that the *p* values for both ALCOHOL and ESTEEM are unchanged. The *p* value associated with treatment 3 is the same as we found for treatment 1 in the earlier model as both are comparing treatments 1 and 3.

However we can now see the *p* value we wanted, that for treatment 2 compared with treatment 1. It is 0.764, certainly not significant. We have now shown that the annual rate of seizures before the treatment did not vary significantly between the high dose/low dose/placebo treatment arms after allowing for ESTEEM and ALCOHOL. Hence, the research team have more confidence that the differences found in the number of seizures in the year after the experiment were related to their intervention and not a difference in the average number of seizures prior to the experiment.

This was our purpose in fitting this Poisson regression model with unequal times, but we will explore it a bit more to enhance your grasp of the unequal times model.

Obtaining expected PRE_TRT_EVENTS rates and totals from the parameter values

As before, we can use the **Save** button to get the predicted value of the linear predictor added to the datasheet. This time the **Predicted value of mean of response** will be the predicted value of the total of PRE_TRT_EVENTS, the predicted number over the time the case was observed. We now show how to use the parameter values in SPSS Output 13.9 to obtain these. You can achieve the same results with the values in SPSS Output 13.10, as we will demonstrate for one case. Note that, as in the case of equal times, your calculations may differ slightly from the predicted values calculated by SPSS using more decimal places than are shown in the output.

First consider a case with ALCOHOL = 1 and TREATMENT = 3. There is such a case in the last but one row of Table 13.2. This participant has an esteem score of 11. Since ALCOHOL = 1 and TREATMENT = 3 are the reference categories for SPSS Output 13.9, we have zeros for these variables. Look in the B column to see that the constant is 2.634 and the parameter for ESTEEM is -0.037. So for this case, the log of predicted *rate* of PRE_TRT_EVENTS is

$$2.634 - 0.037*11 = 2.227$$

For the predicted rate of PRE_TRT_EVENTS per year, use your calculator to find exp(2.227) = 9.272 (look for e^x or the inverse of ln). This is the predicted rate of PRE_TRT_EVENTS per year.

Multiply by 4.77 years to get the predicted total, 44.22 or 44 to the nearest whole number. Alternatively, to get the log of the predicted total count over the 4.77 YEARS this case was observed, we need to add the log of 4.77 YEARS, the OFFSET (1.562), to give us 3.789. For the predicted count of events of the 4.77 years find exp(3.789) = 44.21. Once again we have 44 predicted PRE_TRT_EVENTS to the nearest whole number (the observed number was 35).

If ALCOHOL = 0 but TREATMENT = 3, we need to add the parameter 0.060 from the B column of SPSS Output 13.9. The case on the bottom row of Table 13.2 is an example, and this person has an ESTEEM score of 13, so the log of the predicted rate of PRE_TRT_EVENTS per year is

$$2.634 + 0.060 - 0.037*13 = 2.213$$

This gives a predicted rate of exp(2.213) = 9.143. Multiply by the time this case was observed, 2.04 years, to get the predicted total PRE_TRT_EVENTS, 18.65 or 19 to the nearest whole number. The observed value was 21.

It the TREATMENT category is not 3 then we need to add the appropriate parameter from the B column of SPSS Output 13.9. If we take the first case from the TREATMENT = 2 group in Table 13.2 (this person has ALCOHOL equal to 0 and an ESTEEM score of 16), we get the log of the predicted rate of PRE_TRT_EVENTS per year as

$$2.634 + 0.060 – 0.078 - 0.037*16 = 2.024$$

Find exp(2.024) = 7.569 to get the predicted rate per year. Multiply this by the years observed, 3.07, to get the predicted total of PRE_TRT_EVENTS as 23.24, or 23 to the nearest whole number. The observed count was 28.

To use the values in SPSS Output 13.10 to obtain the same result for the TREATMENT = 2 case we just considered, notice that this time the constant is 2.599 and the parameter for ALCOHOL = 0 is zero because we used the first instead of the last categories as the reference. The parameter for TREATMENT = 2 is 0.017 and the parameter for ESTEEM remains the same at -0.037. So the predicted log of the rate of PRE_TRT_EVENTS per year is

$$2.599 + 0.017 - 0.037*16 = 2.024$$

This is the same as the result from SPSS Output 13.9.