

Categorical Data Analysis in Python

Categorical Data Analysis in Python

Import and test-specific preparation

Dataset Name

Notes

Output Created		16-JUL-2026 14:40:28
Comments		
Input	Data	D:\DATA ANALYSIS\Chi Square and Categorical Data Tests\Categorical Data Analysis in Python\dataset.csv
	Filter	<none>
	Weight	<none>
	Split File	<none>
Syntax		DATASET NAME CategoricalDataAnalysisinPythonData.
Resources	Processor Time	00:00:00.02
	Elapsed Time	00:00:00.01

Warnings

The active dataset will replace the existing dataset named CategoricalDataAnalysisinPythonData.

school into school_n

Old Value	New Value	Value Label
-----------	-----------	-------------

GP	1	GP
----	---	----

MS	2	MS
----	---	----

reason into reason_n

Old Value	New Value	Value Label
-----------	-----------	-------------

course	1	course
--------	---	--------

home	2	home
------	---	------

other	3	other
-------	---	-------

reputation	4	reputation
------------	---	------------

Categorical Data Analysis in Python

Import and test-specific preparation

```
guardian into guardian_n
```

Old Value	New Value	Value Label
father	1	father
mother	2	mother
other	3	other

Categorical workflow cross-check in SPSS

Categorical workflow cross-check in SPSS
School by four-level final-grade band

Crosstabs

Notes

Output Created		16-JUL-2026 14:40:29
Comments		
Input	Data	D:\DATA ANALYSIS\Chi Square and Categorical Data Tests\Categorical Data Analysis in Python\dataset.csv
	Active Dataset	CategoricalDataAnalysisinPythonData
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	649
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing.
	Cases Used	Statistics for each table are based on all the cases with valid data in the specified range(s) for all variables in each table.
Syntax		CROSSTABS /TABLES=school_n BY grade3 /FORMAT=AVALUE TABLES /STATISTICS=CHISQ PHI /CELLS=COUNT ROW COLUMN EXPECTED RESID SRESID /COUNT ROUND CELL.
Resources	Processor Time	00:00:00.03
	Elapsed Time	00:00:00.02
	Dimensions Requested	2
	Cells Available	524245

[CategoricalDataAnalysisinPythonData]

Categorical workflow cross-check in SPSS
School by four-level final-grade band

Case Processing Summary

	Valid		Cases Missing		Total	
	N	Percent	N	Percent	N	Percent
school_n * G3 ordinal grade band	649	100.0%	0	0.0%	649	100.0%

school_n * G3 ordinal grade band Crosstabulation

			G3 ordinal grade band		
			Low below 10	Medium 10-13	High 14-16
school_n	GP	Count	32	245	112
		Expected Count	65.2	231.4	96.5
		% within school_n	7.6%	57.9%	26.5%
		% within G3 ordinal grade band	32.0%	69.0%	75.7%
		Residual	-33.2	13.6	15.5
		Standardized Residual	-4.1	.9	1.6
	MS	Count	68	110	36
		Expected Count	34.8	123.6	51.5
		% within school_n	30.1%	48.7%	15.9%
		% within G3 ordinal grade band	68.0%	31.0%	24.3%
		Residual	33.2	-13.6	-15.5
		Standardized Residual	5.6	-1.2	-2.2
Total		Count	100	355	148
		Expected Count	100.0	355.0	148.0
		% within school_n	15.4%	54.7%	22.8%
		% within G3 ordinal grade band	100.0%	100.0%	100.0%

Categorical workflow cross-check in SPSS
School by four-level final-grade band

school_n * G3 ordinal grade band Crosstabulation

			G3 ordinal grade ...	
			Very high 17+	Total
school_n	GP	Count	34	423
		Expected Count	30.0	423.0
		% within school_n	8.0%	100.0%
		% within G3 ordinal grade band	73.9%	65.2%
		Residual	4.0	
		Standardized Residual	.7	
	MS	Count	12	226
		Expected Count	16.0	226.0
		% within school_n	5.3%	100.0%
		% within G3 ordinal grade band	26.1%	34.8%
		Residual	-4.0	
		Standardized Residual	-1.0	
Total	Count		46	649
	Expected Count		46.0	649.0
	% within school_n		7.1%	100.0%
	% within G3 ordinal grade band		100.0%	100.0%

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	59.534 ^a	3	.000
Likelihood Ratio	57.075	3	.000
Linear-by-Linear Association	35.263	1	.000
N of Valid Cases	649		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 16.02.

