Appendix B Formula Summary

Appendix B
Formula Summary

Frequency: midpoints - proportion, *p* - percentage - percentile - mean from frequency distribution - **Central Tendency**: mean - median - mode - **Variability** - range - sum of deviations - sum of square - range - variance - standard deviation - quartiles - interquartile - semi-interquartile range - **Standard Score** - z-scores - other standard scores - percentile - percent rank - **Normal Curve** - area under curve - NCE -
Correlation - sum of products - covariance, Pearson r - standard error of r - confidence interval of r - Spearman r

Frequency

Midpoint = $\frac{U_a - L_a}{2}$, where $\boldsymbol{U_a}$ is the upper apparent limit and $\boldsymbol{L_a}$ is the lower

$N = \Sigma f$, where f is the frequency of each group, and N is total frequency

proportion = $p = \frac{f}{N}$, where f is the frequency, and N is total frequency

percentage = $p(100) = \frac{f}{N}(100)$, where f is the frequency, and N is sum of frequencies

$$P_{\%} = LL_i + [(\frac{n_p - C_f}{f_i}) \cdot I]$$

where $P_{\%}$ = any specified percentile point
LL_i = the *exact lower limit* or *real lower limit* of the group interval containing the percentile point, $P_{\%}$
n_p = number of scores or cases comprising the specified percentage for n.
C_f = cumulative frequency up to but not including the percentile interval

f_i = frequency within the percentile interval
I = group interval size or class size

Mean from frequency:

mean = $\overline{X} = \frac{\Sigma f_i X_i}{n}$, where $\Sigma f_i X_i$ is product of frequency and score or midpoint

Central Tendency

Population mean: $\mu = \frac{\Sigma X}{N}$, where ΣX is sum of scores, and N is population size

Sample: $\overline{X} = M = \frac{\Sigma X}{n}$, where ΣX is sum of scores, and n is sample size

Median: (a) **odd** N, ,median = middle score of an ordered set (min to max)

(b) **even** N, median = $\frac{\text{two center scores}}{2}$, average of center scores

Mode: the most frequent score(s) from a simple frequency distribution

Variability

Range = $X_{\max} - X_{\min}$, where X_{max} is maximum score, and X_{min} is minimum score

Sum of deviation from mean: $\Sigma(X - \overline{X}) = 0$, where $\overline{X}$ is mean

Sum of square: $SS = \Sigma(X - M)^2$ or $SS = \Sigma(X - \overline{X})^2$ (definition formula)

$$SS = \Sigma X^2 - \frac{(\Sigma X)^2}{N} \text{ (computational formula)}$$

Population Variance = σ^2 (Sigma squared) = $\frac{\Sigma(X - \overline{X})^2}{N}$ (definition)

$$\sigma^2 = \frac{SS}{N} = \frac{\Sigma X^2 - \frac{(\Sigma X)^2}{N}}{N} \text{ (computational)}$$

Sample Variance = $s^2 = \frac{\Sigma(X - \overline{X})^2}{n - 1}$ (definition)

$$s^2 = \frac{SS}{n - 1} \text{ (computational, see } SS \text{ above)}$$

standard deviation = σ or s = $\sqrt{\text{variance}} = \sqrt{\sigma^2} = \sqrt{s^2}$

Population stadard deviation = σ (Sigma) = $\sqrt{\frac{\Sigma(X - \overline{X})^2}{N}}$ (definition)

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{SS}{N}} = \sqrt{\frac{\Sigma X^2 - \frac{(\Sigma X)^2}{N}}{N}} \text{ (computational)}$$

Sample Variance = $s = \sqrt{\dfrac{\Sigma(X-\overline{X})^2}{n-1}}$ (definition)

Degrees of freedom, *df*, for sample variance: $df = n - 1$

$$s = \sqrt{s^2} = \sqrt{\frac{SS}{n-1}} = \sqrt{\frac{SS}{df}} = \sqrt{\frac{\Sigma X^2 - \frac{(\Sigma X)^2}{n}}{n-1}} \text{ (computational)}$$

50% Quartile: Q_2 = Median (50% of scores at or below median)

25% Quartile: Q_1 = Median of first half of dataset (25% of scores at or below Q_1)

75% Quartile: Q_3 = Median of second half of dataset (25% of scores at or above Q_3)

interquartile range: $IQR = Q_3 - Q_1$

semi-interquartile range = $\dfrac{Q_3 - Q_1}{2}$

Standard Score

$z = \dfrac{X-\mu}{\sigma}$ or $z = \dfrac{X-M}{s}$ (sample, *M* is mean)

50% Quartile: Q_{50} =Median (50% of scores at or below median)

Other Standard Scores:

Other Standard Score Systems standard score = z = $\dfrac{X-\mu}{\sigma}$		
Standard Score System	μ	σ
z scores	0	1
T score	50	10
General Aptitude Test Battery (GATB)	100	20
College Entrance Examination Board (CEEB)	500	100
IQ Test	100	16

