In [2]:

```
%run 00_basic.ipynb
```

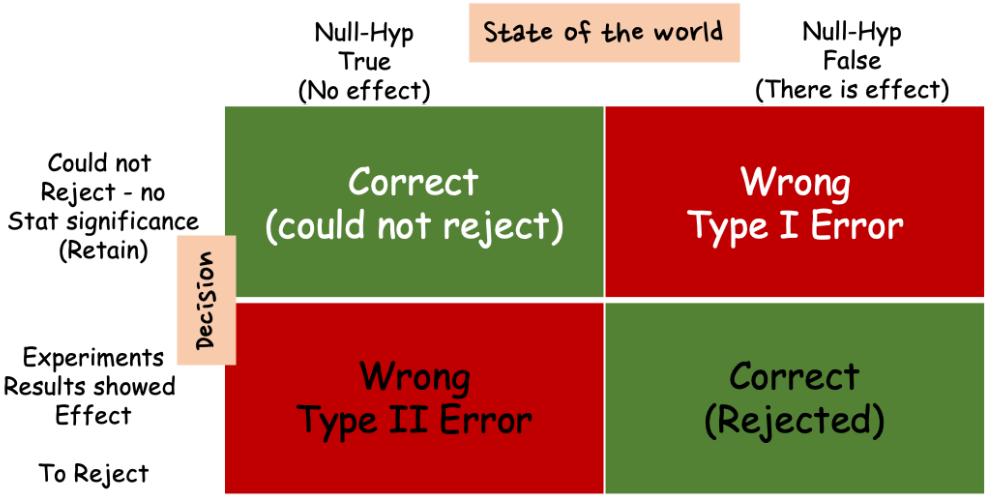
Preliminaries to recall

Type 1 and Type 2 Errors

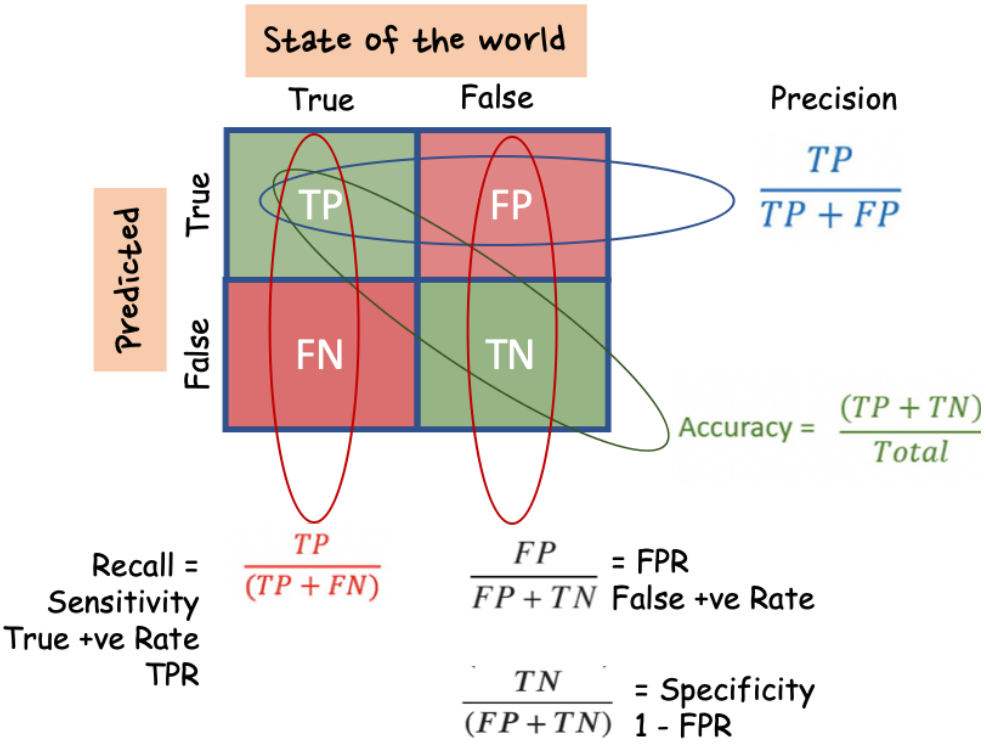
In statistics, when we set up experiments, we state the hypothesis h_0 (AKA null hypothesis) as negative statement that inidicates no relationship to observations and the phenomenon; the reason for stating it as "Null" hypothesus is bit philophical - we can rather prove the absence of the phenomenon. An alternate hypothesis h_1 (which is stated or assumed to be as the opposite of null hypothesis may be true if the experimental results (or evidence) show no significant difference).

