Orange puzzle: Can't have it all

Question: Many companies compete to hire employees using two tests, A and B. The company DataPi manages to attract people with quite high total scores in A and B combined, but they notice something surprising. The performances of their recruits in test A and B seem to have an overall negative relationship: The better they perform in A the worse they perform in B, and vice versa. How can this be explained?

Hints:

- Think about the whole pool of applicants applying for jobs.
- Think about other companies trying to hire the best applicants.
- Consider that some companies are more attractive than others.
- Sketch a plot of (hypothetical) test B scores versus test A scores and highlight the type of area from where *DatPi* recruits.
- Highlight the area in the plot from where the most attractive company would recruit.

Solution: To answer this question we need to find plausible scenarios that leads to such a situation. Represent test scores A and B as values between 0 and 100 corresponding to achieved percentages. *DataPi* offers jobs to applicants with $A + B \geq c_1$. There is, however, the leading company in the same industry also uses these tests and offers jobs to applicants with $A + B \geq c_2$, where $c_2 > c_1$. Given the choice, applicants would prefer the latter. Hence, applicants recruited by *DataPi* have scores with $c_1 \leq A + B < c_2$, while those with recruited by the leading company have scores with $A + B \geq c_2$.

Since *DataPi* refers to their threshold as being "quite high", c_1 could be a value between say 100 and 160, say $c_1 = 120$. The top competitor would go up to say $c_2 = 150$.

A simple scenario assumes that the test scores are independent of each other and equally distributed between 0 and 100. The left figure below shows the relationship between test A and test B performances with the green area indicating which applicants are above the threshold to received a job offer from *DataPi*.

The right figure is based on the same data, but shows which job applicants are successfully recruited by *DataPi* (blue area) and the leading company (red area).

Within the blue area, the performances in test A and test B are in a negative relationship!
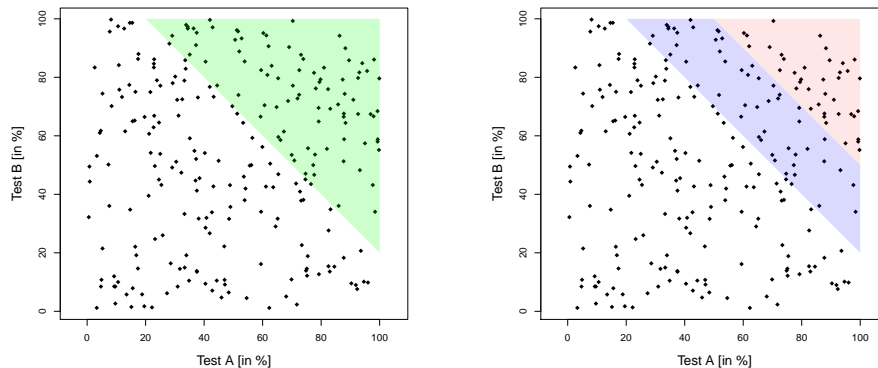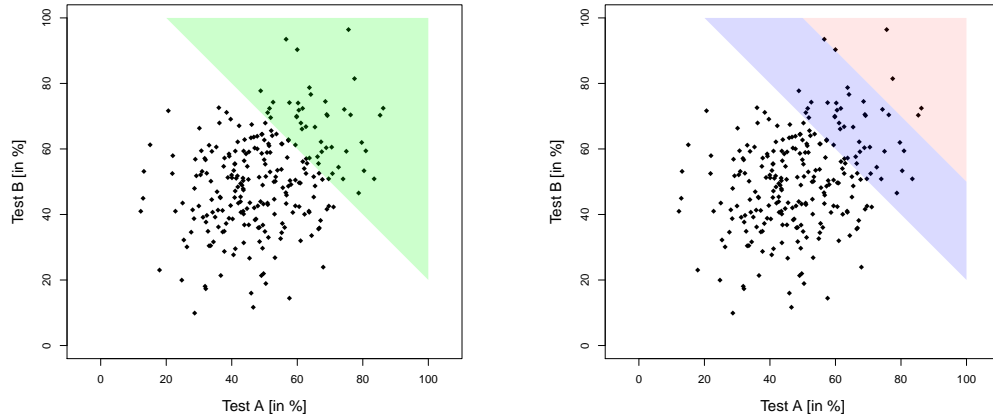

Technical note: The figures are based on data simulated from two independent uniform distributions on $[0, 100]$.

More realistically, however, the test scores have a less simplistic distribution and there is some correlation between them. Often, test scores are normally distributed. The second row of figures shows such a scenario.

Again, within the blue area, the performances in test A and test B are in a negative relationship!

Technical note: The figures are based on data simulated from two correlated normal distributions with mean 50, standard deviation 15 and correlation 0.3.