

QUESTION 1: Which gender seems to have a greater number of stroke cases?

Notes

Output Created		03-SEP-2022 09:18:05
Comments		
Input	Data	/Users/olorunleke.akindo te/Downloads/Brainnest _DA/PROJECT 1/PROJECT 1_GROUP D.sav
	Active Dataset	DataSet2
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	29065
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=stroke BY gender /FORMAT=AVALUE TABLES /CELLS=COUNT ROW COLUMN TOTAL /COUNT ROUND CELL.
Resources	Processor Time	00:00:00.02
	Elapsed Time	00:00:00.00
	Dimensions Requested	2
	Cells Available	524245

Case Processing Summary

	Valid		Cases Missing		Total	
	N	Percent	N	Percent	N	Percent
stroke * gender	29065	100.0%	0	0.0%	29065	100.0%

stroke * gender Crosstabulation

		gender			
		Female	Male	Total	
stroke	0	Count	17539	10978	28517
		% within stroke	61.5%	38.5%	100.0%
		% within gender	98.2%	97.9%	98.1%
		% of Total	60.3%	37.8%	98.1%
	1	Count	313	235	548
		% within stroke	57.1%	42.9%	100.0%
		% within gender	1.8%	2.1%	1.9%
		% of Total	1.1%	0.8%	1.9%
Total	Count	17852	11213	29065	
	% within stroke	61.4%	38.6%	100.0%	
	% within gender	100.0%	100.0%	100.0%	
	% of Total	61.4%	38.6%	100.0%	

There are 548 entries of people who have stroke. Based on the analysis above, 42.9% of the people who have stroke are male while the remaining 57.1% are female. Therefore, females seem to have a greater number of stroke cases.

QUESTION 2: Is stroke related to hypertension and average glucose level?

Notes

Output Created		03-SEP-2022 09:19:13
Comments		
Input	Data	/Users/olorunleke.akindo te/Downloads/Brainnest _DA/PROJECT 1/PROJECT 1_GROUP D.sav
	Active Dataset	DataSet2
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	29065
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=stroke BY hypertension /FORMAT=AVALUE TABLES /CELLS=COUNT ROW COLUMN TOTAL /COUNT ROUND CELL.
Resources	Processor Time	00:00:00.02
	Elapsed Time	00:00:00.00
	Dimensions Requested	2
	Cells Available	524245

Case Processing Summary

	Valid		Cases Missing		Total	
	N	Percent	N	Percent	N	Percent
stroke * hypertension	29065	100.0%	0	0.0%	29065	100.0%

stroke * hypertension Crosstabulation

hypertension		Total
0	1	

stroke	0	Count	25435	3082	28517
		% within stroke	89.2%	10.8%	100.0%
		% within hypertension	98.5%	95.1%	98.1%
		% of Total	87.5%	10.6%	98.1%
	1	Count	389	159	548
		% within stroke	71.0%	29.0%	100.0%
		% within hypertension	1.5%	4.9%	1.9%
		% of Total	1.3%	0.5%	1.9%
Total		Count	25824	3241	29065
		% within stroke	88.8%	11.2%	100.0%
		% within hypertension	100.0%	100.0%	100.0%
		% of Total	88.8%	11.2%	100.0%

Based on the analysis above, only 29.0% of the people who have stroke, also have hypertension while the remaining 71.0% do not have hypertension. We see that there are more people who do not have hypertension but have stroke. Therefore, stroke is not related to hypertension

stroke vs average glucose level

Notes

Output Created		03-SEP-2022 09:23:28
Comments		
Input	Data	/Users/olorunleke.akindo te/Downloads/Brainnest _DA/PROJECT 1/PROJECT 1_GROUP D.sav
	Active Dataset	DataSet2
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	29065
Missing Value Handling	Definition of Missing	User-defined missing values for dependent variables are treated as missing.
	Cases Used	Statistics are based on cases with no missing values for any dependent variable or factor used.
Syntax		EXAMINE VARIABLES=avg_gluco

		se_level BY stroke /PLOT BOXPLOT HISTOGRAM /COMPARE GROUPS /STATISTICS DESCRIPTIVES /INTERVAL 95 /MISSING LISTWISE /NOTOTAL.
Resources	Processor Time	00:00:01.74
	Elapsed Time	00:00:02.00