Categorical workflow cross-check in SPSS
School by four-level final-grade band

Symmetric Measures

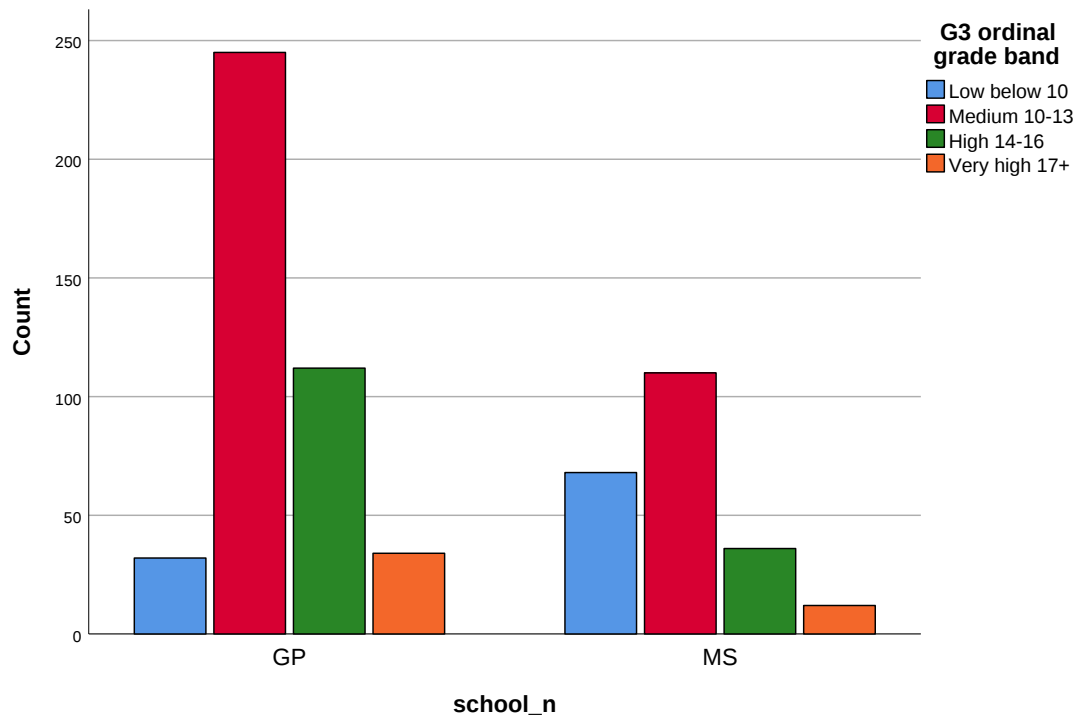
		Value	Approximate Significance
Nominal by Nominal	Phi	.303	.000
	Cramer's V	.303	.000
N of Valid Cases		649	

Graph

Notes

Output Created		16-JUL-2026 14:40:29
Comments		
Input	Data	D:\DATA ANALYSIS\ Chi Square and Categorical Data Tests\Categorical Data Analysis in Python\dataset.csv
	Active Dataset	CategoricalDataAnalysisin PythonData
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	649
Syntax		GRAPH /BAR(GROUPED) =COUNT BY school_n BY grade3.
Resources	Processor Time	00:00:00.45
	Elapsed Time	00:00:00.50

Categorical workflow cross-check in SPSS
School by four-level final-grade band



```
=====
=====
CATEGORICAL DATA ANALYSIS IN PYTHON
=====
=====
Formula:  $X^2 = \sum (O-E)^2/E$ ;  $V = \sqrt{X^2/(N \cdot \min(r-1, c-1))}$ 
Variables: school, G3 grade band
cells_expected_below_5 = 0
cramers_v = 0.30287285267769787
degrees_of_freedom = 3
fisher_freeman_halton_monte_carlo_p = 1.999960000799984e-05
minimum_expected_count = 16.018489984591678
mutual_information_nats = 0.04397165580391293
p_value = 7.392451635150176e-13
statistic = 59.534045213043086
Cross-check status: pass
=====
=====
```