discrete : pmf . prob mass fn ← discrete lines

continuous : pdf : prob density fn ← single curve

Presenting Data

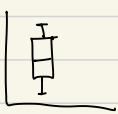
Line Graphs

- Use only when X, Y are quantitative
- better than bar at comparing change over time

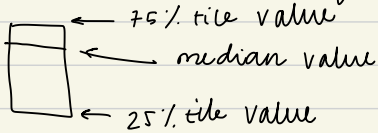
Bar Chart (Histogram for cont. data)

- X : categorical , Y : frequency ← discrete
- can be used to display means (box plot preferred)

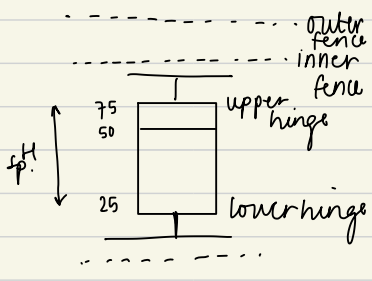
Box Plots



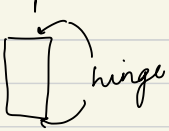
- useful for in dentifying outliers



- mean is indicated with a plus sign.



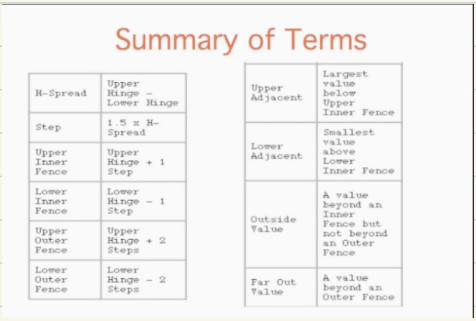
- H spread = 75% tile − 25% tile
- step = 1.5 (H-spread)



- fences above : upper hinge + Astep
  fences below : lower hinge − Astep.
  inner : A = 1
  outer : A = 2
- upper adj value = largest value below the upper inner fence
- lower adj value : smallest value above the lower inner fence

FOR CONTINUOS DATA, NUMBER OF CLASS INTERVALS = $\sqrt{\#\text{ of data}}$

for box plot



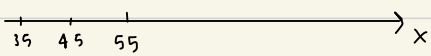
Summary of Terms

H-Spread	Upper Hinge − Lower Hinge
Step	1.5 x H-Spread
Upper Inner Fence	Upper Hinge + 1 Step
Lower Inner Fence	Lower Hinge − 1 Step
Upper Outer Fence	Upper Hinge + 2 Steps
Lower Outer Fence	Lower Hinge − 2 Steps

Upper Adjacent	Largest value below Upper Inner Fence
Lower Adjacent	Smallest value above Lower Inner Fence
Outside Value	A value beyond an Inner Fence but not beyond an Outer Fence
Far Out Value	A value beyond an Outer Fence

Frequency Polygon

- graphical device for understanding shapes of dist
- good for displaying cumulative frequency distribution
  - ↳ add up freq. up till & including that class



choose class int., mark midpoints on axis

Histogram

- cont. classes v/s freq (x ⊄ y)
  - ↓
  - or relative frequencies (0 to 1) $rf = \frac{f}{n}$ ← total obs
- x axis → variable whose distribution is of interest

edward tufty ♡

Stem + Leaf Displays

- stem : 10's digits
  leaves : 1's digit
- helps clarify the shape of the distribution
- stems can be split however it is easier
- suitability of a stem and leaf display depends on whether the data can be rounded w/o loss of info

