**Ôn tập chương I**

**(Các khái niệm: Population, sample, parameter, statistics, observational study, Retrospective study, experiment; quantitative data, qualitative data, discrete data, continuous data).**

1. A city engineering wants to estimate the average weekly water consumption for single-family dwelling units in the city. 50 single-families are chosen randomly. And it is found that 25 families consumpt 30m^3 water per month. What is population and sample?
2. The population is
3. A collection of observations.
4. A collection of methods for planning studies and experiments.
5. The complete collection of all elements.
6. A sub-collection of members drawn from a larger group.
7. Casualty data from the great flu epidemic of 1918 were collected for a study. This represents what type of study?

A. Observational study

B. Retrospective.

C. An experiment.

D. Qualitative

**ÔN TẬP CHƯƠNG II**

**Các quy tắc tính xác suất (Quy tắc cộng, quy tắc nhân, quy tắc xác suất đầy đủ, quy tắc Bayes; khái niệm disjoint events; independent events)**

1. Tossing a six-sided die and a coin. What is the sample space?
2. The Ski Patrol at Criner Mountain Ski Resort has determined the following probability distribution for the number of skiers that are injured each weekend:

Injured Skiers Probability

0 0.05

1 0.15

2 0.40

3 0.30

4 0.10

What is the probability that the number of injuries per week is at most 3?

1. The probability of a New York teenager owning a skateboard is 0.37, of owning a bicycle is 0.81 and of owning both is 0.36.
2. If a New York teenager is chosen at random, what is the probability that the teenager owns a skateboard o a bicycle?
3. If a New York teenager is chosen at random, what is the probability that the teenager owning a skateboard but not owning a bicycle.
4. Find the probability that the teenager owns a bicycle given that the teenager owns a skateboard.
5. Let P(A) = 0.5; P(B) = 0.4; P(AB) = 0. Which of the following statements are true?
6. A and B are disjoint but not independent.
7. A and B are independent but not disjoint
8. A and B are neither independent nor disjoint
9. A and B are both independent and disjoint.
10. The alarm system at a nuclear power plant is not completely reliable. If there is something wrong with the reactor, the probability that the alarm goes off is 0.99. On the other hand, the alarm goes off on 0.01 of the days when nothing is actually wrong. Suppose that something is wrong with the reactor only one day out of 100. What is the probability that something is actually wrong if the alarm goes off?

**Ôn tập Xác suất- Chương III**

Chương III – **Discrete random variable (Biến ngẫu nhiên rời rạc)**

**Cần nhớ các kiến thức trọng tâm:**

* Tính chất của các hàm f(x) (probability mass function) và F(x) (cumulative distribution function).
* Công thức tính mean (expected value) và variance hoặc standard deviation của biến ngẫu nhiên rời rạc nói chung và của các biến ngẫu nhiên rời rạc cụ thể (như Uniform, Binomial, Geometric, Negative Binomial, Hypergeometric, Poisson).
* Công thức tính xác suất của biến ngẫu nhiên rời rạc nói chung và của các biến ngẫu nhiên rời rạc cụ thể (như Uniform, Binomial, Geometric, Negative Binomial, Hypergeometric, Poisson).

1. The Ski Patrol at Criner Mountain Ski Resort has determined the following probability distribution for the number of skiers that are injured each weekend:

Injured Skiers Probability

0 0.05

1 0.15

2 0.40

3 0.30

4 0.10

Based on this information, what is the expected number of injuries per weekend?

A) 2.50 B) 1.00 C) 2.25 D) 3.50

2) The number of customers that arrive at a fast-food business during a one-hour period is known to be Poisson distributed with a mean equal to 8.60. What is the probability that 2 or 3 customers will arrive in one hour?

A) 0.0263 B) 0.1023 C)0.0679 D) none of the other choices is true

3) The following probability distribution has been assessed for the number of accidents that occur in a mid western city each day:

Accidents Probability

0 0.25

1 0.20

2 0.30

3 0.15

4 0.10

Based on this probability distribution, the standard deviation in the number of accidents per day is:

A) None of the others. B) 2.65 C) 2 D) 0.12

4) Let X be a discrete uniform random variable on the interval [2; 20].

a) Find P(X <13).

b) Find the mean and standard deviation of X.

