Chapter 39
Logistic Regression with Categorical Predictors (60 Patients)

1 General Purpose

In the Chap. 8 the effect of categorical predictors on an continuous outcome has been assessed. Linear regression could be used for the purpose. However, the categorical predictor variable had to be restructured prior to the analysis. If your outcome is binary, the analysis of categorical predictors is more easy, because SPSS provides an automatic restructure procedure.

2 Schematic Overview of Type of Data File

Outcome binary	predictor	predictor	predictor categorical
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.
.	.	.	.

3   Primary Scientific Question

Is logistic regression appropriate for assessing categorical predictors with binary outcomes.

4   Data Example

In 60 patients of four races the effect of the race category, age, and gender on the physical strength class was tested. We will use the example of the Chap. 8. The effect of race, gender, and age on physical strength was assessed. Instead of physical strength as continuous outcome, a binary outcome (physical strength $<$ or ≥ 70 points) was applied.

Race	Age	Gender	Strength score
1,00	35,00	1,00	1,00
1,00	55,00	,00	1,00
1,00	70,00	1,00	,00
1,00	55,00	,00	,00
1,00	45,00	1,00	1,00
1,00	47,00	1,00	1,00
1,00	75,00	,00	,00
1,00	83,00	1,00	1,00
1,00	35,00	1,00	1,00
1,00	49,00	1,00	1,00

race 1 = hispanic, 2 = black, 3 = asian, 4 = white
age = years of age
gender 0 = female, 1 = male
strength score 1 = ≥ 70 points, 0 = <70 points

The entire data file is in "chapter39categoricalpredictors", and is in extras. springer.com. We will start by opening the data file in SPSS.

5   Logistic Regression with Categorical Predictors

For analysis the statistical model Binary Logistic Regression in the module Regression is required.

Command:
Analyze....Regression....Binary Logistic Regression....Dependent: strengthbinary.... Covariates: race, gender, age....click Categorical....Categorical Covariates: enter race....Reference Category: mark Last....click Continue....click OK.

Variables in the equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Race			13,140	3	,004	
	Race(1)	2,652	1,285	4,256	1	,039	14,176
	Race(2)	−2,787	1,284	4,715	1	,030	,062
	Race(3)	1,423	1,066	1,782	1	,182	4,149
	Age	−,043	,029	2,199	1	,138	,958
	Gender	1,991	,910	4,791	1	,029	7,323
	Constant	1,104	1,881	,345	1	,557	3,017

[a]Variable(s) entered on step 1: race, age, gender

The above table shows the results of the analysis. As compared to the hispanics (used as reference category),

blacks are significantly more strengthy (at $p = 0,039$)
asians are significantly less strengthy (at $p = 0,030$)
whites are not significantly different from hispanics.

Age is not a significant predictor of the presence of strength.
Gender is a significant predictor of the presence of strength.
The above results are less powerful than those of the continuous outcome data. Obviously with binary outcome procedures some statistical power is lost. Nonetheless they show patterns similar to those with the continuous outcomes.

6 Conclusion

In the Chap. 8 the effect of categorical predictors on an continuous outcome was shown to be applicable for categorical predictors. However, the categorical predictor variable had to be restructured prior to the analysis. If your outcome is binary, the analysis of categorical predictors is more easy, because SPSS provides an automatic restructure procedure. The analysis is presented above.

7 Note

More background, theoretical and mathematical information of categorical predictors is given in the Chap. 21, pp 243–252, in Statistics applied to clinical studies, Springer Heidelberg Germany, 2012, from the same authors.