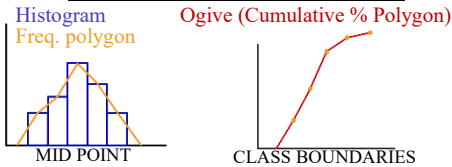


Tabulating and Graphing Numerical Data

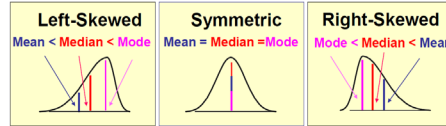
- method
 - 1) $Range = max - min$
 - 2) Define number of class (NOC)
 - 3) Class interval (I) = $\frac{Range}{NOC}$, round up.
 - a. Round up.
 - b. No remain, +1 instead.
 - 4) Class boundaries $\rightarrow [min, min + I), \dots$
 - 5) Class midpoint $\rightarrow \frac{min + (min + I)}{2}, \dots$
 - 6) Assign value to the table.
- Table example

Class	Freq.	Percent.
[32, 45)	4	$\frac{4}{16} = 25\%$
\vdots	\vdots	\vdots



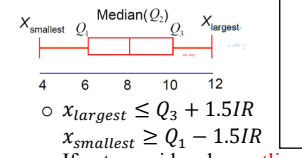
Numerical Descriptive Measure (sample, population)

- Sample size (n).
- Population size (N).
- Central Tendency
 - Mean (\bar{x}, μ)
 - Mode: most common value
 - Median \rightarrow Index of median: $((n + 1)/2)$

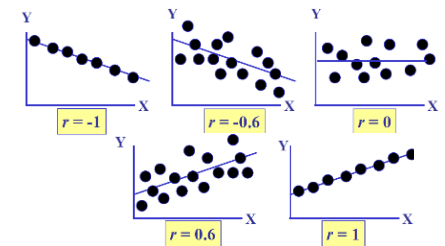


- Quartiles
 - Index of $Q_i: \frac{i}{4}(n + 1)$
 - Interquartile range (IR): $Q_3 - Q_1$
- Variance & Standard deviation (S^2, σ^2)
 - Variance = SD^2
 - $S^2 = \frac{\sum(x_i - \bar{x})^2}{n}$
 - $\sigma^2 = \frac{\sum(x_i - \mu)^2}{N}$
 - Coefficient of Variation (CV) = $\frac{S}{\bar{x}} \times 100\%$

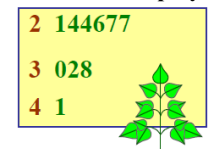
Box n' Whisker plot



- $x_{largest} \leq Q_3 + 1.5IR$
- $x_{smallest} \geq Q_1 - 1.5IR$
- If not, considered as outlier.
- Coefficient of correlation (r)
 - Measure the strength of linear relationship.
 - $r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$



Stem-Leaf Display



Discrete Probability Distribution (DPD)

- Random variable (X): possible value of a random event
- Probability distribution ($P(X = x)$)
 - A set of probability of a scenario when value of random variable is x
 - Add up to 1.0
- Summary Measures
 - Mean: $\mu = \sum x_j \cdot P(X = x_j)$
 - SD: $\sigma^2 = \sum (x_j - \mu)^2 \cdot P(X = x_j) \rightarrow \sum x_j^2 P(X = x_j) - \mu^2$
- Considering population

Binomial Distribution

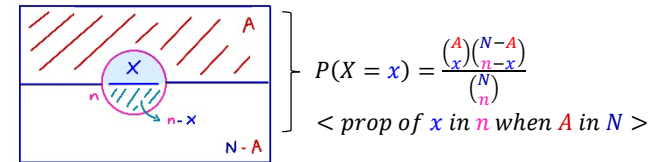
- Condition
 - Have only 2 outcomes: yes/no
 - All trials are INDEPENDENT & IDENTICAL
- Probability of x success.
 - $P(X = x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$
 - $(5, n, p, TRUE) \rightarrow P(X \leq 5)$
 - $(5, n, p, FALSE) \rightarrow P(X = 5)$
 - $n \rightarrow$ number of trial
 - $x \rightarrow$ number of success
 - $p \rightarrow$ probability of success ($q = 1 - p$)
 - $P(X < 5) = P(X = 4) + P(X = 3) + \dots + P(X = 0)$
 - Mean: $\mu = np$
 - SD: $\sigma = \sqrt{np(1-p)}$
 - Excel: **BIONOM.DIST**($x, n, p, \text{cumulative}$) \rightarrow cumulative: TRUE/FALSE

Poisson Distribution

- Describe how likely an event is to occur over a time interval.
- Probability of x success.
 - $P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}$
 - $\lambda \rightarrow$ average NOS in a unit time
 - if $\lambda = 5$ u/hr \rightarrow when 2 hr, $\lambda = 5 \times 2 = 10$ u/hr
- Mean: $\mu = \lambda$
- SD: $\sigma = \sqrt{\lambda}$
- Excel: **POISSON.DIST**($x, \lambda, \text{cumulative}$)

Hypergeometric Distribution

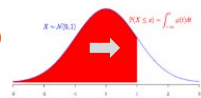
- Describe the success of sample when know the success of population.