A) 0 & 30 B) 11 & 30 C) 11 & 5.477 D) None of the others

5) A total of 12 cells are replicated. Freshly-synthesized DNA cannot be replicated again until mitosis is completed. Two control mechanisms have been identified- one positive and one negative- that are used with equal probability. Assume that each cell independently uses a control mechanism.

What is the mean and variance of the number of cells use a positive control mechanism?

A)5 and 6 B) 5 and 4.64 C)6 and 3 D) 4 and 1.73

6) Bill Price is a sales rep in northern California representing a line of athletic socks. Each day, he makes 10 sales calls. The chance of making sale on each call is thought to be 0.30. What is the probability that he will make exactly two sales?.

A) 0.009 B) 0.5002 C) 0.300 D) 0.2335

7) Bill Price is a sales rep in northern California representing a line of athletic socks. Each day, he makes 10 sales calls. The chance of making sale on each call is thought to be 0.30. Find the probability that the first sale call is the fourth call.

A) 0.1029 B) 0.4116 C) 0.4570 D) None of the others.

8) The Ski Patrol at Criner Mountain Ski Resort has determined the following probability distribution for the number of skiers that are injured each weekend:

Injured Skiers (X) Probability

0 0.05

1 0.15

2 0.40

3 0.30

4 0.10

Based on this information, find F(3).

1. 0.85 B) 0.55 C) 0.45 D) None of the others.

9) A clinical trial involves 30 patients. Ten of the 30 are diabetic. If a researcher selects 6 patients at random, what is the probability that three or more of the 6 are diabetic? **(0.3064)**

**Chương IV) Continuous random variable Biến ngẫu nhiên liên tục**

**Cần nhớ các kiến thức trọng tâm:**

* Tính chất của các hàm f(x) (probability density function) và F(x) (cumulative distribution function).
* Công thức tính mean (expected value) và variance hoặc standard deviation của biến ngẫu nhiên liên tuc nói chung và của các biến ngẫu nhiên liên tục cụ thể (như Uniform, Standard Normal – Normal; Exponential)
* Công thức tính xác suất của biến ngẫu nhiên liên tục nói chung và của các biến ngẫu nhiên liên tục cụ thể (như Uniform, Standard Normal – Normal; Exponential).

1. The time it takes to assemble a children's bicycle by a parent has been shown to be normally distributed with a mean equal to 295 minutes with a standard deviation equal to 45 minutes. Given this information, what is the probability that it will take a randomly selected parent between 300 and 340 minutes?. Let P(Z < 0) = 0.5000, P(Z <0.11 ) = 0.5438, P(Z <1 ) = 0.8413

A) 0.0438 B) 0.2975 C) 0.3413 D) 1.000

1. Let X be a normal distribution with the mean of 4 and the variance of 9. Find the value of x such that P(x < X < 7) = 0.5. Let P(Z < 0) = 0.5, P(Z < 1) = 0.8413, P(Z < -0.4) = 0.3413.

A)0 B) 2.8 C) 7 D) 4

1. If the time it takes for a customer to be served at a fast-food chain business is thought to be uniformly distributed between 3 and 8 minutes,
2. what is the probability that the time it takes for a randomly selected customer will be less than 5 minutes?

A) 0.30 B) 0.80 C) 0.40 D) 0.20

b) Find the mean and standard deviation of the time it takes for a customer to be served.

4) The manager of a computer help desk operation has collected enough data to conclude that the distribution of time per call is normally distributed with a mean equal to 8.21 minutes and a standard deviation of 2.14 minutes. The manager has decided to have a signal system attached to the phone so that after a certain period of time, a sound will occur on her employees' phone if she exceeds the time limit. The manager wants to set the time limit at a level such that it will sound on only 8 percent of all calls. Let P(Z < 1.41) = 0.92, P(Z < -1.41) = 0.08, the time limit should be:

A) approximately 5.19 minutes B) about 14.58 minutes.

C) 10.35 minutes. D) about 11.23 minutes.

5) Let X be a continuous random variable with the probability density function

.

1. Find a
2. 1 B) ½ C) 2 D) None of the others
3. Find F(0.5).
4. Find the mean and standard deviation of X.

6) Suppose that a continuous random variable X has probability density function f(x) = 4x3 (0 < x < 1). Find E(X) & V(X)

A) 0.8 & 0.027 B) 0.2 & 0.16 C) 0.45&0.307 D) None of the others

7) Let  be a cumulative distribution function of a continuous random variable X. Find P( X < 0.7).

