FORMULAS

Expected Count = (Row Total x Column Total) / Grand Total

Chi-Square = (Observed - Expected) / Expected

Degrees of Freedom = (# of rows - 1) * (# of columns -1)

p - value = CHISQ.DIST.RT (chi_sqr_stat, degrees of freedom) p - value < 0.05 is **significant**

Age vs Stroke Chi-Squared Test

Age Category	no stroke		had stroke		Row Total	Chi Contribution	
	observed	expected	observed	expected		no stroke	had stroke
elderly	146	176.9365949	40	9.063405088	186	5.409129215	105.5974985
middle aged adult	1504	1487.789432	60	76.21056751	1564	0.1766261363	3.448111039
pediatric	854	814.2888454	2	41.7111546	856	1.9366295	37.80705221
senior	1049	1132.013699	141	57.98630137	1190	6.087624353	118.8431405
young adult	1308	1249.971429	6	64.02857143	1314	2.693913657	52.59082042
Column Total	4861		249		5110		

DF	4
chi-square stat	334.5905455
p-val	0.000000000

SIGNIFICANT

Age had the highest total chi-square value, with strong contributions from senior and elderly patients, and a notably lower-than-expected stroke rate among young adults.

While no single age group contributed more than 40% of the total statistic, the combined variation across age groups makes age the most robust predictor of stroke risk in the dataset.

Work Type vs Stroke Chi-Squared Test

Work Type	no stroke		had stroke		Row Total	Chi Contribution	
	observed	expected	observed	expected		no stroke	had stroke
children	685	653.5238748	2	33.47612524	687	1.516006528	29.59561339
government job	624	624.9857143	33	32.01428571	657	0.00155464778	0.03034997131
never_worked	22	20.92798434		1.072015656	22	0.05491295993	1.072015656
private	2776	2782.470646	149	142.5293542	2925	0.01504751076	0.2937588345
self-employed	754	779.0917808	65	39.90821918	819	0.8081171953	15.77613529
Column Total	4861		249		5110		

DF	4
chi-square stat	49.16351198
p-val	0.000000001

SIGNIFICANT

The chi-square test confirms a statistically significant association between work type and stroke occurrence.

While no individual category contributed more than 100 to the test statistic, the largest contribution came from the "children" group, which had far fewer strokes than expected.

This likely reflects age-related protection rather than work type itself, suggesting that work type is not a strong standalone predictor of stroke risk in this dataset.

Glucose Level vs Stroke Chi-Squared Test

Glucose Level	no stroke		had stroke		Row Total	Chi Contribution	
	observed	expected	observed	expected		no stroke	had stroke
diabetic	900	951.2720157	100	48.72798434	1000	2.763478318	53.94886788
hypoglycemic	727	717.2590998	27	36.7409002	754	0.1322885086	2.582547954
normal	2292	2261.173581	85	115.8264188	2377	0.4202543772	8.204243083
prediabetic	942	931.2953033	37	47.70469667	979	0.1230442486	2.402080693
Column Total	4861		249		5110		

DF	3
chi-square stat	70.57680506

SIGNIFICANT

The chi-square test confirms a statistically significant association between glucose level and stroke occurrence.

Diabetic patients experienced stroke at a rate far higher than expected, contributing 76.44% of the total chi-square value.

This makes diabetes a strong individual predictor of stroke risk in this dataset.

p-val	0.000000000

BMI vs Stroke Chi-Squared Test

BMI	no stroke		had stroke		Row Total	Chi Contribution	
	observed	expected	observed	expected		no stroke	had stroke
normal weight	1208	1182.431115	35	60.56888454	1243	0.5529014317	10.7937906
obese	1822	1826.44227	98	93.55772994	1920	0.01080448235	0.2109260591
overweight	1495	1531.547945	115	78.45205479	1610	0.8721583304	17.02635198
underweight	336	320.5786693	1	16.42133072	337	0.7418380076	14.48222713
Column Total	4861		249		5110		

This chi-square test confirms a statistically significant association between BMI and stroke occurrence.

No individual category contributed a dominant share to the result (contribution > 100).

