

Lineární a nelineární závislost v GLM

Jako téma svého úkolu jsem si vybrala všeobecný lineární model, a to z toho důvodu, že jsem na bakaláři ve staré akreditace neměla předmět Lineární modely a ráda bych si tuto problematiku více procvičila.

V této práci budu zkoumat tvar závislosti na datech, kdy dobrovolníci reprodukují naučenou sadu slov. Data jsem získala z průběžného úkolu z výše uvedeného předmětu od svých spolužáků.

Řada výzkumů v psychologii se zabývá tím, do jaké míry je vybavnost paměti ovlivněna obsahem toho, co se učíme. Například P. Delin provedl výzkum, jež zjistil, že bizarnost slov má určitý vliv na proces vybavení slov z paměti (Delin, 1969).

Data obsahují: číslo respondenta, čas prodlevy mezi naučením a reprodukcí, počet vybavených dvojic, sada slov označující 0=obvyklá dvojice, 1=bizarní dvojice.

Závislá proměnná:

- počet vybavených dvojic

Regresor:

- bizarní a obvyklá dvojice
- čas prodlevy mezi naučením a reprodukcí

K analýze dat jsem použila software TIBCO Statistica, verze 14.

Model 1

V prvním kroku model formuluji s lineární závislostí. Konkrétně tedy modeluji závislost počtu vybavených dvojic na času a sadě slov.

Tento model má koeficient determinace $R = 0,405$ (adjustovaný $R = 0,399$), tedy vysvětluje 40 % rozptylu, $F(2, 197) = 67,27$, $p < 0,001$.

Jednotlivé odhady koeficientů jsou v tabulce níže:

Tabulka č. 1

		Počet dvojic	t	p
Intercept		14,148	34,436	0,000
Čas		-0,045	-11,545	0,000
Sada slov	1	0,448	1,140	0,256
Sada slov	0	0,000		

Odhady koeficientů říkají, že průměrně si respondent vybaví 14,5 dvojic slov, když dojde k bezprostřední prezentaci (bez časové prodlevy) a bude se jednat o sadu běžných dvojic. Za každou jednotku času navíc klesne v průměru počet vybavených dvojic o -0,045. Počátek a koeficient u proměnné čas jsou statisticky významné, $p < 0,001$. Nicméně regresor sada slov významný není (a to i při testování celé sady proměnných, $p = 0,255$).

Model 2

Nyní model upravím tak, aby zvažoval i vzájemnou interakci. Předpokládám, že jedna sada slov se bude respondentům vybavovat hůře/lépe než ta druhá, přidám tedy vzájemnou interakci sady a času.

Tento model má nyní koeficient determinace $R = 0,474$, adjustovaný $R = 0,465$, $F(3, 196) = 58,95$, $p < 0,001$. Koeficienty **se zvýšili**.

Odhady všech parametrů jsou nyní statisticky významné, pro všechny $p < 0,001$.

Tabulka č. 2

		Počet dvojic	t	p
Intercept		15,572	32,495	0,000
Čas		-0,064	-12,107	0,000
Sada slov	1	-2,277	-3,478	0,001
Sada slov	0	0,000		
Sada slov*Cas	1	0,037	5,048	0,000
Sada slov*Cas	2	0,000		

Vzrůstající čas má opět negativní vliv na počet vybavených dvojic. Bizarní sada slov také, dokonce větší efekt než čas.

Model 3

Pokusím se nyní model upravit a uvažovat kvadratickou funkci času. Nový model má koeficient determinace $R = 0,475$, adjustovaný $R = 0,465$, model vysvětluje nenulové množství rozptylu $F(4, 195) = 44,22$, $p < 0,001$. Hodnoty jsou stejné jako u předchozího modelu.

Odhady koeficientů uvádím v tabulce.

Tabulka č. 3

		Počet dvojic	t	p
Intercept		15,341	26,751	0,000
Čas		-0,054	-3,731	0,000
Sada slov	1	-2,272	-3,466	0,001
Sada slov	0	0,000		
Sada slov*Čas	1	0,037	4,969	0,000
Sada slov*Čas	2	0,000		
Čas ²		0,000	-0,735	0,463

Domnívám se, že důvodem, proč koeficient determinace nevzrostl, je ten, že regresor čas na druhou není statisticky významný ($p=0,463$), tedy nemůžeme zamítnout nulovou hypotézu o tom, že je roven nule. Tedy tato úprava nám ve vysvětlení dalšího rozptylu nepomohla.

Závěr

Z proběhlých transformací je zřejmé, že existuje určitá souvislost mezi slovy, jež jsou bizarní a paměťovými procesy. Tuto tezi mj. potvrzuje výzkum Reifera a Iamayové. Závěrem jejich studie je to, že bizarní výrazy se vynořují z dlouhodobé paměti jednodušeji než obyčejná slova (Reifer, Iamayová 1998).

V tomto úkolu jsem si chtěla přiblížit znalosti lineární regrese a jejích možnosti. Jsem ráda za tu možnost si přiblížit tuto problematiku, kterou jsem na bakalářském studiu neměla. Přijde mi užitečné se takto moci podívat hlouběji na to, jaké regresory mají významný vliv na závislé proměnné a u jakých je vliv naopak nižší. Skrze toto povědomí můžeme tak více porozumět výsledkům daných studií a rozpoznat tak, co je pro danou problematiku důležité a naopak.

Zdroje informací

- P. S. Delin. (2013). The learning to criterion of a serial list with and without mnemonic instructions. *Psychonomic Science*, 16, 169–170.
- Riefer, D. M., & lamay, M. L. (1998). Memory for common and bizarre stimuli: A storage-retrieval analysis. *Psychonomic Bulletin & Review*, 5(2), 312–317.
<https://doi.org/10.3758/bf03212957>
- Data z průběžných úkolu z předmětu Statistické metody 3 jsou k nalezení zde:
<https://dostal.vyzkum-psychologie.cz/stat4?i=40>