Exercise Sheet 7

Exercise 7.1 - Confidence Intervals

A company ships tea in wooden crates, each containing 10 tea packages. The weight of the individual tea packages is assumed to be normally distributed with $\mu = 6$ kg and $\sigma = 0.06$ kg. We also assume that the weight of the empty wooden crate is normally distributed with $\mu = 5$ kg and $\sigma = 0.05$ kg.

(a) Assuming all individual components to be independent, compute an interval, symmetric with respect to the expected value, which contains the gross weight of the ready-to-dispatch wooden crate in 95% of the cases.

Let X be the weight of the ready-to-dispatch wooden crate, Y_i the weight of the i-th tea package and V the weight of the empty wooden crate. From the exercise text, we get:

$$Y_i \sim N(6, 0.0036) \qquad V \sim N(5, 0.0025)$$

For $X = Y_1 + \ldots + Y_{10} + V$ we get a normal distribution due to the independence of all random variables involved:

$$\begin{aligned} \mu_X &= 10 \cdot \mu_{Y_i} + \mu_V = 10 \cdot 6 + 5 = 65 \\ \sigma_X^2 &= 10 \cdot \sigma_{Y_i}^2 + \sigma_V^2 = 10 \cdot 0.0036 + 0.0025 = 0.0385 \end{aligned}$$

so that

$$X \sim N(65, 0.0385)$$

We are trying to find a symmetric interval to μ_X, $[\mu_X - D, \mu_X + D]$, which overlaps the gross weight of the package in 95% of all cases, i.e.

$$P(\mu_X - D \leq X \leq \mu_X + D) = 0.95 = 1 - \alpha$$

has to hold. For the standardised variable Z we have

$$Z = \frac{X - \mu_X}{\sigma_X} \sim N(0, 1)$$

and thus for the $1 - \frac{\alpha}{2}$-quantile of the standard normal

$$P(-z_{1-\frac{\alpha}{2}} \leq Z \leq z_{1-\frac{\alpha}{2}}) = 1 - \alpha$$

Applying some minor transformations yields:

$$\begin{aligned} 1 - \alpha &= P(-z_{1-\frac{\alpha}{2}} \leq Z \leq z_{1-\frac{\alpha}{2}}) \\ &= P\left(-z_{1-\frac{\alpha}{2}} \leq \frac{X - \mu_X}{\sigma_X} \leq z_{1-\frac{\alpha}{2}}\right) \\ &= P\left(-z_{1-\frac{\alpha}{2}} \cdot \sigma_X \leq X - \mu_X \leq z_{1-\frac{\alpha}{2}} \cdot \sigma_X\right) \\ &= P\left(\mu_X - z_{1-\frac{\alpha}{2}} \cdot \sigma_X \leq X \leq \mu_X + z_{1-\frac{\alpha}{2}} \cdot \sigma_X\right) \end{aligned}$$

For the wanted interval with $\alpha = 0.05$ we thus obtain:

$$D = z_{1-\frac{\alpha}{2}} \cdot \sigma_X = z_{0.975} \cdot \sqrt{0.0385} = 1.96 \cdot \sqrt{0.0385} \approx 0.385$$

Therefore, putting things together:

$$[65 - 0.385\,,\, 65 + 0.385] = [64.615\,,\, 65.385]$$

(b) A customer of the above tea company checks whether the tea packages really comply with the specified target value of 6 kg. For this purpose, he takes a sample of $n = 16$ tea packages from a delivery. The arithmetic mean was found to be 5.95 kg. Calculate a 95% confidence interval for the average weight of the individual tea packages in the delivery, and interpret the result.

