Extra exercise 13

Data: Dogs trained to identify smells from breast cancer patients completed 125 trials. In each trial, the dog had to choose between 5 breath samples, of which one originated from a breast cancer patients and the four others were control samples. The dogs correctly identified the breast cancer sample in $X = 110$ trials out of the total $n = 125$.

Model: We assume a binomial setting and hence that $X \sim \mathrm{B}(125, p)$, where p is the probability of correct identification. The observed proportion is $\hat{p} = X/n = 110/125 = 0.88$.

Confidence interval

For a 95% confidence interval we have the choice between our 3 approaches. As $n(1-\hat{p}) = 15$ (the number of negatives), we are exactly on the cut-off for meeting the condition for use of the normal approximation. The best CI is obtained by the plus four method. For comparison purposes we compute all 3 intervals:

- standard error of $\hat{p}$: $\sqrt{\hat{p}(1-\hat{p})/n} = 0.0291$,
  classical approximate 95% CI for $p : 0.88 \pm 1.96 \cdot 0.0291 = (0.823, 0.937)$,
- plus four adjusted sample proportion: $\tilde{p} = (X+2)/(n+4) = 112/129 = 0.868$,
  standard error of $\tilde{p}$: $\sqrt{\tilde{p}(1-\tilde{p})/(n+4)} = 0.0298$,
  plus four approximate 95% CI for p: $0.868 \pm 1.96 \cdot 0.0298 = (0.810, 0.927)$,
- "exact" binomial 95% CI (Minitab 21 & 22, Stata): $(0.810, 0.931)$.

It is seen that the plus four CI and the exact CI are very similar (the latter is a bit wider), whereas the normal approximation CI gives a quite different range that is symmetric around 0.88. As noted above, the normal approximation CI is not very good in this case.

Hypothesis testing I

If the dogs were purely "guessing" (randomly deciding on one of the 5 samples), we would have $p = 1/5 = 0.2$. It is therefore of interest to test this hypothesis with a one-sided alternative:

$$\mathrm{H}_0 : p = 0.2 \quad \text{versus} \quad \mathrm{H}_a : p > 0.2.$$

A one-sided alternative is chosen because the interest is in whether the dogs do better than guessing (it is hard to imagine they could do worse; this is similar to the duo-trio testing example in Lecture 6).

An exact test of the hypothesis has a P-value $P < 0.0005$ or $P < 0.0000005$, from Minitab and Stata, respectively. An approximate test using the normal approximation (conditions ok: $np_0 = 125 \cdot 0.2 = 25 > 10$) is based on

$$z = (0.88 - 0.2)/\sqrt{0.2 \cdot 0.8/125} = 19.01,$$

and has a very small P-value as well. There is no doubt we must reject H_0 and conclude that the dogs do better than guessing. This conclusion was probably pretty obvious from the observed proportion being so much larger than 0.2 anyway, as well as the value 0.2 being so far away from the CIs.

Hypothesis testing II

The paper by McCulloch et al. (2006) report that dogs achieved a detection rate of 99% for breath samples from lung cancer patients. We are interested in assessing whether the dogs do equally well with breast cancer and lung cancer samples. This is really a two-sample situation, but in absence of actual data for lung cancer samples (and thus assuming that the rate of 0.99 was obtained with very small standard error) we will compare the breast cancer data to a rate of 0.99:

$$\mathrm{H}_0 : p = 0.99 \quad \text{versus} \quad \mathrm{H}_a : p \neq 0.99.$$

A two-sided alternative is chosen because we are interested in any differences between the rates for the two types of samples. Recall that it is not allowed to observe that the breast cancer rate is lower than 0.99 and chose the alternative as $\mathrm{H}_a : p < 0.99$ for this reason. The hypotheses must always be chosen independently of the data.

In this case, the z-test based on the normal approximation is not a valid option, because $n(1-p_0) = 125 \cdot (1-0.99) = 1.25 < 10$. Therefore we have to use an exact test based on the binomial distribution. So far all our exact tests in the binomial distribution have been with a one-sided alternative or for the hypothesis $\mathrm{H}_0 : p = 0.5$ where the reference distribution, i.e. the distribution under the null hypothesis – $\mathrm{B}(n, 0.5)$, was symmetrical. It is less clear how to compute P-values for a two-sided alternative when the reference distribution is asymmetrical, and statistical software use different formulae, as we will now explain.

