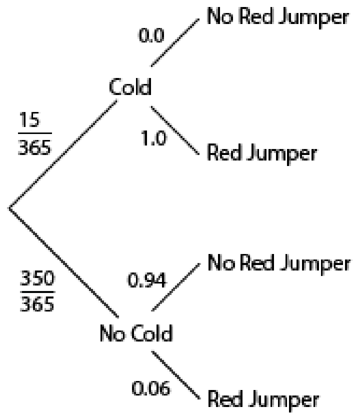
Exercise Set 13 - Solution

1 Refresher on probabilities [basic-normal]

a)



b) P(Jumper) = P(Cold) × P(Jumper|Cold) + P(No Cold) × P(Jumper|No Cold) = $15/365 \times 1.0 + 0.06 \times (1 - 15/365) = 9.9\%$

c) 0 (since Mr Schmitt always wears a red jumper when he has a cold).

d)

$$\text{P(Cold | Red Jumper)} = \frac{15/365 \times 1.0}{15/365 + 0.06 \times (1 - 15/365)} = 42\%$$

e) They are not independent. You can prove this either via $P(Jumper) \times P(Cold) \neq P(Jumper \cap Cold)$ or via $P(Jumper|Cold) \neq P(Jumper)$.

2 Finishing the two-factor ANOVA from Exercise Set 12

We look at the factor interaction. We have shown that both skiing and fondue increase happiness. But do these factors interact, so are people significantly happier if the have both on the same day, or do they just feel the same increase in happiness that fondue and skiing do separately?

We start from $SS_T = \sum_{i,j,k}(X_{i,j,k} - \bar{X}_{\bullet,\bullet,\bullet})^2 = 181.56$. We know that the difference between the total sum of squares and the model with non-interacting factors is just given by the factor interaction. We find: $SS_{Bskiing:fondue} = SS_T - SS_{B,skiing} - SS_{B,fondue} - SS_E \approx 1.0$. Clearly, this is a very small contribution. Constructing the updated ANOVA table:

Source	ν	SS	MS	F
Skiing	1	28.44	28.44	7.16
Fondue	1	25	25	6.29
Fondue:Skiing	1	1	1	0.25
Error	32	127.11	3.97	
Total	35	181.56		

Given this low F-value, there is no statistically significant evidence for factor interaction. It seems people gain just the same happiness from fondue on a ski hill than at any other place. We have to tell the company that marketing on the ski hills will not be much more effective than any place else.

3 The tuna-nutella sandwich

As before, we construct an ANOVA table including factor interaction. First we compute the SSB. Denoting nutella as factor 1 and tuna as factor 2, we write $\bar{X}_{i,j}$ for the *mean* taste scores (we could also write $\bar{X}_{i,j,\bullet}$ to emphasize the fact that each of these is already the mean of a certain dataset.

The partial means are:

$$\bar{X}_{0,\bullet} = 5.39, \bar{X}_{1,\bullet} = 5.285, \bar{X}_{\bullet,0} = 5.285, \bar{X}_{\bullet,1} = 5.39$$

We see that all numbers are very similar, and close to the global mean. It is clear that there will be little effect of the main factors.

$$SSB_{nutella} = JK \sum_i (\bar{X}_{i,\bullet} - \bar{X}_{\bullet,\bullet})^2 = 0.099$$

$$SSB_{tuna} = IK \sum_j (\bar{X}_{\bullet,j} - \bar{X}_{\bullet,\bullet})^2 = 0.099$$

We are given the sum of squares for the errors within each group. The total SSE is their sum, $SSE = 14.67$. This leads to:

$$SSB_{tuna:nutella} = SST - SSB_{tuna} - SSB_{nutella} - SSE = 7.13$$

.

This allows us to construct the ANOVA table:

Source	df	SS	MS	F
Tuna	1	0.099	0.099	0.22
Nutella	1	0.099	0.099	0.22
Tuna:Nutella	1	7.13	7.13	15.55
Error	32	14.67	0.46	
Total	35	22		

We see that just adding Nutella or Tuna to a sandwich has little effect, but adding both has a huge one. As in the exercise above, we compare this to the critical F-value of $F^{crit}_{1,32} = 4.15$. We find that there is no effect of the main factors, but that there is statistically significant effects of the interaction.

The total variability is dominated by this interaction. Given that the mean for having tuna and nutella on the same sandwich is much smaller than the mean, we conclude that having both together tastes worse. Not surprisingly!

4 Python

a) The loop will return the difference between a point and the mean for each elements in the array :

0.0

1.0

-1.0

b) This will return 0.

c) Comparator operations are ill-defined for arrays (should they be applied to each element or to the whole array somehow?). To solve use np.vectorize(myFilter)(myArray) or make a loop (either inside or outside the function) over each element.