Homework exercise

To be solved at home before the exercise session.

1. a. A simple sample size calculation can be performed for binary proportion confidence intervals as follows. We bound the standard deviation estimate from above as $\sqrt{\hat{p}(1-\hat{p})} \leq 0.5$ to obtain the *conservative* confidence interval,

$$\left(\hat{p} - z_{\alpha/2}\frac{0.5}{\sqrt{n}}, \hat{p} + z_{\alpha/2}\frac{0.5}{\sqrt{n}}\right).$$

The half-width of a confidence interval is known as its *margin of error* and for the conservative confidence interval the margin of error does not depend on the proportion of "successes". Thus we can compute a universal sample size for which a certain desired margin of error is reached.

i. Compute the required sample sizes to obtain the margins of error of 0.01, 0.02 and 0.03 for a 95% conservative confidence interval.

ii. Study how much the calculations in part *i* over-estimate the required sample sizes when the proportion of successes is small $\hat{p} = 0.05$. That is, redo part *i* using the regular binary confidence interval in slide 4.6.

b. A manufacturer claims that only 6% of their products are faulty. To investigate this, a customer picks a random sample of size n of products and observes the proportion of faulty ones to be $\hat{p} = 0.09$. He tests the manufacturer's claim using the asymptotic one-sample proportion test in slide 4.9. Is the p-value of the test smaller for sample size $n = 100$ or $n = 200$?

a.i) For 95% conservative CI, alpha would be 5% and thus, $z_{alpha/2}$ = 1.96, from this table

Confidence Level	Alpha	Alpha/2	z alpha/2
90%	10%	5.0%	1.645
95%	5%	2.5%	1.96
98%	2%	1.0%	2.326
99%	1%	0.5%	2.576

Thus, margin of error is $z_{alpha/2}$ * 0.5/sqrt(n)

- For margins of error of 0.01, the sample size is: 1.96 * 0.5/sqrt(n) = 0.01 => n = 9604
- For margins of error of 0.02, the sample size is: 1.96 * 0.5/sqrt(n) = 0.02 => n = 2401
- For margins of error of 0.03, the sample size is: 1.96 * 0.5/sqrt(n) = 0.03 => n = 1067

a.ii) The formula in slide 4.6

For large n, a level $100(1-\alpha)\%$ confidence interval for the success probability p is obtained as

$$\left(\hat{p} - z_{\alpha/2}\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}, \hat{p} + z_{\alpha/2}\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}\right),$$

where $\hat{p}$ is the observed proportion of successes and $z_{\alpha/2}$ is the $(1-\alpha/2)$-quantile of the standard normal distribution.

=> sqrt(p(1-p)) = sqrt(0.05*(1-0.05)) = 0.2179

- For margins of error of 0.01, the sample size is: 1.96 * 0.2179/sqrt(n) = 0.01 => n = 1824
- For margins of error of 0.02, the sample size is: 1.96 * 0.2179/sqrt(n) = 0.02 => n = 456
- For margins of error of 0.03, the sample size is: 1.96 * 0.2179/sqrt(n) = 0.03 => n = 202

=> if the success probability is high at p = 0.5, it overestimate more than low success probability of p = 0.05 by magnitude of around 5

b) The asymptotic one sample-test

Asymptotic one-sample proportion test

If the sample size is large, then under the null hypothesis H_0 the standardized test statistic,

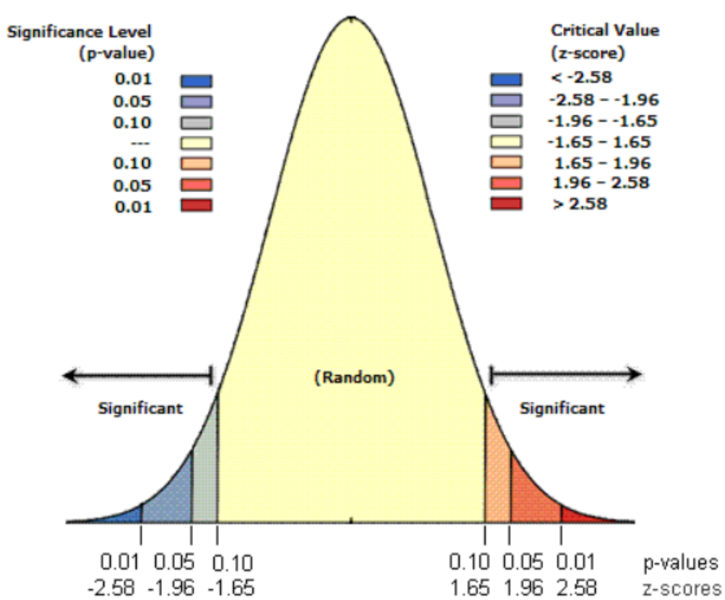
$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}},$$

where $\hat{p}$ is the unbiased estimator $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} x_i$ of the parameter p, follows approximately the standard normal distribution.

Plug in p^ = 0.09 and p_0 = 0.06 into the formula

- For n = 100, Z score is 1.26
- For n = 200, Z score is 1.78

If a z-score is 0, it indicates that the data point's score is identical to the mean score. A z-score of 1.0 would indicate a value that is one standard deviation from the mean. Z-scores may be positive or negative, with a positive value indicating the score is above the mean and a negative score indicating it is below the mean.
p-value indicates how unlikely the statistic is. z-score indicates how far away from the mean it is. Both z-scores and p-values are associated with the standard normal distribution



=> According to the graph, Z = 1.26 or n = 100 will yield a higher p-value