Let us denote our observed count of 110 by $X_{\text{obs}} : X_{\text{obs}} = 110$. The simplest formula for computing the P-value against a two-sided alternative is to double the one-sided P-value, here (with $X \sim B(125, 0.99)$),

$$P = 2 \cdot \min(P(X \leq X_{\text{obs}}), P(X \geq X_{\text{obs}})) = 2 \cdot P(X \leq 110) < 2 * 0.00000005 = 0.0000001.$$

As the reference is asymmetrical, it is however better to compute the probabilities in the two tails directly. Stata and Minitab does this in slightly different ways but in this particular example there is no contribution from the upper tail by either approach. The method in Minitab which is easiest to understand and the slightly worse of the two, computes the expected number of positives under H_0, here $125 \cdot 0.99 = 123.75$, and determines the point from where to include probabilities in the upper tail by symmetry of the observed count: $123.75 + (123.75 - X_{\text{obs}}) = 137.5$. As $137.5 > 125$ there is no contribution from the upper tail. Stata does not determine the point in the upper tail by symmetry but by when the individual probabilities are lower than $P(X = X_{\text{obs}})$. Also here there is no contribution from the upper tail. Hence for both methods,

$$\text{two-sided } P = P(X \leq X_{\text{obs}}) < 0.00000005.$$

We conclude that there is strong evidence that the dogs are better at sniffing out lung cancer patients than breast cancer patients. As noted above, the proper statistical analysis would be from a two-sample test if we had the data for the lung cancer samples.

We conclude with some Minitab commands and output. The last output is from Minitab version 22 and shows the CI calculation by the "Adjusted Blaker exact" method; it turns out that to the first three decimal places it is the same result as with the standard (Clopper-Pearson) "exact" method.

```
POne 125 110;
 Confidence 95.0;
 Alternative 0;
 UseZ.
```

Test and CI for One Proportion

Method

p: event proportion
Normal approximation method is used for this analysis.

Descriptive Statistics

N	Event	Sample p	95% CI for p
125	110	0.880000	(0.823033, 0.936967)

```
POne 125 110;
 Confidence 95.0;
 Alternative 0.
```

Test and CI for One Proportion

Method

p: event proportion
Exact method is used for this analysis.

Descriptive Statistics

N	Event	Sample p	95% CI for p
125	110	0.880000	(0.809811, 0.931257)

```
POne 125 110;
 Test 0.2;
 Confidence 95.0;
 Alternative 1;
 UseZ.
```

Test and CI for One Proportion

Method

p: event proportion
Normal approximation method is used for this analysis.

Descriptive Statistics

N	Event	Sample p	95% Lower Bound for p
125	110	0.880000	0.832192

Test

Null hypothesis	H_0: $p = 0.2$
Alternative hypothesis	H_1: $p > 0.2$

Z-Value	P-Value
19.01	0.000

```
POne 125 110;
 Test 0.99;
 Confidence 95.0;
 Alternative 0.
```

Test and CI for One Proportion

Method

p: event proportion
Exact method is used for this analysis.

Descriptive Statistics

N	Event	Sample p	95% CI for p
125	110	0.880000	(0.809811, 0.931257)

Test

Null hypothesis	H_0: $p = 0.99$
Alternative hypothesis	H_1: $p \neq 0.99$

P-Value
0.000

```
PONE 125 110;
CONFIDENCE 95.0;
ALTERNATIVE 0;
TEST .99;
ABLAKER.
```

WORKSHEET 1

Test and CI for One Proportion

Method

p	Event proportion
Method	Adjusted Blaker's exact method

Descriptive Statistics

N	Event	Sample p	95% CI for p
125	110	0.880000	(0.810325, 0.930581)

Test

Null hypothesis	H_0: p = 0.99
Alternative hypothesis	H_1: p ≠ 0.99

P-Value
0.000