# Chapter 30
Nonnegative Outcomes Assessed with Gamma Distribution (110 Patients)

1 General Purpose

The gamma frequency distribution is suitable for statistical testing of nonnegative data with a continuous outcome variable and fits such data often better than does the normal frequency distribution, particularly when magnitudes of benefits or risks is the outcome, like costs. It is often used in marketing research. This chapter is to assess whether gamma distributions are also helpful for the analysis of medical data, particularly those with outcome scores.
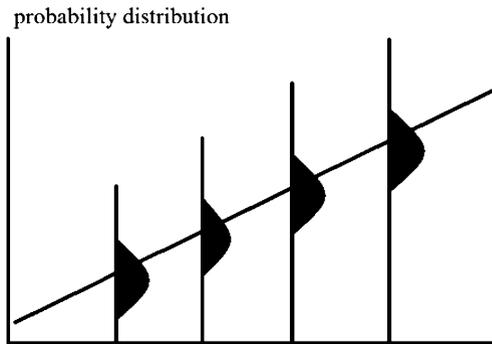
probability distribution

Linear regression where the measured y-values are assumed to have uncertainties in the form of identical normal curves (Gaussian curves)

probability distribution

Linear regression where the measured y-values
are assumed to have uncertainties in the form of
identical gamma frequency distributions instead
of normal curves

The upper graph gives a schematic view of a linear regression using normal
probability distributions around different y-values, the lower graph does equally so,
but uses probability distributions of the gamma type (skewed to the right). Skewed
data like quality of life (QOL) scores in sick populations (that are clustered towards
low QOL scores) better fit gamma distributions, than they do normal distributions.
More background and mathematical information of gamma distributions is given in
Machine learning in medicine a complete overview, Chap. 80, Heidelberg Springer
Germany, 2015, from the same authors.

2   General Overview of Type of Data File

Outcome nonnegative values	predictor	predictor	predictor
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.

3 Primary Scientific Question

Is gamma regression a worthwhile analysis model complementary to linear regression, can it elucidate effects unobserved in the linear models.

4 Data Example

In 110 patients the effects of age class, psychological and social score on health scores were assessed. The first ten patients are underneath. The entire data file is entitled "chapter30gamma", and is in extras.springer.com.

Health score	Age class	Psychologic score	Social score
8	3	5	4
7	1	4	8
4	1	5	13
6	1	4	15
10	1	7	4
6	1	8	8
8	1	9	12
2	1	8	16
6	1	12	4
8	1	13	1

age = age class 1–7
psychologicscore = psychological score 1–20
socialscore = social score 1–20
healthscore = health score 1–20.

Start by opening the data file in SPSS statistical software. We will first perform linear regressions.

5 Linear Regressions

For analysis the statistical model Linear in the module Regression is required.

Command:
Analyze….Regression….Linear….Dependent: enter healthscore….Independent (s): enter socialscore….click OK.

The underneath table gives the result. Social score seems to be a very significant predictor of health score.

Coefficients[a]

Model		Unstandardized coefficients		Standardized coefficients		
		B	Std. error	Beta	t	Sig.
1	(Constant)	9.833	.535		18.388	.000
	Social score	−.334	.050	−.541	−6.690	.000

[a]Dependent Variable: health score

Similarly psychological score and age class are tested.

Coefficients[a]

Model		Unstandardized coefficients		Standardized coefficients		
		B	Std. error	Beta	t	Sig.
1	(Constant)	5.152	.607		8.484	.000
	Psychological score	.140	.054	.241	2.575	.011

[a]Dependent Variable: health score

Coefficients[a]

Model		Unstandardized coefficients		Standardized coefficients		
		B	Std. error	Beta	t	Sig.
1	(Constant)	7.162	.588		12.183	.000
	Age class	−.149	.133	−.107	−1.118	.266

[a]Dependent Variable: health score

Linear regression with the three predictors as independent variables and health scores as outcome suggests that both psychological and social scores are significant predictors of health, but age class is not. In order to assess confounding and interaction a multiple linear regression is performed.

Command:
Analyze....Regression....Linear....Dependent: enter healthscore....Independent (s): enter socialscore, psychologicscore, age....click OK.