When experiments are run, the results are used to evaluate the state of the world. If the experiments are accurate we should be making the correct decisions. Otherwise we commit error:

Type I error is when the null hypothesis is **True**, but it is rejected (False Positive) *Type II* error is when the null hypothesis is **False**, but it could not be rejected (False Negative)



Precision Recall Sensitivity



Precision Among the results predicted as +ve, how many are actually +ve; how precise was the +ve predictions.

$$Precision(+) = \frac{TP}{(TP+FP)}$$

Accuracy is ratio all predictions were correct.

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)}$$

Recall - how many of real true +ve values were recalled.

$$Sensitivity = Recall(+) = \frac{TP}{(TP+FN)}$$

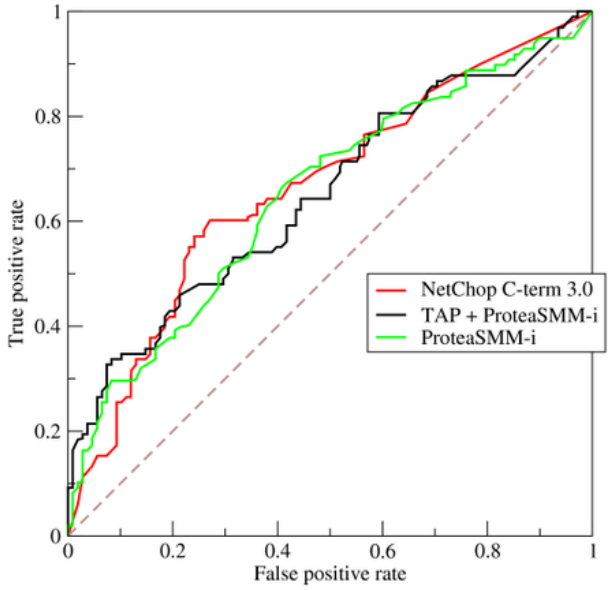
$$Specificity = 1 - FPR = \frac{TN}{(FP+TN)}$$

F1 Score is the harmonic mean of precision and recall.

$$F1_{score} = \frac{2 * precision * recall}{precision + recall}$$

AOC-Area Under the Curve, (ROC- Receiver Operating Characteristic) curve

AOC curve is the plot of TPR Vs FPR for various threshhold settings of parameter of prediction of algorithm.



(Reference Wikipedia)

Ideally we want to have zero False positive rate (FPR) that implies 100% TPR.

AUC of 0.5 says that model is as bad as a random picker.

Hypothesis Testing

Hypothesis testing (or experiement) is done is test the effect of some treatment.

For example: One could test the effect of sleeping pill. Now, if we can make observational study on people using the drug and test if those people are sleeping better. This type os study can show correlation but still it is unclear if the drug is **causing it**.

As we described in the experiemental section, we choose two groups

1. Control
2. Treatment group But the first step is to state the hypothesis

Hypothesis is stated as Null Hypothesis as showing no effect. Null hypothesis H_0.

If we disprove the *Null* hypothesis by showing the probablity of seeing those effects (by observing population parameters) are very low, then we can reject *Null* hypothesis in favor of Alternative hypothesis H_a that states the drug has an effect.

Suppose we know the population mean μ and standard deviation σ;
If we take sample of size n, give them treatment and measure their average $\bar{x}$ and standard deviation = $\frac{\sigma}{\sqrt{n}}$

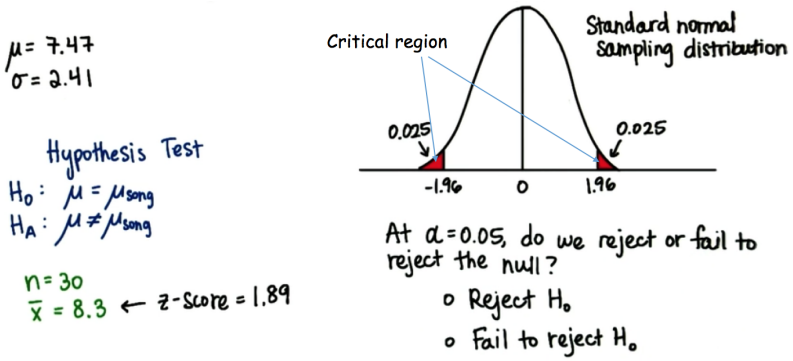
We compute the z-score:

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

and test if the z-score is in the critical region.