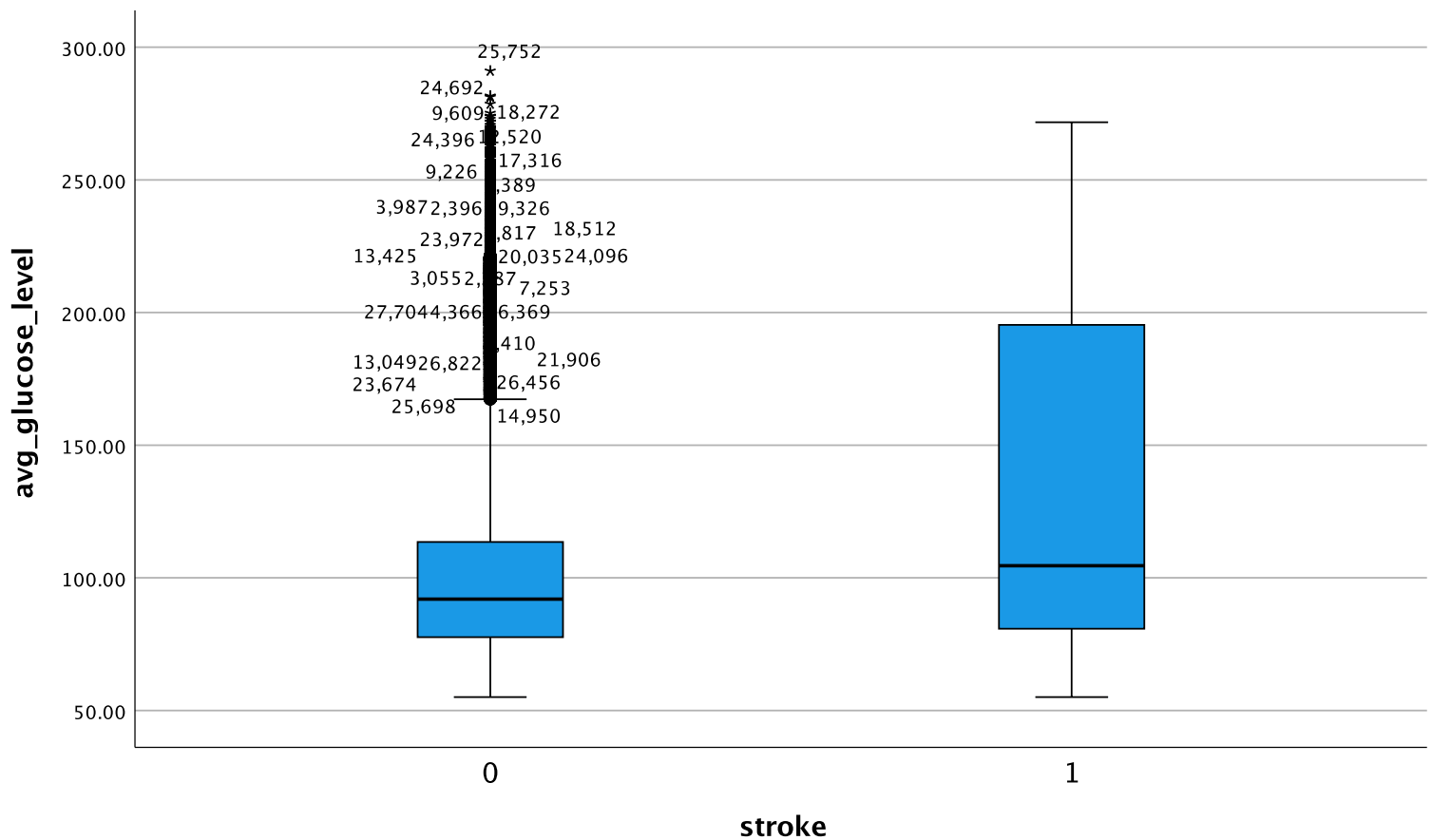
Case Processing Summary

		Valid		Cases Missing		Total	
	stroke	N	Percent	N	Percent	N	Percent
avg_glucose_level	0	28517	100.0%	0	0.0%	28517	100.0%
el	1	548	100.0%	0	0.0%	548	100.0%

Descriptives

		stroke		Statistic	Std. Error
avg_glucose_level	0	Mean		105.9322	.26519
		95% Confidence Interval for Mean		Lower Bound	105.4124
				Upper Bound	106.4519
		5% Trimmed Mean		101.6619	
		Median		91.9800	
		Variance		2005.543	
		Std. Deviation		44.78329	
		Minimum		55.01	
		Maximum		291.05	
		Range		236.04	
		Interquartile Range		35.91	
		Skewness		1.593	.015
		Kurtosis		1.762	.029
	1	Mean		131.0452	2.60892
		95% Confidence Interval for Mean		Lower Bound	125.9204
				Upper Bound	136.1699
		5% Trimmed Mean		128.5128	
		Median		104.5300	
		Variance		3729.950	
		Std. Deviation		61.07332	
		Minimum		55.01	
		Maximum		271.74	
		Range		216.73	
		Interquartile Range		114.65	

Skewness	.616	.104
Kurtosis	-1.159	.208



The box plot shows that most of the people who have stroke (4th quartile), have an average glucose level of between 198-275. The table of descriptives also shows that the mean glucose level of people with stroke is higher than for those without. Therefore stroke is related to average glucose level.

QUESTION 3: How does age affect the occurrence of stroke?

