

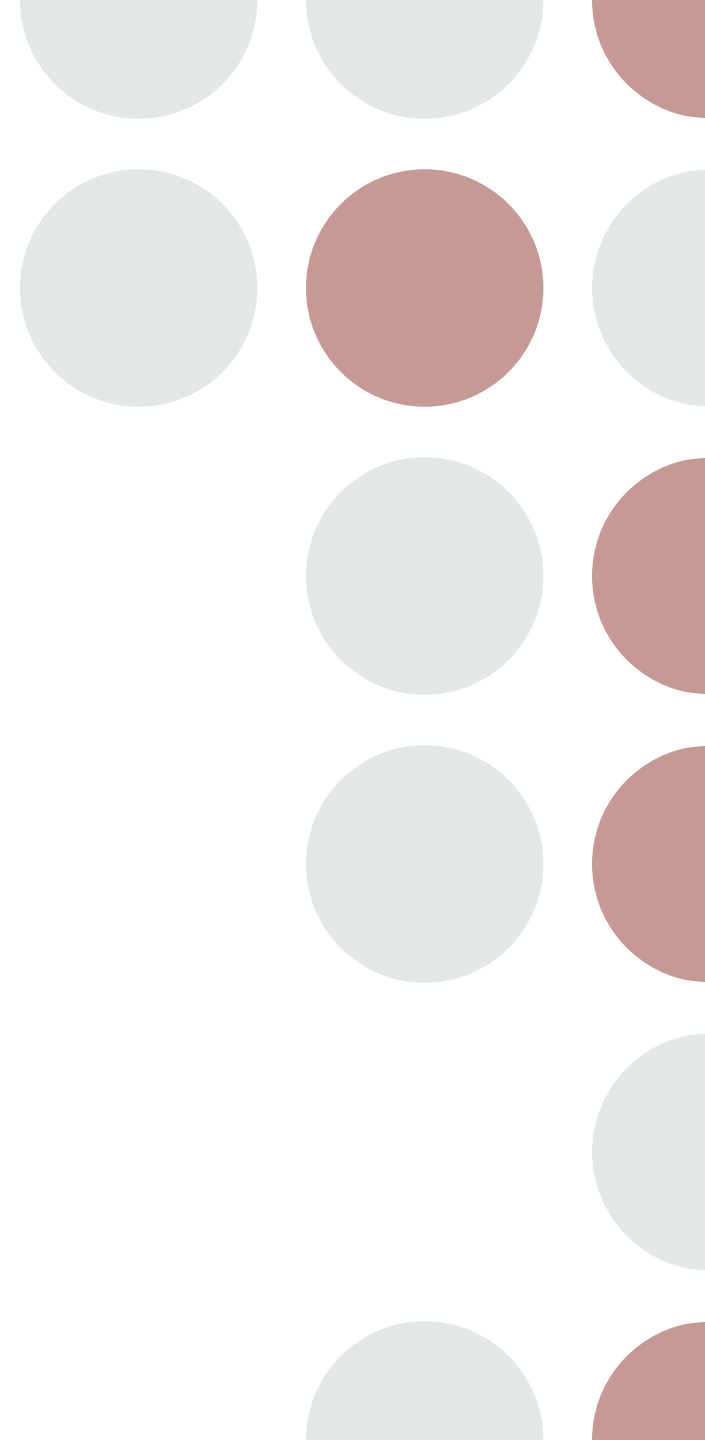


Week 8

*Statistical
Inference*

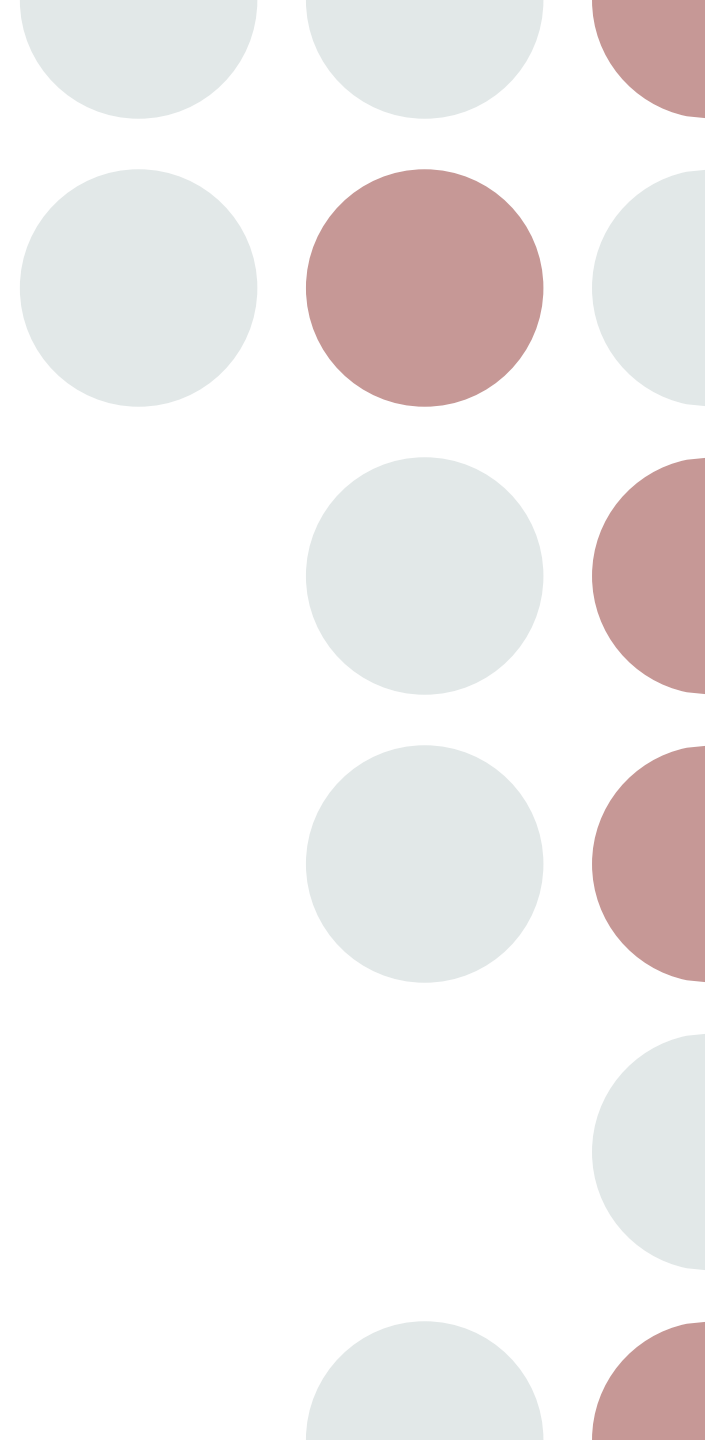
Project

- Set up team member (1 – 3) per team
- Submit your team member



Statistical Inference

- Statistical inference
 - Hypothesis testing
 - Significance test for one sample mean
-

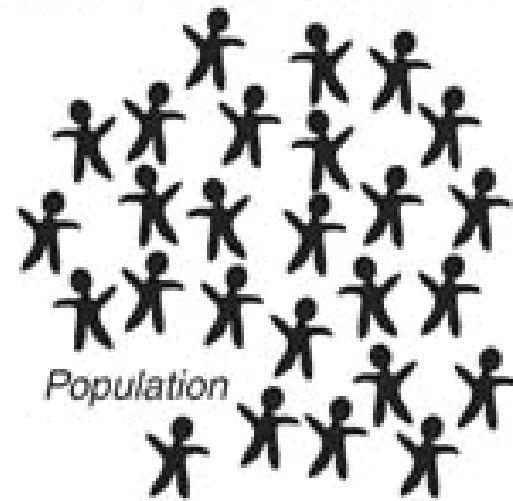


Statistical Inference

- Everyday meaning
 - We infer a conclusion based on **evidence and reasoning**
 - Statistical meaning
 - We infer a property of a **population from a sample**
-

