**Machine learning in python notes**

**with codes and concepts**

**Index**

**Chapter 1**

**Hypothesis testing**

**Terms:**

H0 – Null hypothesis

H1 – Alternate hypothesis

**Tests:**

1. **One T test**

This is used when only 1 column is available and number of samples are 30 or so.

It is used to compare a specified value to the mean of the samples.

T0 = (mean – specified val)/(SD x sqrt(n))

Python code / terms:

Uses function: ttest\_1samp(column, specified value)

Function in: scipy.stats

Returns: t0 and P value. If P >= 0.05, H0 is accepted

1. **Two T test**

This is used when we have 2 independent data columns

Compare the means of both the columns

H0 = means are same

H1 = means are different

T0 = (m1 – m2)/(sqrt((SD1^2/n1)+(SD2^2/n2)))

Python terms:

Uses function: ttest\_ind(col1, col2)

Function in: scipy.stats

Returns: t0 and P value. P>=0.05 – m1=m2

1. **Paired T test**

Used when the data is related in some way. Calculate the difference between both the data and form a new column of difference. Use 1 sample ttest on this new column

H0 => m1-m2=0

H1=> m1-m2!=0

T0 = m1-m2/(SD/sqrt(n))

Python terms:

Uses function: ttest\_rel(col1, col2)

Function in: scipy.stats

Returns t0 and P val