Standardized residuals can be applied to any table with numbers

Column %	18 - 29	30 - 39	40 - 49	50 +
Coca-Cola	55%	53%	28%	38%
Diet Coke	3%	15%	14%	11%
Coke Zero	16%	16%	23%	17%
Pepsi	5%	8%	22%	6%
Diet Pepsi	1%	0%	3%	6%
Pepsi Max	18%	7%	11%	23%

Which do you prefer

Column %	18 - 29	30 - 39	40 - 49	50 +
Coca-Cola	55%	53%	28%	38%
Diet Coke	3%	15%	14%	11%
Coke Zero	16%	16%	23%	17%
Pepsi	5%	8%	22%	6%
Diet Pepsi	1%	0%	3%	6%
Pepsi Max	18%	7%	11%	23%

Column %	18 - 29	30 - 39	40 - 49	50 +
Coca-Cola	55% ↑	53% ↑	28%	38%
Diet Coke	3%	15%	14%	11%
Coke Zero	16%	16%	23%	17%
Pepsi	5%	8%	22% ↑	6%
Diet Pepsi	1%	0%	3%	6%
Pepsi Max	18%	7%	11%	23% ↑

Let's explore residuals manually – to understand what they are

- The coloured cells indicate a standardised residual that is statistically significant ($p<0.05$)
- Look at the 22% for Pepsi and 40 – 49.
- This tells us that people aged 40 -49 are more likely to prefer Pepsi than are people in the other categories.

Column %	18 - 29	30 - 39	40 - 49	50 +	*NET*
Coca-Cola	55%	53%	28%	38%	*44%*
Diet Coke	3%	15%	14%	11%	*11%*
Coke Zero	16%	16%	23%	17%	*18%*
Pepsi	5%	8%	22%	6%	*9%*
Diet Pepsi	1%	0%	3%	6%	*3%*
Pepsi Max	18%	7%	11%	23%	*16%*

sample size n=588

Computing the Expected %

- We start with the raw *Counts*
- Then, we compute the *Total %*
- We start by computing 5 key numbers to understand Pepsi 40 - 49:

Observed % = 4% = 24 / 588

Column Total % = 19%

Row Total % = 9%

N = 588

And then the *Expected %* for that cell.

Counts	18 - 29	30 - 39	40 - 49	50 +	*TOTAL*
Coca-Cola	81	70	31	75	***257***
Diet Coke	5	20	16	21	***62***
Coke Zero	24	21	25	34	***104***
Pepsi	8	11	24	12	*55*
Diet Pepsi	1	0	3	12	***16***
Pepsi Max	27	9	12	46	***94***
TOTAL	***146***	***131***	***111***	***200***	***588***

Total %	18 - 29	30 - 39	40 - 49	50 +	*TOTAL*
Coca-Cola	14%	12%	5%	13%	44%
Diet Coke	1%	3%	3%	4%	11%
Coke Zero	4%	4%	4%	6%	18%
Pepsi	1%	2%	4%	2%	9%
Diet Pepsi	0%	0%	1%	2%	3%
Pepsi Max	5%	2%	2%	8%	17%
TOTAL	25%	23%	19%	35%	100%

What is the Expected %

So on the table on the right, here are the *Expected %* for all the cells on the table.

The Expected % is computed using the row and column %'s.

In this case it's **9.35%** x **18.88%...**
(when all the decimal places are used in the calculation)
….which gives us 1.77%.

Expected %	18 - 29	30 - 39	40 - 49	50 +	*TOTAL*
Coca-Cola	10.85%	9.74%	8.25%	14.87%	43.71%
Diet Coke	2.62%	2.35%	1.99%	3.59%	10.54%
Coke Zero	4.39%	3.94%	3.34%	6.02%	17.69%
Pepsi	2.32%	2.08%	1.77%	3.18%	9.35%
Diet Pepsi	0.68%	0.61%	0.51%	0.93%	2.72%
Pepsi Max	3.97%	3.56%	3.02%	5.44%	15.99%
TOTAL	*24.83%*	*22.28%*	*18.88%*	*34.01%*	100.00%