This suggests that BMI may be a contributing factor, but not a strong standalone predictor of stroke risk in this dataset.

DF	3
chi-square stat	44.69099802
p-val	0.000000001

SIGNIFICANT

Smoking vs Stroke Chi-Squared Test

Smoking	no stroke		had stroke		Row Total	Chi Contribution	
	observed	expected	observed	expected		no stroke	had stroke
formerly smoked	815	841.8757339	70	43.12426614	885	0.8579711247	16.7493881
never smoked	1802	1799.806654	90	92.19334638	1892	0.002672936191	0.05218129648
smokes	747	750.5536204	42	38.44637965	789	0.01682520378	0.3284631147
unknown	1497	1468.763992	47	75.23600783	1544	0.5428184121	10.596949
Column Total	4861		249		5110		

This chi-square test confirms a statistically significant association between smoking status and stroke occurrence.

Patients who formerly smoked showed a moderately higher stroke rate than expected, contributing 57.46% (16.75 / 29.15) of the total chi-square value.

This suggests that smoking history may be a contributing factor, but not a strong standalone predictor of stroke risk in this dataset.

DF	3
chi-square stat	29.14726919
p-val	0.000002085

SIGNIFICANT

Gender vs Stroke Chi-Squared Test

Gender	no stroke		had stroke		Row Total	Chi Contribution	
	observed	expected	observed	expected		no stroke	had stroke
female	2853	2848.079859	141	145.9201409	2994	0.008499686788	0.1658975012
male	2007	2011.920141	108	103.0798591	2115	0.01203218073	0.2348449733
Grand Total	4860		249		5109		

This chi-square test shows that gender is not a significant predictor of stroke, as the p-value exceeds the standard significance threshold of 0.05.

The chi-square statistic is 0.42, indicating a very weak association and minimal contribution to stroke occurrence.

DF	3
chi-square stat	0.421274342
p-val	0.935813486

NOT SIGNIFICANT

Hypertension vs Stroke Chi-Squared Test

Hypertension	no stroke		had stroke		Row Total	Chi Contribution	
	observed	expected	observed	expected		no stroke	had stroke
have hypertension	432	473.7334638	66	24.2665362	498	3.676501944	71.77299578
no hypertension	4429	4387.266536	183	224.7334638	4612	0.3969856826	7.74998957
Grand Total	4861		249		5110		

The chi-square test confirms a statistically significant association between hypertension and stroke occurrence.

Patients with hypertension experienced stroke at a notably higher rate than expected, contributing 86% (71.77 / 83.60) of the total chi-square value.

This makes hypertension one of the stronger individual predictors of stroke risk in this dataset, despite its contribution falling below the dominant threshold of 100.

DF	3
chi-square stat	83.59647297
p-val	0.000000000

SIGNIFICANT

Heart Disease vs Stroke Chi-Squared Test

Heart Disease	no stroke		had stroke		Row Total	Chi Contribution	
	observed	expected	observed	expected		no stroke	had stroke
have heart disease	229	262.5510763	47	13.44892368	276	4.287450419	83.6999859
no heart disease	4632	4598.448924	202	235.5510763	4834	0.2447944385	4.778898657
Grand Total	4861		249		5110		

DF	3
chi-square stat	93.01112941
p-val	0.000000000

SIGNIFICANT

The chi-square test confirms a statistically significant association between heart disease and stroke occurrence.

Patients with heart disease experienced stroke at a much higher rate than expected under independence (χ^2 contribution = 83.70 out of 93.01).

This makes heart disease one of the strongest individual predictors of stroke risk in this dataset.

Ever Married vs Stroke Chi-Squared Test

Ever Married	no stroke		had stroke		Row Total	Chi Contribution	
	observed	expected	observed	expected		no stroke	had stroke
married	3133	3189.615068	220	163.3849315	3353	1.004906834	19.61788
never married	1728	1671.384932	29	85.61506849	1757	1.917730572	37.43810566
Grand Total	4861		249		5110		

DF	3
chi-square stat	59.97862307
p-val	0.000000000

SIGNIFICANT

The chi-square test confirms a statistically significant association between marital status and stroke occurrence ($\chi^2 = 59.98$, $p < 0.000001$).

