EXAMPLE PROJECT

JANE DOE

Preliminary – please do not quote

April 15, 2025

Abstract

Some abstract here.

_____

*EXAMPLE UNIVERSITY. Email: janedoe[at]example-university.de

1 Introduction

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

The data set for the template 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.
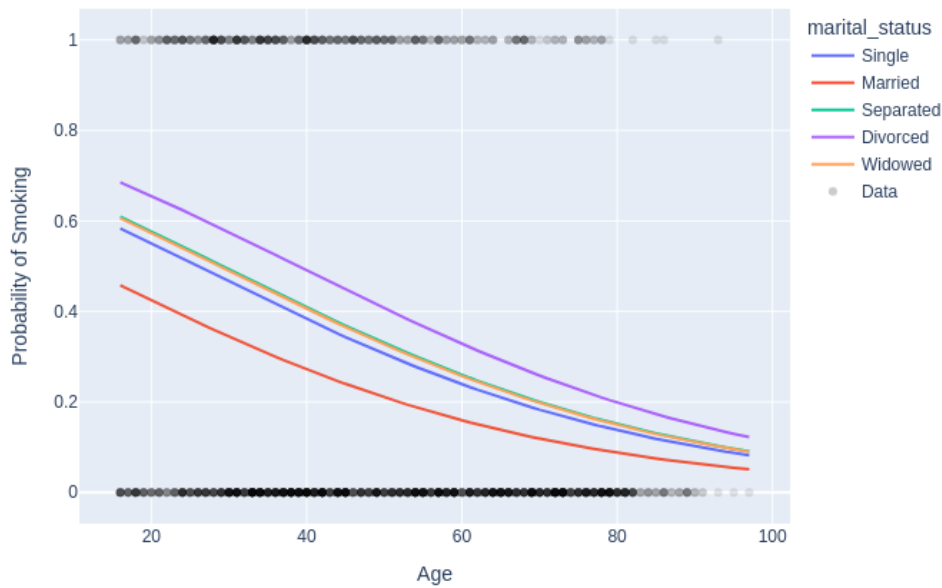

Figure 1: Model predictions of the smoking probability over the lifetime. Each colored line represents a case where marital status is fixed to one of the values present in the data set.

References

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

		coef	std err	z	P> |z|
Dep. Variable:	current_smoker_numerical				
Model:	Logit				
Method:	MLE				
Date:	Tue, 15 Apr 2025				
Time:	15:18:43				
converged:	True				
Covariance Type:	nonrobust				

	No. Observations:	1691
	Df Residuals:	1679
	Df Model:	11
	Pseudo R-squ.:	0.08598
	Log-Likelihood:	-867.39
	LL-Null:	-948.98
	LLR p-value:	2.961e-29

	coef	std err	z	P> |z|
Intercept	0.8786	0.255	3.443	0.001
gender[T.Male]	0.1776	0.122	1.455	0.146
marital_status[T.Married]	-0.5050	0.157	-3.220	0.001
marital_status[T.Separated]	0.1102	0.292	0.378	0.706
marital_status[T.Divorced]	0.4419	0.216	2.050	0.040
marital_status[T.Widowed]	0.0970	0.269	0.360	0.719
highest_qualification[T.GCSE/CSE or GCSE/O Level]	-0.1076	0.168	-0.642	0.521
highest_qualification[T.ONC/BTEC]	-0.3583	0.292	-1.228	0.220
highest_qualification[T.Other/Sub or Higher/Sub Degree]	-0.2999	0.192	-1.560	0.119
highest_qualification[T.A Levels]	-0.9393	0.288	-3.266	0.001
highest_qualification[T.Degree]	-1.1184	0.218	-5.139	0.000
age	-0.0339	0.005	-7.120	0.000

Table 1: Estimation results of the linear Logistic regression.