Notes

Output Created	03-SEP-2022 09:38:10	
Comments		
Input	Data	/Users/olorunleke.akindo te/Downloads/Brainnest _DA/PROJECT 1/PROJECT 1_GROUP D.sav
	Active Dataset	DataSet2
	Filter	<none>
	Weight	<none>

	Split File	<none>
	N of Rows in Working Data File	29065
Missing Value Handling	Definition of Missing	User-defined missing values for dependent variables are treated as missing.
	Cases Used	Statistics are based on cases with no missing values for any dependent variable or factor used.
Syntax		EXAMINE VARIABLES=age BY stroke /PLOT BOXPLOT HISTOGRAM /COMPARE GROUPS /STATISTICS DESCRIPTIVES /CINTERVAL 95 /MISSING LISTWISE /NOTOTAL.
Resources	Processor Time	00:00:01.32
	Elapsed Time	00:00:01.00

stroke vs age

Case Processing Summary

		Valid		Cases Missing		Total	
	stroke	N	Percent	N	Percent	N	Percent
age	0	28517	100.0%	0	0.0%	28517	100.0%
	1	548	100.0%	0	0.0%	548	100.0%

Descriptives

				Statistic	Std. Error
age	0	Mean		47.28	.110
		95% Confidence Interval for Mean	Lower Bound	47.06	
			Upper Bound	47.49	
		5% Trimmed Mean		47.26	
		Median		48.00	
		Variance		346.487	
		Std. Deviation		18.614	
		Minimum		10	
		Maximum		82	
		Range		72	

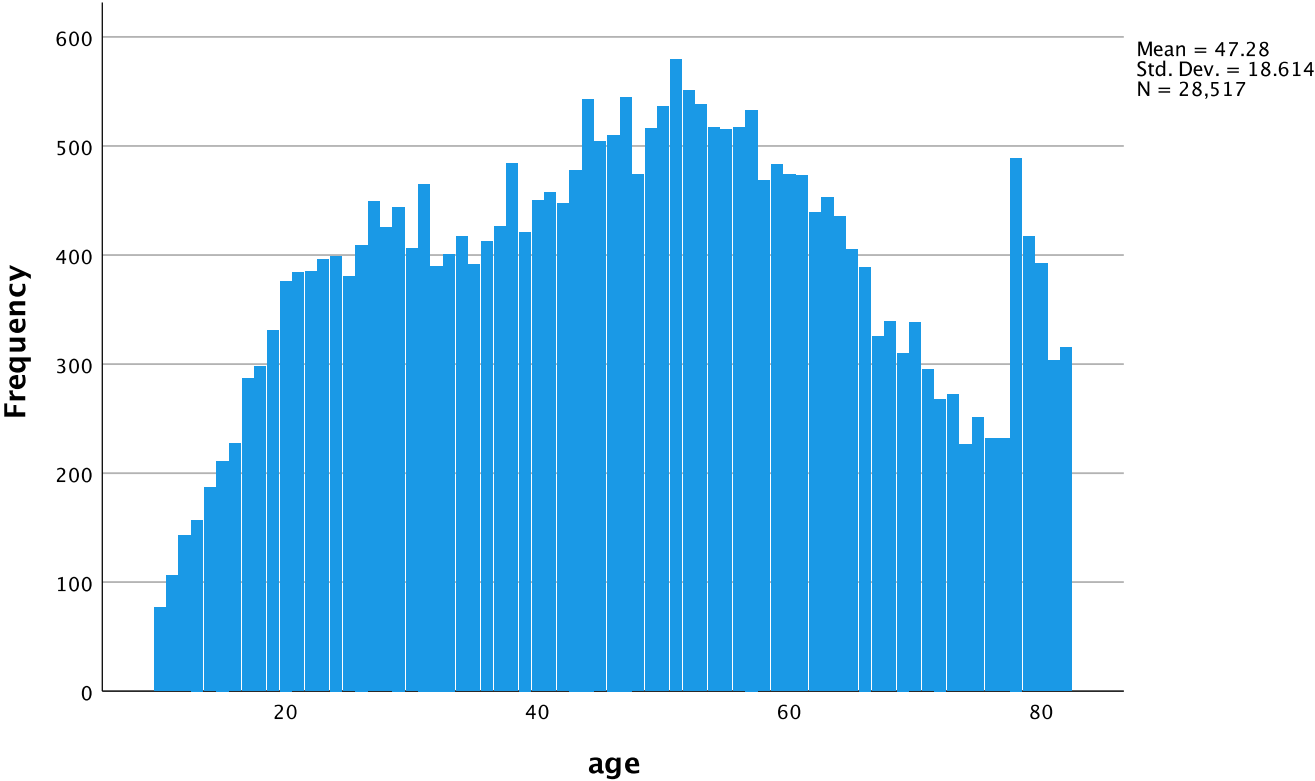
1	Interquartile Range		29	
	Skewness		.013	.015
	Kurtosis		-.955	.029
	Mean		68.49	.506
	95% Confidence Interval for Mean	Lower Bound	67.50	
		Upper Bound	69.49	
	5% Trimmed Mean		69.35	
	Median		72.00	
	Variance		140.265	
	Std. Deviation		11.843	
	Minimum		21	
	Maximum		82	
	Range		61	
	Interquartile Range		19	
	Skewness		-.921	.104
	Kurtosis		.183	.208