Since the weight Y_i of a single tea package is normally distributed with known variance, we know that for an $(1-\alpha)$-confidence interval,

$$[G_u, G_o] = \left[\bar{Y} - z_{1-\frac{\alpha}{2}} \frac{\sigma_{Y_i}}{\sqrt{n}}\,,\, \bar{Y} + z_{1-\frac{\alpha}{2}} \frac{\sigma_{Y_i}}{\sqrt{n}}\right]$$

holds. In the present case, for the wanted 95%-confidence interval with $\bar{y} = 5.95$, $\sigma_{Y_i} = 0.06$ and $z_{0.975} = 1.96$, this results in

$$\begin{aligned}
[g_u, g_o] &= \left[\bar{y} - z_{1-\frac{\alpha}{2}} \frac{\sigma_{Y_i}}{\sqrt{n}}\,,\, \bar{y} + z_{1-\frac{\alpha}{2}} \frac{\sigma_{Y_i}}{\sqrt{n}}\right] \\
&= \left[5.95 - 1.96 \frac{0.06}{4}\,,\, 5.95 + 1.96 \frac{0.06}{4}\right] \\
&= [5.9206\,,\, 5.9794]
\end{aligned}$$

Since the 95%-confidence interval does not contain the specified value of 6 kg, we can conclude at level $\alpha = 0.05$ that the weight of the individual tea packages is statistically significantly too low on average.

Exercise 7.2 - Confidence Intervals for Dichotomous Variables

A brewery is interested in which bottle design is more appealing to potential customers. Of 53 people they ask on the streets to pick between some designs, 42 pick the current blue design, whereas 4 pick a new green design and 7 choose one of some other options.

(a) Why is this way of conducting market research not optimal?

The goal of market research in this case is to find out about potential improvements to the product design that would increase sales. For this to work, participants in a study should not be influenced in their decision. We should try to guarantee that a participant is able to choose in an i.i.d. setting, which we can assume as given here. However, since the existing design, which will likely be known to some participants, is presented amongst other options, there might be a positive or negative bias towards that design, depending on whether a participant likes or dislikes that product. Thus, from the results we cannot conclude whether the old design is actually better or has just been picked more because it has been seen many times before and thus is associated with the product. Many people gravitate towards selecting options and buying products they are already familiar with.

(b) Let us now assume the issues from (a) don't exist and 42 picked the design in blue, whereas 4 picked the green. Calculate approximate 95 percent confidence intervals for the percentage of customers who prefer the blue/green design.

Since we are dealing with a moderate sample size $n = 53 > 30$, an approximate $(1-\alpha)$-confidence interval for the respective proportions is given by

$$[G_u, G_o] = \left[\hat{\pi} - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}\,,\, \hat{\pi} + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}\right]$$

For the concrete calculation of the percentage of participants preferring the blue design π_B and the green design π_G, one first calculates the estimates

$$\hat{\pi}_B = \frac{42}{53} \approx 0.7925 \qquad \hat{\pi}_G = \frac{4}{53} \approx 0.0755$$

and with $z_{1-\frac{\alpha}{2}} = z_{0.975} = 1.96$, for the blue design we get

$$\begin{aligned}
KI_B &= \left[\hat{\pi}_B - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{\pi}_B(1-\hat{\pi}_B)}{n}}\,,\, \hat{\pi}_B + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{\pi}_B(1-\hat{\pi}_B)}{n}}\right] \\
&= \left[0.7925 - 1.96\sqrt{\frac{0.7925 \cdot 0.2075}{53}}\,,\, 0.7925 + 1.96\sqrt{\frac{0.7925 \cdot 0.2075}{53}}\right] \\
&= [0.6833\,,\, 0.9017]
\end{aligned}$$

and for green

$$\begin{aligned}
KI_S &= \left[\hat{\pi}_G - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{\pi}_G(1-\hat{\pi}_G)}{n}}\,,\, \hat{\pi}_G + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{\pi}_G(1-\hat{\pi}_G)}{n}}\right] \\
&= \left[0.0755 - 1.96\sqrt{\frac{0.0755 \cdot 0.9245}{53}}\,,\, 0.0755 + 1.96\sqrt{\frac{0.0755 \cdot 0.9245}{53}}\right] \\
&= [0.0044\,,\, 0.1466]
\end{aligned}$$

