Statistical Concept And Market Returns

Exam Focus：

1. 了解中心趨向的量度(含算術平均、幾何平均、加權平均、中位數、眾數)
2. 了解分散度的量度(含全距、平均絕對離差、母體/樣本變異數、母體/樣本標準差)
3. 中心趨向代表投資報酬、分散度代表投資風險
4. 了解常態分配，及偏離常態的情況(偏態:缺乏對稱姓；峰態:峰的程度)

7.a	Distinguish between descriptive statistics and inferential statistics, between a population and a sample, and among the types of measurement scales;

Statistics is used to refer to data and to the methods we use to analyze data.
Statistical methods fall into two categories：

1. Descriptive statistics (敘述統計):

- 總結大型數據集合的重要特徵(large date sets)
- 將很多數據整合為有用的訊息

2. Inferential statistics (推論統計):

- 主要根據一個小型數據(smaller set)集合的統計特徵，來對一個大的數據(large set)集合進行預測(forecasts)、估計(estimates)或判斷(judgments)

<u>Types of measurement scales:</u>

1. Nominal scales (名目尺度):

- 最不精確
- 類別之間沒有順序分別
- Example：1 = stocks, 2 = bonds...

2. Ordinal scales(順序尺度):

- 比名目尺度更進一步
- 每個觀測值被指定到一個類別中，但這些類別是按照一個給定的特徵排序
- Example：The ranking of 1,000 stocks by performance. 1 = the 100 best performing stocks, 2 = the next 100 best performing stocks, ..., 10 = the 100 worst performing stocks.
- 根據上述例子我們可以說類別3表現比4好，但是還不夠精確，我們不能說3和4的差距有多少，或3和4差距和4與5之間差距有無差別。

3. Interval scales(區間尺度):無真正原點的量

- 和順序尺度類似，但是相鄰尺度間的差別是相同的
- 缺點：0不代表甚麼都不存在，比例也沒有意義

- 如攝氏0度不代表沒有溫度，20度也不代表比10度熱2倍，但10度和20度 VS. 20度和30度中間的差距一樣都是10度

4. Ratio scales (比例尺度):有固定原點的量

- 最精確的度量，有一個真正的0作為起始點
- 順序、區間、比例都有意義，0代表真的甚麼都沒有，4塊錢的購買力比2塊錢多2倍

注意：請按精確順序記住這些尺度 Nominal<Ordinal<Interval<Ratio

【Example 7-1】

Identify the type of scale for each of the following:

a. Stocks ranked as big-size, middle-size, or small-size.

b. Birds divided into categories of songbirds, birds of prey, and game birds.

c. Interest rates on T-bonds each year for 30 years.

d. The weight of each student on the class.

e. The average temperature in January in Taipei.

【Answer】

a. An ordinal scale

b. A nominal scale

c. A ratio scale

d. A ratio scale

e. A interval scale

7.b	Define a parameter, a sample statistic, and a frequency distribution;
7.c	Calculate and interpret relative frequencies and cumulative relative frequencies, given a frequency distribution;

<u>Sample (樣本), Population (母體), and Parameter (參數)</u>

母體(population)被定義為一個研究對象組中所有可能成員的集合，舉例而言紐約證交所 NYSE 中所有交易的股票收益率橫斷面資料(cross-section data)就是一個總體。我們不可能對一個母體的全部個體做研究，因此通常會採取一個樣本(Sample)，樣本是母體中的一個子集合(subset)，例如從 NYSE 中抽取 30 支股票，透過研究樣本特徵來描述母體特徵。

參數(parameter)：用來描述一個總體特徵的量度，投資分析通常只在意幾個參數，例如平均報酬率、報酬率的變異數，我們通常以 μ 表達母體平均數，以 σ^2 表達母體變異數。

類似的，用來描述樣本特徵的量度稱作樣本統計量(sample statistics)，我們通常以上橫線表達樣本平均數（如 $\overline{X}$ ），以 S^2 表達樣本變異數。

Frequency Distribution (次數分配), Relative Frequencies (相對次數分配), and Cumulative Relative Frequencies (累積相對次數分配)

There are three steps to construct a frequency distribution:

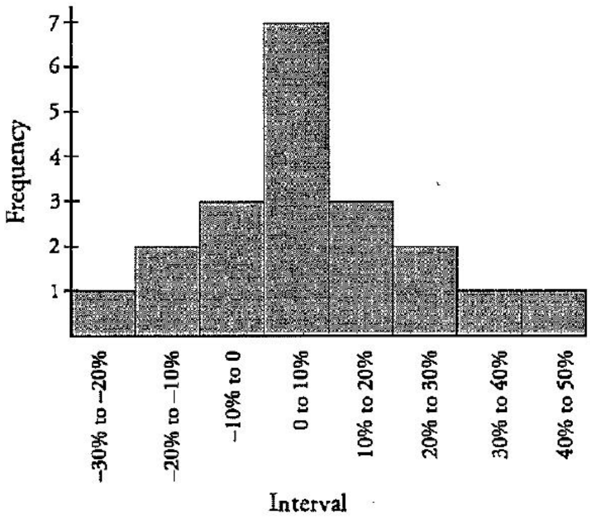
1. 定義區間 Define the intervals.(區間太小，數據可能被高度概括，重要特性可能被遺漏；區間太大，數據無法充分表達)
2. 歸類觀測值 Tally the observations.
3. 對觀測值進行計數 Count the observations.

