Supplementary exercise 8.62 of IPS7e

Data: A randomized trial on aspirin for treatment of stroke, involving a treatment with aspirin group ($n_1 = 78$) and a control group ($n_2 = 77$). After 6 months, each patient's success was evaluated as favourable or not. The statistical model is two independent binomial distributions, one for each of the two groups. Formally, if we let X_1 and X_2 denote the number of favourable outcomes in the treatment and control groups, respectively, we have: $X_1 \sim \mathrm{B}(n_1, p_1)$ and $X_2 \sim \mathrm{B}(n_2, p_2)$.

(a) Estimates and standard errors are given in the table below. The number of positives and negatives exceed 10 in both samples, so the classical (normal distribution approximation) method should work well for a confidence interval for the difference in proportions. For completeness, calculations are included for the plus four interval as well, although there is no pressing need to use it here.

statistic	treatment	control	difference
proportion $\hat{p}_i$	$63/78 = 0.807$	$43/77 = 0.558$	0.249
stand. error $\mathrm{SE}(\hat{p}_i)$	0.0446	0.0566	0.0721
plus four corrected values:			
proportion $\tilde{p}_i$	$64/80 = 0.800$	$44/79 = 0.557$	0.243
stand. error $\mathrm{SE}(\tilde{p}_i)$	0.0447	0.0559	0.0716

where the standard errors are calculated as:

$$\begin{aligned} \mathrm{SE}(\hat{p}_i) &= \sqrt{\hat{p}_i(1-\hat{p}_i)/n_i},\ i = 1,2 \\ \mathrm{SE}(D) &= \sqrt{\mathrm{SE}(\hat{p}_1)^2 + \mathrm{SE}(\hat{p}_2)^2} \end{aligned}$$

(b) An approximate 90% CI for the difference in proportions between treatment and control is (using $z^* = 1.645$),

$$90\% \text{ classical CI}: \quad 0.249 \pm 1.645 \cdot 0.0721 = 0.249 \pm 0.119 = (0.131, 0.368).$$

After augmenting the data with positives and negatives, the corresponding 90% CI becomes

$$90\% \text{ “plus four” CI}: 0.243 \pm 1.645 \cdot 0.0716 = 0.243 \pm 0.118 = (0.125, 0.361).$$

The confidence interval does not include zero (far from close), so there would seem to be strong evidence against the hypothesis of the same proportion of favourable stroke progress in the two groups; the aspirin group is doing much better.

(c) The suggested hypotheses for the test are:

$$\mathrm{H}_0: p_1 = p_2, \quad \text{versus} \quad \mathrm{H}_a: p_1 > p_2,$$

where as before $p_1 \sim$ treatment and $p_2 \sim$ control. The PSLS/IPS condition for use of the classical, large-sample z-test is that both samples have a count of at least 5 of both cases and failures. This condition is easily satisfied. A better condition, discussed below, comes from the equivalence with the X^2-test to be introduced in Session 9.

A first step is to calculate the combined (or pooled) estimate of p across the two groups: $\hat{p} = (63 + 43)/(78 + 77) = 106/155 = 0.6839$. With this pooled estimate of p, the Session 9 condition for the z-test is that expected counts of both positives and negatives in both samples based on this estimate $\hat{p}$ are all greater than 5. For example, the expected count of positives for the treatment group is $\hat{p}n_1 = 0.6839 \cdot 78 = 53$. It is seen that all four conditions are met.

Next we calculate the pooled standard error and the z-statistic:

$$\begin{aligned} \mathrm{SE}(D_p) &= \sqrt{\hat{p}(1-\hat{p})(1/78+1/77)} = 0.0747, \\ z &= D/\mathrm{SE}(D_p) = 0.249/0.0747 = 3.337 \approx 3.34. \end{aligned}$$

A z-value of 3.34 is very clearly significant in N(0, 1). The normal distribution table shows the tail probability to be 0.0004, so $P = 0.0004$. There is strong evidence to reject the hypothesis that the proportion of favourable progress differs between the treatment and control groups. As noted above, the aspirin group has done much better.

We finally give Minitab commands and output for the analysis, including only classical results (i.e., without the "plus four" correction); some notes on the software settings:

- the "pooled" setting only affects the calculation of the z-test and has no impact on the confidence interval,
- the listing also includes a P-value for Fisher's exact test (Session 8),
- if extra counts are being added for the "plus four" CI, the corresponding test statistics (based on the augmented data) are *never* valid!

```
PTwo 78 63 77 43;
 Confidence 90;
 Test 0.0;
 Alternative 0;
 Pooled.
```

Test and CI for Two Proportions

Method

p_1: proportion where Sample 1 = Event
p_2: proportion where Sample 2 = Event
Difference: $p_1 - p_2$

Descriptive Statistics

Sample	N	Event	Sample p
Sample 1	78	63	0.807692
Sample 2	77	43	0.558442

Estimation for Difference

Difference	90% CI for Difference
0.249251	(0.130710, 0.367792)

CI based on normal approximation

Test

Null hypothesis H_0: $p_1 - p_2 = 0$
Alternative hypothesis H_1: $p_1 - p_2 \neq 0$

Method	Z-Value	P-Value
Normal approximation	3.34	0.001
Fisher's exact		0.001

The pooled estimate of the proportion (0.683871) is used for the tests.

```
PTwo 78 63 77 43;
 Confidence 90;
 Test 0.0;
 Alternative 1;
 Pooled.
```

Test and CI for Two Proportions

Method

p_1: proportion where Sample 1 = Event
p_2: proportion where Sample 2 = Event
Difference: $p_1 - p_2$

Descriptive Statistics

Sample	N	Event	Sample p
Sample 1	78	63	0.807692
Sample 2	77	43	0.558442

Estimation for Difference

Difference	90% Lower Bound for Difference
0.249251	0.156892

CI based on normal approximation

Test

Null hypothesis	H_0: $p_1 - p_2 = 0$
Alternative hypothesis	H_1: $p_1 - p_2 > 0$

Method	Z-Value	P-Value
Normal approximation	3.34	0.000
Fisher's exact		0.001

The pooled estimate of the proportion (0.683871) is used for the tests.