From the table of descriptives above, the mean age for stroke patients is higher. This goes to show that stroke occurs more in older people. The fact the minimum age for stroke patients is higher, further reinforces this inference.

age distribution

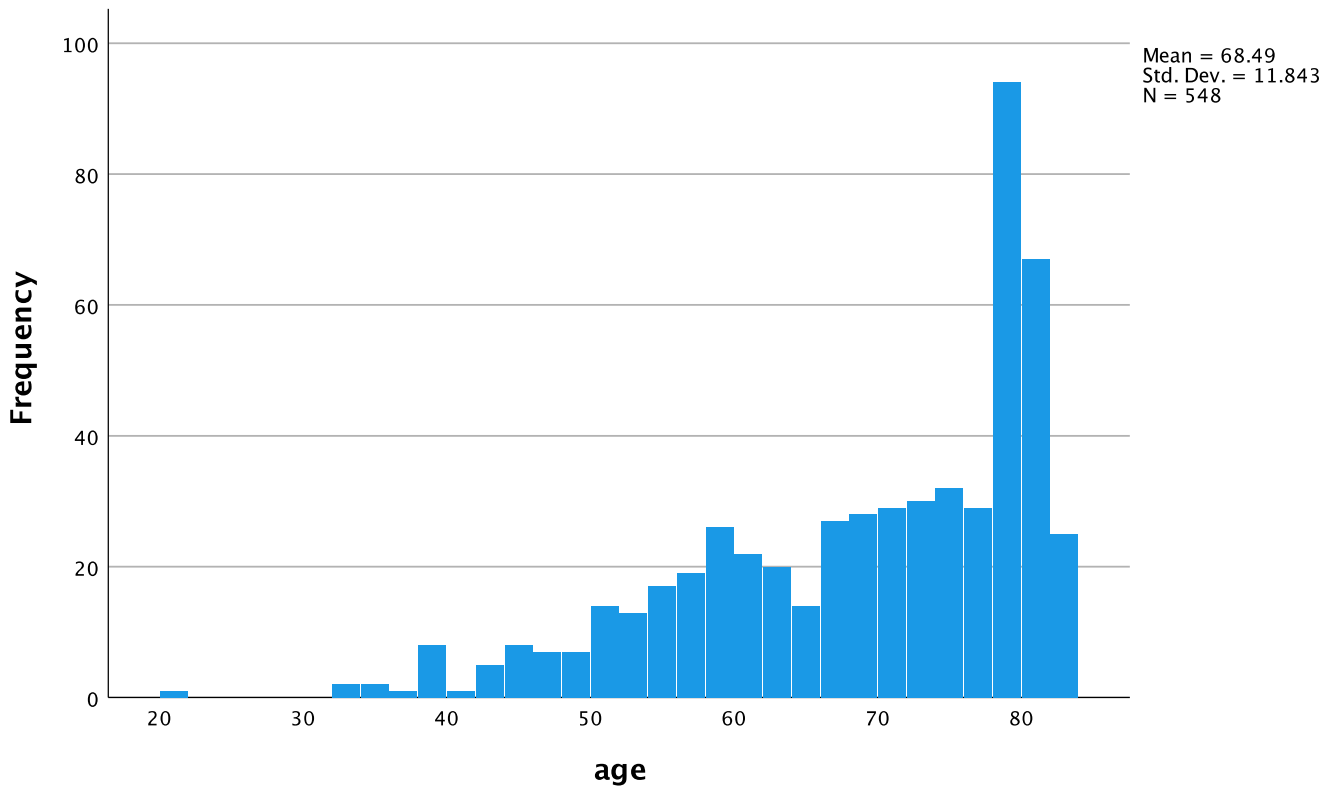
Histograms

Histogram
for stroke= 0

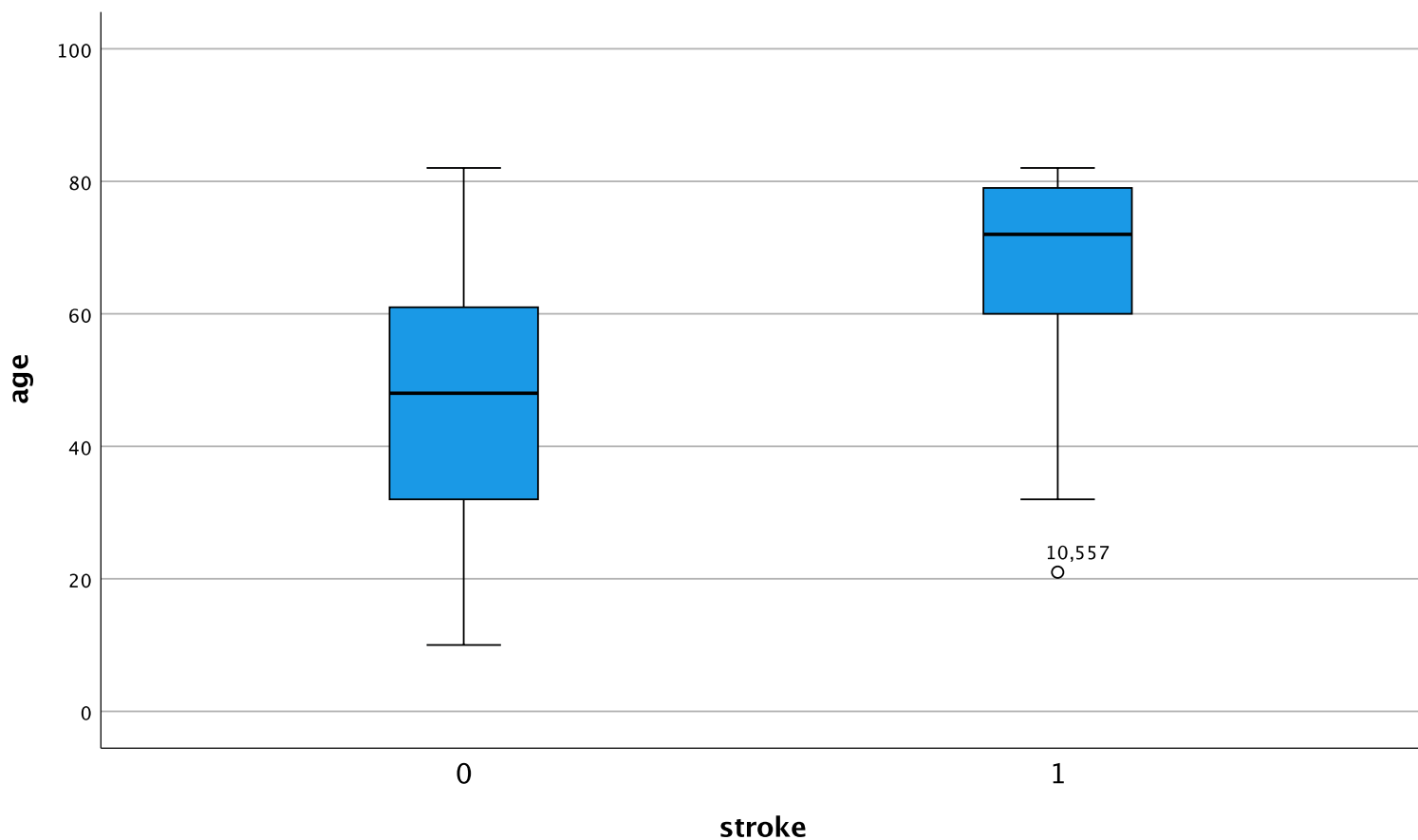


Histogram

for stroke= 1



The histograms above, show the age distribution relative to stroke status. As can be seen for those who have stroke, the highest cases are recorded between 79-81 years of age. Again, older people seem to be more affected.



The box plot gives the age range with the greatest number of stroke cases (upper quartile) to be between 79 and 82 years of age. Therefore, older people are more likely to develop stroke.

QUESTION 4: Are urban residents more prone to stroke than rural ones?

Notes

Output Created		03-SEP-2022 09:48:04
Comments		
Input	Data	/Users/olorunleke.akindo te/Downloads/Brainnest _DA/PROJECT 1/PROJECT 1_GROUP D.sav
	Active Dataset	DataSet2
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	29065