A) 0.2401 B) 0.3560 D) 0.1207 E) None of the others.

8) Let X be a random variable that have exponential distribution with mean 3. Find P(X > 1).

A) 2.718 B) 3.504 C) 1.024 D) None of the others.

II) Statistics

Chapter 6. (Descriptive Statistics )

(**các khái niệm: Mode, median, range, quartiles, mean, standard deviation, variance, stem-and-leaf plot, dot plot, box plot, scatter plot, frequency distribution, relative frequency distribution.** )

1. Find the mean, variance, standard deviation, mode, median, Quartiles, Interquartile range, lower whisker, upper whisker of the following sample

2, 3, 5, 3, 6, 8, 9, 20, 11, 4, 6.

1. You are given the following data: 23 34 11 40 25 47

Assuming that these data are a sample selected from a larger population, the median value for these sample data is ..........

A)34 B) 25.5 C) 29.5 D) 40

3) Suppose a study of houses that have sold recently in your community showed the following frequency distribution for the number of bedrooms:

Bedrooms Frequency

1 1

2 18

3 140

4 57

5 11

Based on this information, determine the mode for the data.

A) 3 B) 140 C) 4 D) 57

4) The Good-Guys Car Dealership has tracked the number of used cars sold at its downtown dealership. Consider the following data as representing the population of cars sold in each of the 8 weeks that the dealership has been open.

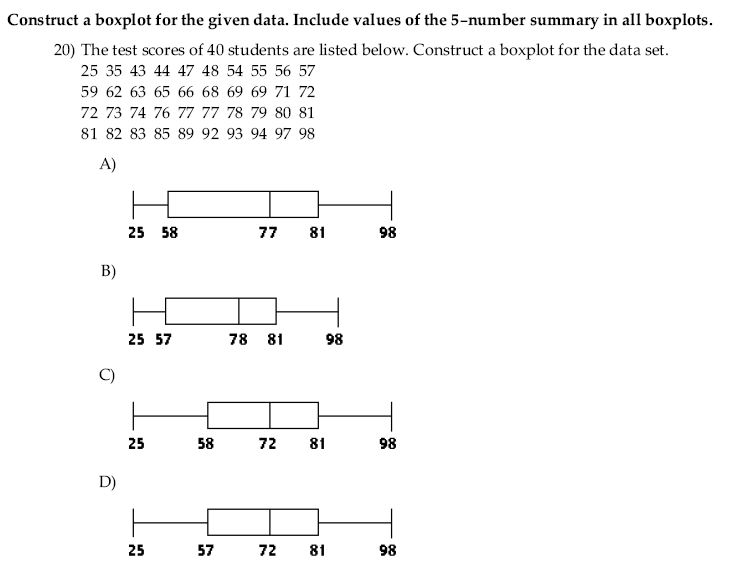
3 5 2 7 7 7 9 0.What is the population standard deviation approximately?

A) 3 cars B) 2.87 cars C) 2.50 cars D) 7 cars

5) You are given the following data: 23 34 11 40 25 47

Assuming that the data reflect a sample from a larger population, what is the sample mean?

A) 30 B) 25 C) 22 D) 32

6) 

**Chapter 7. (The central limit theorem)**

\*) Điều kiện áp dụng: n > = 30 or n bất kì nhưng population có normal distribution.

\*) Nội dung: Nói về phân phối xác suất (sampling distribution) của Statistic : .

\*) Khi n > = 30 thì  có phân phối xấp xỉ chuẩn với mean bằng mean của population, standard deviation = standard deviation of population/√n.

\*) Khi **population có normal distribution** thì **với mọi n**,  có phân phối **chính xác** phân phối chuẩn với mean bằng mean của population, standard deviation = standard deviation of population/√n.

1) if we select a sample with sample size 40 from a population with mean of 20 and standard deviation of 5 then:

A) Sample mean will be approximately normally distributed with mean of 20 and standard deviation of 5.

B) Sample mean will be approximately normally distributed with mean of 20 and standard deviation of 0.79.