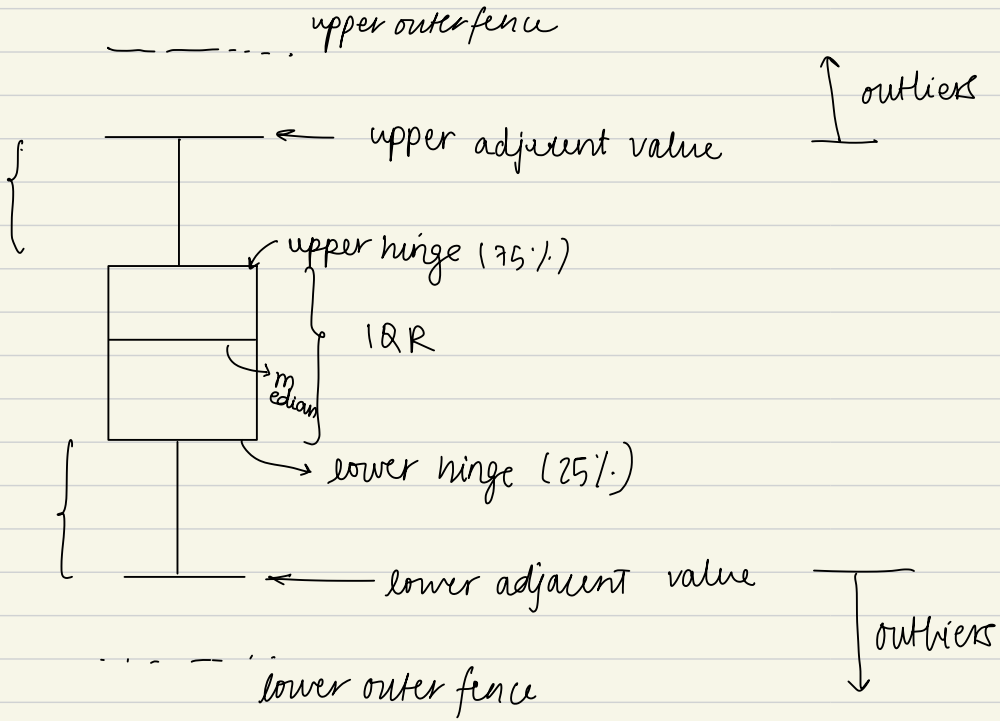
Lecture Examples

Q histogram for IOT salaries

right skew is expected as some (few) people are earning large amounts.

salaries as that is the variable whose dist is of interest

box plot



max: upper hinge + step
↓
upper inner fence
lower inner fence
↓
max lower hinge - step

far out value : beyond outer fence
outside value : between inner and outer fence.

$\frac{1}{3!} \times 3!$

total outcomes: $13 \times 12 \times 11$ — fav.: $10 \times 9 \times 8 \times \frac{1}{3!} \times 3!$

$26^6 - 21^6$ $\quad$ $26^6 - {}^{26}P_6$

$\_\ \_\ \_\ \_$

$8 \times 7 \times 6 \times 1 \times 5$ $\quad$ 1, 3, 5, 7, 9

$({}^8C_3 \times {}^7C_2) \times 3 \times 2$

${}^{15}C_5$ — $\binom{8}{5} - \binom{7}{5}$ $\quad$ $\binom{7}{3}\binom{6}{2}$

↑ total

$\binom{7}{3}\binom{6}{2} + \binom{7}{2}\binom{6}{3} + \binom{7}{1}\binom{6}{4} + \binom{6}{5}$

64

$\binom{13}{5} - \binom{7}{4}\binom{6}{1} - \binom{7}{5}$

H, T, $\quad 2^7$

$\binom{10}{4}\binom{6}{3}\binom{3}{2}$

HH TTT $\quad$ 3H 4T

$\frac{6!}{2!4!}$ $\quad$ $7 \frac{7!}{3!4!} \times 2$ $\quad$ $\left(\frac{10 \times 9 \times 8 \times 7}{4 \times 3 \times 2 \times 1}\right)\left(\frac{6 \times 5 \times 4}{3 \times 2}\right) \times 2$

$\frac{9 \times 8}{2}$ $\quad$ $10 \times \frac{9 \times 8}{2} \times \frac{7 \times 6 \times 5}{3 \times 2}$

$\frac{3 \times 5}{64}$

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21

— — O — OO — OOO — OOOO — — OOOOO —

$\frac{6 \times 5 \times 4 \times 3}{4 \times 3 \times 2 \times 1}$ $\quad$ $\frac{10!}{2 \times 2 \times 3 \times 2}$

choose position of x

$2 + \frac{7 \times 6}{2} + \frac{7 \times 6 \times 5 \times 4}{4 \times 3 \times 2}$ $\quad$ $\frac{1}{2^3}(1 + {}^7C_2 + {}^7C_4 + {}^7C_6)$

$\frac{2 + 21 + 35}{27}$ $\quad$ $\frac{58}{128}$

$\overline{6^3}$

$\overline{5}\ \overline{6}\ \overline{6} \times 3$

$+$

$\overline{6}\ \overline{6 \times 6} \times 3$