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=stroke BY Residence_type /FORMAT=AVALUE TABLES /CELLS=COUNT ROW COLUMN TOTAL /COUNT ROUND CELL.
Resources	Processor Time	00:00:00.03
	Elapsed Time	00:00:00.00
	Dimensions Requested	2
	Cells Available	524245

Case Processing Summary

	Valid		Cases Missing		Total	
	N	Percent	N	Percent	N	Percent
stroke * Residence_type	29065	100.0%	0	0.0%	29065	100.0%

stroke * Residence_type Crosstabulation

Residence_type	Total
----------------	-------

		Rural	Urban	
stroke	0	Count	14204	14313
		% within stroke	49.8%	50.2%
		% within Residence_type	98.1%	98.1%
		% of Total	48.9%	49.2%
	1	Count	269	279
		% within stroke	49.1%	50.9%
		% within Residence_type	1.9%	1.9%
		% of Total	0.9%	1.0%
	Total	Count	14473	14592
		% within stroke	49.8%	50.2%
		% within Residence_type	100.0%	100.0%
		% of Total	49.8%	50.2%

The analysis below shows that 48.9% of people who have stroke, live in rural areas while the other 49.2% live in urban areas. Although urban areas take the lead with respect to the number of stroke cases, there's not much of difference in count of the people who live in urban areas and those in the rural areas. Therefore, one cannot conclude with this data.

