Supplementary exercises 8.84 and 8.85 of IPS7e

<u>Data</u>: A total of $n = 100$ employees from a chain of restaurants were asked whether work stress has a negative impact on personal life. The obvious statistical model is a single binomial distribution, corresponding to assuming a binomial setting. Formally, if we let X denote the number of those employees answering yes, we assume: $X \sim \mathrm{B}(n, p)$.

<u>Estimation</u>: Our estimates based on observing $X = 68$ out of $n = 100$ are:

$$\begin{aligned} \text{sample proportion} &: \quad \hat{p} = X/n = 68/100 = 0.680, \\ \text{standard error of } \hat{p} &: \quad \mathrm{SE}(\hat{p}) = \sqrt{\hat{p}(1-\hat{p})/n} = \sqrt{0.68 \cdot 0.32/100} = 0.04665. \end{aligned}$$

Exercise 8.84
The condition for use of the classical (normal distribution approximation) method for the confidence interval is satisfied here because the data contain more than 15 positives and 15 negatives. For illustration, the other methods for computing a confidence interval are included as well, although in this situation the classical interval is acceptable.

First, the classical approximate 95% CI is computed as:

$$\hat{p} \pm 1.96 \cdot \mathrm{SE}(\hat{p}) = 0.680 \pm 1.96 \cdot 0.04665 = 0.680 \pm 0.0914 = (0.589, 0.771).$$

Next, for the "plus four" method we first compute the adjusted estimate for p and then follow the same steps:

$$\begin{aligned} \text{adjusted proportion} &: \quad \tilde{p} = (X+2)/(n+4) = 70/104 = 0.673, \\ \text{standard error of } \tilde{p} &: \quad \mathrm{SE}(\tilde{p}) = \sqrt{\tilde{p}(1-\tilde{p})/(n+4)} = 0.045998, \\ 95\%\ \text{CI} &: \quad 0.673 \pm 1.96 \cdot 0.045998 = 0.673 \pm 0.0902 = (0.583, 0.763). \end{aligned}$$

Finally, the "exact" CI (based on the binomial distribution) must be obtained by software, and the standard (Clopper-Pearson) method gives the interval: $(0.579, 0.770)$.

These CIs are similar but not exactly the same. The plus four CI is not symmetrical around the estimate (it's symmetrical around $\tilde{p}$). The "exact" interval is seen to be wider than the other two, by approximately 0.01, reflecting that it is a bit conservative. The Minitab (version 21) commands and output for these confidence intervals are shown below.

```
POne 100 68;
 Confidence 95.0;
 Alternative 0;
 UseZ.
```

WORKSHEET 1

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
100	68	0.680000	(0.588572, 0.771428)

```
POne 104 70;
 Confidence 95.0;
 Alternative 0;
 UseZ.
```

WORKSHEET 1

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
104	70	0.673077	(0.582923, 0.763231)

```
POne 100 70;
 Confidence 95.0;
 Alternative 0.
```

WORKSHEET 1

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
100	68	0.680000	(0.579233, 0.769780)

Minitab version 22 has a new menu for 1 Proportion inference, and the methods available for confidence intervals are different. First, the default "exact" method (for two-sided CIs) is new and termed "Adjusted Blaker's exact" method, which is supposed to be less conservative than the Clopper-Pearson method. The Minitab version 22 listing below gives this interval as: $(0.581, 0.767)$; it is indeed a bit shorter than the Clopper-Pearson interval above.

The normal approximation CI is no longer available, but the interval labelled as the "Agresti-Coull method" is very close to the "plus four" interval from the notes (and textbooks). The difference is due to the use of $\tilde{p} = (X + (z^\star)^2/2)/(n + (z^\star)^2)$, which with $z^\star = 1.96$ (instead 2 in the "plus four" formula) becomes 0.67334 instead of 0.67308 above with a corresponding small change in $\mathrm{SE}(\tilde{p})$, 0.046023. The difference in these methods is negligible in practice, so for all practical purposes the "Agresti-Coull" method can be considered the same as the "plus four" method.

