Basic Statistics

基本统计

➢ Statistics

- *Descriptive statistics* are used to summarize the important characteristics of large data sets.
- *Inferential statistics* pertain to the procedures used to make forecasts, estimates, or judgments about a large set of data on the basis of the statistical characteristics of a smaller set (a *sample*).
- A *population* is defined as the set of all possible members of a stated group.

Mean

- **Arithmetic mean**
  - Population mean $$\mu = \frac{\sum_{i=1}^{N} X_i}{N}$$
  - Sample mean $$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$
- **Geometric mean**
- **Median**
- **Mode**

Expectations

- **Expected Value**: A Measure of Central Tendency

$$E(X)=\sum P(x_i)x_i=P(x_1)x_1+P(x_2)x_2+...+P(x_n)x_n$$

Properties of Expectation

1. If b is a constant, $E(b)=b$
2. $E(X+Y)=E(X)+E(Y)$
3. In general, $E(XY) \neq E(X)E(Y)$; If X and Y **are *independent*** *random variables*, then $E(XY) = E(X)E(Y)$
4. $E(X^2) \neq [E(X)]^2$
5. If a is a constant, $E(aX)=aE(X)$
6. If a and b are constants, then $E(aX+b)=aE(X)+E(b)=aE(X)+b$

Variance and Standard Deviation

- **Variance**: the measures of dispersion around the mean

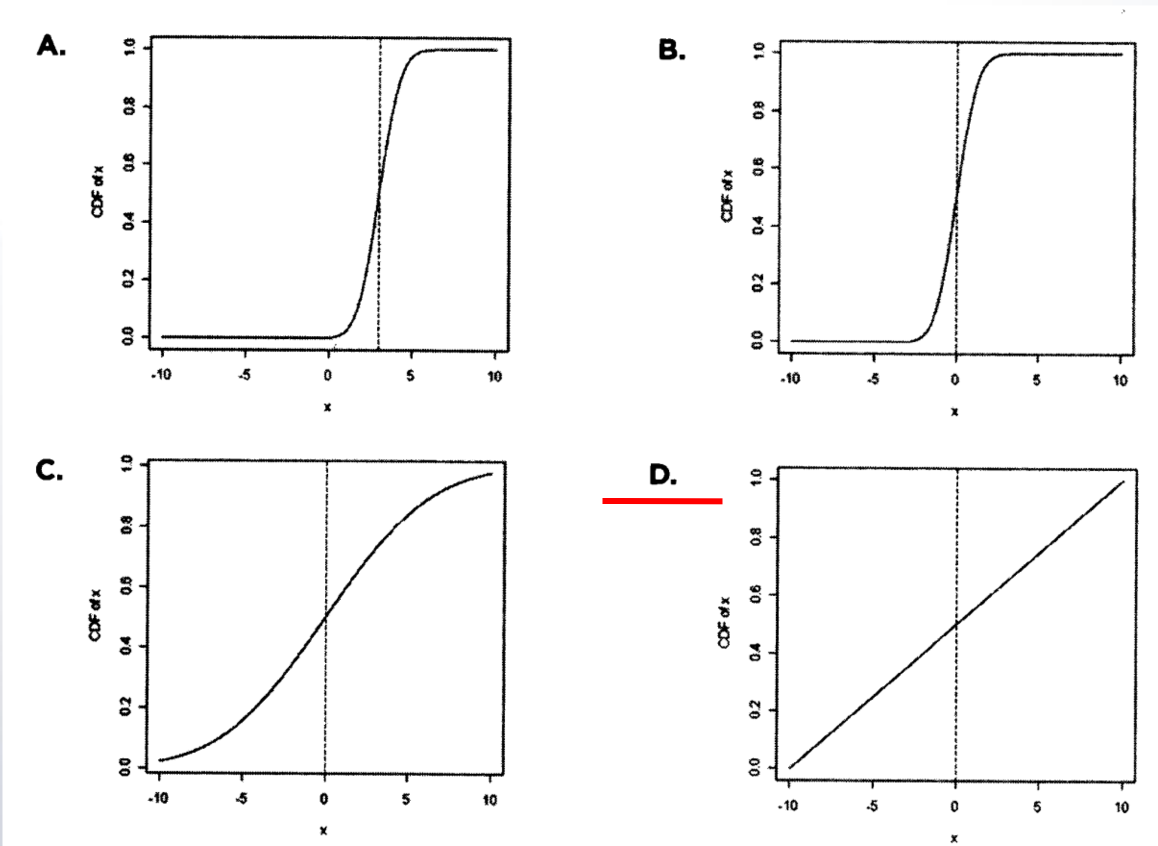
$$Var(X)=\sigma^2=E[(X-\mu)^2]$$

- **Properties of variance**:

1. if c is any constant, then Var(c)=0.
2. If X and Y are two independent random variables, then $Var(X+Y)=Var(X)+Var(Y)$ and $Var(X-Y)=Var(X)+Var(Y)$
3. If c is a constant, then: $Var(X+c)=Var(X)$
4. If c is a constant, then: $Var(cX)=c^2Var(X)$
5. If a and b are constant, then: $Var(aX+b)=a^2Var(X)$
6. If X and Y are independent random variables and a and b are constants, then $Var(aX+bY)=a^2Var(X)+b^2Var(Y)$
7. For computational convenience, we can get: $Var(X)=E(X^2)-[E(X)]^2$

真题回顾

- The following graphs show the cumulative distribution function (CDF) of four different random variables. The dotted vertical line indicates the mean of the distribution. Assuming each random variable can only be values between -10 and 10, which distribution has the highest variance?



Covariance

$$\text{Cov}(X,Y) = E[(X\text{-}E(X))(Y\text{-}E(Y))] = E(XY)\text{-}E(X)E(Y)$$

- Covariance measures how one random variable moves with another random variable.
- Covariance ranges from negative infinity to positive infinity.

Properties of Covariance

1. If X and Y are independent random variables, their covariance is zero.
2. $\text{Cov}(X, X) = E[(X\text{-}E(X))(X\text{-}E(X))] = \sigma^2(X)$
3. $\text{Cov}(a+bX, c+dY)=b\times d\times \text{Cov}(X,Y)$
4. If X and Y are NOT independent, then:

$$\text{Var}(X\pm Y) = \text{Var}(X) + \text{Var}(Y) \pm 2\text{Cov}(X,Y)$$

真题回顾

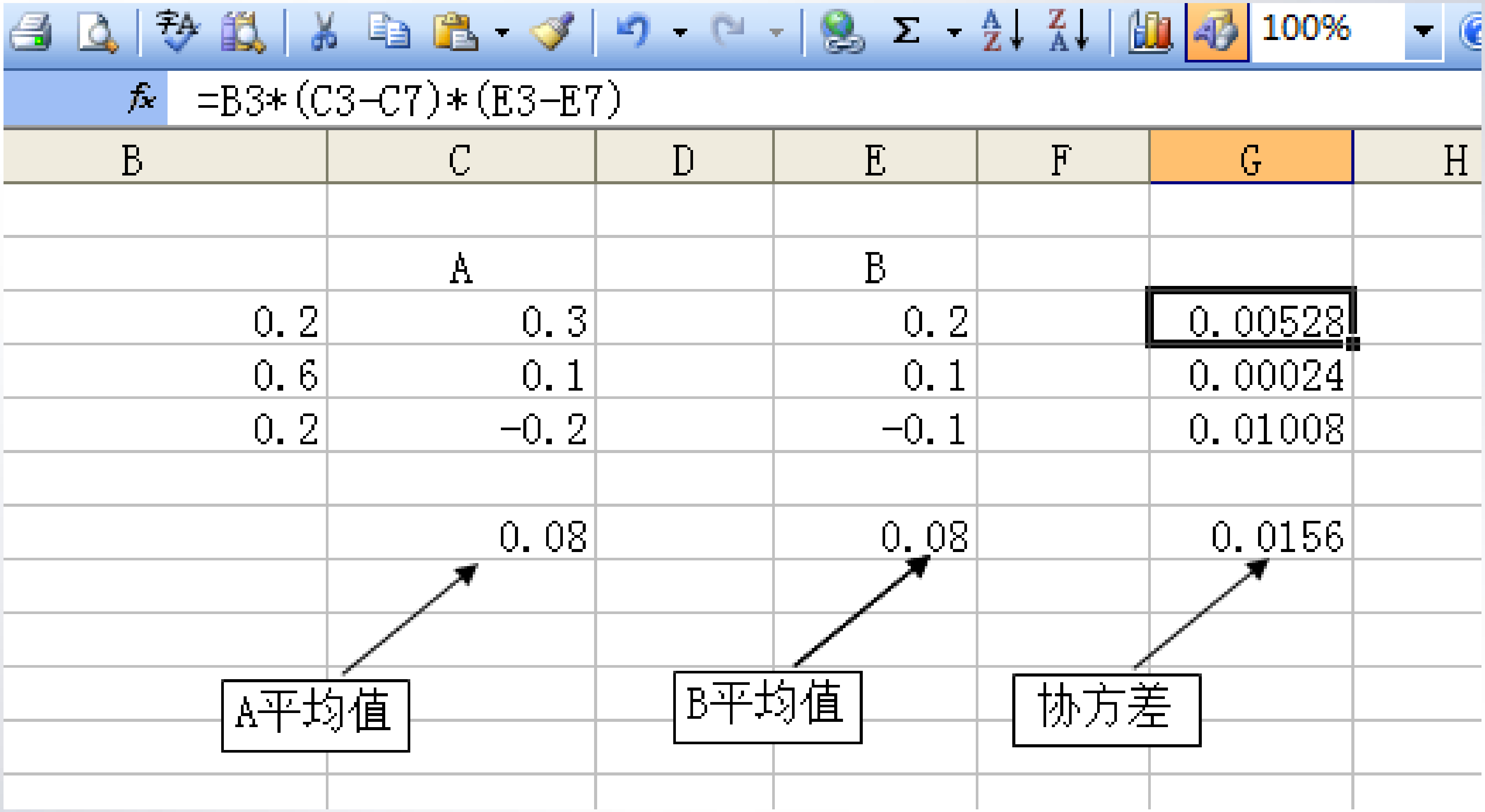
34. An economic analyst has calculated the probabilities of three possible states for the economy next year: growth, normal, and recession. A bank analyst has estimated the possible returns on two stocks, A and B, in each of the three scenarios shown in the following table:

State	Probability	Return of Stock A	Return of Stock B
Growth	0.20	0.30	0.20
Normal	0.60	0.10	0.10
Recession	0.20	-0.20	-0.10

Given that the standard deviations of the estimated returns on stocks A and B are 16.0% and 9.8%, respectively, what is the covariance of the estimated returns on stocks A and B?

A. -0.0187
B. -0.0156
C. 0.0156
D. 0.0178

真题解答


100%
=B3*(C3-C7)*(E3-E7)
B
C
D
E
F
G
H
A
B
0.2
0.3
0.2
0.00528
0.6
0.1
0.1
0.00024
0.2
-0.2
-0.1
0.01008
0.08
0.08
0.0156
A平均值
B平均值
协方差

EXAMPLE 2.4: FRM EXAM 2002—QUESTION 70

Given that x and y are random variables, and a, b, c and d are constant, which one of the following definitions is *wrong*.

a. $E(ax + by + c) = aE(x) + bE(y) + c$, if x and y are correlated.

b. $V(ax + by + c) = V(ax + by) + c$, if x and y are correlated.

c. $Cov(ax + by, cx + dy) = acV(x) + bdV(y) + (ad + bc)Cov(x, y)$, if x and y are correlated.

d. $V(x - y) = V(x + y) = V(x) + V(y)$, if x and y are uncorrelated.