QUESTION 5: Does BMI and smoking status affect stroke rates?

Notes		
Output Created		03-SEP-2022 09:53:20
Comments		
Input	Data	/Users/olorunleke.akindo te/Downloads/Brainnest _DA/PROJECT 1/PROJECT 1_GROUP D.sav
	Active Dataset	DataSet2
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	29065
Missing Value Handling	Definition of Missing	User-defined missing values for dependent variables are treated as missing.
	Cases Used	Statistics are based on cases with no missing values for any dependent variable or factor used.
Syntax		EXAMINE VARIABLES=bmi BY stroke /PLOT BOXPLOT HISTOGRAM /COMPARE GROUPS /STATISTICS DESCRIPTIVES /INTERVAL 95 /MISSING LISTWISE /NOTOTAL.
Resources	Processor Time	00:00:01.27
	Elapsed Time	00:00:02.00

stroke vs bmi

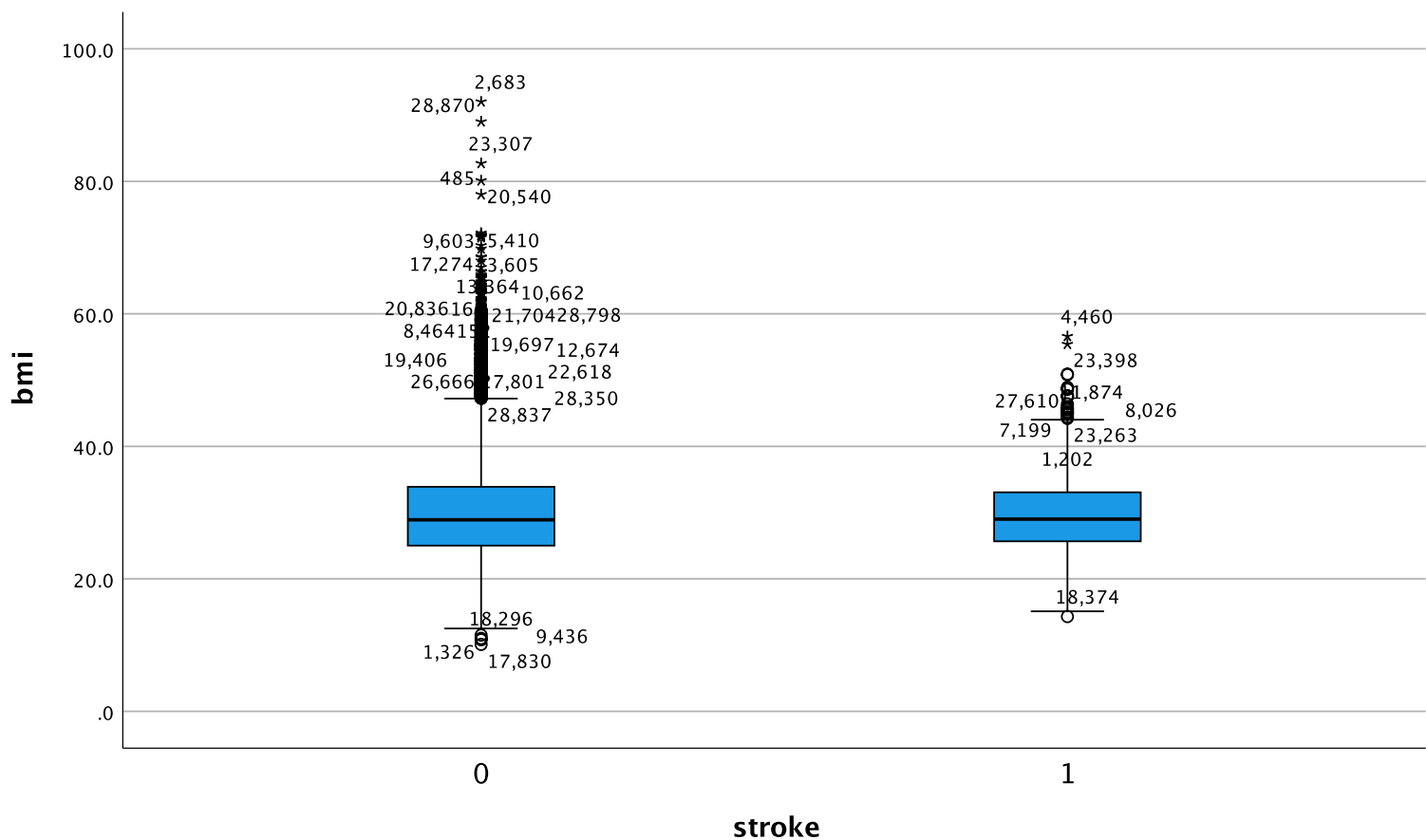
Case Processing Summary

		Valid		Cases Missing		Total	
	stroke	N	Percent	N	Percent	N	Percent
bmi	0	28517	100.0%	0	0.0%	28517	100.0%

