

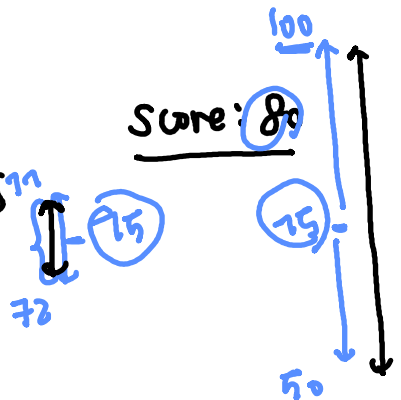
Week 2

1. Specify a claim $\begin{cases} H_0: \\ H_a: \end{cases}$
2. choose the proper test and calculate Statistic
distribution check \rightarrow parametric
nonparametric
3. check p-values
 $\alpha = 0.05$ 0.0001
4. Make a conclusion

Exploratory Analysis

\Rightarrow location, dispersion, skewness

skewness



Visualization

: histogram, Boxplot, scatter plot

Descriptive Statistic

: mean median

* location

: Variance, interquartile, range - dispersion
 $Q_3 - Q_1$ max - min

\leftarrow $\begin{bmatrix} 7 \\ 6 \\ 5 \end{bmatrix}$

\leftarrow : Skewness

\leftarrow $\begin{bmatrix} Q_2 \end{bmatrix}$

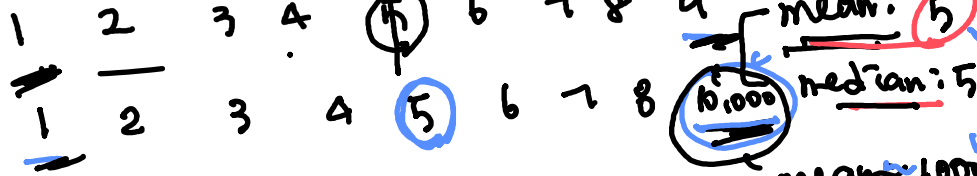
$n=100$
number

"Quantile"

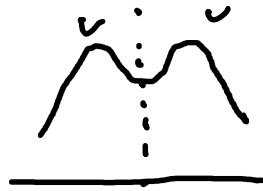
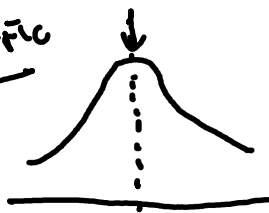


set 1

set 2



Symmetric



mean = median

mean > median



right skewed. (long right tail)

mean < median

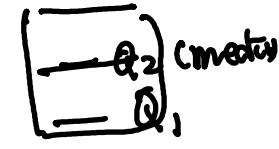
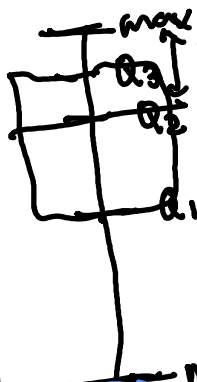
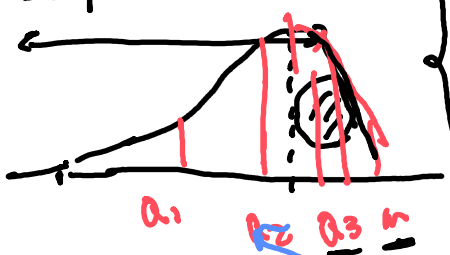


left skewed

skewness

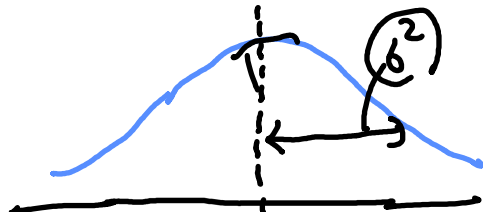
mean < median

Boxplot



* Normal Distribution

$$\rightarrow (\mu, \sigma^2)$$



Symmetric.

Bell Shape

35

Normal

Symm. bell shape

Right tail

less chance to have extreme values

123-10

■ Distributional check

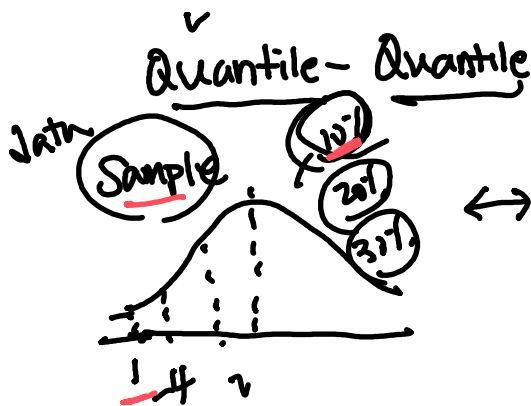
Qualitative : Boxplot, Histogram, QQ-plot

Quantitative : test

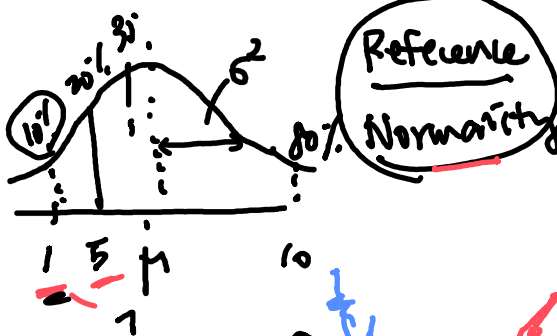
H_0 : Data \sim Normal
 H_a : Data $\not\sim$ Normal.

