Statistics Tutorial (Part 1 of 2)

The mark of a truly educated man is to be moved deeply by statistics.

-- George Bernard Shaw (1856 – 1950)
Irish playwright and winner of the Nobel Prize for Literature (1925)

(Just about) all the statistics you'll need

- In this lecture. . .
  - discrete and continuous probability distributions
  - expected value, variance, standard deviation, covariance, and correlation
  - Numerical examples – expected returns and risks for 2-asset portfolios

Discrete & Continuous Distributions

- There are two types of random variables: discrete and continuous.
  - **Discrete random variables** can only take on a finite number of "countable" values; e.g., time and temperature rounded to the nearest minute or degree.
  - **Continuous random variables** can take on an infinite number of possible values which lie along a continuum; e.g., the non-rounded versions of time and temperature.

Discrete Probability Distribution

A discrete probability function is a function that satisfies the following properties:

1. The probability that the random variable X can take a specific state contingent value X_s is $p(X_s)$; that is, $\Pr[X = X_s] = p(X_s) = p_s$.
2. p_s is non-negative for all possible values of X_s.
3. The sum of p_s over all possible states is 1; i.e.,

$$\sum_{s=1}^{n} p_s = 1.$$

A consequence of properties 2 and 3 is that $0 \leq p_s \leq 1$ for all s.

Continuous Probability Distribution

The mathematical definition of a continuous probability function, $f(x)$, is a function that satisfies the following properties.

1. The probability that x is between two points a and b is $\Pr[a \le x \le b] = \int_a^b f(x)dx$.
2. $f(x)$ is non-negative for all possible values of x.
3. The integral of the probability function is one; that is, $\int_{-\infty}^{\infty} f(x)dx = 1$.

Expected value

- Expected value is also known as the mean, or average value for a random variable; it represents the *central value* about which variable observations scatter.
- Suppose a random variable X exists which can take on $s = 1, 2, \ldots, n$ possible state-contingent values (each with probability p_s). Then the expected value of X is

$$E(X) = \sum_{s=1}^{n} p_s X_s.$$

Properties of expected values

- Expected values have the following properties:
  - $E(c) = c$. (The expected value of a constant is the constant).
  - $E(cX) = cE(X)$ (The expected value of a constant times a random variable is equal to the constant multiplied by the expected value of the random variable).
  - $E[X + Y] = E[X] + E[Y]$ (The expected value of a sum of random variables is equal to the sum of the expected values of the random variables).

Variance and Standard Deviation

- Variance is the expected value of the squared deviation of the random variable from its mean.
- Standard Deviation is the square root of the variance, and it measures how far most of the variable observations scatter about the mean.
- Standard Deviation is commonly used in finance and risk management as a definition for risk (although as a risk measure, it has several shortcomings which we'll discuss at a later date!).

Variance and Covariance

- The variance $Var(X)$ is computed as follows:

$$Var(X) = \sigma_X^2 = E[(X - E(X))^2] = \sum_{s=1}^{n} p_s (X_s - E(X))^2.$$

- Variances have the following properties:

$$Var(cX) = E[(cX - cE(X))^2] = c^2 \sum_{s=1}^{n} p_s (X_s - E(X))^2 = c^2 \sigma_X^2.$$

  - If X and Y are statistically *independent*, $Var(X + Y) = Var(X) + Var(Y)$. (Total variance is the sum of variances)
  - If X and Y are statistically *dependent*, $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X,Y)$. (Total variance is sum of variances & covariances)

Covariance and Correlation

- Covariance between X and Y is computed as follows:

$$Cov(X,Y) = \sigma_{xy} = E[(X - E(X))(Y - E(Y))]$$

$$= \sum_{s=1}^{n} p_s (X_s - E(X))(Y_s - E(Y)).$$

- Note that variance is a "special case" of covariance, where $X = Y$!
- Correlation coefficient: $\rho_{XY} = Cov(X,Y) / \sigma_X \sigma_Y .$
  - Correlation is a "standardized" covariance
  - Covariance is defined over the open interval $(-\infty, +\infty)$, whereas correlation is defined over the closed interval [-1, +1].

Risk Management Application

- Suppose we wish to determine expected returns and risks for individual assets and portfolios comprising such assets.
- Expected returns, standard deviations, and covariances for individual assets are calculated as follows:

$$E(r_i) = \sum_{s=1}^{n} p_s r_{i,s}$$

$$\sigma_i = \sqrt{\sum_{s=1}^{n} p_s (r_{i,s} - E(r_i))^2}$$

$$\sigma_{i,j} = \sum_{s=1}^{n} p_s (r_{i,s} - E(r_i))(r_{j,s} - E(r_j))$$

Risk Management Application

Portfolio expected returns and standard deviations are calculated as follows:

$$E(r_p) = \sum_{i=1}^{n} w_i E(r_i)$$

$$\sigma_p = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \sigma_{i,j}}$$

where w_i represents the proportion of the portfolio allocated to asset i, σ_i and σ_i^2 correspond to asset i's standard deviation and variance respectively, and σ_{ij} corresponds the covariance between the returns on assets i and j.

Risk Management Application

For a two-asset portfolio, portfolio variance is written as:

$$\sigma_p^2 = w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\sigma_{12}$$
$$= w_1^2\sigma_1^2 + (1-w_1)^2\sigma_2^2 + 2w_1(1-w_1)\sigma_{12}$$
$$= w_1^2(\sigma_1^2 + \sigma_2^2) + 2w_1(\sigma_{12} - \sigma_2^2) + \sigma_2^2 - 2w_1^2\sigma_{12}.$$

Suppose we wish to determine the values for w_1 and w_2 which *minimize* σ_p^2.

Risk Management Application

Since $\sigma_p^2 = w_1^2(\sigma_1^2 + \sigma_2^2) + 2w_1(\sigma_{12} - \sigma_2^2) + \sigma_2^2 - 2w_1^2\sigma_{12}$,

$\sigma_p{}^2$ is minimized by differentiating it with respect to w_1, setting the result equal to zero, and solving for w_1

$$\frac{d\sigma_p^2}{dw_1} = 2w_1(\sigma_1^2 + \sigma_2^2) + 2(\sigma_{12} - \sigma_2^2) - 4w_1\sigma_{12}$$

$$= w_1(\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}) + \sigma_{12} - \sigma_2^2 = 0.$$

Solving the equation above for w_1, we find that

$$w_1 = \frac{\sigma_2^2 - \sigma_{12}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}.$$

Statistics Class Problem

Suppose the return distributions for two risky assets are as follows:

State	p_s	$r_{a,s}$	$r_{b,s}$
1	1/3	-3%	36%
2	1/3	9%	-12%
3	1/3	21%	12%

1. Calculate the expected returns for assets *a* and *b*.
2. Calculate the variances and standard deviations for assets *a* and *b*.
3. Calculate the covariance and correlation between assets *a* and *b*.
4. Calculate the expected return and standard deviation for an equally weighted portfolio consisting of asset *a* and *b*.
5. Determine the least risky combination of assets *a* and *b* and calculate the expected return and standard deviation for such a portfolio.