Correlation Coefficient

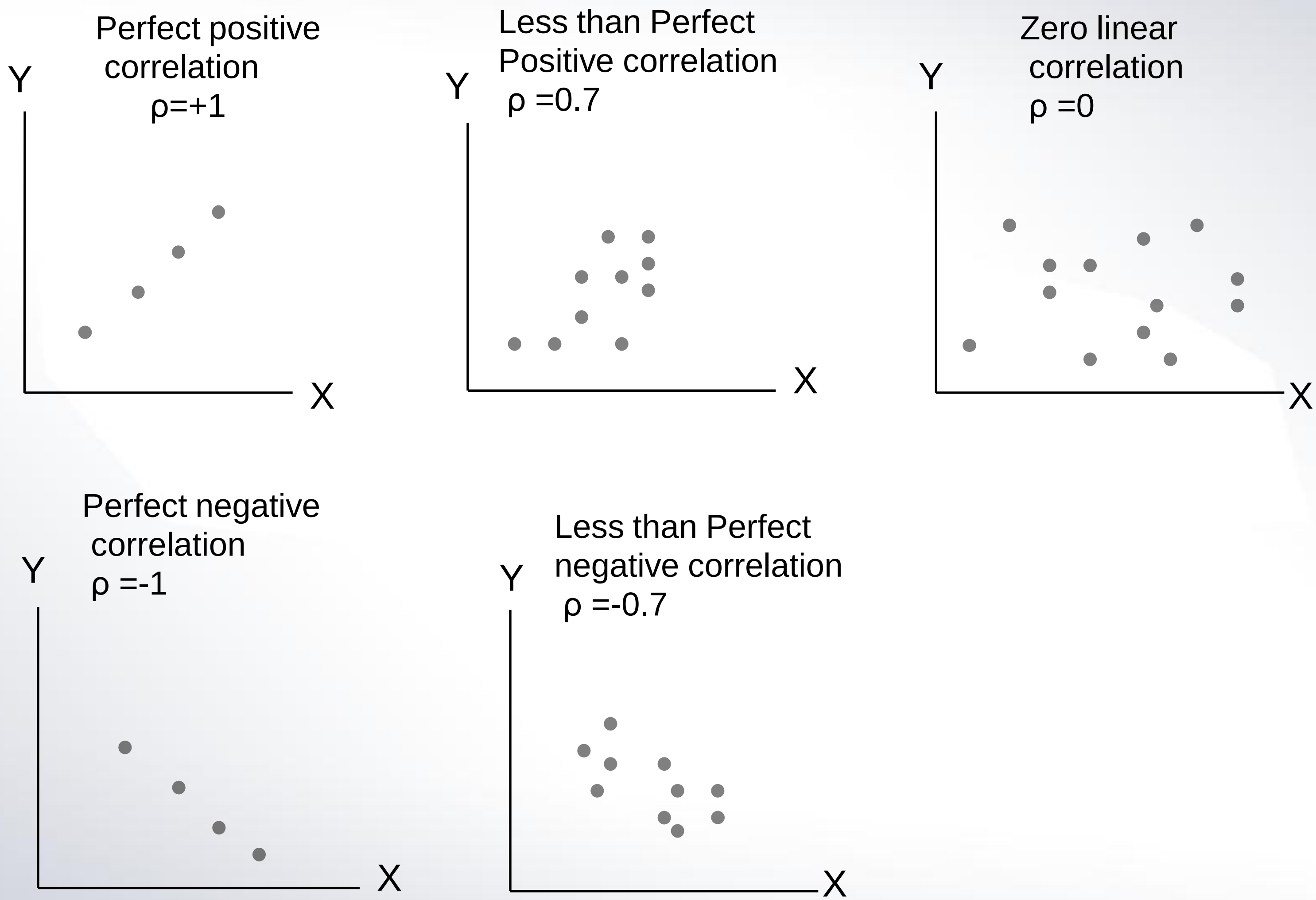
- **Correlation coefficient**

$$\rho_{XY} = \frac{Cov(X,Y)}{\sigma_x \sigma_y}$$

- **Properties of Correlation coefficient**
  - ✓ Correlation measures the *linear relationship* between two random variables.
  - ✓ Correlation has no units, ranges from –1 to +1.
  - ✓ If two variables are independent, their covariance is zero, therefore, the correlation coefficient will be zero. The converse, however, is not true. For example, $Y=X^2$
  - ✓ Variances of correlated Variables.

