

13

Weitere Regressionsarten

Dominic Schmitz & Janina Esser

Regression für jeden Geschmack

... Regression	abhängige Variable	Beispiel
Lineare	numerisch	Reaktionszeiten, Dauern
Logistische/Binominale	binär kategorisch	richtig/falsch
Ordinale	geordnet kategorisch	Likert-Skala
Multinominale	kategorisch	Wortarten
Quantil	numerisch mit Autokorrelation	Mouse-Tracking-Daten
Poisson	integer	Zähldaten
Polynomielle		
Ridge		
Lasso		
...		

Regression für jeden Geschmack

... Regression	abhängige Variable	Beispiel
Lineare	numerisch	Reaktionszeiten, Dauern
Logistische/Binominale	binär kategorisch	richtig/falsch
Ordinale	geordnet kategorisch	Likert-Skala
Multinominale	kategorisch	Wortarten
Quantil	numerisch mit Autokorrelation	Mouse-Tracking-Daten
Poisson	integer	Zähldaten
Polynomielle		
Ridge		
Lasso		
...		

Regression für jeden Geschmack

... Regression	abhängige Variable	Beispiel
Lineare	numerisch	Reaktionszeiten, Dauern
Logistische/Binominale	binär kategorisch	richtig/falsch
Ordinale	geordnet kategorisch	Likert-Skala
Multinominale	kategorisch	Wortarten
Quantil	numerisch mit Autokorrelation	Mouse-Tracking-Daten
Poisson	integer	Zählraten
Polynomielle		
Ridge		
Lasso		
...		

Regression für jeden Geschmack

... Regression	abhängige Variable	Beispiel
Lineare	numerisch	Reaktionszeiten, Dauern
Logistische/Binominale	binär kategorisch	richtig/falsch
Ordinale	geordnet kategorisch	Likert-Skala
Multinominale	kategorisch	Wortarten
Quantil	numerisch mit Autokorrelation	Mouse-Tracking-Daten
Poisson	integer	Zähldaten
Polynomielle		
Ridge		
Lasso		
...		

Logistische Regression

- Modellierung von Zusammenhängen zwischen einer binären abhängigen Variablen und einer oder mehreren unabhängigen Variablen
- Schätzt die Wahrscheinlichkeit oder die Odds Ratio für das Eintreten des Ereignisses als Funktion der unabhängigen Variablen
- Basiert auf der logistischen Verteilungsfunktion, die eine S-förmige Kurve bildet und die Wahrscheinlichkeit für das Eintreten eines Ereignisses im Bereich von 0 bis 1 abbildet

Logistische Regression

Beispiel: Same-Different Task

VPs sollen beantworten, ob sie 2x das gleiche Wort oder beide Wörter eines englischen Homophonpaares gehört haben, z. B. *freeze* vs. *frees*

korrekt	Differenz	Wortart	Komplexität	Wortlänge	VP
0	5	Nomen	simplex	9	VP1
0	5	Verb	komplex	21	VP1
0	25	Nomen	simplex	13	VP1
1	25	Verb	komplex	12	VP1
1	45	Nomen	simplex	4	VP1
1	45	Verb	komplex	25	VP1
...

Logistische Regression

Beispiel: Same-Different Task

VPs sollen beantworten, ob sie 2x das gleiche Wort oder beide Wörter eines englischen Homophonpaares gehört haben, z. B. *freeze* vs. *frees*

korrekt ~

Differenz +

Wortart +

Komplexität +

Wortlänge +

(1 | versuchsperson)

Logistische Regression

Beispiel: Same-Different Task

VPs sollen beantworten, ob sie 2x das gleiche Wort oder beide Wörter eines englischen Homophonpaares gehört haben, z. B. *freeze* vs. *frees*

	Chisq	Df	p-value
Differenz	173.04	2	0.000 ***
Wortart	0.35	1	0.553
Komplexität	0.02	1	0.880
Wortlänge	0.74	1	0.391

Logistische Regression

Beispiel: Same-Different Task

VPs sollen beantworten, ob sie 2x das gleiche Wort oder beide Wörter eines englischen Homophonpaares gehört haben, z. B. *freeze* vs. *frees*

	Estimate	Std. Error	z-value	p-value	
Intercept	-3.246	0.532	-6.096	0.000	***
Differenz_25	2.355	0.343	6.864	0.000	***
Differenz_45	5.604	0.432	12.985	0.000	***
wortart_Verb	0.161	0.271	0.593	0.553	
Komplexität_komplex	0.039	0.261	0.150	0.880	
wortlänge	0.018	0.021	0.858	0.391	

Regression für jeden Geschmack

... Regression	abhängige Variable	Beispiel
Lineare	numerisch	Reaktionszeiten, Dauern
Logistische/Binominale	binär kategorisch	richtig/falsch
Ordinale	geordnet kategorisch	Likert-Skala
Multinominale	kategorisch	Wortarten
Quantil	numerisch mit Autokorrelation	Mouse-Tracking-Daten
Poisson	integer	Zählraten
Polynomielle		
Ridge		
Lasso		
...		