Coefficients[a]

Model		Unstandardized coefficients		Standardized coefficients		
		B	Std. error	Beta	t	Sig.
1	(Constant)	9.388	.870		10.788	.000
	Social score	−.329	.049	−.533	−6.764	.000
	Psychological score	.111	.046	.190	2.418	0.17
	Age class	−.184	.109	−.132	−1.681	.096

[a]Dependent Variable: health score

Social score is again very significant. Psychological score also, but after Bonferroni adjustment (rejection p-value = 0,05/4 = 0,0125) it would be no more so, because p = 0,017 is larger than 0,0125. Age class is again not significant. Health score is here a continuous variable of nonnegative values, and, perhaps, better fit of these data might be obtainable by a gamma regression. We will use SPSS statistical software again.

For analysis the module Generalized Linear Models is required. It consists of two submodules: Generalized Linear Models and Generalized Estimation Models. The first submodule covers many statistical models like gamma regression (current chapter), Tweedie regression (Chap. 31), Poisson regression (Chaps. 21 and 47), and the analysis of paired outcomes with predictors (Chap. 3). The second is for analyzing binary outcomes (Chap. 42). We will use the statistical model Gamma Distribution in the submodule Generalized Linear Models.

6 Gamma Regression

Command:

Analyze….click Generalized Linear Models….click once again Generalized Linear Models….mark Custom….Distribution: select Gamma….Link function: select Power….Power: type −1….click Response….Dependent Variable: enter healthscore click Predictors….Factors: enter socialscore, psychologicscore, age….Model: enter socialscore, psychologicscore, age….Estimation: Scale Parameter Method: select Pearson chi-square….click EM Means: Displays Means for: enter age, psychologicscore, socialscore….click Save….mark Predict value of linear predictor….Standardize deviance residual….click OK.

Tests of model effects

	Type III		
Source	Wald Chi-Square	df	Sig.
(Intercept)	216.725	1	.000
Ageclass	8.838	6	.183
Psychologicscore	18.542	13	.138
Socialscore	61.207	13	.000

Dependent Variable: health score
Model: (Intercept), ageclass, psychologicscore, socialscore

The above table give the overall result: it is comparable with that of the multiple linear regression with only social score as significant independent predictor.

Parameter estimates

Parameter	B	Std. error	95% Wald confidence interval		Hypothesis test		
			Lower	Upper	Wald Chi-square	df	Sig.
(Intercept)	.188	.0796	.032	.344	5.566	1	.018
[ageclass = 1]	−.017	.0166	−.050	.015	1.105	1	.293
[ageclass = 2]	−.002	.0175	−.036	.032	.010	1	.919
[ageclass = 3]	−.015	.0162	−.047	.017	.839	1	.360
[ageclass = 4]	.014	.0176	−.020	.049	.658	1	.417
[ageclass = 5]	.025	.0190	−.012	.062	1.723	1	.189
[ageclass = 6]	.005	.0173	−.029	.039	.087	1	.767
[ageclass = 7]	0[a]	.	.	.	.	.	.
[psychologies core = 3]	.057	.0409	−.023	.137	1.930	1	.165
[psychologies core = 4]	.057	.0220	.014	.100	6.754	1	.009
[psychologies core = 5]	.066	.0263	.015	.118	6.352	1	.012
[psychologies core = 7]	.060	.0311	−.001	.121	3.684	1	.055
[psychologies core = 8]	.061	.0213	.019	.102	8.119	1	.004
[psychologies core = 9]	.035	.0301	−.024	.094	1.381	1	.240
[psychologies core = 11]	.057	.0325	−.007	.120	3.059	1	.080
[psychologies core = 12]	.060	.0219	.017	.103	7492	1	.006
[psychologies core = 13]	.040	.0266	−.012	.092	2.267	1	.132
[psychologies core = 14]	.090	.0986	−.103	.283	.835	1	.361
[psychologies core = 15]	.121	.0639	−.004	.247	3.610	1	.057
[psychologies core = 16]	.041	.0212	−.001	.082	3.698	1	.054
[psychologies core = 17]	.022	.0241	−.025	.069	.841	1	.359
[psychologies core = 18]	0[a]	.	.	.	.	.	.
[socialscore = 4]	−.120	.0761	−.269	.029	2.492	1	.114
[socialscore = 6]	−.028	.0986	−.221	.165	.079	1	.778
[socialscore = 8]	−.100	.0761	−.249	.050	1.712	1	.191
[socialscore = 9]	.002	.1076	−.209	.213	.000	1	.988
[socialscore = 10]	−.123	.0864	−.293	.046	2.042	1	.153
[socialscore = 11]	.015	.0870	−.156	.185	.029	1	.865

