

P hacking.

it is a risk when we choose null hypothesis significance testing.

example. jelly bean & acne.

normally \rightarrow P value > 0.05 . \rightarrow alpha cutoff
 \rightarrow fail to reject the null that is 0.05.
jelly bean are not associated with breakout.

\downarrow however. jelly bean has different color.

(maybe there are a color linked with acne).

\rightarrow test 20 colors \rightarrow green has high P value.

However, there is a 5% chance that P value less than 0.05. ($\frac{1}{20}$ chance of getting type II error).

when we have 20 times experiment.

$$0.95^{20} \approx 0.358 = 35.8\%$$

$$1 - 35\% = 65\%$$

means that 65% of time. at least one experiment might show significant falsely.

compare to 5% chance. 65% is higher.

\rightarrow inflated error rate \rightarrow family wise error rate.
(type I).

how to fix? Bonferroni correction.

divide the threshold by # test we are doing.