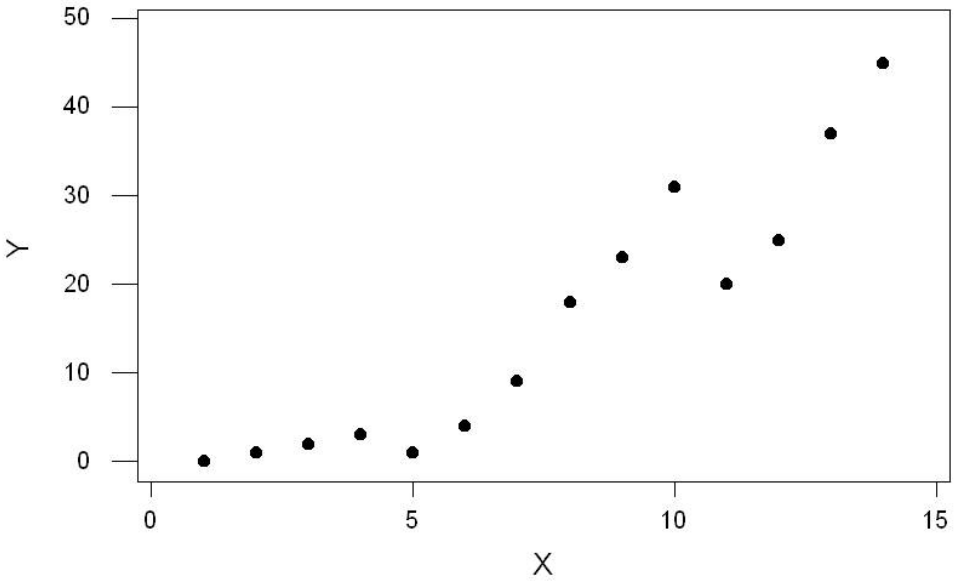
Handout: Poisson Regression

Example 1
Classic Example: Whyte, et al 1987 (Dobson, 1990) reported the number of deaths due to AIDS in Australia *per 3 month period* from January 1983 June 1986.
The data: X = time (measured in multiple of 3 month after January 1983
Y = # of deaths in Australia due to AIDS

xi	yi	xi	yi
1	0	8	18
2	1	9	23
3	2	10	31
4	3	11	20
5	1	12	25
6	4	13	37
7	9	14	45



The graph shows a nonlinear relationship and a loglinear model seems to be a reasonable choice.
Use SPSS for the analysis:

Parameter Estimates

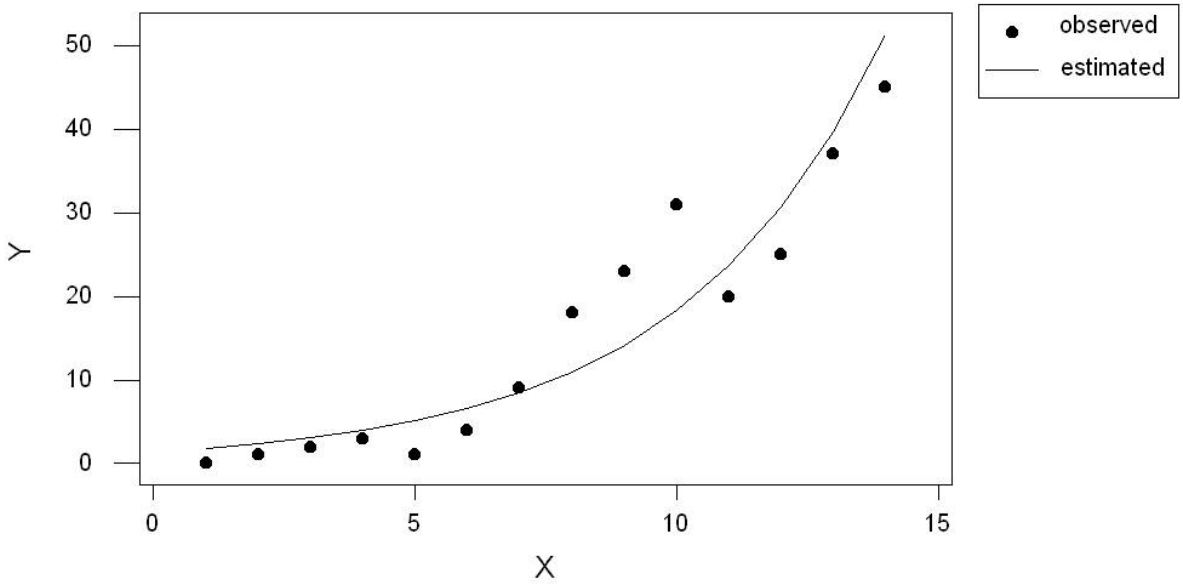
Parameter	B	Std. Error	95% Wald Confidence Interval		Hypothesis Test		
			Lower	Upper	Wald Chi-Square	df	Sig.
(Intercept)	.340	.2512	-.153	.832	1.828	1	.176
X	.257	.0220	.213	.300	135.477	1	.000
(Scale)	1[a]						

