Biostats VHM 801 Course
Fall 2019, Atlantic Veterinary College, PEI
Henrik Stryhn

Solution to home assignment II

The assignment is based on a competitive ELISA test for Bluetongue antibody detection. Bluetongue is a viral disease that mainly affects ruminants. Details about the test and its application to a cattle population in Corsica (France) are given in Biteau-Coroller et al. (2006), *Vet. Microbiol.* **118**, 57–66. The solution is more detailed than required for a 100% mark, by including answers to all six questions, and by offering multiple answers and detailed to several of the questions.

1. Test sensitivity and specificity

Let X be the test result for a randomly selected non-diseased animal; we assume that $X \sim \mathrm{N}(9, 17)$. The specificity is the probability to test negative, that is for the test result to not exceed the cut-off at 50:

$$\mathrm{Sp} = \mathrm{P}(X < 50) = \mathrm{P}\left(\frac{X-9}{17} < \frac{50-9}{17}\right) = \mathrm{P}(Z < 2.412) = 0.9921 \approx 0.992.$$

Similarly, let Y be the test result for a randomly selected diseased animal; we assume that $Y \sim \mathrm{N}(75, 12)$. The sensitivity is the probability to test positive, that is for the test result to exceed the cut-off at 50:

$$\mathrm{Se} = \mathrm{P}(Y > 50) = \mathrm{P}\left(\frac{Y-75}{12} > \frac{50-75}{12}\right) = \mathrm{P}(Z > -2.083) = 0.9814 \approx 0.981.$$

2. Choice of cut-off value

The 1% percentile in a standard normal distribution is -2.33 (Table B of PSLS), or more precisely -2.326 (Table C). Therefore, in order to achieve a sensitivity of 0.99 at a cut-off value x, the corresponding z-score of x must equal -2.326; that is, we solve the equation

$$z = \frac{x-75}{12} = -2.326, \quad \text{or} \quad x = 75 - 2.326 \cdot 12 = 47.09 \approx 47.$$

At a cut-off of 47%, the sensitivity of the test equals approximately 0.99.

A calculation similar to that of Question 1 shows that the specificity at a cut-off of 47% would be 0.987. The desired cut-off to make the sensitivity and specificity equal must therefore be between 47% and 50%. One approach to solving the question is to try some values between 47% and 50%, and choose the one with the best agreement between sensitivity and specificity. An analytical solution is obtained by noting that the sensitivity and specificity are both computed from z-scores, and if these are equal then the probabilities will be equal as well. At a cut-off value of x, the z-score for the sensitivity is $-(x-75)/12$ (the minus sign is because the probability is for values greater than x), and the z-score for the specificity is $(x-9)/17$. We therefore solve

$$-\frac{x-75}{12} = \frac{x-9}{17} \quad \Rightarrow \quad x = \frac{17 \cdot 75 + 12 \cdot 9}{12 + 17} = \frac{1383}{29} = 47.7.$$

Any cut-off value between 47.5% and 48% will give approximately equal sensitivity and specificity. Note that this does *not* equal the point where the two density curves intersect. The requirement is about the areas under the curve, and as the two standard deviations are different these areas will be different to the left and right, respectively, of where the curves intersect (that value being ≈ 46.6).

3. Classification for a sample of animals

When splitting the test results into two groups at the cut-off of 50%, the diseased group comprises 26 animals and the non-diseased group the remaining 9 animals. Assuming that these animals represent a population, we can use these results for inference about the proportion of diseased animals in the population, as determined by the test. Because the test is not perfect, this is not the same as the true proportion of diseased animals (assuming a perfect classification would exist), and to stress the difference the parameter is often called the *apparent* prevalence (as opposed to a true prevalence). We assume the count of (apparently) diseased animals to follow a binomial distribution, $X \sim \text{Bin}(35, p)$. In order to use a binomial distribution for this sampling from a finite population, we would need the sample size to not exceed 5% of the population size (N), i.e. $N \geq 20 \cdot 35 = 700$. Our estimated proportion equals $\hat{p} = 26/35 = 0.743$, and for a 95% confidence interval we need to use either the "plus four" or the "exact" method, because the "classical" method does not meet the condition of at least 15 negatives. The latter interval is obtained from Minitab as $(0.567, 0.875)$. For the plus four interval we compute the adjusted proportion of the augmented (plus four) data as $\tilde{p} = (26+2)/(35+4) = 0.718$, and proceed as for the classicial interval:

