<u>Confidence Intervals and Hypothesis Testing</u>

Statistical reasoning provides a framework for making informed decisions in the presence of risk and uncertainty. Confidence intervals and hypothesis tests are fundamental tools in this framework, allowing reliable evaluation of claims and consequences. Understanding the potential errors in these calculations is essential for the correct interpretation of such results. Where confidence intervals might estimate the average weight of a chocolate bar lies somewhere between 99g and 101g, hypothesis testing would instead check if the claim that the average weight is 100g is true. Together, these concepts offer a structured approach to balancing risk with results and ensuring that conclusions are drawn correctly and efficiently.

Definition 1: (hypothesis testing)
$H_0: \mu = \mu_0$
$H_1: \mu \neq \mu_0$

Here, μ represents the population parameter, usually the mean, and μ_0 is the hypothesised value. H_0 is the null hypothesis – our original assumption – and H_1 is the alternative hypothesis.

A significance level is also chosen to determine the threshold for rejecting the null hypothesis. If, for example H_0 is rejected, H_1 is accepted. evidence from the sample is then used to decide whether the original assumption should be rejected or not.

Definition 2: (confidence interval estimate for sample mean)
$[\bar{x} - \frac{\sigma_{\bar{x}} z_{0.025}}{\sqrt{n}}, \bar{x} + \frac{\sigma_{\bar{x}} z_{0.025}}{\sqrt{n}}]$

Assuming the variance (σ^2), sample mean ($\bar{x}$), and the sample size (n) are given, a confidence interval can be calculated. Here, z represents the critical value from the normal distribution, determined using the significance level. The confidence interval, therefore, gives the range of values for which we can be confident that the true mean lies. While the sample mean is our best estimate, the true mean could be some amount higher or lower.

For example, it is unlikely that the exact average score on an exam is 75. But it is more likely that the true average falls within a range, such as from 73 to 77.

Proposition 2:
Values outside of the confidence interval provide evidence to reject a null hypothesis.

While hypothesis testing tests a critical value against a test statistic, confidence intervals calculate a range of potential true means. In hypothesis testing, if the test statistic is outside of the critical value, we reject the null hypothesis. Similarly, in confidence intervals, if the value we are testing lies outside of the confidence interval, the null hypothesis is rejected.

Conversely, if the value calculated lies within the interval, the sample taken is proven to be consistent with that value. Thus, the alternative hypothesis is rejected, and the null hypothesis is accepted. This is evidenced in the following application and represented in the later code sample.

Application:

In food manufacturing quality control, a chocolate bar may be labelled as 100g. However, the company may want to check that the average weight is consistent with what we expect.

Suppose a sample of 50 chocolate bars is taken, and the sample mean is a weight of 99.45g with a standard deviation of 1.27g. A significance level of 5%, with 2.5% on each tail, gives a confidence interval of 95%. We can then use the defined formula for the interval estimate of a sample mean with $\bar{x} = 99.45$g, $\sigma = 1.27$, $n = 50$, and $z = 1.96$ (of which the latter is found using R, as shown in the later code). This gives the interval $99.1 \leq \mu \leq 99.8$.

Now, using hypothesis testing, the null hypothesis is that the true mean is 100g, and the alternative hypothesis is that it is not 100g. With a significance level of 5%, we find that 100 lies outside of the confidence interval $99.1 \leq \mu \leq 99.8$. Therefore, we would reject the null hypothesis. This makes sense, as it would be unlikely that the true mean is exactly 100g.

This can be visualised with a normal distribution curve, found using R. First, we define all our parameters and begin plotting the sample distribution.

```
# parameters
mu0 <- 100    # null hypothesis mean
xbar <- 99.45 # sample mean
sd <- 1.27    # standard deviation
n <- 50       # sample size
alpha <- 0.05 # significance level

# calculate spread of sample mean
sigmaX <- sd/sqrt(n)

# graph bell curve
x <- seq(98.5, 101.5, length=1000)
plot(x, dnorm(x, mean=xbar, sd=sigmaX), type="l", lwd=2,
     xlab = "Mean weight (g)\nFigure 1", ylab="Density",
     main = "Sampling Distribution of Sample Mean")
```

Further, we can calculate our critical value using R, previously stated as 1.96.

```
# critical value calculation

zsc <- qnorm(1 – alpha)
```

After this, our confidence interval can be plotted.

```
# 95% confidence interval
lower <- xbar – zsc*sigmaX
upper <- xbar + zsc*sigmaX
abline(v=c(lower, upper), col="blue", lty=3, lwd=2)

# sample mean and null hypothesis
abline(v=xbar, col="darkgreen", lwd=2) # sample mean
abline(v=mu0, col="red", lwd=2)        # null hypothesis (100g)
```

