STATISTICAL REVIEW

Statistics is a fundamental aspect of modern finance. Fortunately, only a few statistical concepts are important for us, and they are fairly easy to understand if you take the time to work through a couple of example problems. This appendix is intended to get you up to speed even if you have never had a course in statistics.

For the study of finance you will find it is useful to be comfortable with the following statistical terms and concepts:

mean

variance

standard deviation

covariance

correlation

regression

This review includes definitions, examples, and exercises that should be adequate for anything you encounter in this text.

Mean

The mean of a sample or population is a measure of central tendency. It is simply the average or expected value of the distribution.

You may encounter two different types of data: sampling data (where the observations are equally likely) or probabilistic data (which may be based on judgment or past experience).

If you are working with sampling data, then the mean is calculated as a simple (unweighted) average of the observations. If you look at a sample of stock returns (holding period returns as discussed in the chapter), the mean is calculated as:

$$\bar{r} = \sum_t r_t / n$$

where $\bar{r}$ is the mean return, r_t is the observed return for time t and n is the number of observations in the sample.

Example 1 You observe the following five annual returns (percentage change in price plus dividend yield) for the stock of your company:

20%, 5%, -10%, 15%, 20%

The mean return is:

(20 + 5 - 10 + 15 + 20)/5 = 50/5 = 10%

If you are working with probabilistic data, then the mean return is calculated as a probability weighted average:

$$\bar{r} = p_s r_s$$

where p_s is the probability of state of nature s and r_s is the return on an asset that is conditional on the occurrence of state s.

Example 2 The probability of the bad state is 20 percent, in which case the return will be minus 10 percent. The probability of the good state is 30 percent, with a return of 30 percent. The probability of the intermediate state is 50 percent, with a return of 10 percent.

The mean return is:

$$(.2\times(-10) + .3\times 30 + .5\times 10) = -2 + 9 + 5 = 12\%$$

Variance

Variance is a measure of dispersion around the mean. It is equal to the expectation of the squared deviations about the mean.

For sampling data, the variance is calculated by figuring the deviations from the mean, squaring them, and then taking their average.

$$\sigma_r^2 = \sum_t (r_t - \bar{r})^2/n$$

where σ_r^2 is the variance of r.

Note that σ_r is the notation for the standard deviation of r. The variance is equal to the square of the standard deviation, hence, σ_r^2. However, it is necessary to calculate the variance first in order to determine the standard deviation.

For small samples a better estimate of the population variance is found by correcting for sampling degrees of freedom—dividing by n _ 1 instead of *n*. The statistical functions in many computer programs do not correct for degrees of freedom. Also, the text normally does not make the correction because, in practice, in finance we normally work with large numbers of observations. Whether you correct for sampling is not critical in this text, but it may be in other settings.

Example 3 Find the variance from Example 1.

Deviation from Mean $(r_t - \bar{r})$	Deviation Squared $(r_t - \bar{r})^2$	Mean Squared Deviation
(20 – 0) = 10	100	
(5 – 10) = -5	25	
(-10 – 10) = -20	400	
(15 – 10) = 5	25	
(20 – 10) = 10	100	
	sum = 650	sum/n = 650/5 = 130 (650/4 = 162.5 corrected)

For probabilistic data, the variance is calculated as a probability weighted average of the conditional squared deviations.

$$\sigma_r^2 = p_s(r_s - \bar{r})^2$$

Example 4 Find the variance from Example 2.

Deviation from Mean $(r_t - \bar{r})$	Deviation Squared $(r_t - \bar{r})^2$	Probability (decimal form)	Probability × Squared Deviation
(-10 – 12) = 22	484	.20	96.8
(30 – 12) = 18	324	.30	97.2
(10 – 10) = -2	4	.50	2.0
		1.00	Mean = 196.0

Standard Deviation

The standard deviation of a distribution is the most commonly used measure of dispersion. It is calculated by simply taking the square root of the variance. The reason we need the standard deviation in addition to the variance is that the standard deviation has some useful mathematical properties. For example, when two distributions are perfectly positively correlated with each other, the standard deviation of a weighted average of the two distributions will equal the weighted average of the standard deviations. This is not true for the variance.

$$\sigma_r = \sqrt{\sigma_r^2}$$

Example 5 Find the standard deviation of the distribution in Example 1.

$$\sqrt{130} = 11.4\%$$

Example 6 Find the standard deviation of the distribution in Example 2.

$$\sqrt{196} = 14\%$$

Covariance

The covariance is a dispersion measure of how two distributions relate to each other. Unlike the variance, a covariance can be either positive or negative and be of any magnitude. The absolute magnitude increases with the underlying variances of the two distributions and with the degree of positive or negative association.