$$\begin{aligned} \text{95\% CI for } p\text{:} \quad \tilde{p} \pm z^* \sqrt{\tilde{p}(1-\tilde{p})/(n+4)} \quad &= 0.718 \pm 1.96\sqrt{0.718 \cdot 0.282/39} \\ &= 0.718 \pm 1.96 \cdot 0.072 = 0.718 \pm 0.141 = (0.577, 0.859). \end{aligned}$$

We are 95% confident that the apparent prevalence is within any of these (quite similar) intervals.

The one concern in the classification relates to the test result of 50.3. It is very close to the cut-off, and it seems uncertain whether the animal is really diseased. We know from the calculations in (A) and (B) that such a value would be quite unlikely in both the diseased and non-diseased populations, and we could exactly calculate the probability of getting a value of 50.3 (or more extreme) for both diseased and non-diseased animals. However, as 50.3 is so close to the cut-off, we would get values very close to the already computed sensitivity and specificity. With the sensitivity being somewhat lower than the specificity, we could argue that the result most likely belongs to the diseased group (so our classification being correct). One option that might exist is to retest the animal (multiple times), to see if the test result tends to increase or decrease in value. Alternatively, another test might be used to help classifying the animal, as outlined in Part 5.

4. Test results for a sample of animals

If the population consists of two separate subpopulations (i.e., the non-diseased and diseased animals), we would expect the distribution of test results to be bimodal, and this is easily confirmed with descriptive statistics for the full sample. It is more interesting to look at the distributions in the two subpopulations separately. For the animals in each group, the test results are assumed to be independent and follow the same distribution (i.i.d.). Specifically we want to compare the distributions to the respective normal distributions referred to in Part 1. We estimate the mean, standard deviation and skewness, compute a P-value for A-D test for the normal distribution, and give a confidence interval and a statistical test for the mean parameter. As a population standard deviation is available, we have a choice between using this more precise value (if the sample matches the previous data) or using the sample standard deviation, leading to z- or t-distribution inference, respectively. Both would be acceptable with a reasonable justification. The results are shown in the table below.

statistic	diseased animals	non-disased animals
no. of obs. (n)	26	9
mean ($\hat{\mu}$)	74.57	8.27
std.dev. (s)	9.54	14.94
skewness	-0.86	-0.42
normality test (P)	0.153	0.959
95% CI for mean:	$\pm 1.96 \cdot 12/\sqrt{26} = \pm 4.61$	$\pm 1.96 \cdot 17/\sqrt{9} = \pm 11.1$
(t-distrib.)	$\pm 2.06 \cdot 9.54/\sqrt{26} = \pm 3.85$	$\pm 2.306 \cdot 14.94/\sqrt{9} = \pm 11.5$
hypotheses for mean	$\mathrm{H}_0 : \mu = 75,\ \mathrm{H}_a : \mu \neq 75$	$\mathrm{H}_0 : \mu = 9,\ \mathrm{H}_a : \mu \neq 9$
test for mean:	$z = -0.18$ ($P = 0.86$)	$z = -0.13$ ($P = 0.90$)
(t-distrib.)	$t = -0.23$ ($P = 0.82$)	$t = -0.15$ ($P = 0.89$)

The sample means are very close to the assumed population means. The sample standard deviations are somewhat larger than the population values, but this may be acceptable in view of the small sample size. The normal probability plots (not shown) look fine, and the A-D tests give no evidence of non-normality of the distributions, even if the distribution for diseased animals has moderate left-skewness. We should be able to use inference based on normal distribution assumptions for both samples. The tests for the population means are totally non-significant. In conclusion, there is nothing in these data to indicate discrepancies with the assumed distributions.

