Hypothesis Testing

Coin flip

- Toss a 10p coin ten times in a row. Make a note of what you get each time.
- What probability are we expecting to find here?
- What would make us think that the coin is biased?
- What would make us almost certain that the coin is biased?

Statistics and Probability

- Statistical analysis considers the probability of an event being due to chance.
- We can never be 100% certain that something is going to work. For example, saying that one medical treatment is better than another.
- We can use probability to say mathematically how sure we are that a result is true.

Hypothesis Testing

- A hypothesis is a prediction about data.
- The **null hypothesis** ($\boldsymbol{H_0}$) is a statement that indicates nothing unusual is happening i.e. the status quo, for example 'this coin produces exactly 50% heads'
- The **alternative hypothesis** ($\boldsymbol{H_1}$) is a statement about the data that might be true, for example 'this coin produces more than 50% heads'.
- Rather than try and prove the alternative hypothesis, which is generally not possible, the process is instead to show the null hypothesis is unlikely.

Hypothesis test

- We first start off by stating the null and alternative hypotheses.
- For example, we test the effect of two painkillers on blood glucose levels.

Analgesic A	44	51	52	55	60	62	66	68	69	71	71	76	82	91	108
Analgesic B	52	64	68	77	79	83	84	88	95	97	107	116			

Table 1: Blood glucose levels after administration of two analgesics

- Null hypothesis: There is no difference in the blood glucose levels between the two groups.
- Alternative Hypothesis: There is some difference in the blood glucose levels between the two groups.

Hypothesis test

- We then check the normality of our distribution.

```
Descriptive Statistics: Analgesic A, Analgesic B

Variable      N   Mean  StDev  Minimum     Q1  Median     Q3  Maximum
Analgesic A  15  68.40  16.47    44.00  55.00   68.00  76.00   108.00
Analgesic B  12  84.17  18.12    52.00  70.25   83.50  96.50   116.00
```

Hypothesis test

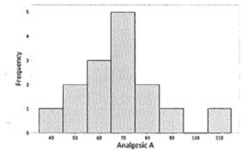

Figure 23: Distribution of blood glucose levels in patients receiving analgesic A

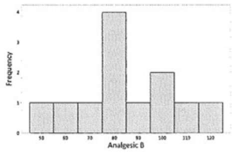

Figure 24: Distribution of blood glucose levels in patients receiving analgesic B

Hypothesis test

- Could the difference in mean values be just down to chance?
- We can check this by working out the probability that the variation in mean values is not just down to chance.
- The number we look at is known as the p-value.
- The conventional cut-off for significance is 5% or less. If the p-value is 5% (0.05) or less, there is evidence to suggest that the null hypothesis is false and the alternative hypothesis is of interest.
- https://www.youtube.com/watch?v=-MKT3yLDkqk

T-Test

- The t test (also called Student's t-Test) compares two averages (means) and tells you if they are different from each other. The t-test also tells you how significant the differences are; In other words it lets you know if those differences could have happened by chance.

Two sample t-test (Independent or unpaired)

- If we are examining the difference between the mean values of two normally distributed data sets, we perform a **two sample t-test**. We must also ensure that people are randomly allocated into the two groups.

```
Two-sample T for Analgesic A vs Analgesic B

             N  Mean  StDev  SE Mean
Analgesic A  15  68.4   16.5      4.3
Analgesic B  12  84.2   18.1      5.2

Difference =  (Analgesic A) -  (Analgesic B)
Estimate for difference:  -15.77
95% CI for difference:  (-29.75, -1.78)
T-Test of difference = 0 (vs ): T-Value = -2.34  P-Value = 0.029  DF = 22
```

One Sample t – test (Paired)

- We use a paired t-test if we are recording the data from the same group of individuals twice.
- E.g. testing their heart rate and then testing their heart rate 3 minutes later.
- There tends to be less variability in paired t-tests as we are carrying out the observations on the same individuals reducing natural variability.

One tailed and two tailed t-tests

- A one tailed t-test is a test where you are only looking at the significance in difference of the means in one direction.
- E.g. comparing two sets of heart rates and seeing if one is lower than the other.
- A two tailed t-test looks at both directions, e.g. comparing two sets of heart rates and seeing if there is a difference between the two going in either direction (higher or lower).

Confidence Interval

- A sample mean is only an estimate of the population mean.
- A plausible range of values for the true mean can be developed. This is known as the confidence interval. Usually a confidence interval is quoted at 95% certainty. For example, looking at the example from earlier:

```
95% CI for difference:  (-29.75, -1.78)
```

- If the confidence interval contained zero, this means that would be no significant difference between the two groups.
- If the C.I doesn't contain zero, it will correspond to a p-value less than 0.05

Z - test

- A z – test is another type of hypothesis test. We can use it when we are working with categorical data and want to test if two populations have the same proportions.
- For example:
- The data shown in Table 1 is taken from the work of Katkiciet al. and shows results of their survey of defence wounds observed during the post-mortem examination of 195 Turkish victims of all forms of stabbing. Of the 195 victims, 162 were male and 33 female. It is of interest to determine whether there is any evidence of a difference between the behaviour of males and females during the course of a fatal stabbing (are they trying to defend themselves whilst being stabbed?).

Gender	Present	Absent	Total
Male	57	105	162
Female	18	15	33
Total	75	120	195

```
    Test and CI for Two Proportions

Sample   X    N  Sample p
1       57  162  0.351852
2       18   33  0.545455

Difference = p (1) - p (2)
Estimate for difference:  -0.193603
95% CI for difference:  (-0.378722, -0.00848332)
Test for difference = 0 (vs  0):  Z = -2.05  P-Value = 0.040
```

- Interpret the 95% confidence interval in the context of the problem.
- Explain how the interpretation of the confidence interval agrees with the results of the hypothesis test.

- We are 95% confident that the true difference in population proportion lies between 0.00848332 and 0.378722. We can be 95% sure that the proportion of women who defended themselves was at least .085% more than men and at most 37.9%.
- The p – value is < 0.05 and the confidence interval does not contain zero. Therefore there is evidence to suggest that there is significant difference between the two proportions.