Z test

Two group comparison : Do women have a greater risk of developing tennis elbow?

Null hypothesis, H_0 → $P_{women} : P_{men}$.

In USA.

$\hat{P}_{women} = 0.11$

$\hat{P}_{men} = 0.11$

USA

	Tennis Elbow (TE)	Non-TE	Total
Women	11	89	100
Men	20	90	100
Total	21	179	200

AUSTRALIA

	TE	Non-TE	Total
Women	20	80	100
Men	6	94	100
Total	26	174	200

No enough evidence.

But does this mean, we have enough evidence. No.

In Australia sample, is the difference in the SAMPLE above big enough to infer a difference in the population? $\hat{P} = 0.20$ $\hat{P} = 0.06$

→ use Z test, $n_{women} = 100$ $\hat{P}_{women} = 0.20$ $\hat{P} = 0.13$

$n_{men} = 100$ $\hat{P}_{men} = 0.06$

$$z = \frac{\hat{P}_{women} - \hat{P}_{men}}{\sqrt{\hat{P} \times (1-\hat{P}) \times \left(\frac{1}{n_{women}} + \frac{1}{n_{men}}\right)}}$$

, $z = 2.943$, pvalue $= 0.0032$.

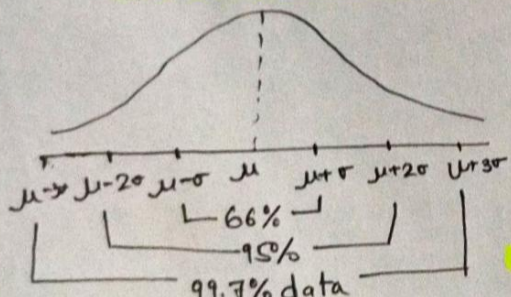
So, z score is ~~accepting based on~~ rejecting null hypothesis as P value is less than 0.05.

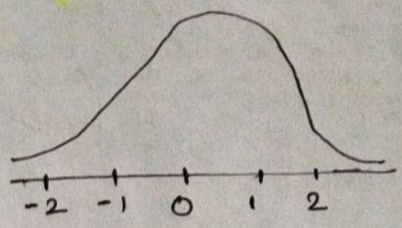
Another method, Chi-Square test, chi-sq $= 8.664$, p-value $= 0.0032$.

So, by seeing $p = 0.0032$, yes there is a significant difference between the men & women.

Z Score

Gaussian / Normal Distribution → ($\mu = 0$, $\sigma = 1$ convert) → Standard Normal Distribution





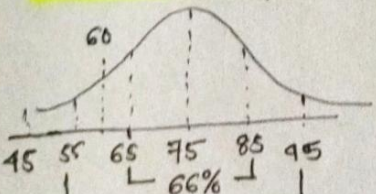
With help of Zscore we will be able to convert this part.

$$Z\text{score} = \frac{x_i - \mu}{\sigma}$$

In feature engineering we use standard normilization which is also a z score.

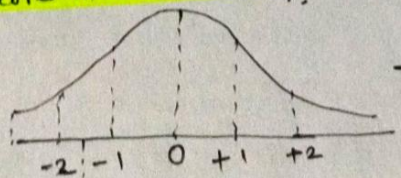
Suppose in a class, avg student marks is $\mu = 75$ and standard deviation $\sigma = 10$.

What is the probability student will score more than 60? $P(x > 60)$



But by this we cannot get exactly at 60.

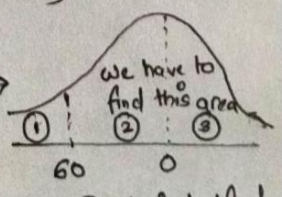
convert to standard normal dist →



60, which is equal to -1.5

So through z score we will find area < -1.5 which is area ① that is around 6% we found the Z table / Z chart.

So, 44% is the probability student will score >60.



For ③, it is half of chart that is 50% of distribution.

Area = ① + ② + ③

100% = 6% + ② + 50%

② = 44%.