Homework Week 10 (Real Data)

Simon Chen
Apr 5 2022

Attention: *For the remainder of the homework assignments this semester, it is suggested that you work with the same dataset and group that you plan to use for your final project. This gives you an opportunity to get to know your data and to do preliminary analyses (with feedback).*

This Homework builds on Homework Week 9.

1) Use the same dataset selected in Homework Week 9, explore the development of different question types and run some models.

   (a) What is one **descriptive question** you could ask with this dataset? Write out a model that could be used to answer this question, highlighting why you included the particular covariates you selected and which coefficients would answer your question.

   Describe how news clicks changed over time.
   News clicks = $\alpha + \beta_1 Time_{1i} + \beta_2 Time_{2i} + ... + \beta_j Time_{i} + \varepsilon$
   Clicks # , news, Yes # Users, # Clicks, News ID as particular covariates, coefficient ε would answer the question.

   (b) What is one **causal question** you could ask with this dataset? Write out a model that could be used to answer this question, highlighting why you included the particular covariates you selected and which coefficients would answer your question.

   I would like to know the effect of the yes # users and news ID to clicks.
 $Y_i = \alpha + \gamma TRT_i + \varepsilon$
   Particular covariates: User gender, age, race, religions.
   Coefficient γ would answer the question.

   (c) What is one **predictive question** you could ask with this dataset? Write out a model that could be used to answer this question, highlighting why you included the particular covariates you selected and which coefficients would answer your question.

   What are factors that lead young people to news and clicks?
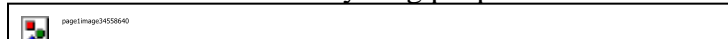
   Particular covariates: Race, gender, age, religions
   ε would answer the question.

(d) Select one of the models provided in (a) – (c). Estimate the model in SPSS. Provide your table of coefficient estimates.

Information Criteria[a]

−2 Log Likelihood	6138.197
Akaike's Information Criterion (AIC)	6190.197
Hurvich and Tsai's Criterion (AICC)	6190.685
Bozdogan's Criterion (CAIC)	6371.508
Schwarz's Bayesian Criterion (BIC)	6345.508

The information criteria are displayed in smaller−is−better form.

Type III Tests of Fixed Effects[a]

Source	Numerator df	Denominator df	F	Sig.
Intercept	1	2903	592.430	.000
V1_50	5	2903.000	15.045	.000
V1_52	5	2903.000	3.763	.002
V1_54	5	2903.000	12.218	.000
V1_56	5	2903.000	1.909	.090
race	1	2903	15.453	.000
sex	1	2903	37.497	.000
age	1	2903.000	1.412	.235
religion	1	2903.000	4.006	.045

Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Intercept	2.636481	.109365	2903.000	24.107	.000	2.422041	2.850921
[V1_50=1]	.372341	.046249	2903	8.051	.000	.281656	.463025
[V1_50=2]	.338443	.064457	2903.000	5.251	.000	.212057	.464828
[V1_50=3]	.128886	.063372	2903.000	2.034	.042	.004627	.253144
[V1_50=4]	.137754	.058164	2903.000	2.368	.018	.023707	.251801
[V1_50=5]	.064439	.034833	2903	1.850	.064	-.003861	.132739
[V1_50=6]	0[b]	0	.	.	.	.	.
[V1_52=1]	-.215824	.215536	2903.000	-1.001	.317	-.638443	.206796
[V1_52=2]	-.115668	.072112	2903.000	-1.604	.109	-.257064	.025728
[V1_52=3]	-.175894	.047268	2903	-3.721	.000	-.268576	-.083211
[V1_52=4]	-.131539	.041116	2903.000	-3.199	.001	-.212159	-.050919
[V1_52=5]	-.047165	.035951	2903.000	-1.312	.190	-.117658	.023328
[V1_52=6]	0[b]	0	.	.	.	.	.
[V1_54=1]	.477624	.065811	2903	7.258	.000	.348583	.606664
[V1_54=2]	.292037	.057140	2903	5.111	.000	.179997	.404077
[V1_54=3]	.228961	.057358	2903	3.992	.000	.116493	.341428
[V1_54=4]	.184444	.061522	2903	2.998	.003	.063813	.305076
[V1_54=5]	.122135	.041706	2903.000	2.928	.003	.040359	.203911
[V1_54=6]	0[b]	0	.	.	.	.	.
[V1_56=1]	-.265547	.233875	2903	-1.135	.256	-.724124	.193030
[V1_56=2]	.139023	.146171	2903	.951	.342	-.147587	.425632
[V1_56=3]	.288381	.130914	2903.000	2.203	.028	.031686	.545075
[V1_56=4]	.108867	.096275	2903	1.131	.258	-.079908	.297642
[V1_56=5]	.095675	.060002	2903.000	1.595	.111	-.021975	.213326
[V1_56=6]	0[b]	0	.	.	.	.	.
race	-.176060	.044788	2903	-3.931	.000	-.263879	-.088241
sex	-.163162	.026645	2903	-6.123	.000	-.215408	-.110916
age	-.003012	.002534	2903.000	-1.188	.235	-.007981	.001958
religion	-.008379	.004186	2903.000	-2.002	.045	-.016588	-.000171

Covariance Parameters

Estimates of Covariance Parameters[a]

Parameter	Estimate	Std. Error
Residual	.485078	.012732

(e) Interpret the coefficient(s) of interest, being sure to answer the question you asked.

The γ is the average difference in Y between those in news types and age. So the news types and age can lead a downward impact on news clicks.