Examination for UE3.7-8 « Brain-computer interfaces: from modeling to engineering » of master BME-BIN: Statistics Applied to Biology

Isabelle Rivals, Équipe de Statistique Appliquée, ESPCI Paris Tech

Duration: 50 minutes. Pocket calculator and lecture notes are authorized.

1 Thyroid cancer

We consider a large industrial country where the annual thyroid cancer probability for children under 15 years of age equals $p_0 = 2\ 10^{-5}$.

a) Compute the expected number of cancers for a population of $n = 200000$ children under 15 years of age.

b) Establish the probability density function of the number of cancers in such population.

c) We now consider a population of $n = 200000$ children under 15 years of age in a region of the country that was polluted 28 years ago by a major nuclear accident, and where 10 new cancer cases where observed in 2014. Assuming that the annual probability of thyroid cancer in this region is the same as the probability in the whole country, compute the probability of observing 10 thyroid cancer cases or more in one year. *Hint: justify and use an approximation of the probability density function established in b).*

d) We wonder whether the number of thyroid cancers observed in the polluted region is abnormally large. Propose an adequate hypothesis test. Explicit the null and alternative hypotheses, and use a first type error risk of $\alpha = 5\%$. What is the conclusion of the test? Draw the probabilities of interest on a graph.

2 Cardiovascular drug

In order to evaluate the efficiency of a cardiovascular drug supposed to slow down the cardiac rhythm, the heart beats of 21 patients with the same cardiac pathology were measured at rest, 12 of them being treated with the drug (sample 1), and the remaining 9 with a placebo (sample 2).

sample	sample size	heart beat values (beats per minute)
1 (drug)	12	68 70 75 63 64 63 70 63 66 67 63 67
2 (placebo)	9	67 81 78 67 76 69 70 68 82

A biostatistician uses Matlab and obtains the following results:

```
[hv, pv, civ, statsv] = vartest2(x1,x2)
hv =
     0
pv =
    0.1337
civ =
    0.0882    1.3719
statsv =
    fstat: 0.3744
      df1: 11
      df2: 8
```

```
[he, pe, cie, statse] = ttest2(x1,x2)
he =
     1
pe =
    0.0070
cie =
  -11.0483   -2.0073
statse =
    tstat: -3.0224
       df: 19
       sd: 4.8979
```

a) Give the literal expressions of the results of « vartest2 »: fstat, df1, df2.

b) What is tested by « vartest2 » (explicit the null and alternative hypotheses) and why? What are the unquestioned hypotheses? What is the conclusion of this test if conducted with a first type error risk $\alpha = 5\%$? With what probability of error?

c) Give the literal expressions of the results of « ttest2 »: tstat, df, sd.

d) Is the effect of the drug the expected one?

e) What is tested by « ttest2 » (explicit the null and alternative hypotheses)? What are the unquestioned hypotheses? What is the conclusion of this test if conducted with a first type error risk $\alpha = 5\%$? With what probability of error?

f) If meaningful (explain why), provide a 95%confidence for the mean heart beat difference between the two groups.