Ordinale Regression

- Modellierung von Zusammenhängen zwischen einer ordinalen abhängigen Variablen und einer oder mehreren unabhängigen Variablen
- Schätzt (ähnlich der logistischen Regression) die Wahrscheinlichkeit oder die Odds Ratio für das Eintreten des Ereignisses als Funktion der unabhängigen Variablen
- Es wird angenommen, dass die Kategorien der abhängigen Variablen in einer bestimmten Reihenfolge angeordnet sind, wobei der Abstand zwischen den Kategorien gleichmäßig ist

Ordinale Regression

Beispiel: Likert-Skala

VPs sollen auf einer Skala von 1 bis 5 angeben, wie groß ein gezeigtes Alien ist, dessen Namen (= ein Pseudowort) sie vorgespielt bekommen

Größe	Vokal	Onset1	Onset2	Alter
5	a	k	d	19
3	e	f	j	19
4	o	r	f	19
1	i	j	r	37
2	u	d	k	37
...

Ordinale Regression

Beispiel: Likert-Skala

VPs sollen auf einer Skala von 1 bis 5 angeben, wie groß ein gezeigtes Alien ist, dessen Namen (= ein Pseudowort) sie vorgespielt bekommen

Größe ~
vokal +
onset1 +
onset2 +
Alter

Ordinale Regression

Beispiel: Likert-Skala

VPs sollen auf einer Skala von 1 bis 5 angeben, wie groß ein gezeigtes Alien ist, dessen Namen (= ein Pseudowort) sie vorgespielt bekommen

	Chisq	Df	p-value
Differenz	689.11	7	0.000 ***
Wortart	74.49	4	0.000 ***
Komplexität	70.14	4	0.000 ***
Wortlänge	3.49	1	0.062

Ordinale Regression

Beispiel: Likert-Skala

VPs sollen auf einer Skala von 1 bis 5 angeben, wie groß ein gezeigtes Alien ist, dessen Namen (= ein Pseudowort) sie vorgespielt bekommen

	Estimate	Std. Error	z-value	p-value
vokal_A	-0.561	0.066	-8.464	0.000 ***
vokal_e	-0.622	0.067	-9.332	0.000 ***
vokal_i	-1.637	0.069	-23.601	0.000 ***
...
onset1_f	-0.105	0.066	-1.575	0.115
...
onset2_f	0.016	0.066	0.248	0.804
...
Alter	-0.002	0.001	-1.869	0.062

Regression für jeden Geschmack

... Regression	abhängige Variable	Beispiel
Lineare	numerisch	Reaktionszeiten, Dauern
Logistische/Binominale	binär kategorisch	richtig/falsch
Ordinale	geordnet kategorisch	Likert-Skala
Multinominale	kategorisch	Wortarten
Quantil	numerisch mit Autokorrelation	Mouse-Tracking-Daten
Poisson	integer	Zählraten
Polynomielle		
Ridge		
Lasso		
...		

Multinominale Regression

- Modellierung von Zusammenhängen zwischen einer nominalen abhängigen Variablen und einer oder mehreren unabhängigen Variablen
- Schätzt (ähnlich der logistischen Regression) die Wahrscheinlichkeit oder die Odds Ratio für das Eintreten des Ereignisses als Funktion der unabhängigen Variablen
- Es wird angenommen, dass die Kategorien der abhängigen Variablen in keiner bestimmten Reihenfolge angeordnet sind

Multinominale Regression

Beispiel: Generische Maskulina vs. Spezifische Maskulina & Feminina

Anhand von semantischen Messwerten soll vorhergesagt werden, ob es sich um ein generisches Maskulinum oder um ein spezifisches Maskulinum/Femininum handelt

Form	Stereotyp- wertung	Verständnis- qualität	Nachbarschafts- dichte	Aktivierungs- vielfalt
GM	74.87	0.811	0.918	5.041
SM	74.87	0.809	0.917	5.042
SF	74.87	0.980	0.991	8.310
GM	24.01	0.901	0.681	4.312
SM	24.01	0.923	0.682	4.313
...