$$Var(X \pm Y) = Var(X) + Var(Y) \pm 2\rho\sigma_x\sigma_y$$

Correlation Coefficient



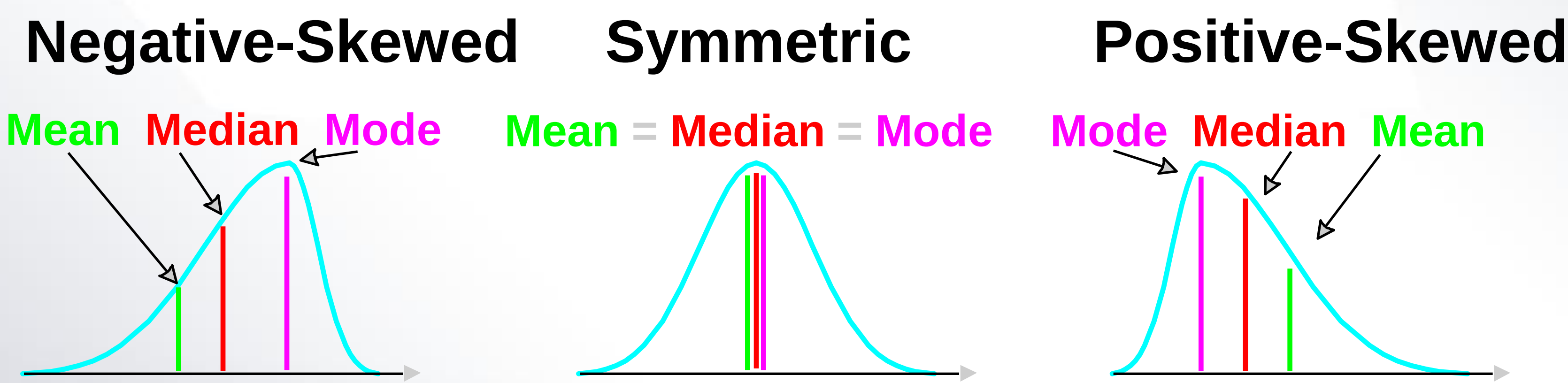
真题回顾

- Which one of the following statements about the correlation coefficient is false?
  - A. It always ranges from −1 to +1.
  - B. A correlation coefficient of zero means that two random variables are independent.
  - C. It is a measure of linear relationship between two random variables.
  - D. It can be calculated by scaling the covariance between two random variables.

Skewness

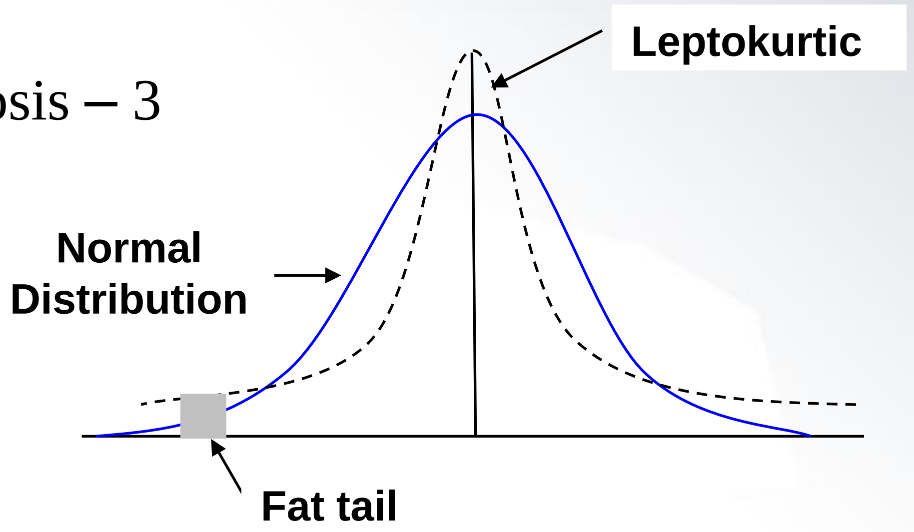
- Symmetrical and nonsymmetrical distributions
- Positively skewed and negatively skewed
- Sample skewness:

$$S = \frac{\frac{1}{n} \times \sum_{i=1}^{n} (X_i - \overline{X})^3}{s^3}$$



Kurtosis

- **Sample kurtosis**: $$K = \frac{\frac{1}{n} \times \sum_{i=1}^{n} (X_i - \overline{X})^4}{s^4}$$
- Excess kurtosis = sample kurtosis – 3



	leptokurtic	Mesokurtic (normal distribution)	platykurtic
Sample kurtosis	>3	=3	<3
Excess kurtosis	>0	=0	<0

真题回顾

- An analyst gathered the following information about the return distributions for two portfolios during the same time period:

Portfolio	Skewness	Kurtosis
A	-1.6	1.9
B	0.8	3.2

The analyst states that the distribution for Portfolio A is more peaked than a normal distribution and that the distribution for Portfolio B has a long tail on the left side of the distribution. Which of the following is correct?

A. The analyst's assessment is correct.

B. The analyst's assessment is correct for Portfolio A and incorrect for portfolio B.

C. The analyst's assessment is incorrect for Portfolio A but is correct for portfolio B.

D. The analyst is incorrect in his assessment for both portfolios.

恭祝大家

FRM学习愉快！

顺利通过考试！