Elementary statistics

Expected values

Mean or *expected value* of a random variable X is weighted average of the possible outcomes

$$\mu_X = E(X_t) = p_1 X_1 + p_2 X_2 + \ldots + p_N X_N = \sum_{i=1}^{N} p_i X_i$$

where p_i is the probability that X_i occurs, $\sum p_i = 1$ and $E(.)$ is the expectations operator. The expected value should be distinguished from the *sample mean* which tells us the average of the outcomes obtained in a sample in which a number of observations are chosen from the underlying probability distribution. We denote sample mean of a set of outcomes on X by $\bar{X}$ (see next section).

The *variance* provides a mesures of the spread around the mean

$$\text{var}(X) = \sigma_X^2 = \sum_{i=1}^{N} p_i [X_i - E(X)]^2$$

or equivalently

$$\sigma_X^2 = E[X - E(X)]^2$$

The square root of the variance is called the *standard deviation*

$$\text{std}(X) = \sqrt{\sigma_X^2} = \sigma_X$$

Covariance is a measure of the linear association between X and Y

$$\begin{aligned} \text{cov}(X, Y) = & \quad E[(X - E(X))(Y - E(Y))] \\ = & \quad \textstyle\sum_{i=1}^{N} p_i \sum_{j=1}^{N} p_j (X_i - E(X))(Y_j - E(Y)) \end{aligned}$$

Correlation coefficient measures also linear relationship between X and Y but unlike covarince it is normalized and is *scale-free*. It always lie between -1 and + 1.

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

where σ_X and σ_Y are standard deviations of X and Y.

Estimators of Mean, Variance and Covariance

Usually, we make inferences about the above mentioned characteristics from the sample (we take a sample of N data points). We cannot know the true mean and variance of a random variable, or the true covariance between two random variables, we use the sample information to obtaing the best possible estimates. The estimator $\bar{X}$ for random variable X with mean μ_X is defined by

$$\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$$

$\bar{X}$ is a random variable whose values will vary from sample to sample even thought the corresponding (true mean) μ_X remains unchanged.
The (unbiased) estimator for variance

$$\widehat{\text{var}}(X) = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})^2$$

The sample variance is distinguished from the true variance by placin a hat hat ($\hat{}$) above the cov.
The (unbiased) estimator for covariance

$$\widehat{\text{cov}}(XY)\frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})(Y_j - \bar{Y})$$

The sample correlation coefficient is

$$r_{XY} = \frac{\sum_{i=1}^{N}(X_i - \bar{X})(Y_j - \bar{Y})}{\sqrt{\sum_{i=1}^{N}(X_i - \bar{X})^2 \sum_{i=1}^{N}(Y_i - \bar{Y})^2}}$$

or

$$r_{XY} = \frac{\widehat{\text{cov}}(XY)}{\sqrt{\widehat{\text{var}}(X)\widehat{\text{var}}(Y)}}$$

Some useful properties of the expectations operators

Rule 1

$$E(aX + b) = aE(X) + b$$

where X is a random variable and a and b are constants

Rule 2

$$E(aX)^2 = a^2E(X^2)$$

Rule 3

$$\mathrm{var}(aX + b) = a^2\mathrm{var}(X)$$

Rule 4

If X and Y are random variables then

$$E(X + Y) = E(X) + E(Y)$$

Rule 5

$$\mathrm{var}(X + Y) = \mathrm{var}(X) + \mathrm{var}(Y) + 2\mathrm{cov}(X, Y)$$

Rule 6

If X and Y are independent then

$$E(XY) = E(X)E(Y)$$

Rule 7

If X and Y are independent then

$$\mathrm{cov}(XY) = 0$$

References

Pindyck, R. S., Rubinfeld, D. L. *Econometric Models and Economic Forecasts,* Irwin/McGraw-Hill, Fourth edition, 1998, Chapter 1 and 2, STM-126.