Married patients experienced more strokes than expected, while never married patients experienced far fewer.

The large chi-square contribution from the never married group reflects a protective effect, not an increased risk. This distinction is important—high contribution doesn't always mean high risk.

These results support earlier findings that being married or ever married is associated with a higher risk of stroke, while never married individuals appear significantly less likely to experience stroke.

Residence Type vs Stroke Chi-Squared Test

Residence Type	no stroke		had stroke		Row Total	Chi Contribution	
	observed	expected	observed	expected		no stroke	had stroke
rural	2400	2391.497847	114	122.5021526	2514	0.03022649576	0.5900843209
urban	2461	2469.502153	135	126.4978474	2596	0.02927172972	0.5714452938
Grand Total	4861		249		5110		

DF	3
chi-square stat	1.22102784
p-val	0.747965511

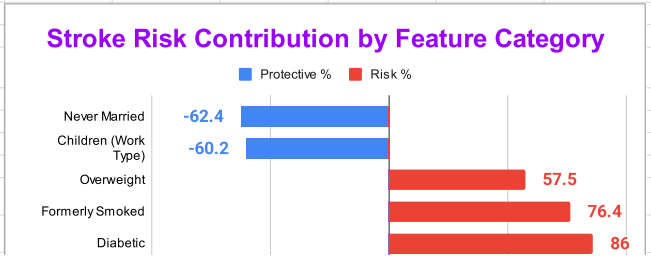
NOT SIGNIFICANT

The chi-square test shows that residence type is not significantly associated with stroke occurrence ($\chi^2 = 1.22$, $p = 0.748$).

Stroke rates were nearly identical between urban and rural patients, suggesting that residency type is not a meaningful predictor of stroke risk in this dataset.

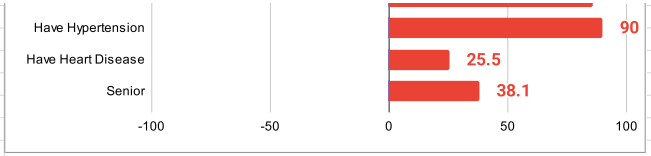
Chi-Squared Test Summary Table

Feature	Chi-Square Stat	p-value	Top Contributing Group	Top Contribution	% of Total χ^2	Interpretation
Age	334.59	< 0.000001	Senior (Stroke)	118.84	35.50%	Strongest Predictor
Heart Disease	93.01	< 0.000001	Heart Disease (Stroke)	83.7	90.00%	Strong Predictor
Hypertension	83.6	< 0.000001	Hypertension (Stroke)	71.77	86.00%	Strong Predictor
Glucose Level	70.58	< 0.000001	Diabetes (Stroke)	53.95	76.40%	Strong Predictor
Ever Married	59.98	< 0.000001	Never Married (Protective)	37.44	62.40%	Protective – significantly fewer strokes than expected
Work Type	49.16	< 0.000001	Children (Protective)	29.6	60.20%	Not a standalone predictor (age confounded)



BMI	44.69	< 0.000001	Overweight (Stroke)	17.03	38.10%	Moderate Contributor
Smoking Status	29.15	< 0.000002	Former Smoker (Stroke)	16.75	57.50%	Moderate Contributor
Gender	0.42	-	-	-	-	Not Significant
Residence Type	1.22	-	-	-	-	Not Significant

Feature	Protective %	Risk %
Never Married	-62.4	
Children (Work Type)	-60.2	
Overweight		57.5
Formerly Smoked		76.4
Diabetic		86
Have Hypertension		90
Have Heart Disease		25.5
Senior		38.1



The chart above highlights the most influential feature categories in relation to stroke occurrence. Features like diabetes, hypertension, and heart disease contribute heavily to the overall chi-square value, confirming their strong association with increased stroke risk.

In contrast, categories like never married and children (work type) contribute in a protective direction, with stroke rates significantly lower than expected.
These insights help prioritize which risk factors deserve the most attention in awareness campaigns and future predictive modeling.