For the widths of the confidence intervals we get

$$b_B = 0.9017 - 0.6833 = 0.2184 \qquad b_G = 0.1466 - 0.0044 = 0.1422$$

In general, the width of a $(1-\alpha)$ confidence interval is as follows (as long as the lower or upper bound don't exceed [0, 1]):

$$b = \hat{\pi} + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}} - \left(\hat{\pi} - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}\right) = 2 \cdot z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}$$

For fixed n and α, the width depends only on the product $\hat{\pi}(1-\hat{\pi})$ and thus on the estimate $\hat{\pi}$. In subtask (a) the following holds

$$\hat{\pi}_B(1-\hat{\pi}_B) = 0.7925 \cdot 0.2075 = 0.1644 \qquad \hat{\pi}_G(1-\hat{\pi}_G) = 0.0755 \cdot 0.9245 = 0.0698$$

which leads to the different widths of the CIs.

(c) Determine the minimum sample size n required to ensure that the estimated 95% confidence interval has a maximum width of 6 percentage points. For now, assume that the estimates $\hat{\pi}$ from (a) are known.

Solving the general equation for the width b from (a) for n, one gets a formula for the required sample size for a given maximum width b_{max}:

$$
\begin{aligned}
& b_{max} \geq b = 2 \cdot z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}} \\
\overset{b,z,\hat{\pi},n \geq 0}{\iff} \quad & b_{max}^2 \geq b^2 = 4 \cdot z_{1-\frac{\alpha}{2}}^2 \cdot \frac{\hat{\pi}(1-\hat{\pi})}{n} \\
\implies \quad & n \geq \frac{4 \cdot z_{1-\frac{\alpha}{2}}^2 \cdot \hat{\pi}(1-\hat{\pi})}{b_{max}^2}
\end{aligned}
$$

We now assume we have prior knowledge from the initial study with 53 participants. We can base the calculation of the sample size required to obtain a CI width of 6% on the expected proportion of customers liking each design in the larger sample, given our previous results. For this, we plug in the existing estimates:

$$
z_{1-\frac{\alpha}{2}} = 1.96 \qquad \hat{\pi}_B(1-\hat{\pi}_B) = 0.1644 \qquad \hat{\pi}_G(1-\hat{\pi}_G) = 0.0698 \qquad b_{max}^2 = 0.06^2 = 0.0036
$$

Thus, when rounding up to the next integer, for the blue design we get a minimum sample size of

$$
n_B = 702 > \frac{4 \cdot 1.96^2 \cdot 0.1644}{0.0036} \approx 701.7
$$

and for the green design

$$
n_G = 298 > \frac{4 \cdot 1.96^2 \cdot 0.0698}{0.0036} \approx 297.9
$$

(d) Let's now assume you don't yet know the estimates $\hat{\pi}$ from (a) (i.e. you are planning the survey and how many people you need to ask). How large would the respective sample size at least have to be if you are trying to reach the requested estimation accuracy from (c) for any possible value of π?

If there is no prior knowledge about $\hat{\pi}$, the product $\hat{\pi}(1-\hat{\pi})$ has to be estimated in a different way in the formula for n. We saw in previous exercises that for any value of $\hat{\pi}$, $\hat{\pi}(1-\hat{\pi}) \leq 0.25$, which is the 'worst case', and therefore we can estimate the minimum sample size without prior knowledge about $\hat{\pi}$ with:

$$
n > \frac{4 \cdot z_{1-\frac{\alpha}{2}}^2 \cdot 0.25}{b^2} = \frac{z_{1-\frac{\alpha}{2}}^2}{b^2}
$$

Note: Halving b results in quadrupling of n for fixed α.