Multinominale Regression

Beispiel: Generische Maskulina vs. Spezifische Maskulina & Feminina

Anhand von semantischen Messwerten soll vorhergesagt werden, ob es sich um ein generisches Maskulinum oder um ein spezifisches Maskulinum/Femininum handelt

Form ~

Stereotypwertung +
Verständnisqualität +
Nachbarschaftsdichte +
Aktivierungsvielfalt

Multinominale Regression

Beispiel: Generische Maskulina vs. Spezifische Maskulina & Feminina

Anhand von semantischen Messwerten soll vorhergesagt werden, ob es sich um ein generisches Maskulinum oder um ein spezifisches Maskulinum/Femininum handelt

	Chisq	Df	p-Value
Stereotypwertung	3.734	2	0.155
Verständnisqualität	42.669	2	0.000 ***
Nachbarschaftsdichte	38.951	2	0.000 ***
Aktivierungsvielfalt	10.128	2	0.006 **

Multinominale Regression

Beispiel: Generische Maskulina vs. Spezifische Maskulina & Feminina

Anhand von semantischen Messwerten soll vorhergesagt werden, ob es sich um ein generisches Maskulinum oder um ein spezifisches Maskulinum/Femininum handelt

	Estimate	Std. Error	z-value	p-value
(Intercept)	49.923	3.761	13.275	0.000 ***
Stereotypwertung	0.019	0.011	1.754	0.079
Verständnisqualität	36.542	5.756	6.349	0.000 ***
Nachbarschaftsdichte	-81.108	4.123	-19.674	0.000 ***
Aktivierungsvielfalt	-1.037	0.335	-3.092	0.002 **

Regression für jeden Geschmack

... Regression	abhängige Variable	Beispiel
Lineare	numerisch	Reaktionszeiten, Dauern
Logistische/Binominale	binär kategorisch	richtig/falsch
Ordinale	geordnet kategorisch	Likert-Skala
Multinominale	kategorisch	Wortarten
Quantil	numerisch mit Autokorrelation	Mouse-Tracking-Daten
Poisson	integer	Zähldaten
Polynomielle		
Ridge		
Lasso		
...		

Quantil-Regression

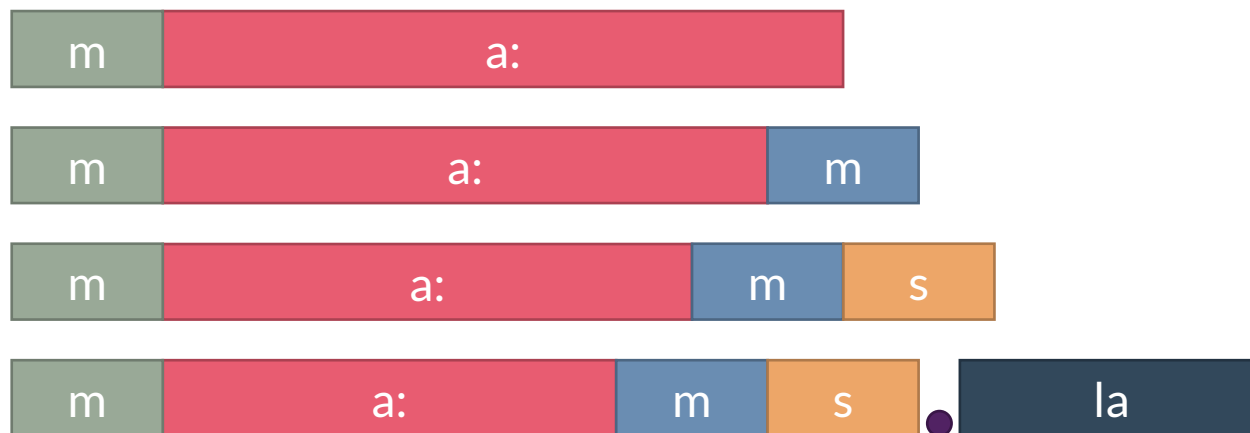
- Modellierung von Zusammenhängen zwischen einer numerischen abhängigen Variablen und einer oder mehreren unabhängigen Variablen
- Funktioniert prinzipiell wie lineare Regression, nutzt aber die Daten spezifizierter konditionaler Quantile (statt aller Daten)
- Bietet eine robuste Schätzung der Bedingungswahrscheinlichkeiten, insbesondere wenn die Daten nicht normal verteilt, heteroskedastisch und/oder autokorreliert sind

Achtung: Bei Quantil-Regression befinden wir uns im Reich der nicht-linearen Regression – Effekte können also nicht-linearer Natur sein!