C) Sample mean will be exactly normally distributed with mean of 20 and standard deviation of 5.

D) Sample mean will be exactly normally distributed with mean of 20 and standard deviation of 0.79.

2) The monthly electrical utility bills of all customers for the Far East Power and Light Company are known to be distributed as a normal distribution with mean equal to $87 a month and standard deviation of $36. If a statistical sample of n = 100 customers is selected at random, what is the probability that the mean bill for those sampled will exceed $75? Let P(Z < -3.33) = 0, P(Z < 0.33) = 0.63 and P(Z < -0.44) = 0.33.

A) 0.33 B) Approximately 0.63 C) About 1.00 D) None of the others.

3) Two different box-filling machines (A and B)are used to fill cereal boxes on an assembly line. The critical measurement influenced by these machines is the weight of the product in the boxes. Engineers are quite certain that the variance of the weight of product is 1 ounce. Experiments are conducted using both machines with sample sizes of 36 each. Given that the mean of two population are equal. What is the distribution of ?

**Chapter 8: Confidence Interval on population parameters (µ; σ; p).**

1) A major tire manufacturer wishes to estimate the mean tread life in miles for one of their tires. They wish to develop a confidence interval estimate that would have a maximum sampling error of 500 miles with 90 percent confidence. Let population standard deviation equal to 4,000 miles. Based on this information and let z0.05 = 1.645, the required sample size is:

A) 196. B) 124. C) 246. D) 174.

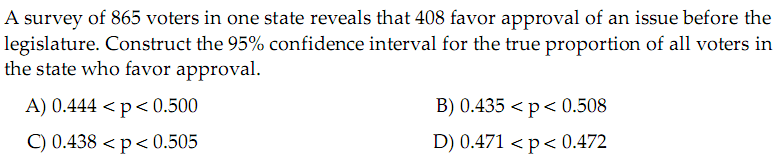
2) Given  = 15.3, s = 4.7, and n = 18, form a 99% confidence interval for σ2. Let 

A) (13.61, 43.30) B) (10.51, 65.88) C) (2.24, 14.02) D) (11.13, 69.79)

3) In an application to estimate the mean number of miles that downtown employees commute to work roundtrip each day, the following information is given: n = 20;  = 4.33; s = 3.50. Based on this information and let t0.025,19 = 2.09, the upper limit for a 95 percent confidence interval estimate for the true population mean is:

A) about 5.97 miles. B) nearly 12.0 miles.

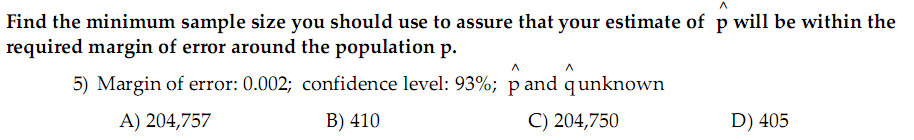
C) about 7.83 miles. D) None of the above.

4) 

Key: C

5) In an application to estimate the mean number of miles that downtown employees commute to work roundtrip each day, the following information is given: n = 20;  = 4.33; s = 3.50; the population is normally distributed. The Confidence Interval on the true population mean with the confident level of 94% is:

A) (2.50; 5.56). B) (2.34; 5.12) C) (2.76; 5.90) D) None of the above.

6) 

Key: C

**Chapter 9: Test of hypothesis on population parameters**

1. Your statistics instructor claims that 60 percent of the students who take her Elementary Statisticsclass go through life feeling more enriched. For some reason that she can't quite figure out, most people don't believe her. You decide to check this out on your own. You randomly survey 64 of her past Elementary Statistics students and find that 34 feel more enriched as a result of her class.

Assume that significance level of 0.05 (z0.025 = 1.96, z0.05 = 1.65). Which of the following states is true?

A)The value of the test statistic is 1.123. There is sufficient evidence to support your statistic instructor's claim

B)The value of the test statistic is -2.97. There is not sufficient evidence to support your statistic instructor's claim

C) The value of the test statistic is -1.123. There is sufficient evidence to support your statistic instructor's claim

D)The value of the test statistic is 2.97. There is not sufficient evidence to support your statistic instructor's claim.