Statistical inference

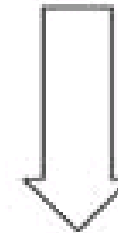
We want to know about these



Parameter μ

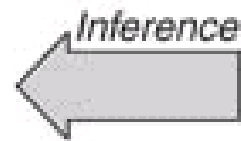
(Population mean)

We have these to work with



\bar{x} Statistic

(Sample mean)

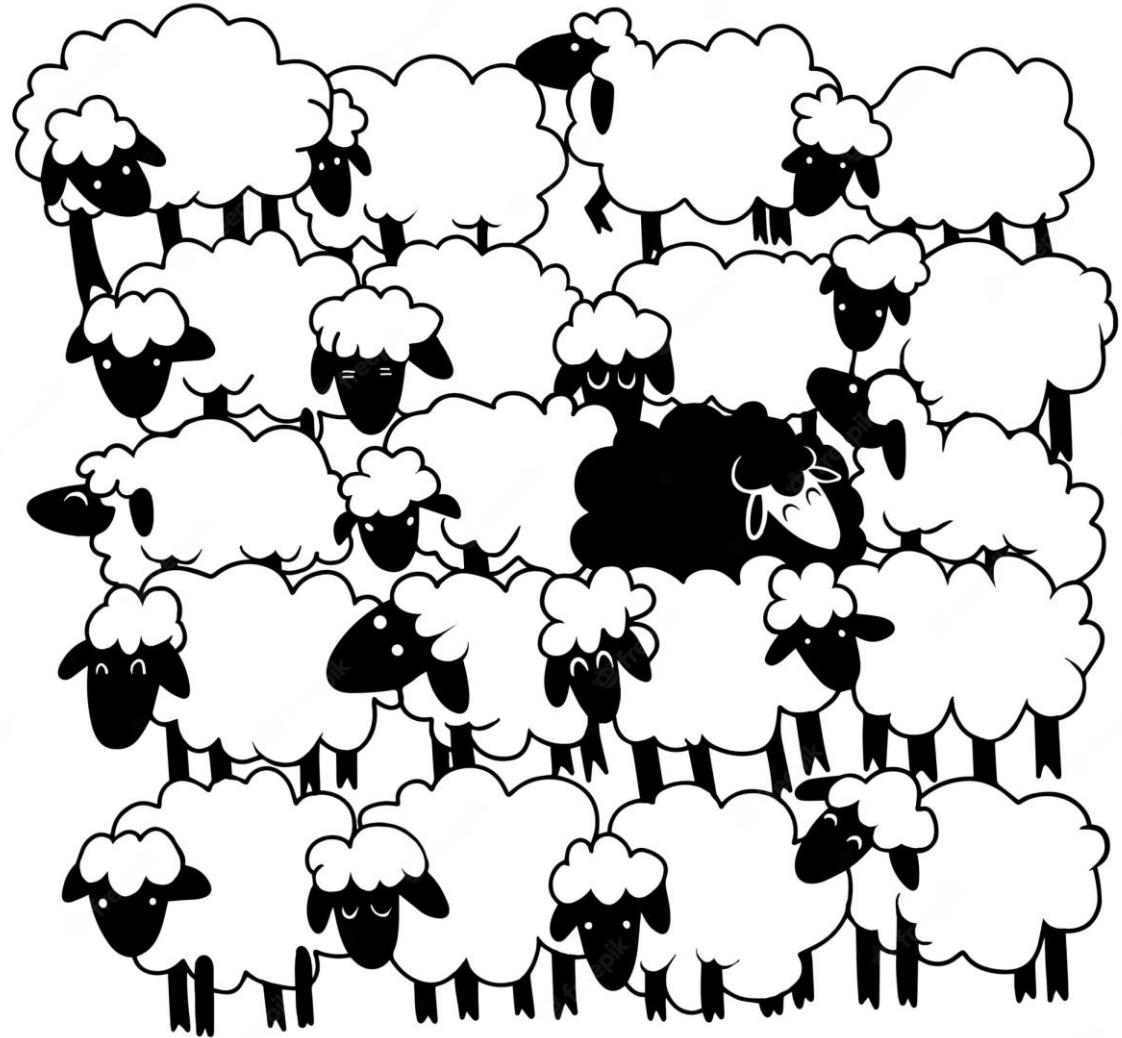


Hypothesis testing

- It is a claim or statement about a parameter of a population
 - It has two components are Null hypothesis (H_0) and Alternative hypothesis (H_a)
 - Null hypothesis – Statement regarding the value(s) of unknown parameter(s). It is always contained an equality
 - Alternative hypothesis – The contradiction statement to the null hypothesis. It is always contained an inequality
-

Hypothesis testing

- There are two type of error
 - Type I error (α)
 - Type II error (β)



Hypothesis testing

State the null and alternative hypothesis

- a) **Climate change:** Thai Meteorological Department state that the average temperature in Bangkok is 30 degree, however a researcher suspect that the temperature in Bangkok over 30 degree

H_0 : The average temperature in Bangkok equal to 30 degree