For sampling data, the covariance is calculated as the average of the products of the deviations between the two distributions.

$$Cov_{r_1,r_2} = \sigma_{1,2} = \sum_t (r_{1t} - \bar{r}_1)(r_{2t} - \bar{r}_2)/n$$

Example 7 The distribution for stock 1 is shown in Example 1. The distribution for stock 2 (ordered in the same way—first annual return listed first) is as follows: 12%, 20%, 8%, 5%, 0%. Calculate the covariance between the returns of the two stocks. (Note that the mean return for stock 2 is 9%.)

Deviation from Mean for Stock 1	Deviation Squared for Stock 2	Product of Deviations for Stocks 1 and 3	
(20 – 10) = 10	(12 – 9) = 3	10 × 3 = 30	
(5 – 10) = -5	(20 – 9) = 11	(-5) × 11 = -55	
(-10 – 10) = -20	(8 – 9) = -1	(-20) × -1 = 20	
(15 – 10) = 5	(5 – 9) = -5	5 × (-4) = -20	
(20 – 10) = 10	(0 – 9) = -9	10 × (-9) = -90	
		sum = - 115	sum/n = -115/5 = -23

Example 8 Distribution data for stock 1 are in Example 2: The states of nature are the same for stock 2, and the conditional returns are 0%, 10% , and 4%, respectively (with the mean 5%).

Deviation from Mean for Stock 1	Deviation from Mean for Stock 2	Product of Deviations for Stocks 1 and 2	Probability	Probability Times Product of Deviations
(-10 – 12) = 22	(0 – 5) = -5	(-22) × (-5) = 110	.20	22
(30 – 12) = 18	(10 – 5) = 5	18 × 5 = 90	.30	27
(10 – 10) = -2	(4 – 5) = -1	(-2) × (-1) = 2	.50	1
		sum = 202	1.00	Cov = 50

Correlation

The correlation coefficient is a standardized measure of relatedness between two distributions. As a result of the standardization, it has a maximum value of one (perfect positive correlation) and a minimum value of minus one (perfect negative correlation).

The correlation coefficient is calculated as follows:

$$\rho_{1,2} = Cov_{1,2}/\sigma_1\sigma_2 \qquad or \qquad Cov_{1,2} = \rho_{1,2}\sigma_1\sigma_2$$

where $\rho_{1,2}$ is the symbol for the correlation between distributions 1 and 2.

Example 9 Find the correlation between the two distributions in Example 7. (*Note:* The standard deviation of distribution 2 is 6.88.)

$$\rho_{1,2} = -\frac{23}{11.4 \times 6.88} = -.293 \quad \text{(moderate negative correlation)}$$

Example 10 Find the correlation between the two distributions in Example 8 (*Note:* The standard deviation of distribution 2 is 3.61).

Regression

Regression is a statistical measure of the relation between one (dependent) variable and one or more independent variables that allows you to predict how much the dependent variable will change in response to a change in one of the independent variables. There is more to regression analysis than this, but it is not necessary to go into it in this review.

The regression coefficient between two variables depends on which one you are calling independent. For example, if variable 1 is dependent and variable 2 is independent, then the regression coefficient is calculated as follows:

$$\beta_1 = Cov_{1,2}/Var_2$$

where _1 is referred to as the slope coefficient that describes how variable 1 is expected to change in response to a change in variable 2.

Example 11 Calculate the regression coefficient for the distributions in Example 7 if the first distribution is treated as dependent.

$$\beta_2 = -\frac{23}{47.3} = -.486$$

Note that 47.3 is the variance of distribution 2.

Example 12 Calculate the regression coefficient for the distributions in Example 8 if the second distribution is treated as dependent.

$$\beta_2 = \frac{50}{196} = .225$$

Why do you think beta 2 is less than beta 1 in absolute value, even though the correlation in Example 8 is much higher (in absolute value) than that in Example 7?

Questions and Problems

1. Here is a sample of monthly returns: 5%, 7%, 2%, -3%, 0%, 5%. Find the mean monthly return.

2. Here are some probabilistic daily returns: a 10% chance of -1%, a 20% chance of 1%, a 30% chance of 0% and a 40% chance of 0.5%. Find the expected (mean) daily return. It is a good idea to also try these problems using a spreadsheet program on your computer.

3. Calculate the variance of monthly returns from problem 1.

4. Calculate the variance of daily returns from problem 2.

5. Find the standard deviation of the distribution in problem 1.

6. Find the standard deviation of the distribution in problem 2.

7. Make up some returns for a second security to use in conjunction with the distribution in problem 1 and calculate the covariance. Try to select returns that make the covariance positive.

8. Make up some returns for a second security to use in conjunction with the distribution in problem 2 and calculate the covariance. Try to select returns that make the covariance negative.

9. Calculate the correlation coefficient of the distributions from problem 7.

10. Calculate the correlation coefficient of the distributions from problem 8.

11. Calculate the regression coefficient for the distributions in problem 7 if the second variable is dependent.

12. Calculate the regression coefficient for the distributions in problem 8 if the first variable is dependent.