Please note that this is an advanced session. Strong working knowleged of Tableau and a knowledge of fundamental statistics is necessary.

Let us know if you need assistance with developing your lab. We are always happy to help!

We will begin by setting up the environment as this will be a necessary step for the remainder of the Lab.

Anaconda will be used to run TabPy in a **Virtual Environment**. While it is possible, and preferable, to run TabPy in a Server we are using a Virtual Environment to show additional functionality and since we do not have a TabPy server set up for this session. To run TabPy in a Virtual Environment, please take the following steps: (1) In the Search bar, search for **Anadonda Prompt** (2) Select the Anaconda Prompt App (3) Type the following: **conda activate "Adv_Stats_tabpy"** (4) Type the following: **tapby** When Prompted, enter y (note that we will not be publishing so the user name and password requirements will not be necessary) For full info on a setting up a TabPy Server or TabPy in a virtual environment, go to our github page: https://tableau.github.io/TabPy/docs/tabpy-virtualenv.html

Now that we have the tabpy environment ready to go, we will be able to pass commands from Tableau to our Python virtual environment and get results of our models visualized in Tableau. But first, some work on the Tableau side: (1) Open Tableau Desktop (2) go to Help and select 'Manage Analytics Extensions' (3) Choose Python (4) For Server type "locahost" (not quotation marks) and for Port type 9004 Note that 9004 is the port our TabPy instance is on in Anaconda

Now, onto the modelling!

For the first exercise we will be using the SPRT_TC24 dataset SPRT_TC24.txt

(1) Ensure it is saved to Desktop and then open Tableau and Load as a Text file. (2) Go to **Help**, then **Manage Analytics Extensions** then **TabPy** Hostname is **localhost** Port is **9004** (3) From below Files (lower left) in Data Source, drag New Table Extension and create a Logical Join to the SPRT_TC24 data you just loaded (4) Drag and drop the SPRT_TC24.txt Table in the Table Extension (to the Left of the code block which is open on the Right) (5) Now for the code:

Import your Libraries

```
import pandas as pd

import numpy as np

import statistics as st
```

import sprt

Create our Dataframe

df = pd.DataFrame(_arg1) #Note that _arg1 references the table that we dragged into the Table Extension

Set our a and b thresholds

a = 0.05 (*what does this mean in terms of our hypotheses?) b = 0.01 (*what happens if we just set a and b to equal? Try it later and find out --> you will not break antyhing but your interpretation may change.*)

Now detemine the elements we want to compare for our hypotheses

h0 = st.mean(df['Cumulative Control']) #this is the threshold we are looking to be below to fail to reject the null h1 = st.mean(df['Cumulative Test']) #this is the threshold we are looking to ve above to reject the null values = np.array(df['Cumulative Test']) #this is the array whose likelihood we are tracking

Formal Stuff that is Probably Overkill: *In a Poisson distribution, the mean of the distribution is represented by λ and e aka Euler's number is constant (approximately equal to 2.71828). Then, the Poisson probability is: P(x, λ) =(e− λ λx)/x! In Poisson distribution, the mean is represented as E(X) = λ.*

Now we run our SPRT test = sprt.SPRTPoisson(h0=h0, h1=h1, alpha=a, beta=b, values = values) *#Note that this is for a Poisson distribution. The dataset also includes a binomial distribution that we can track with SPRTBinomial. Try this later to see if you come to any different conclusions.*

lower = {"yl": test.*yl} #_Gets the lower threshold* Note that this looks italicized but it should be test._yl

upper = {"yu": test.*yu} #_Gets the upper threshold* same, should be test._yu

cumulative = {"cum_val": test.cum_values} *#Gets the cumulative log likelihoods that we are concerned with*

df1 = pd.DataFrame(lower)

df2 = pd.DataFrame(upper)

df3 = pd.DataFrame(cumulative)

df5 = df1.combine_first(df2)

df6 = df5.combine_first(df3)

df7 = df.combine_first(df6)

return df7.to_dict(orient = 'list')

Joine the Table Extension and the SPRT_TC24 data by **Observation Count** (*Check… do you need to change the number format?*) **We have our results, now what?**

Drag **cum_val** to rows and **Date** (from the initial SPRT_TC24 data) to Columns. Switch cum_val to AVG() switch Date to DAY() (Continuous not Discrete)

We are now looking at our Cumulative Log Likelihoods over time, but when should we have stopped the test?

Let's see… Drag **yu** to Details. Switch it to AVG().

Right Click the Y Axis (the Column axis) and **Add Reference Line** Make the line reference AVG(yu). *We can now see the point at which the cumulative log likelihood crossed the threshold for rejecting the null. But what date was it, and how could be notified when it happened?*

We will need to create a calculated field to identify the exact date that the threshold is crossed.

IF MIN({ FIXED [Date]: AVG([cum_val])}) >= MIN({ FIXED [Date]: AVG([yu])}) THEN MIN([Date]) ELSE NULL END *#we are specifying the earliest (MIN) date that the cumulative LLR becomes greater that the threshold. If it does exceed it, then we know the time to stop. _How could you be notified of this in Tableau?*

Thank you for your time and attention!