1	548	100.0%	0	0.0%	548	100.0%
---	-----	--------	---	------	-----	--------

Descriptives

stroke				Statistic	Std. Error
bmi	0	Mean		30.059	.0427
		95% Confidence Interval for Mean	Lower Bound	29.975	
			Upper Bound	30.142	
		5% Trimmed Mean		29.623	
		Median		28.900	
		Variance		51.964	
		Std. Deviation		7.2086	
		Minimum		10.1	
		Maximum		92.0	
		Range		81.9	
		Interquartile Range		8.9	
		Skewness		1.075	.015
		Kurtosis		2.235	.029
	1	Mean		29.845	.2729
		95% Confidence Interval for Mean	Lower Bound	29.309	
			Upper Bound	30.381	
		5% Trimmed Mean		29.517	
		Median		29.000	
		Variance		40.820	
		Std. Deviation		6.3890	
		Minimum		14.3	
		Maximum		56.6	
		Range		42.3	
		Interquartile Range		7.5	
		Skewness		.838	.104
		Kurtosis		1.230	.208



From the table of descriptives, the mean bmi for people who have stroke is lower than those who don't. The box plot also shows a much shorter bmi range for stroke patients, meaning that bmi doesn't necessarily affect stroke rates.

smoking status vs stroke

Notes		
Output Created	03-SEP-2022 09:58:43	
Comments		
Input	Data	/Users/olorunleke.akindo te/Downloads/Brainnest _DA/PROJECT 1/PROJECT 1_GROUP D.sav
	Active Dataset	DataSet2
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	29065
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=stroke BY smoking_status /FORMAT=AVALUE TABLES /CELLS=COUNT ROW COLUMN TOTAL /COUNT ROUND CELL.
Resources	Processor Time	00:00:00.03
	Elapsed Time	00:00:00.00
	Dimensions Requested	2
	Cells Available	524245

Case Processing Summary

	Valid		Cases Missing		Total	
	N	Percent	N	Percent	N	Percent
stroke * smoking_status	29065	100.0%	0	0.0%	29065	100.0%

stroke * smoking_status Crosstabulation

		smoking_status			Total
		formerly smoked	never smoked	smokes	
stroke 0	Count	6913	15490	6114	28517
	% within stroke	24.2%	54.3%	21.4%	100.0%
	% within smoking_status	97.5%	98.4%	98.2%	98.1%
	% of Total	23.8%	53.3%	21.0%	98.1%
1	Count	180	256	112	548
	% within stroke	32.8%	46.7%	20.4%	100.0%
	% within smoking_status	2.5%	1.6%	1.8%	1.9%
	% of Total	0.6%	0.9%	0.4%	1.9%
Total	Count	7093	15746	6226	29065
	% within stroke	24.4%	54.2%	21.4%	100.0%
	% within smoking_status	100.0%	100.0%	100.0%	100.0%
	% of Total	24.4%	54.2%	21.4%	100.0%

From the cross tabulation above, 46.7% of the people who have stroke, have never smoked and 32.8% used to smoke in the past. Active smokers make up only 20.4%. We could say that smoking status does not influence the occurrence of stroke but the distribution of entries is biased towards

the never smoked category which might explained the percentage distribution of stroke patients based on smoking status.

QUESTION 6: Does heart disease influence the occurrence of stroke? QUESTION 6: Does heart disease influence the occurrence of stroke?

Notes

Output Created		03-SEP-2022 10:07:02
Comments		
Input	Data	/Users/olorunleke.akindo te/Downloads/Brainnest _DA/PROJECT 1/PROJECT 1_GROUP D.sav
	Active Dataset	DataSet2
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	29065
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=stroke BY heart_disease /FORMAT=AVALUE TABLES /CELLS=COUNT ROW COLUMN TOTAL /COUNT ROUND CELL.
Resources	Processor Time	00:00:00.03
	Elapsed Time	00:00:00.00
	Dimensions Requested	2
	Cells Available	524245

Case Processing Summary

	Valid		Cases Missing		Total	
	N	Percent	N	Percent	N	Percent
stroke *	29065	100.0%	0	0.0%	29065	100.0%

heart_disease						
---------------	--	--	--	--	--	--

stroke * heart_disease Crosstabulation

			heart_disease		
			0	1	Total
stroke	0	Count	27122	1395	28517
		% within stroke	95.1%	4.9%	100.0%
		% within heart_disease	98.5%	92.0%	98.1%
		% of Total	93.3%	4.8%	98.1%
	1	Count	427	121	548
		% within stroke	77.9%	22.1%	100.0%
		% within heart_disease	1.5%	8.0%	1.9%
		% of Total	1.5%	0.4%	1.9%
Total	Count	27549	1516	29065	
	% within stroke	94.8%	5.2%	100.0%	
	% within heart_disease	100.0%	100.0%	100.0%	
	% of Total	94.8%	5.2%	100.0%	

The analysis above shows that 77.9% of people who have stroke, do not have heart disease. Therefore, heart disease doesn't influence the occurrence of stroke.