5. Additional test

Consider first the combined test with test A the initial test and test B the confirmatory test. For an animal to test positive, it must test positive on both tests, and when the animal's disease status is known (either diseased or non-diseased), this occurs independently of each other (this is actually an assumption of the tests, called *conditional independence*). Let us use A^+ and A^- to denote a positive and a negative result by test A, respectively, and similarly for test B. For a diseased animal, we obtain the sensitivity as (by the multiplication rule),

$$\mathrm{Se}_{AB} = \mathrm{P}(A^+ \text{ and } B^+) = \mathrm{Se}_A \cdot \mathrm{Se}_B = 0.9 \cdot 0.95 = 0.855.$$

An animal will be deemed negative either by testing negative on test A, or by testing positive on test A and subsequently testing negative by test B. Again, for a truly non-diseased animal, the latter two events are independent. Therefore, the specificity is

$$\mathrm{Sp}_{AB} = \mathrm{P}(A^-) + P(A^+) \cdot P(B^-) = \mathrm{Sp}_A + (1 - \mathrm{Sp}_A) \cdot \mathrm{Sp}_B = 0.99 + (1 - 0.99) \cdot 0.95 = 0.9995.$$

Somewhat less intuitively, because by the testing strategy the event (A^- and B^-) never occurs, we can also obtain the probability as

$$\mathrm{P}(A^- \text{ or } B^-) = \mathrm{P}(A^-) + \mathrm{P}(B^-) - \mathrm{P}(A^- \text{ and } B^-) = \mathrm{Sp}_A + \mathrm{Sp}_B - \mathrm{Sp}_A \cdot \mathrm{Sp}_B,$$

which is exactly the same formula. With the roles of test A and B reversed, we calculate

$$\begin{aligned} \mathrm{Se}_{BA} &= 0.95 \cdot 0.9 = 0.855, \\ \mathrm{Sp}_{BA} &= 0.95 + (1 - 0.95) \cdot 0.99 = 0.9995. \end{aligned}$$

The performance of the combined test does not depend on the order in which the two tests are applied! Therefore, one can safely use the cheapest test first. When two tests are combined in this way, usually termed a *series* interpretation (i.e., positive results are required on both tests), there is no need to test samples with the second test if they tested negative on the first test.

6. Group testing

We can use a binomial distribution for inference about the sample of 50 fish from a finite population of size 10 000 fish, because the population is much larger than the sample ($50 \cdot 20 \ll 10\,000$). The number of fish testing positive in the sample is therefore assumed to follow a binomial distribution with $n = 50$; let $X \sim \text{Bin}(50, p)$. With no diseased fish in the cage, the probability of each fish testing (false) positive is 1 minus the specificity, thus $p_A = 1 - 0.99 = 0.01$ for test A. Then we have

$$\begin{aligned}
\text{test A}: \quad \text{P}(X \geq 1) &= 1 - \text{P}(X = 0) = 1 - (1 - p_A)^{50} = 1 - 0.99^{50} = 1 - 0.605 = 0.395, \\
\text{test B}: \quad \text{P}(X \geq 1) &= 1 - \text{P}(X = 0) = 1 - (1 - p_B)^{50} = 1 - 0.95^{50} = 1 - 0.077 = 0.923.
\end{aligned}$$

With test A it is possible but not most likely to get one or more false positive test results. Due to the lower specificity of test B, this test with high probability gives at least one false positive result when applied to 50 fish.

Finally, a similar approach will apply also if the cage contains both diseased and non-diseased fish, once we modify our calculation of the probability of samples testing positive. A sample can test positive in two says: either for a diseased fish that correctly tests positive or for a non-diseased that falsely tests positive. In order to calculate the probabilities for these events, we will need to know the proportion of diseased fish in the cage. If we call this parameter p_d, then the binomial proportion becomes $p = p_d \cdot \text{sensitivity} + (1 - p_d) \cdot \text{specificity}$, and calculations from there on proceed as above.