1. According to an article in Newsweek, the natural ratio of girls to boys is 100:105. In Vietnam, the birth ratio is 100: 114 (46.7% girls). Suppose you don't believe the reported figures of the percent of girls born in Vietnam. You think that the percent of girls born in Vietnam is less than 46.7%. You conduct a study. In this study, you count the number of girls and boys born in 150 randomly chosen recent births. There are 60 girls and 90 boys born of the 150. Based on the results, draw your conclusion. Use α = 2% (z0.01 = 2.33 and z0.02 = 2.05).

A) The percent of girls born in Vietnam is more than 46.7%

B) The percent of girls born in Vietnam is equal 46.7%

C) The percent of girls born in Vietnam is less than 46.7%

D) None of the others

1. When a new drug is created, the pharmaceutical company must subject it to testing before receiving the necessary permission from the Food and Drug Administration (FDA) to market the drug. Suppose the null hypothesis is "the drug is unsafe." What is the Type II Error?

A) To claim the drug is safe when, in fact, it is unsafe

B) To claim the drug is unsafe when, in fact, it is unsafe

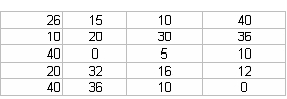
C) To claim the drug is safe when, in fact, it is safe

D) To claim the drug is unsafe when, in fact, it is safe.

1. An assembly line produces widgets with a mean weight of 10 and a standard deviation of 0.2. A new process supposedly will produce widgets with the same mean and a smaller standard deviation. A sample of 20 widgets produced by the new method has a sample standard deviation of 0.126. At a significance level of 10%, what is the value of the test statistic ?

A) 0.234 B)5.77 C) 7.54 D)none of them

1. The cost of a college education has increased at a much faster rate than costs in general over the past twenty years. In order to compensate for this, many students work part- or full-time in addition to attending classes. At one university, it is believed that the average hours students work per week exceeds 20. To test this at a significance level of 0.05 (t0.025,19 = 2.09 and t0.05,19 = 1.73), a random sample of n = 20 students was selected and the following values were observed:



Based on these sample data, the critical value:

A) is equal to 1.73.

B) cannot be determined without knowing the population standard deviation.

C) is approximately equal to 2.09.

D) None of the others.

1. A soft drink company has a filling machine that can be set at different levels to produce different average fill amounts. The company sets the machine to provide a mean fill of 15 ounces. The standard deviation on the machine is known to be 0.20 ounces. Assuming that the hypothesis test is to be performed using a random sample of n = 100 cans, which of the following would be the correct formulation of the null and alternative?

A)H0 : µ = 15 H1 : µ ≠15 ounces

B) H0 :  ≠ 15 H1 :  > 15 ounces

C)H0 : µ ≠ 15 H1 : µ = 15 ounces

D) None of the others.

**Chương 11 Regression and Correlation (**Hồi quy và tương quan)

**(viết phương trình đường hồi quy, tính hệ số tương quan, ý nghĩa của hệ số tương quan, dự báo giá trị của Y khi biết X, kiểm định giả thiết về ba tham số ).**

1) A bank is interested in determining whether their customers' checking balances are linearly related to their savings balances. A sample of n = 20 customers was selected and the correlation was calculated to be +0.40. If the bank is interested in testing to see whether there is a significant linear relationship between the two variables using a significance level of 0.05, what is the value of the test statistic?

A) 1.96 B) 1.8516 C) 1.645 D) 2.438

2) The following regression model has been computed based on a sample of twenty observations: = 34.2 + 19.3x. The first observations in the sample for y and x were 300 and 18, respectively. Given this, the residual value for the first observation is approximately ….

A) 34.2 B) 381.6 C) -81.6 D) -300

3) State University recently randomly sampled seven students and analyzed grade point average (GPA) and number of hours worked off-campus per week. The following data were observed:

y-GPA : 3 2.8 3.7 2.5

x-Hours: 25 30 11 22

Find the simple linear regression equation based on these sample data..

A)  = 4.05 - 0.05x B)  = 3.25 - 0.016x C)  = 7.25 - 0.216x D) None of them

4) Over a period of one year, a greengrocer sells tomatoes at six different prices (x pence per kilogram). He calculates the average number of kilograms, y, sold per day at each of the six different prices. From these data the following are calculated .

Estimate the correlation coefficient.

A) 0.055 B) 0.962 C) -0.962 D) -0.055