

Topic 26: Independence

We have a population and each element of that population has two kinds of characteristics. Perhaps we are looking at packaged food. Each package has a kind of food (**Fruit, Veggy, Bread, Dairy**) and a predominant color (**Red, Blue, Green**). These two characteristics are said to be **Independent** if knowing one characteristic of an item in the population does not change the probabilities of the item having particular values of the other characteristic.

Expressed in math that means $P(A|B) = P(A)$ and $P(B|A) = P(B)$. In our example, if I take a random item from the population, and if I then tell you the predominant color of the package for that item, does knowing the color of the package change the probability that the item is a certain kind of food? If it does not, then the item kind and the item color are independent.

It is rarely the case that a specific population has independent characteristics, but here is one such example.

Observed		Columns		
		Red	Blue	Green
R o w s	Fruit	24	20	12
	Veggy	36	30	18
	Bread	30	25	15
	Dairy	24	20	12

		Columns			Totals
		Red	Blue	Green	
R o w s	Fruit	24	20	12	56
	Veggy	36	30	18	84
	Bread	30	25	15	70
	Dairy	24	20	12	56
Totals		114	95	57	266

It is more often the case that we have something that is possibly close to having independent characteristics. Here is one such example.

Observed		Columns		
		Red	Blue	Green
R o w s	Fruit	26	23	13
	Veggy	27	40	16
	Bread	35	19	10
	Dairy	21	13	14

		Columns			Totals
		Red	Blue	Green	
R o w s	Fruit	26	23	13	62
	Veggy	27	40	16	83
	Bread	35	19	10	64
	Dairy	21	13	14	48
Totals		109	95	53	257

Row %		Columns			
		Red	Blue	Green	Totals
R o w s	Fruit	42.86%	35.71%	21.43%	100.00%
	Veggy	42.86%	35.71%	21.43%	100.00%
	Bread	42.86%	35.71%	21.43%	100.00%
	Dairy	42.86%	35.71%	21.43%	100.00%
	Totals	42.86%	35.71%	21.43%	100.00%

Col %		Columns			
		Red	Blue	Green	Totals
R o w s	Fruit	21.05%	21.05%	21.05%	21.05%
	Veggy	31.58%	31.58%	31.58%	31.58%
	Bread	26.32%	26.32%	26.32%	26.32%
	Dairy	21.05%	21.05%	21.05%	21.05%
	Totals	100.00%	100.00%	100.00%	100.00%

Tot %		Columns			
		Red	Blue	Green	Totals
R o w s	Fruit	9.02%	7.52%	4.51%	21.05%
	Veggy	13.53%	11.28%	6.77%	31.58%
	Bread	11.28%	9.40%	5.64%	26.32%
	Dairy	9.02%	7.52%	4.51%	21.05%
	Totals	42.86%	35.71%	21.43%	100.00%

Expected		Columns			
		Red	Blue	Green	Totals
R o w s	Fruit	24	20	12	56
	Veggy	36	30	18	84
	Bread	30	25	15	70
	Dairy	24	20	12	56
	Totals	114	95	57	266

Row %		Columns			
		Red	Blue	Green	Totals
R o w s	Fruit	41.94%	37.10%	20.97%	100.00%
	Veggy	32.53%	48.19%	19.28%	100.00%
	Bread	54.69%	29.69%	15.63%	100.00%
	Dairy	43.75%	27.08%	29.17%	100.00%
	Totals	42.41%	36.96%	20.62%	100.00%

Col %		Columns			
		Red	Blue	Green	Totals
R o w s	Fruit	23.85%	24.21%	24.53%	24.12%
	Veggy	24.77%	42.11%	30.19%	32.30%
	Bread	32.11%	20.00%	18.87%	24.90%
	Dairy	19.27%	13.68%	26.42%	18.68%
	Totals	100.00%	100.00%	100.00%	100.00%

Tot %		Columns			
		Red	Blue	Green	Totals
R o w s	Fruit	10.12%	8.95%	5.06%	24.12%
	Veggy	10.51%	15.56%	6.23%	32.30%
	Bread	13.62%	7.39%	3.89%	24.90%
	Dairy	8.17%	5.06%	5.45%	18.68%
	Totals	42.41%	36.96%	20.62%	100.00%

