Prof. Andreas Orth
High Integrity Systems, 2nd semester: Data Mining
Assignment 2 week 2

Exercise 1:
(a) How many possible ways are there of choosing 3 days of one week?
(b) **If** the probability of a medicament being effective for one person X is p=0,9, **and if** it can be assumed that the same p holds for all other persons, **and if** it can be assumed, that whether it works on one person is independent of whether it works for another person (.... how would you formulate this in your capacity as a professional data analyst?), **then**: what's the probability of the medicament being ineffective for more than 2 people, i. e. for two or more people out of a (small) population of seven? Write the formula down, then use R and the binom( )- distribution to calculate the above.

Exercise 2:
Assume you tested the medicament on 8 people, 4 of whom received it, 4 others received a placebo (something that looks like a medicament, but contains no active ingredient). Assume 3 of 4 med-receivers improved their illness after 1 week. But also 3 of 4 plac-receivers improved their illness after one week, ... hmm: obviously this medicament you don't really need!

result / treatment	ok	ill	Σ
med	x		4
plac			4
Σ	6	2	8

(a) Let x be the number of med-receivers that get healthy given the "marginal sums", 4, 4, 6, 2 (as in the table). Assuming independence, calculate the probability that x = 3 (careful: this has nothing to do with Ex. 1.) Fill out the rest of the table. Do you think independence is true? Do you think independence is wanted?
(b) Fill out the table for x = 2, 3, 4 (more is not possible) and calculate the corresponding probabilities. Do they sum up to 1?

Exercise 3:
Use R and Fisher's exact test to check out the example above. Get help with `help(fisher.test)`. Use the function `BDtest( )` in the `bdpv`-package to calculate confidence intervals for *sensitivty* and *specificity*. Enlargen the numbers to 20 in each group and try different effects. Try to really understand what's going on!

Exercise 4:
For the enlargened table use the following formula and χ^2-distribution for 1 degree of freedom to establish how a table has to look to be significant.

$$\chi^2_{corr} = \frac{n \cdot \left(\left| n_{1,1} \cdot n_{2,2} - n_{2,1} \cdot n_{1,2} \right| - n/2 \right)^2}{\left(n_{.,1} \cdot n_{.,2} \cdot n_{1,.} \cdot n_{2,.} \right)}$$