Interval	Absolute Frequency	Relative Frequency	Cumulative Absolute Frequency	Cumulative Relative Frequency
-30% ~ -20%	1	5%=1/20	1	5%
-20% ~ -10%	2	10%=2/20	3	15%
-10% ~ 0%	3	15%=3/20	6	30%
0% ~ 10%	7	35%=7/20	13	65%
10% ~ 20%	3	15%=3/20	16	80%
20% ~ 30%	2	10%=2/20	18	90%
30% ~ 40%	1	5%=1/20	19	95%
40% ~ 50%	1	5%=1/20	20	100%
Total	20	100%		

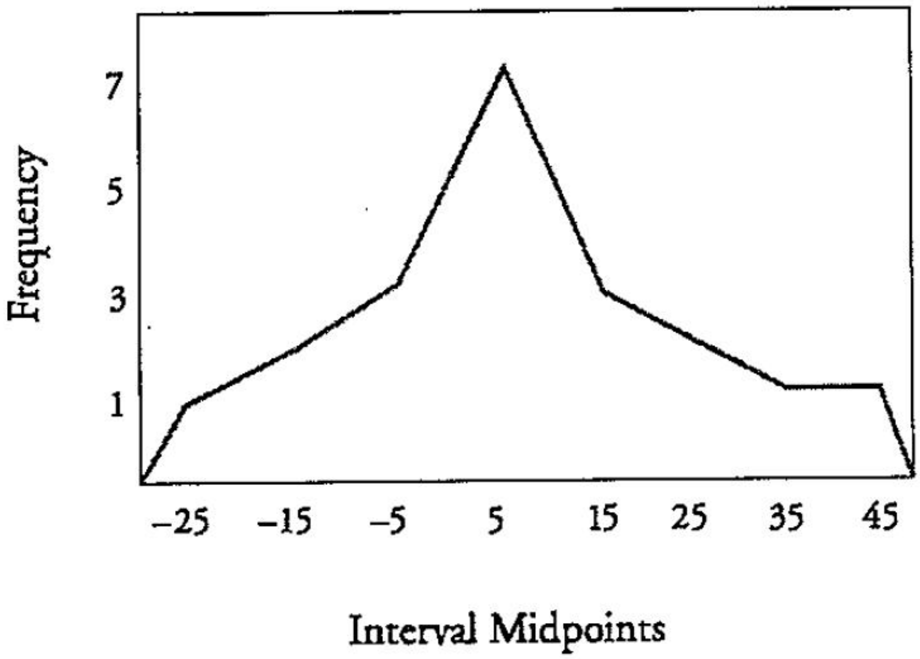
7.d Describe the properties of a data set presented as a histogram or a frequency polygon;

Histogram (直方圖) and Frequency Polygon (次數多邊圖或折線圖)

直方圖(Histogram)是絕對次數分配的圖形表示。



頻率多邊形(**frequency polygon**), 每個區間中點被標示在橫軸，絕對次數的數值標示在縱軸，再把每一點連起來。



7.e	Calculate and interpret measures of central tendency, including the population mean, sample mean, arithmetic mean, weighted average or mean, geometric mean, harmonic mean, median, and mode;
7. m	compare the use of arithmetic and geometric means when analyzing investment returns.

<u>Measures of Central Tendency(中心趨向)</u>

1. Population Mean (母體平均數)： $\mu = \dfrac{\sum_{i=1}^{N} X_i}{N}$

2. Sample Mean (樣本平均數)： $\overline{X}=\dfrac{\sum_{i=1}^{n}X_i}{n}$

3. Arithmetic Mean (算術平均數)：

The arithmetic mean return is simply the sum of each period's asset return divided by the number of periods:

$$AR=\frac{\sum_{i=1}^{n}R_i}{n}$$

4. Geometric Mean (幾何平均數)：

The *geometric mean return* is used when calculating investment returns over multiple periods or to measure compound growth rates:

$$GR=\sqrt[n]{(1+R_1)\times(1+R_2)\times\cdots\times(1+R_n)}-1$$

幾何平均數永遠小於等於算術平均數，差異隨觀測值的分散程度而增大。The geometric mean return approximately equal the arithmetic returns minus half the variance of return. The only time the arithmetic and geometric means are equal is when there is no variability in the observations (i.e., all observations are equal). 只有當所有觀測值大小都一樣時，幾何平均才會等於算術平均。