This generates a bell curve as shown below:

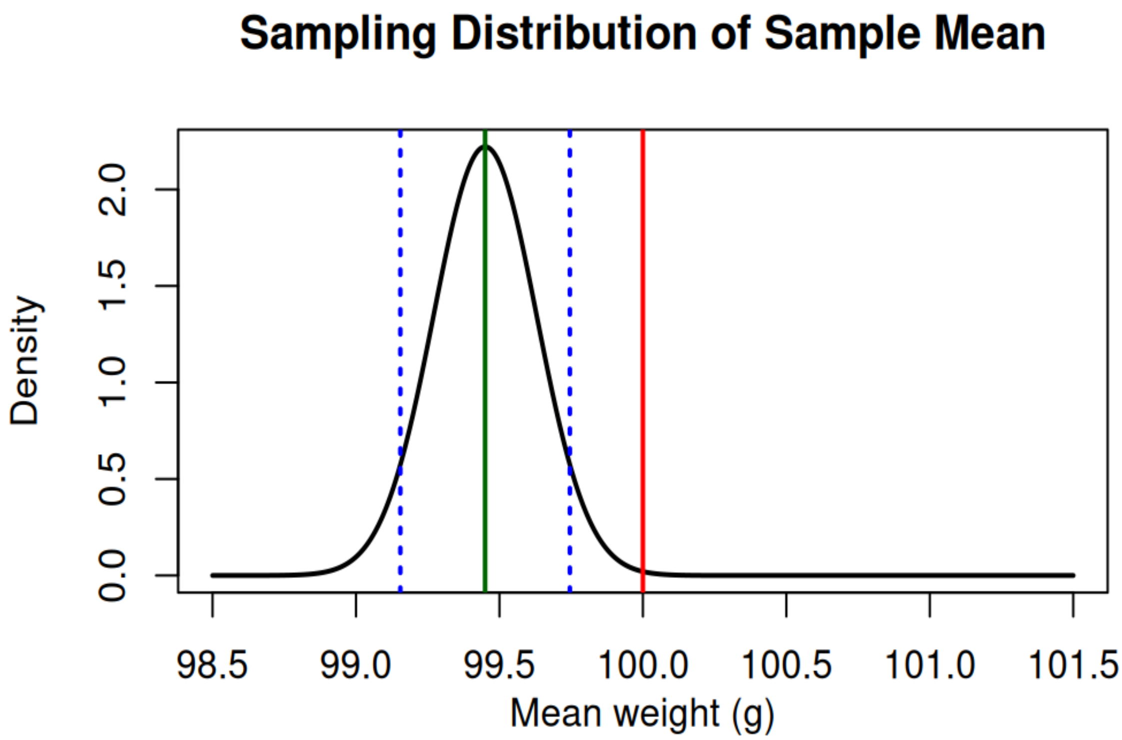

Figure 1

The blue represents the confidence interval, and the red line represents our predicted mean, the null hypothesis. Here, it is also clear that the null hypothesis of the mean being 100g must be rejected. Thus, R can be used to confirm such results.

Now, while the confidence interval provides a range of plausible values for the true mean, the decision to reject or accept the null hypothesis is not certainly correct. This uncertainty leads to type-I and type-II errors, which quantify the risks associated with hypothesis testing.

Definition 3: (Power)
A type-I error occurs when a true hypothesis is incorrectly rejected. A type-II error occurs when a false hypothesis is incorrectly accepted.

In short, a type-I error is the case in which we reject a statement that turns out to be true, whereas a type-II error is failing to reject a statement that is false. They can be better understood as a false positive and a false negative respectively. The table below summarises this:

	H_o is true	H_o is false
H_o is accepted	Correctly concluded	type-II error
H_o is rejected	type-I error	Correctly concluded

Proposition 3:
Increasing the likelihood of type-I errors decreases the likelihood of type-II errors.

Typically, by increasing the chance of a false positive (which is our significance level when hypothesis testing), we decrease the chance of a false negative. This increases the power of a test.

A higher power means that we are more likely to detect real anomalies from a null hypothesis. This is useful in the case that incorrectly concluding a positive proves dangerous and risky. The choice of our significance level is important in order to balance these consequences accordingly. In quality control, as aforementioned, as well as in medicine and science, understanding this balance helps increase the reliability of tests and reduce excessive risk.

Thus, hypothesis testing can be strengthened by both confidence intervals and power tests. These tests are particularly useful and applicable in real-world situations, where high-risk situations rely on this data analysis. These tests can be used to make more reliable and informed conclusions.

Declaration: This mini project is my own work. It may be based on my lecture notes, but I have written it in my own words, and I have not taken examples or explanations from any book, webpage or other source. I have not used ChatGPT or any other AI in writing this project.