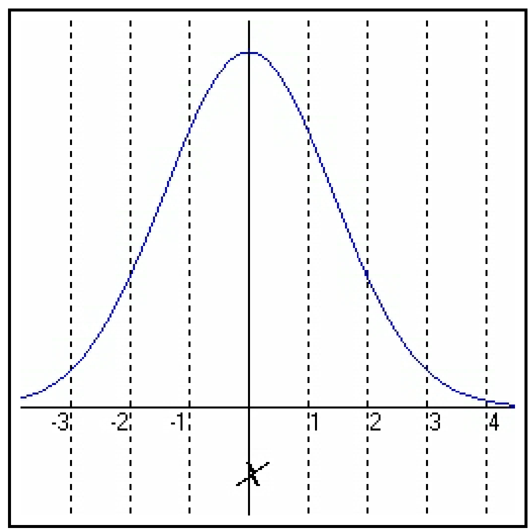
Percentile, $P_{\%}$ raw score, $X = \sigma(z) + \mu$

Example: $\boldsymbol{P_{75}}$ is **X = 87.39** [given σ= 2.5 and μ= 85.7; z = 0.675 for Pr(≤0.75)]
The raw score that 75% of distribution is below: 75 percentile

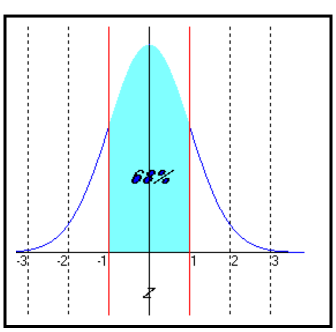
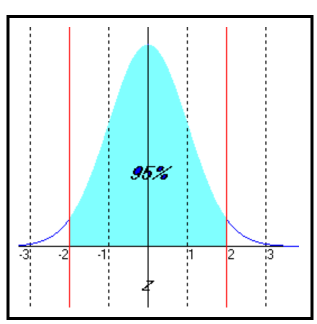
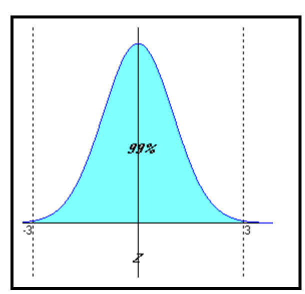
Percent Rank, the percentile of a given raw score

Example: Pr_{89} (percent rank of 89), X = 89 (σ= 2.5 and μ= 85.7; z = 1.32)
Pr(z ≤ 1.32) = 90.66 or Pr_{89} = **90.66%**

Normal Curve

Normal Curve formula	Normal Curve Graph
$y = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{X-\mu^2}{2\sigma^2}}$ y = height of normal curve σ = standard deviation π= pi, or about 3.14 e = natural logarithm constant, about 2.718 X = any score μ = population mean	

Area under curve between various z-scores

Pr between z = -1 and +1	Pr between z = -2 and +2	Pr between z = -3 and +3
		
Area = 68% between z = -1 & +1	Area = 95% between z = -2 & +2	Area = 99% between z = -3 & +3

Normal Curve Equivalence (NCE)

$NCE = \mathbf{21}z + \mathbf{50}$

Correlation

Sum of square of X: $SS_x = \Sigma(X - M_x)^2$ or $SS_x = \Sigma(X - \overline{X})^2$ (definition formula)

Sum of square of Y: $SS_y = \Sigma(Y - M_y)^2$ or $SS_y = \Sigma(Y - \overline{Y})^2$ (definition formula)

Sum of products: $SP = \Sigma[(X - M_x)(Y - M_y)]$ (definition formula)

$$SP = \Sigma XY - \frac{(\Sigma X)(\Sigma Y)}{n} \text{ (computational formula)}$$

Covariance, $S_{xy} = \dfrac{\Sigma[(X - M_x)(Y - M_y)]}{n - 1}$

Pearson correlation coefficient, $r_{xy} = \dfrac{S_{xy}}{S_x S_y} = \dfrac{SP}{\sqrt{SS_x SS_y}}$

where S_{xy} is covariance, S_x and S_y are standard deviations of X & Y
SS_x and SS_y are sum of square of X and Y

Pearson $r = \dfrac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{N\Sigma X^2 - (\Sigma X)^2} \cdot \sqrt{N\Sigma Y^2 - (\Sigma Y)^2}}$ (computational formula)

standard error of r

$s_r = \dfrac{1}{\sqrt{n - 1}}$ (estimate of standard deviation of r)

Confidence interval for r

95% of all sample $r's$ fall between $0 \pm 1.96 s_r$
99% of all sample $r's$ fall between $0 \pm 2.58 s_r$

Spearman correlation coefficient, r_s

Spearman rank-Difference Method

$$r_s = 1 - \frac{6\Sigma D^2}{n(n^2 - 1)}$$

where r_s is correlation (population symbol is *pho* or ρ), ΣD^2 is sum of square difference between ranks, and n is number of pairs of ranks