10^{-8}
 10^{-8}

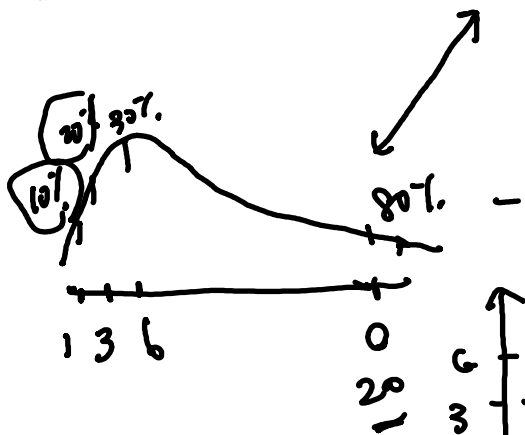
0.....



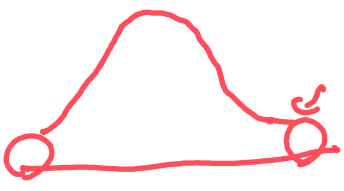
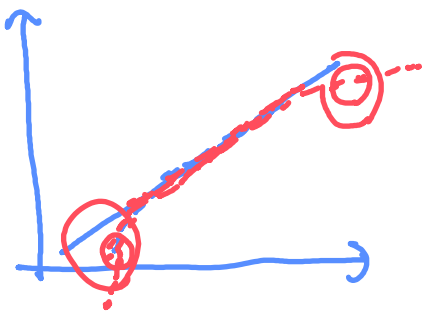
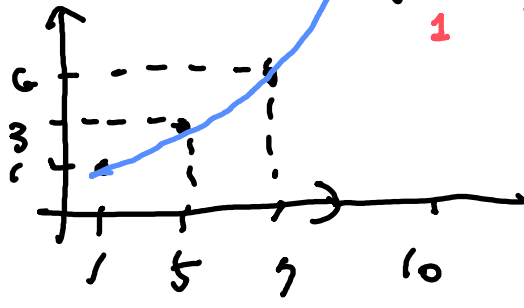
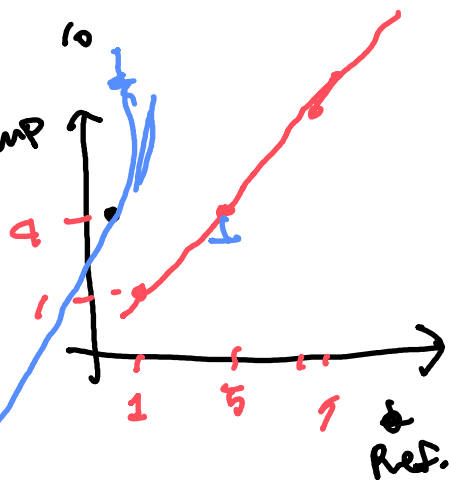
Plot



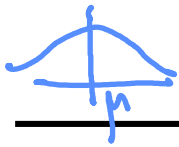
Reference Normality



SAMP



Inferential test (location)



Normality Assumption

mean



to...

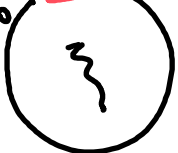
median



One population

Salary = 50,000

n=100



$$\begin{aligned} H_0: \mu &= 50,000 \\ H_a: \mu &\neq 50,000 \end{aligned}$$

$$t\text{-test} = \frac{\bar{x} - \mu_0}{S/\sqrt{n}} = \frac{50,000 - 50,000}{S/\sqrt{n}} = 0$$

p-value

0.001

$$= \sqrt{n} \left(\frac{\bar{x} - \mu_0}{S} \right)$$

Both group ~ Normal

$$\begin{aligned} H_0: M &= 50,000 \\ H_a: M &> 50,000 \end{aligned}$$

Sign test
Signed rank test

At least one

Two populations

Salary F

SS

X

Salary M

M



$$\begin{aligned} H_0: \mu_F &= \mu_M \\ H_a: \mu_F &\neq \mu_M \end{aligned}$$

t-test (i) Equal var
(ii) unequal

$$\begin{aligned} H_0: &\text{two group ~ same dist} \\ H_a: & \end{aligned}$$

Wilcoxon rank-sum test

parametric

nonparametric

(i)



(ii)



μF

$$\begin{aligned} H_0: \sigma_F^2 &= \sigma_M^2 \\ H_a: \sigma_F^2 &\neq \sigma_M^2 \end{aligned}$$

Sign test [1 2 3 5 6 10.000] H₀: M = 4 median
 (-1 -1 -1 +1 +1 +1) H_a: M > 4
 ~ 0

Wilcoxon rank-sum test

Group A: 3 5 6 9 12 ⊕

Group B: 2 7 8 10 99 ⊖

sort: 2 3 5 6 7 8 9 10 12 99
 rank: -1 +2 +3 +4 -5 -6 +7 -8 +9 -10
 = ~ 0

$0 \leq p\text{-value} \leq 1$

Group A: 1 2 3 4 5 ⊕
 Group B: 6 7 8 9 100 ⊖

sort: 1 2 3 4 5 6 7 8 9 100
 rank: +1 +2 +3 +4 +5 -6 -7 -8 -9 -10

- □