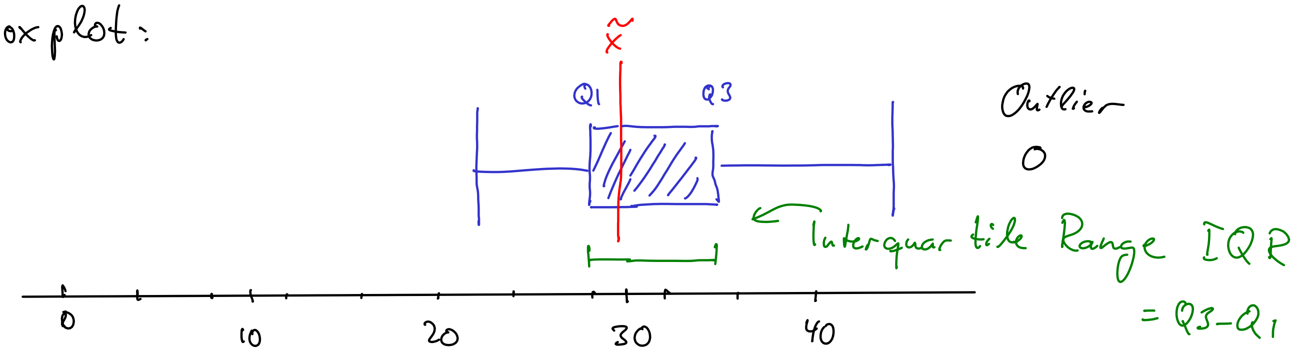
Lecture 22 Revision

Stats:
- histogram, stem & leaf plot
- median $\tilde{x}$ - the $\frac{n+1}{2}$ st observation
- 5 number summary = { min, Q_1, median, Q_3, max }

where Q_1 = median of data belo the median,
Q_3 = " " " above " "

- Boxplot:



The median cuts the box into two pieces.

- If fewer data plot to the right, the data is skewed right,
- " " " " " " left, " " " " left.

So in our example the box plot is left-skewed.

The shape of the box plot helps to choose the best measure of centrality or variance.

- The median / IQR are the best measure of centrality / variance for skewed distributions or data with strong outliers.

- The mean / standard deviation are the best measure of centrality / variance for symmetric distributions with no outliers.

→ Example of data set where the mean is a better measure of the central tendency of the data (than the median)

$\{1, 2, 3, 4, 200\}$

median $\tilde{x} = 3$

mean $\bar{x} = \dfrac{1+2+3+4+200}{5} = 42$

Here $\tilde{x} = 3$ is a better representative of the data.

- (arithmetic) mean $\bar{x} = \dfrac{\sum x_i}{n}$

here: $\bar{x} = \dfrac{30+22+44+31+28+28}{6}$

$= 30.5$

- harmonic mean $= \left(\dfrac{\sum \frac{1}{x_i}}{n}\right)^{-1}$

(used for average of rates)

- geometric mean $= \left(\prod x_i\right)^{\frac{1}{n}}$

(used when comparing different items)

- weighted mean

- standard deviation (another measure of spread, often better than midrange or IQR)

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

- coefficient of variation (cv) $= \dfrac{\sigma}{\bar{x}} \cdot 100\,\%$

Chebyshev's inequality:

At least $1 - \frac{1}{k^2}$ of the distribution's values are within k standard deviations of the mean, $k \geq 1$.

Example: The amount of soft drink (in ounces) to be filled in bottles has a mean of μ ounces and a standard deviation of σ ounces.
The quality control engineer at the bottling plant desires the amount of soft drink X to be within 1 ounce of the mean at least 90% of the time.
If the quality control engineer's goals are to be met, what is the largest value of σ that can be tolerated?

Solution: 90% in $(\mu - k\sigma, \mu + k\sigma)$

$$= (\mu - 1, \mu + 1)$$

so $k\sigma = 1$

$$1 - \frac{1}{k^2} = \frac{9}{10}$$

$$\frac{k^2 - 1}{k^2} = \frac{9}{10}$$

$$k^2 - 1 = \frac{9}{10} k^2$$

$$\frac{1}{10} k^2 = 1 \quad \Rightarrow \quad k^2 = 10 \quad \Rightarrow \quad k = \sqrt{10}$$

$$k\,\sigma = 1 \quad \Rightarrow \quad \sigma = \frac{1}{\sqrt{10}} = \underline{\underline{0.31623}}$$

Probability

Exam 2013/14, Q C.

(a) X = number of heads

2 possible outcomes T/F

binomial distribution, so

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x},$$

with $n = 5$, $p = \frac{1}{2}$

$$P(X=0) = \binom{5}{0} \left(\tfrac{1}{2}\right)^0 \left(\tfrac{1}{2}\right)^5 = \left(\tfrac{1}{2}\right)^5 = \frac{1}{32}$$

$$P(X=1) = \binom{5}{1} \left(\tfrac{1}{2}\right)^1 \left(\tfrac{1}{2}\right)^4 = 5 \cdot \tfrac{1}{2} \cdot \tfrac{1}{2^4} = \frac{5}{32}$$

$$P(X=2) = \binom{5}{2} \left(\tfrac{1}{2}\right)^2 \left(\tfrac{1}{2}\right)^3 = 10 \cdot \tfrac{1}{2^2} \cdot \tfrac{1}{2^3} = \frac{10}{32}$$

$$P(X=3) = \binom{5}{3} \left(\tfrac{1}{2}\right)^3 \left(\tfrac{1}{2}\right)^2 = 10 \cdot \tfrac{1}{2^3} \cdot \tfrac{1}{2^2} = \frac{10}{32}$$

$$P(X=4) = \binom{5}{4} \left(\tfrac{1}{2}\right)^4 \left(\tfrac{1}{2}\right)^1 = 5 \cdot \tfrac{1}{2^4} \cdot \tfrac{1}{2} = \frac{5}{32}$$

$$P(X=5) = \binom{5}{5} \left(\tfrac{1}{2}\right)^5 \left(\tfrac{1}{2}\right)^0 = 1 \cdot \tfrac{1}{2^5} = \frac{1}{32}$$

(b) If even number of coins show heads $(X = 2, 4)$

$\Rightarrow$ win €X^2+X

If odd number of coins show heads $(X = 1, 3, 5)$

$\Rightarrow$ lose €X^2+X

$X=0 \Rightarrow$ no win $(X^2+X=0)$

So expected winnings are

$$P(X=0)\cdot 0 + P(X=1)\left(-(1^2+1)\right) + P(X=2)\left(+(2^2+2)\right)$$

$$+ P(X=3)\left(-(3^2+3)\right) + P(X=4)\left(+(4^2+4)\right) + P(X=5)\left(-(5^2+5)\right)$$

$$= \tfrac{1}{32}(0) + \tfrac{5}{32}(-2) + \tfrac{10}{32}(6) + \tfrac{10}{32}(-12) + \tfrac{5}{32}(20) + \tfrac{1}{32}(-30)$$

$$= -\frac{10}{32} + \frac{60}{32} - \frac{120}{32} + \frac{100}{32} - \frac{30}{32}$$

$$= 0$$

So expected winnings $= 0$

$\Rightarrow$ No point to play the game if you are interested in winning money!