H_a : The average temperature in Bangkok is not equal to 30 degree

- b) **Covid-19 pandemic:** Ministry of Public Health claimed that there are only 1% of people who infect the covid-19 will have a serious condition. However, World Health Organization (WHO) suspect that the number of serious condition lower than 1%.

H_0 : Only 1% of people who infect the covid-19 has a serious condition

H_a : Less than 1% of people who infect the covid-19 has a serious condition

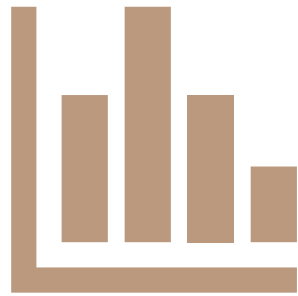
Hypothesis testing examples

State the null and alternative hypothesis

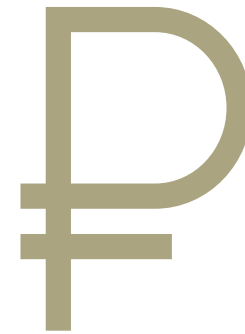
- a) **Food delivery service:** A food delivery service advertises that the average delivery time not higher than 30 minutes. A random sample of times for 20 deliveries were recorded.

 - b) **University registrar:** The university registrar claimed that 80% of Bachelor students will graduated in 4 years, however head department think that the number of bachelor student which graduated in 4 years higher than 80%
-