For this exercise, we obtain:

$$
n_{min} = 1068 > \frac{1.96^2}{0.0036} \approx 1067.1
$$

A different view on this result: At the calculated sample size of 1068, we can be 95% confident that our estimated proportion of customers liking each design deviates less than 3 percentage points from the true value.

Exercise 7.3 - Testing

Someone offers you the following game: If 'tails' appears on a single coin toss, you win; if 'heads' is on top, your opponent wins. You suspect that your opponent is cheating you and is trying to play the game with a rigged coin with a higher probability of 'heads' appearing than 'tails'. You persuade your opponent that you want to check the coin first before you get involved in the game. To do this, you flip the coin 100 times, and based on the number of 'heads', you make a decision whether to stick to your suspicions or not.

(a) How can the hypotheses of this test be captured statistically?

For the random variable X: "We obtain heads on a single coin toss", we have:

$$X \sim B(1, \pi),$$

where π denotes the probability of obtaining heads. For the one-sided test the following hypotheses can be formulated:

$$H_0 : \pi \leq 0.5 \quad \text{versus} \quad H_1 : \pi > 0.5$$

(b) Which statistic, based on the data you obtain when flipping the coin 100 times, is useful for testing the hypothesis defined in (a)? What is the distribution of this test statistic under H_0?

Test statistic T : Number of heads when throwing the coin 100 times

$$T = \sum_{i=1}^{100} X_i \sim B(100, \pi)$$

This test statistic is sensitive to the test situation because the higher the probability π, the more often we expect heads to come up.

Test idea: If the test quantity assumes such a high value that even a critical limit c is exceeded ($T > c$), then this is strong evidence in favour of H_1.

Question: How should c be chosen?

(c) Based on this test statistic or a p-value calculated from it, we may arrive at a test decision, i.e. we either decide to believe that the coin is fair, or we do not. Which mistakes can be made here?

The following 4 situations are possible:

	decision for H_0 (H_0 is not rejected)	decision for H_1 (H_0 rejected)
H_0 is true	ok	Type I error
H_1 is true	Type II error	ok

(d) Keeping (c) in mind, think about when a test is a good test and when it is not. What problem does one face in finding a best test? How can we get a grip on this problem?

A good test is characterised by low probabilities for type I **and** type II error.

Problem: The one error can only be kept small at the expense of the other error:

The lower c, the easier it is to reject H_0, i.e.

- the greater the probability that one wrongly rejects H_0
- the smaller is the probability that one keeps H_0 although H_0 is not is true.

Solution: Asymmetric approach, ie: Limit the type I error by the so-called significance level α (typically $\alpha = 0.05$) and then find the test with the (uniformly) lowest type II error among all tests that meet this α condition.

Here, H_0 range: $\pi \in [0, 0.5]$. In the following, the distribution of the test statistic under the value closest to the H_1-range value $\pi = 0.5$ is considered. The α-condition is fulfilled for $c = 58, \ldots, 99$:

$$\begin{aligned}
P_{\pi=0.5}(T > 60) &= 0.0176 < 0.05 = \alpha \\
P_{\pi=0.5}(T > 59) &= 0.0284 < 0.05 \\
P_{\pi=0.5}(T > 58) &= 0.0443 < 0.05 \\
P_{\pi=0.5}(T > 57) &= 0.0666 > 0.05
\end{aligned}$$

According to test theory, the optimal bound is $c = 58$, since it is the highest value for c where the α-condition is still fulfilled.

If a test constructed in this way leads to the rejection of H_0, this speaks strongly in favour of H_1, i.e. H_1 is accepted at significance level $\alpha = 0.05$.

$\Rightarrow$ Choose the hypotheses H_0 and H_1 such that

- the statement to be statistically verified is in H_1
- or that the type I error represents the 'worst' of the two errors (e.g. in medical tests, where not detecting a disease is more harmful than falsely diagnosing one)

Exception: two-sided test