

BSTA 200 – Midterm FORMULA SHEET

Use a calculator or software where necessary instead of the formulas below.
Excel formulas are shown in **GREEN**.

Nature of data: Qualitative (categorical) data, quantitative (numeric) data (continuous or discrete)

Levels of Measurement: nominal level, ordinal level, interval level, ratio level

Organization of data: Frequency table, relative frequency, cumulative frequency

$$\text{Relative Frequency} = \frac{\text{class frequency}}{\text{sum of frequencies}} = \frac{f}{n}$$

Number of intervals - use the 2^k rule; $2^k > n$ (sample size) where k = number of intervals

$$\text{Class width} = \frac{\text{range}}{k} \quad \text{or} \quad \frac{\text{range}}{\# \text{ of intervals}}$$

$$\text{Class Midpoint} = \frac{\text{lower limit} + \text{upper limit}}{2}$$

Measures of Location (Measures of Central Tendency)

	Population data	Sample data
Arithmetic Mean	$\mu = \frac{\sum X}{N}$	$\bar{x} = \frac{\sum x}{n} = \text{=AVERAGE()}$
Position of Median	$\frac{N + 1}{2}$	$\frac{n + 1}{2} = \text{=MEDIAN()}$
Mode	Most frequent observation = MODE()	
Weighted Mean	$\frac{\sum w(x)}{\sum w}$ where w =weighting factor x = individual value	

Measures of position: deciles, quartiles, percentiles

Location of percentile $L_p = (n + 1) \frac{p}{100}$

n = number of observations in the data set
 L = position of the percentile in the data set

p = desired percentile

P -th percentile =**PERCENTILE.EXC()**

BSTA 200 – Midterm FORMULA SHEET

Measures of Variation (dispersion)

	Population data	Sample data
Range	Highest value – lowest value	=MAX() - =MIN()
Standard deviation	$\sigma = \sqrt{\frac{\sum (X - \mu)^2}{N}}$ =STDEV.P()	$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}}$ =STDEV.S()
Variance	σ^2 =VAR.P()	s^2 =VAR.S()
Coefficient of Variation (CV)	$CV = \frac{\sigma}{\mu}$	$CV = \frac{s}{\bar{x}}$

Correlation and Regression

Coefficient of Correlation (r) $r = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{[n(\sum x^2) - (\sum x)^2][n(\sum y^2) - (\sum y)^2]}}$ **=CORREL(Y, X)**

Regression Equation $\hat{y} = a + bx$ where $b = \text{slope}$ **=SLOPE(Y, X)**

$a = \text{y-intercept}$ **=INTERCEPT(Y, X)**

Coefficient of Determination (r²) $r^2 = \frac{\text{explained variation}}{\text{total variation}}$ or $\frac{SSR}{SST}$

Standard Error of Estimate $s_e = \sqrt{\frac{\sum y^2 - a(\sum y) - b(\sum xy)}{n - 2}}$ or $\sqrt{\frac{SSE}{n - 2}}$

Counting Rules

Factorial	Permutation Rule	Combination Rule
Number of arrangements of n different objects = $n!$ =FACT(n) When objects repeat $n_1, n_2, \dots, n_k$ times, number of arrangements is: $\frac{n!}{n_1! \cdot n_2! \cdots n_k!}$	Selecting r items from n different items when order is important . ${}_nP_r = \frac{n!}{(n - r)!}$ =PERMUT(n, r)	Selecting r items from n different items when order is not important . ${}_nC_r = \frac{n!}{(n - r)! r!}$ =COMBIN(n, r)