Finally, the added "Wilson" method can be viewed as a modification of the "Agresti-Coull" method, but also here the difference in results is typically small (unless sample sizes are small). Minitab version 22 commands and output for these confidence intervals are shown on the next page.

```
PONE 100 68;
CONFIDENCE 95.0;
ALTERNATIVE 0;
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
100	68	0.680000	(0.580738, 0.767345)

```
PONE 100 68;
CONFIDENCE 95.0;
ALTERNATIVE 0;
ACOULL.
```

WORKSHEET 1

Test and CI for One Proportion

Method

p Event proportion
Method Agresti-Coull method

Descriptive Statistics

N	Event	Sample p	95% CI for p
100	68	0.680000	(0.583137, 0.763546)

```
PONE 100 68;
CONFIDENCE 95.0;
ALTERNATIVE 0;
WILSON.
```

WORKSHEET 1

Test and CI for One Proportion

Method

p Event proportion
Method Wilson-score method

Descriptive Statistics

N	Event	Sample p	95% CI for p
100	68	0.680000	(0.583374, 0.763309)

Exercise 8.85
The national survey had 75% of respondents answering yes. Because that survey was large, it may be acceptable to assume there is no error associated with its estimate, and hence treat it as a fixed value. This means we will be testing the hypotheses:

$$\mathrm{H}_0 : p = 0.75, \quad \text{and} \quad \mathrm{H}_a : p \neq 0.75$$

based on the single sample. If data from the large survey were available, it would have been appropriate to use methods for two independent samples (proportions).

Before we start the calculations of hypothesis tests it is worth noting that with the confidence intervals already computed, we can carry out a significance test from those confidence intervals, by simply looking at whether the target value (0.75) is inside the interval or not (slide 5L−19). All three intervals include the value 0.75, so we know that the corresponding significance tests will have $P > 0.05$. The plus-four CI does not have a corresponding significance test, but the other two CIs have roughly matching tests, so we can be quite confident in the significance assessment from the CIs alone. In this case, the target value is rather close to the upper endpoint of the CIs, so we might want to know the precise P-value. That could be one reason for computing a test.

The conditions for use of the classical z-test are met here because $100 \cdot 0.75 = 75 \geq 10$ and $100 \cdot (1 - 0.75) = 25 \geq 10$. The z-test calculation goes as follows:

$$\begin{aligned} z &= (0.68 - 0.75)/\sqrt{0.75 \cdot 0.25/100} = -1.62, \\ P &= 2 \cdot P(Z > 1.62) = 0.106. \end{aligned}$$

If we were using the table for $N(0,1)$, we would get $P(Z > 1.62) = 0.0526$. In any case, we conclude that there is not sufficient evidence to reject the null hypothesis, and the proportion of stressed employees at the chain of restaurants could very well match the nationwide level.

For illustration purposes, we also compute P-values for an exact test based on the binomial distribution. This method is generally preferable and truly exact, but unless the tested proportion equals 0.5 there is no uniform rule for how to compute two-sided P-values. The simplest rule is to double the one-sided P-value, that is, for $X \sim B(100, 0.75)$:

$$P = 2 \cdot P(X \leq 68) = 2 \cdot 0.0693 = 0.139.$$

Minitab computes the two-sided P-value by adding the probabilities for outcomes equally far from the hypothesized value (corresponding to an expected count of $100 \cdot 0.75 = 75$) on both sides, that is:

$$P = P(X \leq 68) + P(X \geq 82) = 0.0693 + 0.0630 = 0.132.$$

Stata and R compute the two-sided P-value by adding all probabilities not larger than that of the observed count ($P(X = 68) = 0.0248$) on both sides, which in our case becomes:

$$P = P(X \leq 68) + P(X \geq 83) = 0.0693 + 0.0376 = 0.107.$$

It is seen that among the exact P-values, the method used in Stata and R (which is considered more accurate) agrees well with the z-test, due to the fairly large sample size. Finally, we show on the next page the Minitab commands for these tests; note that the z-test is no longer available in version 22.

```
POne 100 68;
 Test 0.75;
 Confidence 95.0;
 Alternative 0;
 UseZ.
```

WORKSHEET 1

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
100	68	0.680000	(0.588572, 0.771428)

Test

Null hypothesis H_0: p = 0.75
Alternative hypothesis H_1: p ≠ 0.75

Z-Value	P-Value
-1.62	0.106

```
POne 100 68;
 Test 0.75;
 Confidence 95.0;
 Alternative 0.
```

WORKSHEET 1

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
100	68	0.680000	(0.579233, 0.769780)

Test

Null hypothesis H_0: p = 0.75
Alternative hypothesis H_1: p ≠ 0.75

P-Value
0.132