Total %	18 - 29	30 - 39	40 - 49	50 +	*TOTAL*
Coca-Cola	14%	12%	5%	13%	44%
Diet Coke	1%	3%	3%	4%	11%
Coke Zero	4%	4%	4%	6%	18%
Pepsi	1%	2%	4%	2%	9%
Diet Pepsi	0%	0%	1%	2%	3%
Pepsi Max	5%	2%	2%	8%	17%
TOTAL	25%	23%	19%	35%	100%

Expected %	18 - 29	30 - 39	40 - 49	50 +	*TOTAL*
Coca-Cola	10.85%	9.74%	8.25%	14.87%	43.71%
Diet Coke	2.62%	2.35%	1.99%	3.59%	10.54%
Coke Zero	4.39%	3.94%	3.34%	6.02%	17.69%
Pepsi	2.32%	2.08%	1.77%	3.18%	9.35%
Diet Pepsi	0.68%	0.61%	0.51%	0.93%	2.72%
Pepsi Max	3.97%	3.56%	3.02%	5.44%	15.99%
TOTAL	*24.83%*	*22.28%*	*18.88%*	*34.01%*	100.00%

But if you recall, the total-% for this cell was 4%.
So the actual total-% is higher than the expectation. – 1.77%
So we can say that 40 to 49 year olds are more likely to prefer Pepsi than are other age groups in our sample.
We can quantify these differences by computing the *residuals:*

The residuals are easy to compute.

Residual % = Observed % - Expected %

= 4% - 1.77%

= 2.32%(after correcting for rounding error)

The full table of residuals is shown to the right.

But residuals *on their own are* not enough.

➢ How big is the residual of 2.32%?

Residuals	18 - 29	30 - 39	40 - 49	50 +	*TOTAL*
Coca-Cola	2.92%	2.17%	-2.98%	-2.11%	0.00%
Diet Coke	-1.77%	1.05%	0.73%	-0.02%	0.00%
Coke Zero	-0.31%	-0.37%	0.91%	-0.23%	0.00%
Pepsi	-0.96%	-0.21%	2.32%	-1.14%	0.00%
Diet Pepsi	-0.51%	-0.61%	0.00%	1.12%	0.00%
Pepsi Max	0.62%	-2.03%	-0.98%	2.39%	0.00%
TOTAL	0.00%	0.00%	0.00%	0.00%	0.00%

Better to computer a standardized residual:

$$Standardized\ Residual(Z) = \frac{\text{Observed \%} - \text{Expected \%}}{\sqrt{\text{Expected \%}(1 - \text{Column Total \%})(1 - \text{Row Total \%})/\text{N}}}$$

We can colour code significant standardized residuals

What we can do with these *standardized residuals?*

We can use them for statistical tests.

If the z-statistic is higher than ±1.96, it is considered significant at the 0.05 level.

The significant residuals therefore are highlighted in blue (significantly higher) and red (significantly lower).

z-Statistic	18 - 29	30 - 39	40 - 49	50 +
Coca-Cola	3.3	2.5	-3.7	-2.2
Diet Coke	-3.2	2.0	1.5	.0
Coke Zero	-.5	-.6	1.5	-.3
Pepsi	-1.9	-.4	4.9	-2.0
Diet Pepsi	-1.7	-2.2	.0	3.5
Pepsi Max	1.0	-3.2	-1.7	3.3

Formatting can be applied them to the original tables.

We can then apply that same colour coding back to the original column-% table.

z-Statistic	18 - 29	30 - 39	40 - 49	50 +
Coca-Cola	3.3	2.5	-3.7	-2.2
Diet Coke	-3.2	2.0	1.5	.0
Coke Zero	-.5	-.6	1.5	-.3
Pepsi	-1.9	-.4	4.9	-2.0
Diet Pepsi	-1.7	-2.2	.0	3.5
Pepsi Max	1.0	-3.2	-1.7	3.3

column-%	18 - 29	30 - 39	40 - 49	50 +
Coca-Cola	55%	53%	28%	38%
Diet Coke	3%	15%	14%	11%
Coke Zero	16%	16%	23%	17%
Pepsi	5%	8%	22%	6%
Diet Pepsi	1%	0%	3%	6%
Pepsi Max	18%	7%	11%	23%