Significance test



Test statistic



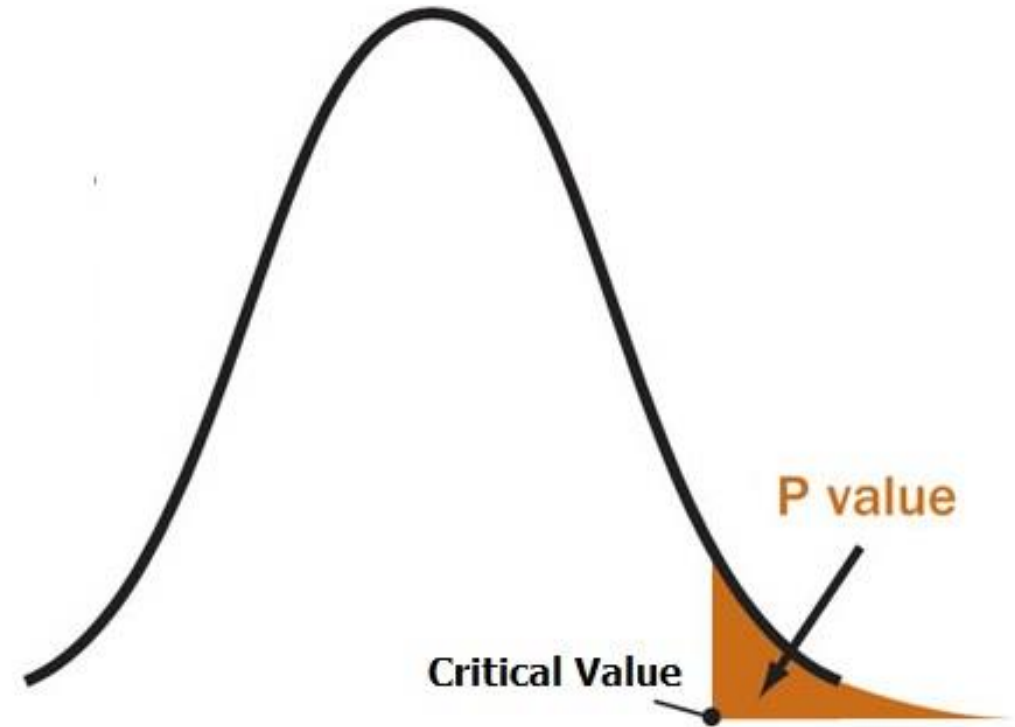
P – value

Significance test – Test statistic

- One sample Z-Test (Known population variance, large sample)
 - One sample T-Test (Unknown population variance, small sample)
 - One population proportion
 - Independent sample T-Test (2 samples T-Test)
-

P-value

- P-value is the probability to validate a hypothesis against observed data
 - If **p-value < significance level**, then **reject Null Hypothesis**
 - If **p-value > significance level**, then **fail to reject Null Hypothesis**
-



P-value example

- Assume the researcher concluded the result of p-value for customer satisfaction at 0.07

- Make a conclusion at 5% significance level

Fail to reject Null Hypothesis because $p\text{-value} > \text{significance level}$ ($0.07 > 0.05$)

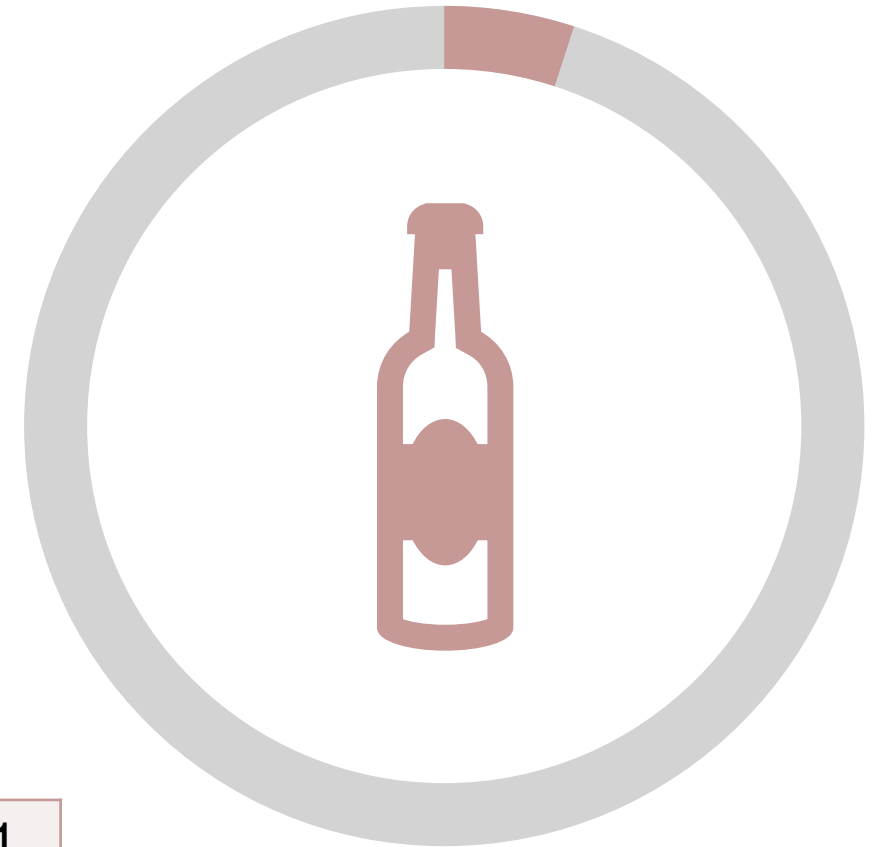
- Make a conclusion at 10% significance level

Reject Null Hypothesis because $p\text{-value} < \text{significance level}$ ($0.07 < 0.10$)

One sample Z-Test

- **Advertisement case:** The soft-drink industry requires that models in soft-drink ads must appear to be at most 30 years. Also recall that a sample of 20 people is randomly selected at a shopping mall. Use these provided data to test the hypothesis at 5% significance level.

