Text

Description automatically generated

Theo J Symonds

Python Test

SC-FACT

Logo, company name

Description automatically generatedLogo, company name

Description automatically generatedA close up of a logo

Description automatically generated with low confidence

**Step 1 Global Template Excel**

File: Python-Test\_Step1\_Theo.py

Commands for terminal: cd “C:\Python Test” > python -m pip install > python Python-Test\_Step1\_Theo.py

**Step 2 Descriptive Statistics**

1. **Data**

The python print-out code for these descriptive statistics can be found on lines #55-65 in Python-Test\_Step2and3\_Theo.py.

1. **Measures of center by region**
   1. *Mean MOS by region*

Shire: 4.15

Polombia: 2.99

Sangala: 3.56

Urkesh: 3.95

Molvania: 8.69

Westerose: 3.51

These are average monthly stock remaining for all products by respective region. These statistics suggest that these regions have 3 months or more of monthly stocks for all products. However, we should look at the median to see how skewed the data is.

* 1. *Median MOS by region*

Shire: 1.026

Polombia: 1.005

Sangala: 1.083

Turgistan: 1.065

Urkesh: 1.035

Nuku’la Atoll: 0.99

Molvania: 0.93

Hogwarts: 1.026

Falsenthurm: 1.003

Westerose: 1.040

These median statistics tell us that half of the MOS observations for each respective region fall below the median value, and the other half of the MOS observations fall above the median value. Regions with MOS of 0.99 and below mean that half of their medical products have less than a month of stock left.

1. **Measures of Spread**
   1. *Standard deviation*

Shire: 25.84

Polombia:

Sangala: 3.56

Urkesh: 3.95

Molvania: 8.69

Westerose: 3.51

* 1. *Outliers*

Given the mean is far higher than the median in all regions, we know that the distribution of MOS data is heavily skewed right, meaning that we have a lot of outliers on the right-side of the distribution (ie. higher MOS). We can remove all MOS outliers, which would be outside of the 3 standard deviation ranges above the mean.

**Step 3 Inferential Statistics**

The simplest set of hypotheses we could make involve average MOS depending on the region. Say we want to explore if region x has better monthly supplies than region y.

Looking at the mean MOS by region, it would appear that Molvania has excellent stock levels, whereas Polombia has on average a much shorter months of supply in comparison. But inferential statistics would allow us to explore deeper to see if this is truly the case within each regional data set, or if the mean values are deceiving, which is a strong possibility given the heavy right skew in each region.

To set this up, I would conduct a z-test. What this does, is it analyzes the means of two different samples (regions’ MOS), to see if the samples are representative of the population (countrywide MOS) – or find out if they are so statistically significant in their differences to where one region’s MOS is not indicative of the the population MOS. In other words, this will allow us to see if the different means between two regions is statistically significant, or if there is actually no real difference within their distributions (even though the mean might nominally suggest a difference).

Therefore, my hypothesis for Molvania and Polombia would be:

Null Hypothesis: The average MOS in Molvania is not significantly different than the average MOS in Polombia

Alt Hypothesis: The average MOS in Molvania is significantly better than the average MOS in Polombia.

To determine significance in the difference of MOS means, we set a stringent level of significance at 1%. The result of our z-test will provide us with a z-value and corresponding p-value. If the p-value falls above our set level of significance (in our case 1%, or 0.01), then we conclude that the null hypothesis is true and that there is no statistically significant difference between. If the p-value falls below our set level of significance (p<0.01), then we reject the null hypothesis, and accept that Molvania has significantly higher MOS than Polombia.