Expected		Columns			
		Red	Blue	Green	Totals
R o w s	Fruit	26.29572	22.91829	12.78599	62
	Veggy	35.20233	30.68093	17.11673	83
	Bread	27.14397	23.65759	13.19844	64
	Dairy	20.35798	17.74319	9.898833	48
	Totals	109	95	53	257

Obs. - Exp		Columns		
		Red	Blue	Green
R	Fruit	-0.29572	0.081712	0.214008
o	Veggy	-8.20233	9.319066	-1.11673
w	Bread	7.856031	-4.65759	-3.19844
s	Dairy	0.642023	-4.74319	4.101167

(Obs. - Exp)^2		Columns		
		Red	Blue	Green
R	Fruit	0.08745	0.006677	0.045799
o	Veggy	67.27829	86.84499	1.247089
w	Bread	61.71723	21.69312	10.23004
s	Dairy	0.412194	22.49786	16.81957

(Obs. - Exp)^2/Exp		Columns		
		Red	Blue	Green
R	Fruit	0.003326	0.000291	0.003582
o	Veggy	1.911188	2.830585	0.072858
w	Bread	2.273699	0.916963	0.775095
s	Dairy	0.020247	1.267971	1.699147

sum of (Obs-Exp)^2/Exp	11.77495
------------------------	----------

Expanded Sample Data with Percents

		C o l u m n s			
		1	2	3	Row Total
R o w s	1	obs: 26 r%: 41.935 c%: 23.853 t%: 10.117	obs: 23 r%: 37.097 c%: 24.211 t%: 8.949	obs: 13 r%: 20.968 c%: 24.528 t%: 5.058	62 c%: 24.125
	2	obs: 27 r%: 32.530 c%: 24.771 t%: 10.506	obs: 40 r%: 48.193 c%: 42.105 t%: 15.564	obs: 16 r%: 19.277 c%: 30.189 t%: 6.226	83 c%: 32.296
	3	obs: 35 r%: 54.688 c%: 32.110 t%: 13.619	obs: 19 r%: 29.688 c%: 20.000 t%: 7.393	obs: 10 r%: 15.625 c%: 18.868 t%: 3.891	64 c%: 24.903
	4	obs: 21 r%: 43.750 c%: 19.266 t%: 8.171	obs: 13 r%: 27.083 c%: 13.684 t%: 5.058	obs: 14 r%: 29.167 c%: 26.415 t%: 5.447	48 c%: 18.677
	Column Total	109 r%: 42.412	95 r%: 36.965	53 r%: 20.623	257

Expanded Sample Data χ^2

		C o l u m n s			
		1	2	3	Row Total
R o w s	1	obs: 26 exp: 26.296 dif: -0.296 sqr: 0.087 quo: 0.00333	obs: 23 exp: 22.918 dif: 0.082 sqr: 0.007 quo: 0.00029	obs: 13 exp: 12.786 dif: 0.214 sqr: 0.046 quo: 0.00358	62
	2	obs: 27 exp: 35.202 dif: -8.202 sqr: 67.278 quo: 1.91119	obs: 40 exp: 30.681 dif: 9.319 sqr: 86.845 quo: 2.83059	obs: 16 exp: 17.117 dif: -1.117 sqr: 1.247 quo: 0.07286	83
	3	obs: 35 exp: 27.144 dif: 7.856 sqr: 61.717 quo: 2.27370	obs: 19 exp: 23.658 dif: -4.658 sqr: 21.693 quo: 0.91696	obs: 10 exp: 13.198 dif: -3.198 sqr: 10.230 quo: 0.77509	64
	4	obs: 21 exp: 20.358 dif: 0.642 sqr: 0.412 quo: 0.02025	obs: 13 exp: 17.743 dif: -4.743 sqr: 22.498 quo: 1.26797	obs: 14 exp: 9.899 dif: 4.101 sqr: 16.820 quo: 1.69915	48
	Column Total	109	95	53	257

χ^2 sum: 11.775

For 6 degrees of freedom, attained significance = 0.0672