MEAN AND VARIANCE OF LINEAR COMBINATIONS OF OBSERVATIONS (APPENDIX 2A)

Fundamental Formulas

Suppose $y_1$, $y_2$, $y_3$ are random variables (not necessarily normal) with means $\eta_1$, $\eta_2$, $\eta_3$, variances $\sigma_1$, $\sigma_1$, $\sigma_1$, and correlation coefficients $\rho_{12}$, $\rho_{13}$, $\rho_{23}$. The mean and the variance of the linear combination $Y = a_1y_1 + a_2y_2 + a_3y_3$, where $a_1$, $a_2$, $a_3$ are any positive or negative constants, are respectively,

$$E(Y) = a_1\eta_1 + a_2\eta_2 + a_3\eta_3. \quad (1)$$

and

$$V(Y) = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + a_3^2\sigma_3^2 + 2a_1a_2\sigma_1\sigma_2\rho_{12} + 2a_1a_3\sigma_1\sigma_3\rho_{13} + 2a_2a_3\sigma_2\sigma_3\rho_{23}. \quad (2)$$

These formulas generalize in an obvious way. For a linear combination $Y = \sum_{i=1}^{n} a_iy_i$ of $n$ random variables

$$E(Y) = \sum_{i=1}^{n} a_i\eta_i. \quad (3)$$

and, $V(Y)$ will have $n$ "squared" terms like $a_i^2\sigma_i^2$ and $n(n-1)/2$ cross-product terms like $2a_ia_j\sigma_i\sigma_j\rho_{ij}$. Specifically

$$V(Y) = \sum_{i=1}^{n} a_i^2\sigma_i^2 + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 2a_ia_j\sigma_i\sigma_j\rho_{ij}. \quad (4)$$

Applications

- Variance of a Sum and a Difference of Two Correlated Random Variables

$$V(y_1 + y_2) = \sigma_1^2 + \sigma_2^2 + 2\rho_{12}\sigma_1\sigma_2. \quad (5)$$

$$V(y_1 - y_2) = \sigma_1^2 + \sigma_2^2 - 2\rho_{12}\sigma_1\sigma_2. \quad (6)$$
• **No Correlation**  
  Consider the statistics \( Y \),  
  \[
  Y = a_1 y_1 + a_2 y_2 + \cdots + a_n y_n.  \tag{7}
  \]
  Then  
  \[
  V(Y) = a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2 + \cdots + a_n^2 \sigma_n^2.  \tag{8}
  \]

• **No Correlation, Equal Variances**  
  If, in addition, all the variances are equal, \( E(Y) \) remains as before and  
  \[
  V(Y) = (a_1^2 + a_2^2 + \cdots + a_n^2) \sigma^2.  \tag{9}
  \]

• **Variance of an Average of \( n \) Random Variables with All Means Equal to \( \eta \) and All Variances Equal to \( \sigma^2 \)**  
  \[
  \bar{y} = \frac{y_1 + y_2 + \cdots + y_n}{n} = \frac{1}{n} y_1 + \frac{1}{n} y_2 + \cdots + \frac{1}{n} y_n.  \tag{10}
  \]
  So,  
  \[
  V(\bar{y}) = \left( \frac{1}{n^2} + \frac{1}{n^2} + \cdots + \frac{1}{n^2} \right) \sigma^2 = \frac{n}{n^2} \sigma^2 = \frac{\sigma^2}{n}.  \tag{11}
  \]

• **Variances of the Average \( \bar{y} \) for Observations That Are Autocorrelated at Lag 1**  
  Suppose now that all the observations \( y_1, y_2, \cdots, y_n \) have a constant variance \( \sigma^2 \) and the same lag 1 autocorrelation \( \rho_{i,i+1} = \rho_1 \). Suppose further that all correlations at greater lags are zero. Then  
  \[
  Y = n \bar{y} = y_1 + y_2 + \cdots + y_n.  \tag{12}
  \]
  and making the necessary substitutions, you get  
  \[
  V(\bar{y}) = C \times \frac{\sigma^2}{n}  \tag{13}
  \]
  where  
  \[
  C = \left[ 1 + \frac{2(n-1)}{n} \rho_1 \right].  \tag{14}
  \]