QUESTION 7: Which employment type is more prone to stroke and hypertension?

Notes

Output Created		03-SEP-2022 10:09:01
Comments		
Input	Data	/Users/olorunleke.akindo te/Downloads/Brainnest _DA/PROJECT 1/PROJECT 1_GROUP D.sav
	Active Dataset	DataSet2
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	29065
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=stroke BY work_type /FORMAT=AVALUE TABLES /CELLS=COUNT ROW COLUMN TOTAL /COUNT ROUND CELL.
Resources	Processor Time	00:00:00.03
	Elapsed Time	00:00:00.00
	Dimensions Requested	2
	Cells Available	524245

Case Processing Summary

	Valid		Cases Missing		Total	
	N	Percent	N	Percent	N	Percent
stroke * work_type	29065	100.0%	0	0.0%	29065	100.0%

stroke * work_type Crosstabulation

work_type					Total
children	Govt_jo	Never_work	Private	Self-	

			b	ed		employed	
stroke 0	Count	615	4129	101	18639	5033	28517
	% within stroke	2.2%	14.5%	0.4%	65.4%	17.6%	100.0%
	% within work_type	100.0%	98.4%	100.0%	98.4%	96.7%	98.1%
	% of Total	2.1%	14.2%	0.3%	64.1%	17.3%	98.1%
1	Count	0	66	0	311	171	548
	% within stroke	0.0%	12.0%	0.0%	56.8%	31.2%	100.0%
	% within work_type	0.0%	1.6%	0.0%	1.6%	3.3%	1.9%
	% of Total	0.0%	0.2%	0.0%	1.1%	0.6%	1.9%
Total	Count	615	4195	101	18950	5204	29065
	% within stroke	2.1%	14.4%	0.3%	65.2%	17.9%	100.0%
	% within work_type	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
	% of Total	2.1%	14.4%	0.3%	65.2%	17.9%	100.0%

work type vs hypertension

work_type * hypertension Crosstabulation

Count

		hypertension		Total
		0	1	
work_type	children	614	1	615
	Govt_job	3722	473	4195
	Never_worked	100	1	101
	Private	17047	1903	18950
	Self-employed	4341	863	5204
Total		25824	3241	29065

The analysis above, shows that 56.8% of the people who have stroke, work in the private sector. Self-employed individuals and government workers make up 31.2% and 12% respectively. Zero stroke found in children and never worked category. Also, it is evident from the next tabulation that private and self-employed category workers have high blood pressure. Seeing that 1,903 people who have stroke are private sector workers while self-employed and government workers are 863 and 473 respectively. Therefore, private sector employees are more likely to develop stroke.

QUESTION 8: How does marital status influence the occurrence of stroke?

Notes		
Output Created		03-SEP-2022 10:11:54
Comments		
Input	Data	/Users/olorunleke.akindo te/Downloads/Brainnest _DA/PROJECT 1/PROJECT 1_GROUP D.sav
	Active Dataset	DataSet2
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	29065
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=stroke BY ever_married /FORMAT=AVALUE TABLES /CELLS=COUNT ROW COLUMN TOTAL /COUNT ROUND CELL.
Resources	Processor Time	00:00:00.02
	Elapsed Time	00:00:00.00
	Dimensions Requested	2
	Cells Available	524245

Case Processing Summary

	Valid		Cases Missing		Total	
	N	Percent	N	Percent	N	Percent
stroke * ever_married	29065	100.0%	0	0.0%	29065	100.0%

stroke * ever_married Crosstabulation

ever_married	Total
--------------	-------

		No	Yes	
stroke 0	Count	7321	21196	28517
	% within stroke	25.7%	74.3%	100.0%
	% within ever_married	99.2%	97.7%	98.1%
	% of Total	25.2%	72.9%	98.1%
1	Count	57	491	548
	% within stroke	10.4%	89.6%	100.0%
	% within ever_married	0.8%	2.3%	1.9%
	% of Total	0.2%	1.7%	1.9%
Total	Count	7378	21687	29065
	% within stroke	25.4%	74.6%	100.0%
	% within ever_married	100.0%	100.0%	100.0%
	% of Total	25.4%	74.6%	100.0%

The analysis above, shows that 89.6% of people who have stroke are married. This is however biased as it can be seen from the total distribution that there are way more entries for married that never married.