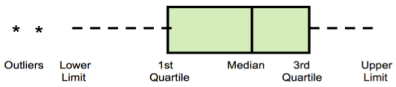
Boxplots

Displays the five-number summary (median, Q1 & Q3, max and min)



Upper Limit = Q3 + 1.5(IQR)
Lower Limit = Q1-1.5(IQR)

Process Control Chart (XBAR Chart)

$$UCL_{\bar{x}} = \bar{\bar{X}} + z\sigma_{\bar{x}}$$
$$LCL_{\bar{x}} = \bar{\bar{X}} - z\sigma_{\bar{x}}$$

Where

$\sigma_{\bar{x}}$ = Standard deviation of sampling distribution of sample means = $\frac{\sigma}{\sqrt{n}}$
σ = Process standard deviation
n = Sample size
z = Standard Normal deviate (usually z = 3)
$\bar{\bar{X}}$ = Average of sample means = grand mean

		Sample				
		1	2	3	4	5
Observation	1	12.11	12.15	12.09	12.12	12.09
	2	12.10	12.12	12.09	12.10	12.14
	3	12.11	12.10	12.11	12.08	12.13
	4	12.08	12.11	12.15	12.10	12.12
	$\bar{x}$	12.10	12.12	12.11	12.10	12.12

If Standard Deviation IS NOT given use this method:

$$UCL_{\bar{x}} = \bar{\bar{X}} + A_2\bar{R}$$
$$LCL_{\bar{x}} = \bar{\bar{X}} - A_2\bar{R}$$

where
Sample range = maximum value – minimum value in the sample
A_2 can be obtained from Shewhart Table *(next page)*
$\bar{R}$ = Average of sample ranges
$\bar{\bar{X}}$ = Average of sample means = grand mean

R CHART

$$UCL_R = D_4\bar{R}$$
$$LCL_R = D_3\bar{R}$$

D_3 and D_4 are obtained from table

Process is **uncontrollable** if the R-Chart has:
- One or more points outside the control limits
- Nine or more points in a row on one side of the center line
- Six or more points moving in the same direction.
- 14 or more points alternating above and below the center line.

Process is **out of control** is this happens in XBAR chart.

Example of Random Variables

- 8 oz box: mean = 9.6 SD = 0.8
- Empty Carton: mean=24 SD=2.2

Weight of Filled Carton when a Carton holds 24 boxes?
T= X1 +……X24 (this is just the weight of 24 boxes without the carton)
Mean = 9.6(24) =230.4
SD = $\sqrt{(0.8)(24)}$
W = T + empty carton
Mean = 230.4 + 24
SD = $\sqrt{4.38^2 + 2.2^2}$

Combining Correlated Random Variables

Expected Value: amount(EX) + amount (EY)
Variance: Var(aX+bY) = a^2 Var(X) + b^2 Var(Y) + 2ab r SD(X) SD(Y)
SD: $\sqrt{a^2(Varx) + b^2(Vary)}$
SD: $\sqrt{a^2SDX^2 + b^2SDY^2}$
Covariance: Cov(x,y) = r (SDX)(SDY)

Number of Classes in a Histogram

- $2^k > n$
- K = number of classes
- N = number of data points
- Once k is known the width of each class can be found:
- W = $\frac{largest\ value - smallest\ value}{number\ of\ classes}$

Inverse Normal

- Probability of 0.99.
- Find 0.99 in the Z-table → 2.33
- Therefore, = mean+ 2.33 x SD

Determine approximate values for lower and upper quartiles (Q1 and Q3)

- Q1 = median – (IQR/2)
- Q3 = median + (IQR/2)

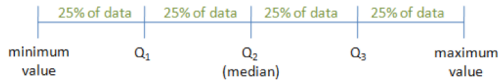
To find any OUTLIERS:
1. Calculate upper and lower fences (REFER TO TOP OF PAGE)

IF VALUES INCREASE:
- Mean increases by the same %
- Variance increases by double
- CV is unchanged
- IQR increase by the same %
- Median would be unchanged.
- Range would decrease
- Standard deviation would decrease

Measures of Relative Position

Percentiles: a value below which a given percentage of data lies, for instance, 10% of data is below the 10th percentile.
To find the kth percentile:
- Arrange data in ascending order
- I = (p/100)(n)
- If your answer is 3, then below the 3rd data value is the percentile.

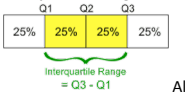
Decile: Splits data into 10% groups. (1st decile is the 10th percentile, 2nd decile is 20th percentile)
Quartile: Splits the data into 4 equal sized groups.



- =(p/100)(n)
- using p as only either 25,50,75

Variability

Describe the spread or dispersion of a set of data.
Range: difference between the max and min values.
Interquartile Range (IQR): describes the middle 50% of range.



Always use IQR if median was used.
Variance (population): (x-mean)^2/N
SD: The square root of variance
Variance (sample) = (x-mean)^2/n-1
Coefficient of Variation: measure relative variation in data
- SD/mean(100)

Population = ALL individuals
Sample = a subset of the population
- We used a sample statistic to estimate a population parameter
- The best way to make sure a sample is representative is to be random.

1. **Simple Random Sampling (SRS)** – each set of n individuals in the population has an equal chance of selection. (With or without replacement)
2. **Stratified Sampling (STR)** – the population is divided into several homogeneous subpopulations, or strata, and random samples are then drawn from each stratum.
   - Stratum could be by age, sex, geography, income etc…
3. **Cluster Sampling** – groups, or clusters, representative of the population are chosen at random and a census is taken of each.
4. **Systematic Sampling (SYS)** – a sample id drawn by selecting individuals systematically from a sampling frame. (sample every 3rd person)
5. **Convenience Sampling** – selects people who are conveniently available.

Sampling Distribution of Proportion

- **P**, which is the percentage of entire population.
- **p**, which is the percentage of the sample.
- **n**, which is the size of the sample
- $\sigma = \sqrt{\frac{p\,(p-1)}{n}}$
- $Z = \frac{\hat{p}-p}{\sigma}$

Sampling Distribution of Means Infinite Population (CLT)

- **μp**, which is the mean of the population
- **σp**, which is the standard deviation of the population
- **n**, which is the size of the sample
- **μs**, which is the mean of the sample.
- $\sigma = \frac{\sigma}{\sqrt{n}}$
- $Z = \frac{\mu s - up}{\sigma}$

Sampling Distribution of Means Finite Population

- **N**, which is the size of the population
- **μp**, which is the mean of the population
- **σp** , which is the standard deviation of the population
- **n**, which is the size of the sample
- **μs**, which is the mean of the sample
- $\sigma = \frac{\sigma p}{\sqrt{n}} \times \frac{N-n}{N-1}$
- $Z = \frac{\mu s - \mu p}{\sigma}$

T-Distribution

- **μp**, which is the mean of the population
- **n,** which is the size of the sample
- **μs**, which is the mean of the sample.
- **σs**, which is the standard deviation of the sample.
- $T = \frac{\mu s - \mu p}{\frac{\sigma s}{\sqrt{n}}}$