Here is a complete example.

Concrete Example:



In this example, the population parameters (taken from udacity stats class):

(The actual experiement details does not matter - it is a hypothetical test to test the engagement of student a class if teacher sings of song :))

We want to test the treatment has an effect (positive or negative); therefore we set the alternate hypothesis is $\mu \neq \mu_{song}$; we choose a α level of 5%

Assume that the populative have the following parameters:

$$\mu = 7.47$$
$$\sigma = 2.41$$

Null Hypothesis: H_0 is: SONG HAS NO EFFECT

Alt Hypothesis: H_a is: SONG HAS Some EFFECT (Notice two tailed test)

By running a hypothesis test on a sample of $n = 30$, we found the sample mean

$$\bar{x} = 8.3$$
$$n = 30$$

Question is - does the song had an effect? First, we compute the $Z - score$ and compare to see if it falls in the critical region.

$$Z = \frac{8.3 - 7.47}{\frac{2.41}{\sqrt{30}}} = 1.89$$

All of the following statements are equivalent

- Z is less than -ve z or greater than +ve z
- the p value falls in the critical region

Now, we see that $Z = 1.89$ is not in the critical region.
The probability of getting this result is $1 - .9706 = 0.0294 > 0.025$
We found 0.9706 by looking in Z-table corresponding to Z-score

Therefore, we cannot reject the null hypothesis; In other words, we got this chance.

$H_0: \mu_{song} = \mu$

$H_A: \mu_{song} \neq \mu$

$\mu = 7.47$

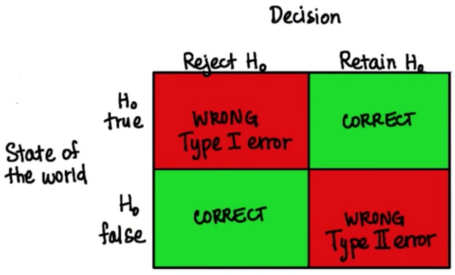
$\sigma = 2.41$

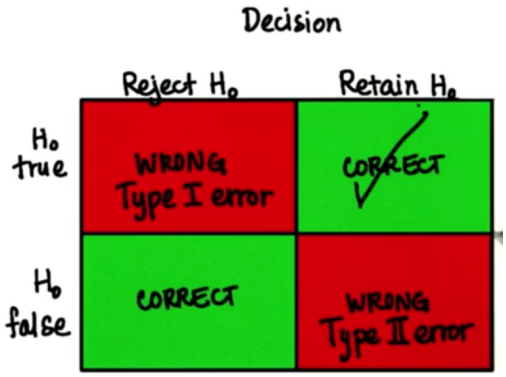
$n = 30 \quad \bar{x} = 8.3$

$\mu_{song} = 7.8$

$\alpha = 0.05$

two-tailed test





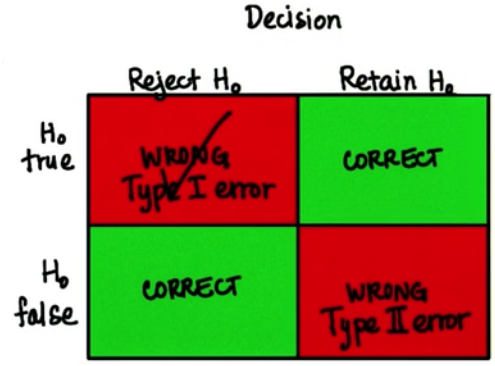
In this example, we got $\bar{x} = 8.3$ with p value was close to 0.025.

Suppose if the true population new $\mu_{song} = 7.8$ then,

$$z = \frac{7.8 - 7.47}{\frac{2.41}{\sqrt{30}}} = .75$$

0.75 is not the critical region, therefore, we correctly retained Null hypothesis;

Although our sample was misleadingly high at 1.89



Now, what if we get the same mean $\bar{x} = 8.3$ with $n = 50$.

Suppose if the true population new $\mu_{song} = 7.8$ then,

$$z = \frac{7.8 - 7.47}{\frac{2.41}{\sqrt{50}}} = .9682$$

0.9682 is not the critical region, therefore, song had no significant effect;

However based on our sample statistic:

$$z = \frac{8.3 - 7.47}{\frac{2.41}{\sqrt{50}}} = 2.43$$

we would reject the Null. making a type I error with probability:

**Type 1 Error* *: H_0 is true - we reject it

**Type 2 Error* *: H_0 is false - we fail to reject it because the sample gave good results