※算術平均適合預測未來單一期間報酬，幾何平均則適合預測未來多期的複利報酬。

【Example 7-2】

Calculate the geometric and arithmetic mean returns for the S&P 500 for the period 1990 to 1992. The relevant data for the S&P500 are reproduced below

Total Returns for the S&P 500: 1990-1992

Year	Return
1990	-3.17%
1991	0.55%
1992	7.67%

【Answer】

As noted above, return data can be negative. Adding 1.0 to the three returns in the above table results in 0.9683, 1.3055 and 1.0767. We use the formula to solve:

$$1+R_G=\sqrt{0.9683\times 1.3055\times 1.0767}=\sqrt[3]{0.9683\times 1.3055\times 1.0767}=\sqrt[3]{1.3610733}=1.108223$$

$$R_G=10.82\%$$

Now we can compare the geometric mean to the arithmetic mean. The arithmetic mean return for the S&P500 is (-3.17%+30.55%+7.67%)/3=11.68%. Notice that the arithmetic mean is larger than the geometric mean.

【Example 7-3】

A hypothetical investment in a single stock initially costs $100. One year later, the stock is trading at $200. At the end of the second year, the stock price falls back to the original purchase price of $100. No dividends are paid during the two-year period. Calculate the arithmetic and geometric mean annual returns.

【Answer】

The arithmetic mean of the annual returns is (100%-50%)/2=25%

The geometric mean of the annual return is $\sqrt{2.0\times 0.5}-1=0\%$

Because returns are variable by period, the arithmetic mean is greater than the geometric mean. In addition, the geometric mean return of 0 percent accurately reflects that the ending value of the investment in year 2 equals the starting value in Year 1. The compound rate of return on the investment is 0 percent. In general, when we need to calculate a compound rate of return or growth rate, we use the geometric mean.

【Example 7-4】

A portfolio had an initial market value of $100,000. At the end of next two years, the market value was $110,000 and $88,000 respectively. The portfolio's arithmetic and geometric mean returns, respectively, excluding any dividends, were:

a. -12.00%   -6.00%

b. -6.00%   -6.19%

c. -5.00%   -6.19%

【Answer】

Arithmetic mean return =

$$\frac{\left(\frac{110000-100000}{100000}\right)+\left(\frac{88000-110000}{110000}\right)}{2}=\frac{10\%+(-20\%)}{2}=-5\%$$

Geometric mean return= $\sqrt{1.1*0.8}-1=$ -0.0619 = -6.19%

5. Weighted Mean (加權平均數)：通常用來計算投資組合(portfolio)的回報

$$w\overline{X}=\sum_{i=1}^{n}w_iX_i=(w_1X_1+w_2X_2+...+w_nX_n)$$

【Example 7-5】

A portfolio consists of 50% big-size stocks, 40% small-size stocks, and 10% bonds. If the return on big-size stocks is 12%, the return on small-size stocks is 7%, and the return on bonds is 3%, what is the portfolio return?

【Answer】

$w\overline{X}=\sum_{i=1}^{n}w_iX_i=(w_1X_1+w_2X_2+...+w_nX_n)$ = 0.5*0.12 + 0.4*0.07 + 0.1*0.03 = 9.1%

6. Harmonic Mean (調和平均數)：用於求不同時期買入股票的成本

$$\overline{X}_H = \frac{n}{\sum_{i=1}^{n}(1/X_i)}$$

※For values that are not all equal：

<u>Harmonic Mean < Geometric Mean < Arithmetic Mean</u>

【Example 7-6】

Suppose an investor purchases $1,000 of a security each month for 2 months. The share prices are $10 and $15 at the two purchase dates. What is the average price paid for the security?

【Answer】

$$\overline{X}_H = \frac{n}{\sum_{i=1}^{n}(1/X_i)} = \frac{2}{(1/10)+(1/15)} = 12$$

The average is adjusted for the weight each period represents. Because in the first period, the investor can buy 100 shares, whereas the second period he can only buy 66.67 shares. Conceptually easier for you to capture is the following equation

share weighted (cost averaging) share price = $= \frac{100}{166.67} \times \$10 + \frac{66.67}{166.67} \times \$15 = \$12$

7. **Median (中位數)：**

當數據升序或降序排列時，中間點對應的數就是中位數。(若數據為偶數，則取中間兩數的算術平均)。

算術平均容易受極端值影響，此時衡量中心趨勢時，中位數比算術平均數好。

【Example 7-7】

Compute the Median for a data set described as:

Data set：[3000, 21, 15, 25, 23]

【Answer】

First, arrange the returns. → 3000, 25, 23, 21, 15

∴ The median is 23.