Dependent Variable: Y
Model: (Intercept), X
a. Fixed at the displayed value.

The estimated equation is

$$\hat{\mu} = \exp(.340 + .257x)$$

In this period the number of deaths in a year was in average $\exp(.257) = 1.29$ times higher than in the 3 month period before.



The graph shows that in the beginning and the end of the time period the residuals are negative, but in the center period they are positive. This points to the fact that the model is not entirely appropriate. It can be improved by using $\ln(x)$ instead of x as an explanatory variable. But the interpretation of those results is less intuitive, though.

Also, for $x = 9$ and $x = 10$ the model gives not a very close fit.

For the Aids in Australia example, test if time has an effect on the number of death due to Aids.

1. $H_0 : \beta = 0$ versus $H_a : \beta \neq 0$, $\alpha = 0.05$.
2. Assumptions: Random samples, sample size is ok
3. Test statistic: Wald χ^2: $\chi_0^2 = 135.477$, $df = 1$ (from the output)
4. P-value: Wald χ^2: $P = P(\chi^2 > 135) < 0.001$.
5. Conclusion: Reject H_0.
6. Context: At significance level of 5% the data provide sufficient evidence that the time has an effect on the number of death due to Aids in Australia.

Goodness of Fit[b]

	Value	df	Value/df
Deviance	29.654	12	2.471
Scaled Deviance	29.654	12	
Pearson Chi-Square	28.847	12	2.404
Scaled Pearson Chi-Square	28.847	12	
Log Likelihood[a]	-41.290		
Akaike's Information Criterion (AIC)	86.581		
Finite Sample Corrected AIC (AICC)	87.672		
Bayesian Information Criterion (BIC)	87.859		
Consistent AIC (CAIC)	89.859		

Dependent Variable: Y
Model: (Intercept), x

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

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

According to the SPSS output the Deviance for the loglinear model for the number of deaths due to Aids in Australia equals
Deviance = 29.654, df=12(=n-2=n-(number of parameters in the model)). It is hard to judge this value, without knowing the distribution of the deviance. A better measure is deviance/df=2.471, measures "close" to one indicate good model fit. Here the score is not close to one and can be interpreted as lack in model fit.

For comparing the model including time with the model only including the intercept:

1. H_0: The intercept only model fits as well as the model including time versus H_a: H_0 is not true

2. Random sample

3. $$Deviance_0 - Deviance_1 = 177.619, \quad df = 2 - 1$$

4. $P - value < 0.001$

5. Reject H_0

6. At significance level of 5% the data provide sufficient evidence that the model including the time fits better than the intercept only model

	x	Y	est	PearsonResidual	DevianceResidual	StdPearsonResidual
1	1.00	.00	1.82	-1.347	-1.905	-1.417
2	2.00	1.00	2.35	-.879	-.993	-.928
3	3.00	2.00	3.04	-.593	-.632	-.627
4	4.00	3.00	3.93	-.464	-.484	-.492
5	5.00	1.00	5.08	-1.806	-2.210	-1.916
6	6.00	4.00	6.57	-.995	-1.073	-1.054
7	7.00	9.00	8.49	.186	.184	.196
8	8.00	18.00	10.98	2.137	1.953	2.249
9	9.00	23.00	14.20	2.359	2.161	2.475
10	10.00	31.00	18.36	2.980	2.708	3.128
11	11.00	20.00	23.74	-.742	-.762	-.785
12	12.00	25.00	30.69	-.997	-1.029	-1.084
13	13.00	37.00	39.69	-.386	-.391	-.449
14	14.00	45.00	51.32	-.834	-.851	-1.134

The residuals for $x = 9$ and $x = 10$ are quite large, and indicate that at that time another factor influenced the increase in Aids in Australia.