Question 2

1. Accuracy of Prediction

We often want to **predict** the value of a variate (the *response* variate) given the values of one or more *explanatory* variates.

We build a *response model* that encodes how the prediction is to be carried out:

$$y = \mu(x) + error$$

- The explanatory variates x = $(x_1, ..., x_p)$ are used to explain or predict the values of the response.
- We use our observed data to estimate the function $\mu(x)$, yielding the **predictor function** $\hat{\mu}(x)$.
- This predictor function $\hat{\mu}(x)$ is then used to predict y at any given value.

One question that we would want to ask ourselves if how do we know if *our predictions are any good*?

- This is where Accuracy of Prediction comes into play. To measure the accuracy of a model we measure the **inaccuracy** of the model's predictions.
- One measure of inaccuracy over the population $\mathcal{P}$ (of size N) is the **average prediction squared error**(APSE)

Why is it important?

Accuracy of prediction is extremely important because of two main reasons:

- Firstly because prediction is important to find out the value of a variate and accuracy of prediction helps in understanding how good is our prediction
- And secondly because measures of inaccuracy such as APSE quantifies the distance between true and predicted values. This helps us to choose between competing models.

2. Significance Test

We use significance test to provide **numerical evidence** in *favour of/or against* the notion that two *sub-populations are similar* to a randomly mixed sub-population. The steps required to gather such evidence include:

1. Stating the null hypothesis: H_0
2. Constructing a measure of discrepancy $D = D(\mathcal{P}_1, \mathcal{P}_2)$ where large values indicate **evidence against the null hypothesis**
3. Calculating the **observed discrepancy** $d_{obs} = D(\mathcal{P}_1, \mathcal{P}_2)$
4. Shuffing the sub-populations M times and calculating the observed $p - values$

Why is it important?

We use Significance test to quantify, *numerically*, how unusual the difference between $a(\mathcal{P}_1)$ and $a(\mathcal{P}_2)$ is relative to the randomly mixed sub-populations. If the two sub-populations are similar/different we want

to provide evidence in favour of/against the notion that the two sub-populations are similar to a randomly mixed sub-population.

The significance test helps to do gather such evidence, hence it is of vital importance. Statistical significance helps you determine the level of risk you're willing to accept, and you can balance the desire for accuracy with the resources you have available.

3. Bootstrap Method

We use the boostrap to **estimate** quantities of interest from the *sampling distribution*.

The distribution of any attribute over the bootstrap samples $\mathcal{S}_i^*$ from $\mathcal{P}_i^*$ is a bootstrap estimate of the distribution of the same attribute over all possible samples $\mathcal{S}_i$ from $\mathcal{P}$

The bootstrap distribution gives a sense of how an attribute varies. We use can the bootstrap to estimate the *standard error, standard deviation of the sampling distribution, for any attribute*. We can also estimate the *sampling bias*.

Why is it important?

With a single sample, we can construct an estimate of the sampling distribution of an attribute that *does not* rely on any assumptions about the *form of the attribute*. Bootstrap works with any attribute. No matter how complicated it is, we simply follow the same standard procedure in terms of Bootstrap.

4. Power Transformations

For any variate y, it is sometimes helpful to re-express the values in a non-linear way via a transformation $T(y)$ so that on the transformed scale location/scale attributes are easier to define, to understand, or simply to determine. This is where we use power transformations.

A commonly used transformation when $y > 0$ is the family of *power transformations* which is indexed by a power α. The general form is:

$$T_\alpha(y) = \begin{cases} y^\alpha & \alpha \neq 0 \\ \log(y) & \alpha = 0 \end{cases}$$

These transformations are monotonic, in the sense that

$$y_u < y_v \longleftrightarrow T_\alpha(y_u) < T_\alpha(y_v)$$

i.e they preserve the order of the variate values with units u and v. They however, change the relative positions of the variate values.

How to pick α?

Two different effects of transformation are often of interest:

- First, producing a more symmetric looking histogram
- Second, producing roughly linear scatter-plots

There are 2 rules to picking power transformation:

- Bump Rule #1: Making historgrams more symmetric
- Bump Rule #2: Straightening Scatter-plots

Why is it important?

Power Transformation are extremely important as they can re-express the values in a non-linear way so that on the transformed scale location/scale attributes are easier to define, to understand, or simply to determine.