BSTA 200 - TEST 1 FORMULA SHEET

Nature of data

Qualitative data, quantitative data (continuous or discrete)

Levels of Measurement

nominal level, ordinal level, interval level, ratio level

Organization of data

Frequency table, relative frequency, cumulative frequency

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

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

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

Class Midpoint = $\frac{sum\ of\ two\ consecutive\ lower\ limits}{2}$

Measures of Location (Measures of Central Tendency)

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

Measures of Variation (spread)

	Population data	Sample data
Range	Highest value – lowest value	
Mean Absolute Deviation	$\frac{\sum \lvert x-\mu \rvert}{N}$	$\frac{\sum \lvert x-\bar{x} \rvert}{n}$
Standard deviation	$\sigma = \sqrt{\frac{\sum (x-\mu)^2}{N}}$	$s = \sqrt{\frac{\sum (x-\bar{x})^2}{n-1}}$ $s = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}}$
Variance	σ^2	s^2
Coefficient of Variation (CV)	$CV = \frac{\sigma}{\mu}$	$CV = \frac{s}{\bar{x}}$

Measures of position: deciles, quartiles, percentiles

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

where n = number of observations in the data set, L = position of designated percentile in data set and P= desired percentile

Correlation and Regression

Total Variation = Regression Sum of Squares + Error Sum of Squares

SST = SSR + SSE

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

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

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

Regression Equation $\hat{y} = a + bx$

where $b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$ and $a = \frac{\sum y}{n} - b\frac{\sum x}{n}$