23	26	32	33	21	29	28	36	34	41
21	30	29	37	16	25	24	23	38	19



One sample Z-Test by Python

```
from statsmodel.stats.weightstats import ztest
data = [23, 26, .... , 38, 19]
alpha = 0.05
zscore, pvalue = ztest(data, value = 30)
print("Z statistics: ", zscore)
print("P value: ", pvalue)
if( pvalue < alpha):
    print("Reject Null Hypothesis")
else:
    print("Fail to reject Null Hypothesis")
```

One sample T-Test

- **Climate change:** Thai Meteorological Department state that the average temperature in Bangkok is 30 degree. The researcher collect data for 12 months during the year of 2021.

Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
26.5	28.1	29.4	30.5	30.1	29.4	29	28.8	28.6	28.3	27.5	26.1

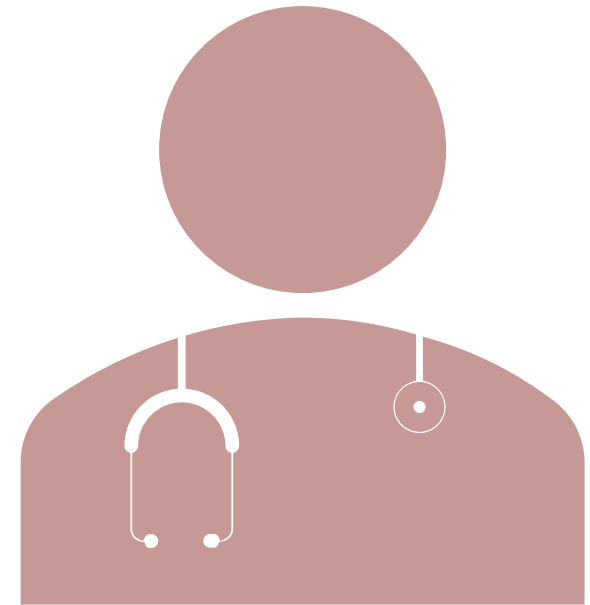
Test at 5% significance level

One sample T-Test by Python

```
from scipy.stats import ttest_1samp
weather = [26.5, 28.1, .... , 27.5, 26.1]
pop_mean = 30
alpha = 0.05
tscore, pvalue = ttest_1samp(weather, pop_mean)
print("T statistics: ", tscore)
print("P value: ", pvalue)
if( pvalue < alpha):
    print("Reject Null Hypothesis")
else:
    print("Fail to reject Null Hypothesis")
```