Quantil-Regression

Beispiel: Compensatory Vowel Shortening in German

Stressed Vowels sind kürzer je nachdem wie viele Konsonanten ihnen folgen



Quantil-Regression

Beispiel: Compensatory Vowel Shortening in German

Stressed Vowels sind kürzer je nachdem wie viele Konsonanten ihnen folgen

Dauer \sim

Sprechgeschwindigkeit +

Silbenstruktur +

vokal +

(1 | Wort) +

(1 | VP),

qu = c(0.3, 0.6)

Quantil-Regression

Beispiel: Compensatory Vowel Shortening in German

Stressed Vowels sind kürzer je nachdem wie viele Konsonanten ihnen folgen

	0.3 Quantil		0.6 Quantil	
	Estimate	p-value	Estimate	p-value
(Intercept)	0.157	0.000 ***	0.172	0.000 ***
vokal_e	-0.025	0.000 ***	-0.025	0.000 ***
vokal_i	-0.038	0.000 ***	-0.037	0.000 ***
vokal_o	-0.023	0.000 ***	-0.021	0.000 ***
vokal_u	-0.041	0.000 ***	-0.038	0.000 ***
Struktur_open	0.060	0.000 ***	0.075	0.000 ***
Struktur_single	0.016	0.000 ***	0.017	0.000 ***
	edf	p-value	edf	p-value
Geschwindigkeit	1.001	0.038 *	1.052	0.636
Wort	4.141	0.057	3.031	0.044 *
VP	8.922	0.000 ***	8.907	0.000 ***

Regression für jeden Geschmack

... Regression	abhängige Variable	Beispiel
Lineare	numerisch	Reaktionszeiten, Dauern
Logistische/Binominale	binär kategorisch	richtig/falsch
Ordinale	geordnet kategorisch	Likert-Skala
Multinominale	kategorisch	Wortarten
Quantil	numerisch mit Autokorrelation	Mouse-Tracking-Daten
Poisson	integer	Zählraten
Polynomielle		
Ridge		
Lasso		
...		

Poisson-Regression

- Modellierung von Zusammenhängen zwischen einer integeren abhängigen Variablen und einer oder mehreren unabhängigen Variablen
- Basiert auf der Poisson-Verteilung, die zur Modellierung von diskreten Zählvariablen verwendet wird
- Nützliches Werkzeug zur Modellierung von Zählvariablen zur Identifikation der Faktoren, die die Häufigkeit von Ereignissen beeinflussen

Poisson-Regression

Beispiel: Auszeichnungen an High Schools

Anhand des Kurrikulums und der Mathenoten soll die Menge der Auszeichnungen modelliert werden

Auszeichnungen	Kurrikulum	Mathenote
0	allgemein	41
0	allgemein	42
0	Ausbildung	47
1	allgemein	53
2	akademisch	67
...

Poisson-Regression

Beispiel: Auszeichnungen an High Schools

Anhand des Kurrikulums und der Mathenoten soll die Menge der Auszeichnungen modelliert werden

Auszeichnungen \sim
Kurrikulum +
Mathenote

Poisson-Regression

Beispiel: Auszeichnungen an High Schools

Anhand des Kurrikulums und der Mathenoten soll die Menge der Auszeichnungen modelliert werden

	Chisq	Df	p-value
Kurrikulum	14.572	2	0.000 ***
Mathenote	45.010	1	0.000 ***

Poisson-Regression

Beispiel: Auszeichnungen an High Schools

Anhand des Kurrikulums und der Mathenoten soll die Menge der Auszeichnungen modelliert werden

	Estimate	Std. Error	z-value	p-value
(Intercept)	-5.247	0.658	-7.969	0.000 ***
kurrikulum_aka	1.084	0.358	3.025	0.002 **
kurrikulum_aus	0.370	0.441	0.838	0.402
Mathenote	0.070	0.011	6.619	0.000 ***

Regression für jeden Geschmack

... Regression	abhängige Variable	Beispiel
Lineare	numerisch	Reaktionszeiten, Dauern
Logistische/Binominale	binär kategorisch	richtig/falsch
Ordinale	geordnet kategorisch	Likert-Skala
Multinominale	kategorisch	Wortarten
Quantil	numerisch mit Autokorrelation	Mouse-Tracking-Daten
Poisson	integer	Zählraten
Polynomielle		
Ridge		
Lasso		
...		