| Parameter | B | Std. error | 95% Wald confidence interval | | Hypothesis test | | |
			Lower	Upper	Wald Chi-square	df	Sig.
[socialscore = 12]	−.064	.0772	−.215	.088	.682	1	.409
[socialscore = 13]	−.065	.0773	−.216	.087	.703	1	.402
[socialseore = 14]	.008	.0875	−.163	.180	.009	1	.925
[socialscore = 15]	−.051	.0793	−.207	.104	.420	1	.517
[socialscore = 16]	.026	.0796	−.130	.182	.107	1	.744
[socialscore = 17]	−.109	.0862	−.277	.060	1.587	1	.208
[socialscore = 18]	−.053	.0986	−.246	.141	.285	1	.593
[socialscore = 19]	0[a]	.	.	.	.	.	.
(Scale)	.088[b]						

Dependent Variable: health score
Model: (Intercept), ageclass, psychologicscore, socialscore
[a]Set to zero because this parameter is redundant
[b]Computed based on the Pearson chi-square

However, as shown above, gamma regression enables to test various levels of the predictors separately. Age class was not a significant predictor. Of the psychological scores, however, no less than 8 scores produced pretty small p-values, even as small as 0,004–0,009. Of the social scores now none were significant.

In order to better understand what is going on, SPSS provides marginal means analysis here.

Estimates

| Age class | Mean | Std. error | 95% Wald confidence interval | |
			Lower	Upper
1	5.62	.531	4.58	6.66
2	5.17	.461	4.27	6.07
3	5.54	.489	4.59	6.50
4	4.77	.402	3.98	5.56
5	4.54	.391	3.78	5.31
6	4.99	.439	4.13	5.85
7	5.12	.453	4.23	6.01

The mean health scores of the different age classes were, indeed, hardly different.

Estimates

Psychological score	Mean	Std. error	95% Wald confidence interval	
			Lower	Upper
3	5.03	.997	3.08	6.99
4	5.02	.404	4.23	5.81
5	4.80	.541	3.74	5.86
7	4.96	.695	3.60	6.32
8	4.94	.359	4.23	5.64
9	5.64	.809	4.05	7.22
11	5.03	.752	3.56	6.51
12	4.95	.435	4.10	5.81
13	5.49	.586	4.34	6.64
14	4.31	1.752	.88	7.74
15	3.80	.898	2.04	5.56
16	5.48	.493	4.51	6.44
17	6.10	.681	4.76	7.43
18	7.05	1.075	4.94	9.15

However, increasing psychological scores seem to be associated with increasing levels of health.

Estimates

Social score	Mean	Std. error	95% Wald confidence interval	
			Lower	Upper
4	8.07	.789	6.52	9.62
6	4.63	1.345	1.99	7.26
8	6.93	.606	5.74	8.11
9	4.07	1.266	1.59	6.55
10	8.29	2.838	2.73	13.86
11	3.87	.634	2.62	5.11
12	5.55	.529	4.51	6.59
13	5.58	.558	4.49	6.68
14	3.96	.711	2.57	5.36
15	5.19	.707	3.81	6.58
16	3.70	.371	2.98	4.43
17	7.39	2.256	2.96	11.81
18	5.23	1.616	2.06	8.40
19	4.10	1.280	1.59	6.61

In contrast, increasing social scores are, obviously, associated with decreasing levels of health, with mean health scores close to 3 in the higher social score patients, and close to 10 in the lower social score patients.

7 Conclusion

Gamma regression is a worthwhile analysis model complementary to linear regression, ands may elucidate effects unobserved in the linear models. The marginal means procedure readily enables to observe trends in the data, e.g., decreasing outcome score with increasing predictor scores.

8 Note

More background, theoretical and mathematical information of gamma regression is given in Machine learning in medicine a complete overview, Chap. 80, Heidelberg Springer Germany, 2015, from the same authors.