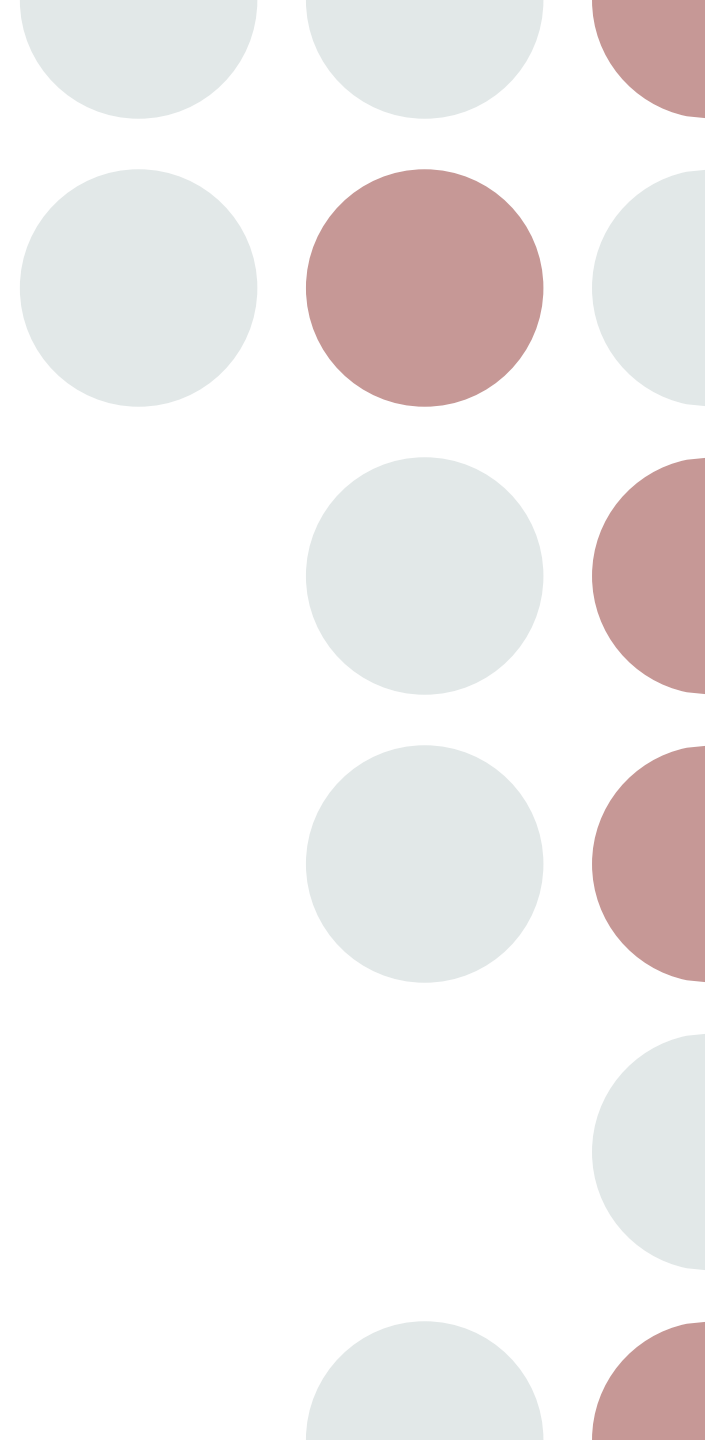
One population proportion

- **Covid-19 pandemic:** Ministry of Public Health indicated that only 1% of patient who infect covid-19 will have a serious condition. However, the researcher collect 1000 patients and it found that 22 patients have a serious condition. Test at 10% significance level



One population proportion by Python

```
from statsmodel.stats.proportion import proportions_ztest
count = 22
sample = 1000
value = 0.01
alpha = 0.1
zscore, pvalue = proportions_ztest(count, sample, value)
print("Z statistics: ", zscore)
print("P value: ", pvalue)
if( pvalue < alpha):
    print("Reject Null Hypothesis")
else:
    print("Fail to reject Null Hypothesis")
```



Independence sample T-Test (2 samples T-Test)

- **Minimart case:** A marketing research would like to compare the average of purchasing bill by two minimart chains – Spar and Hema. The purchasing bill of each minimart chain is provided as the following:

Spar: 54, 48, 39, 99, 48, 71, 82, 37, 63, 43

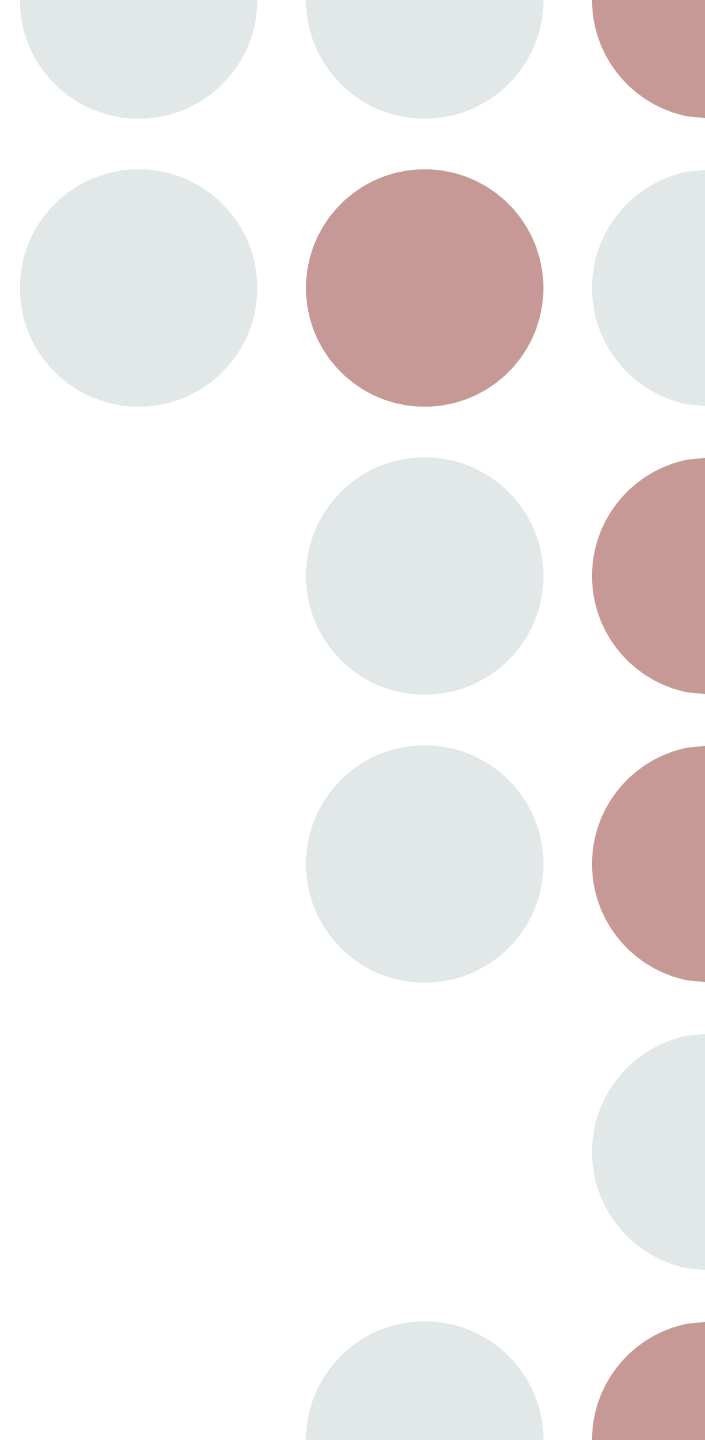
Hema: 67, 23, 50, 79, 45, 55

The researcher believe that the purchasing bill of both minimart are not different. Test the hypothesis at 10% significance level



Independence sample T-Test by Python

```
from scipy.stats import ttest_ind
spar = [54, 48, .... , 63, 43]
Hema = [67, 23, ... , 45, 55]
alpha = 0.1
tscore, pvalue = ttest_ind(spar, hema)
print("T statistics: ", tscore)
print("P value: ", pvalue)
if( pvalue < alpha):
    print("Reject Null Hypothesis")
else:
    print("Fail to reject Null Hypothesis")
```



Practical questions

Q1) **Traffic case:** Thai people spend at least 60 minutes on the road everyday. However, Ministry of transport believe that BTS and MRT can reduce the time spending on the road. Then they decide to collect the data from 10 randomly people and test the hypothesis at 5% significance level.

55	85	90	30	45	65	25	72	103	35
----	----	----	----	----	----	----	----	-----	----

Q2) **Airline case:** An airline industry indicate that only 40% of customer proceed the online check-in. To increase the percentage the airline is consider to use ads to communicate with customer. A random sample of 400 customers and 225 customers check-in online. Use 1% significance level to test the hypothesis
