Reproducible Research Template*

EB

Preliminary – please do not quote

January 25, 2024

Abstract

Some abstract here.

*EB, Nowhere University. Email: eb[at]nowhere[dot]edu.

1 Introduction

If you are using this template, please cite this item from the references: von Gaudecker (2023).

The data set for the example project is taken from https://www.stem.org.uk/resources/elibrary/resource/28452/large-datasets-stats4schools. It contains data on smoking habits in the UK, with 1691 observations and 12 variables. We consider only 4 of the 12 features for the prediction of the variable `smoking`: `marital_status`, `highest_qualification`, `gender` and `age`. We model the dependence using a Logistic model. All numerical features are included linearly, while categorical features are expanded into dummy variables. Figures below illustrate the model predictions over the lifetime. You will find one figure and one estimation summary table for each installed programming language.

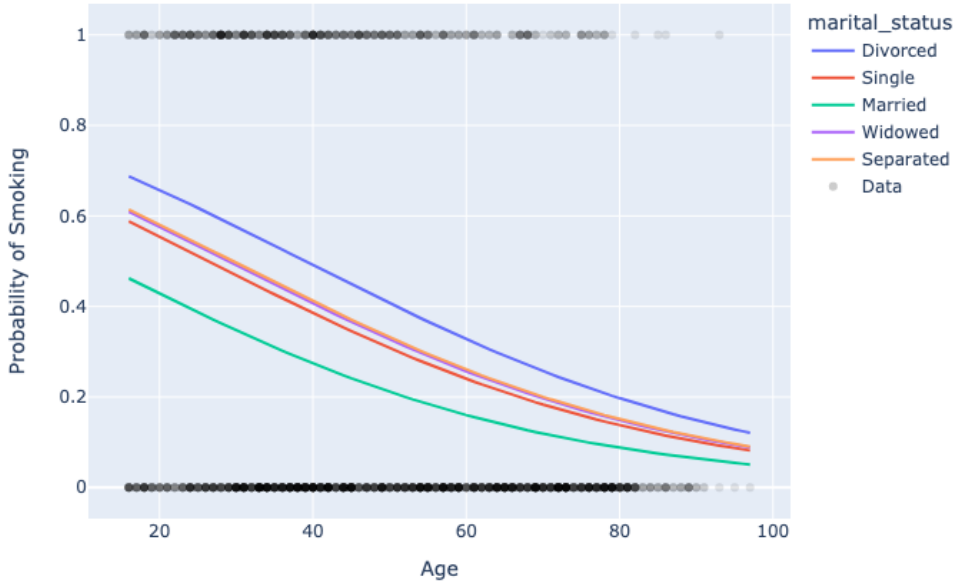

Figure 1: *Python:* Model predictions of the smoking probability over the lifetime. Each colored line represents a case where marital status is fixed to one of the values present in the data set.

References

Gaudecker, Hans-Martin von (2023). "Templates for Reproducible Research Projects in Economics". https://doi.org/10.5281/zenodo.7780520.

Dep. Variable:	smoke_numerical	**No. Observations:**	1691
Model:	Logit	**Df Residuals:**	1677
Method:	MLE	**Df Model:**	13
Date:	Thu, 25 Jan 2024	**Pseudo R-squ.:**	0.08683
Time:	09:25:50	**Log-Likelihood:**	-866.58
converged:	True	**LL-Null:**	-948.98
Covariance Type:	nonrobust	**LLR p-value:**	2.103e-28

	coef	**std err**	**z**	**P> |z|**	**[0.025**	**0.975]**
Intercept	0.3872	0.363	1.068	0.286	-0.324	1.098
qualification[T.Degree]	-0.1773	0.309	-0.574	0.566	-0.783	0.428
qualification[T.GCSE/CSE]	0.8614	0.330	2.609	0.009	0.214	1.508
qualification[T.GCSE/O Level]	0.8233	0.283	2.909	0.004	0.269	1.378
qualification[T.Higher/Sub Degree]	0.4487	0.340	1.318	0.187	-0.218	1.116
qualification[T.No Qualification]	0.9490	0.288	3.297	0.001	0.385	1.513
qualification[T.ONC/BTEC]	0.5820	0.362	1.607	0.108	-0.128	1.292
qualification[T.Other/Sub Degree]	0.8425	0.339	2.485	0.013	0.178	1.507
gender[T.Male]	0.1731	0.122	1.417	0.157	-0.066	0.413
marital_status[T.Married]	-0.9404	0.196	-4.806	0.000	-1.324	-0.557
marital_status[T.Separated]	-0.3234	0.318	-1.019	0.308	-0.946	0.299
marital_status[T.Single]	-0.4332	0.216	-2.008	0.045	-0.856	-0.010
marital_status[T.Widowed]	-0.3441	0.262	-1.312	0.189	-0.858	0.170
age	-0.0342	0.005	-7.179	0.000	-0.044	-0.025

Table 1: *Python:* Estimation results of the linear Logistic regression.