- Mean: $\mu = \frac{nA}{N}$
- SD: $\sigma = \sqrt{\frac{NA(N-A)}{N^2} \cdot \frac{N-n}{N-1}}$ (Finite population correction factor)
- Excel: **HYPGEOM.DIST**(x, n, A, N)

Continuous Probability Distribution

- Calculate Z and use the area under the normal distribution (ND).
 - Density func. of ND: $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$
 - $P(X < x) = P(Z_i < Z) \rightarrow Z = \frac{x - \mu}{\sigma}$
 - Z : Standard normal random variable
- Make sure the distribution is normal before calculating. (method)
 - Construct chart:
 - Small: Stem-leaf, box & whisker \rightarrow symmetric
 - Large: histogram, polygon \rightarrow bell-shaped
 - Find summary measure.
 - $mean \approx med \approx mode \mid IR \approx 1.33\sigma \mid Range \approx 6\sigma$
 - Observe dataset: 2/3 of data \rightarrow lie between $\pm 1\sigma$
 - 4/5 $\rightarrow \pm 1.28\sigma$, 19/20 $\rightarrow \pm 2\sigma$
 - Plot (z, x) must be a straight line.
- Excel: **NORM.DIST**($x, \mu, \sigma, \text{cumulative}$)
 - Give value count from the right.
- Normal approximation (Binomial Dist.)
 - When $n \rightarrow \infty, p \rightarrow 0.5, np \text{ and } nq > 5$
 - Mean: $\mu = np$
 - Variance: $\sigma^2 = npq$
 - $Z = \frac{x - np}{\sqrt{npq}}$
- Normal approximation (Poisson Dist.)
 - When $\lambda \rightarrow \infty$
 - Mean: $\mu = \lambda$
 - Variance: $\sigma^2 = \lambda$
 - $Z = \frac{x - \lambda}{\sqrt{\lambda}}$
- Uniform distribution
 - Each value is equally likely to occur anywhere.
 - Range: between $[a, b]$
 - graph: horizontal line
 - Density function $f(x) = \begin{cases} \frac{1}{b-a}; & a \leq x \leq b \\ 0; & \text{elsewhere} \end{cases}$
 - Mean: $\mu = \frac{a+b}{2}$
 - Variance: $\sigma^2 = \frac{(b-a)^2}{12}$
 - $P(a \leq x \leq b) = \int_a^b f(x) dx$
- Continuity correction (Binomial & Poisson)
 - Use when convert discrete dist. to continuous dist.
 - ± 0.5 depends on the situation.



$$\begin{aligned} \mu &= \int x f(x) dx \\ \sigma^2 &= \int (x - \mu)^2 f(x) dx \end{aligned}$$

Sampling Distribution

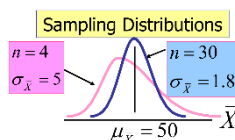
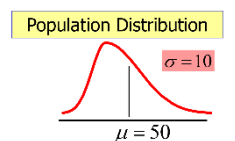
- Statistical inference: estimate character of population from info of sample.
- Terminology
 - Parameter: character of population (μ, σ, p)
 - Sample statistics: character of sample ($\bar{x}, s, \bar{p}/p_s/\bar{p}$)
 - Point estimate: numerical value of (\bar{x}, s, \bar{p})
 - Sampling error: absolute difference $|\mu - \bar{x}|$
- Large sample gives more accurate estimation.
- Sampling distribution of mean: Prop-dist. of \bar{x}
 - Take multiple groups of n data from N and find \bar{x} , and plot a distribution of that multiple \bar{x}
 - Graphs will look like normal distribution.
 - When $N \rightarrow \infty$, graph look like population distribution (PD).
 - When $N \rightarrow \infty$ & large n , graph approach the normal distribution. Even when PD is not normal.
 - Mean: $\mu_{\bar{x}} = \mu \rightarrow \bar{x}$ is unbiased
 - SD: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \rightarrow \sigma_{\bar{x}} < \sigma$ always
 - $Z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}$
 - Size of n to get normal sample mean distribution.
 - Most dist: $n > 30$ | Symmetric dist: $n > 15$ | Normal dist: n can be any value.

Sampling distribution of sample proportions (p_s)

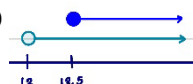
- population proportions (p): A measure of how common a characteristic is in a population.
 - $p = \frac{x}{N} \rightarrow x$ is the number of successes | $p_s = \frac{x}{n} \rightarrow x$ success in sample group.
- Normally focus on x with two outcomes (have/not have) \rightarrow binomial distribution.
- Mean: $\mu_{p_s} = p$
- SD: $\sigma_{p_s} = \sqrt{\frac{p(1-p)}{n}}$
- $Z = \frac{p_s - \mu_{p_s}}{\sigma_{p_s}}$

Finite population correction factor (FPC)

- Used when n is very large: $n/N > 0.05$
- $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$
- $\sigma_{p_s} = \sqrt{\frac{p(1-p)}{n}} \cdot \sqrt{\frac{N-n}{N-1}}$



$$P(X > 12) \rightarrow P(X \geq 12.5)$$



$$P(X \leq 25) \rightarrow P(X < 25.5)$$



