Statistical Methods Exercises 6 Autumn 2020

Return your solutions by 12.00 Finnish time on Thursday 29.10.2020 to Moodle course page: `https://moodle.helsinki.fi/course/view.php?id=30207`

1. **Pearson's χ^2 test.** Pearson's χ^2 test can be used to test whether two distributions are compatible with each other. A group of particle physicists are looking for new particles that are not part of the Standard Model, our best current model of elementary particles and their interactions. Before anybody can pick up his/her Nobel price for a discovery, the simulation needs to be validated to describe the real data.

   The file `real_mass.dat` contain mass measurements of real data events: first two columns are the bin boundaries (lower and upper), and third column the number of entries in the corresponding bin $n_i, i = 1, ..., 20$. The files `MC1_mass.dat` and `MC2_mass.dat` contains simulated events of two types of backgrounds to the new particle production process of interest. The files are found on Moodle along with the exercise paper.

   (i) Test first if data can be described only by the first background (`MC1_mass.dat`). Test then if data can be described only by the second background (`MC2_mass.dat`). Calculate the χ^2-value and the corresponding P-value for both backgrounds. Values of cumulative chi-square distribution (= P-value) can be found from computer programs (e.g. google "chi square applet"). What are the conclusions?

   (ii) A few bins have a small $E[n_i]$ value and hence one doesn't expect that the test statistic above follows exactly the theoretical chi-square distribution. Write a program to determine the true distribution assuming the two hypotheses `MC1_mass.dat` and `MC2_mass.dat` (hint: use a Poisson generator to produce a "pseudodata" distribution according to the $E[n_i]$'s of each bin of each prediction, calculate the χ^2-value for the generated "pseudodata" distribution w.r.t. to the hypothesis and repeat until you get a χ^2-distribution for your "pseudoexperiments"). Determine the P-value for both backgrounds based on the generated distribution of χ^2-values by counting the fraction of "pseudoexperiments" giving a χ^2-value equal or larger than the one of the direct comparison between MC1(2)_mass and data. Any change in the conclusion about the validity of the two hypotheses?

   (iii) Find constant a, so that the combination "$a \cdot$MC1 + $(1-a)\cdot$MC2" best models the data. Give optimal value of a and the corresponding P-value both using the cumulative chi-square distribution and pseudoexperiments. Can the data be fully described by this optimal background combination? *Exercise gives max 12 points instead of usual 6.*

2. **AnOVa**. Some teachers tested whether the usage of more modern

methods of teaching physics compared to traditional lecturing (TL) method gave any increased ability to solve physics problems among high school students. To measure their ability, the students passed a test before the course (“Pre-test”) and after the course (“Post-test”). The teaching methods tried were self-directed problem solving (SDPS) and teacher-directed problem solving (TDPS) techniques. 90 students were exposed to each technique and the test results are shown below.

Technique	N	Pre-test average	Pre-test variance	Post-test average	Post-test variance
TDPS	90	3.02	1.54	4.85	0.59
SDPS	90	3.12	1.39	3.72	0.97
TL	90	3.01	1.42	3.22	1.21

(i) Determine whether the physics problem solving ability of the three student sets was similar before being exposed to any teaching. Use Analysis of Variance (AnOVa) techniques to calculate the P-value from the results of the “Pre-test”. Compute first the variation between the three sets by determining the “sum of squares between sets” (SS_b) i.e. $\sum_{j=1}^{r} n_j(m_j - m)^2$, where r is the number of sets, n_j the number of students in set j, m_j the average result in set j and m the overall average result. Compute then the “sum of squares within the sets” (SS_w) i.e. $\sum_{j=1}^{r} \left[\sum_{k=1}^{n_j}(x_{k,j} - m_j)^2\right] = \sum_{j=1}^{r}(n_j - 1) \cdot V_j[x]$, where $V_j[x]$ is the variance of set j. Then finally compute $F = [SS_b/(r-1)] \,/\, [SS_w/(N-r)]$, where $N = \sum_{j=1}^{r} n_j$. Compute from F the P-value taking into account the degrees of freedom in the sums. Google “F-distribution applet” to find P-value calculators.

(ii) What about the problem solving abilities of the 3 student sets after being exposed to teaching? Hint: use “post-test” results. Any change?

(iii) If the conclusion of (i) is that they are similar and of (ii) that they are different then one should make a AnOVa post hoc Scheffe test between each pair of methods using the ”Post-test” results. Estimate first the expected maximal statistical difference at a P-value of 0.05 and compare whether the observed differences are larger. The expected maximal statistical difference according to the Scheffe test is $\sqrt{(r-1) \cdot F_{P=0.05} \cdot \overline{V}[x] \cdot (1/n_i + 1/n_j)}$, where $F_{P=0.05}$ is the F-value corresponding to a P-value $= 0.05$ in AnOVa, $\overline{V}[x]$ is the mean variance within sets and $n_{i(j)}$ the number of students in set $i(j)$. Is the problem solving ability of any of the three student sets significantly different with respect to the others? And if so in which way?