

※ 考生請注意：本試題不可使用計算機。請於答案卷(卡)作答，於本試題紙上作答者，不予計分。

Part A Choose the answer (80 points, 5pts each)

- The following data are a sample of body weight (with kg) taken from 6 employees, which are 70, 50, 45, 65, 55, and 45 kg. The mean, variance, and median are
 - mean= 55, variance= 110, median= 50
 - mean= 55, variance= 110, median= 52.5
 - mean= 55, variance= 105, median= 50
 - mean= 55, variance= 105, median= 52.5
- Which of the following description is correct?
 - mean, median, and range are as possible choices for measures of location
 - mean, mode, and standard deviation are as possible choices for measures of location
 - mode, quantiles, and range are as possible choices for measures of spread
 - standard deviation, quantiles, and range are as possible choices for measures of spread
- Which of the following description is *incorrect*?
 - If A and B are independent events, then $\Pr(B | A) = \Pr(B) = \Pr(B | \bar{A})$.
 - Two events A and B are mutual exclusive if $\Pr(A \cap B) = \Pr(A) \times \Pr(B)$.
 - If A and B are any events, then $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$.
 - If two events A and B are independent, then $\Pr(A \cup B) = \Pr(A) + \Pr(B) \times [1 - \Pr(A)]$.
- Chi-squared statistic is often used to test if two categorical variables are associated. Larger value of chi-squared indicates more association. Which of the following statement is *incorrect*?
 - The percentages of cases in the first column within each row of a contingency table are the same if the variables are not associated.
 - The value of chi-squared depends on the number of observations in a contingency table.
 - The value of chi-squared depends on which of two categorical variables defines the rows and which of the two variables defines the columns of the contingency table.
 - All the above statements are correct.

Please use the following information to answer Question 5 to 7

A basketball player attempts 5 shots from the field during a game. This player generally hits about 20% of these shots.

[These values may help you get the answer: $0.2^2=0.04$, $0.2^3=0.008$, $0.2^4=0.0016$, $0.2^5=0.00032$, $0.8^2=0.640$, $0.8^3=0.512$, $0.8^4=0.4096$, $0.8^5=0.32768$]

- How many baskets would you expect this player to make in the game?
 - 1
 - 2
 - 3
 - 4
- What is the probability that this player makes less than 3 (not including 3) shots successfully?
 - 0.328

- B. 0.410
C. 0.737
D. 0.942
7. If this player randomly takes half of the shots from 3-point range and half from 2-point range and makes both with 20% chance, how many points would you expect the player to score?
A. 2.5
B. 3.0
C. 4.5
D. 5.0
8. The current age (in years) of 500 employees at a Hospital is normally distributed with mean 40 and standard deviation 8. Which of the following statement is *incorrect*?
A. Most of the employees at this hospital are older than 32, which roughly take 84% of the whole employees.
B. If the company was to convert these ages from years to days, then the ages in days would also be normally distributed.
C. More employees at this hospital are order than 48 than between 40 and 48.
D. A training program for employees under the age of 35 at this hospital would be expected to attract about 133 employees.
9. A manager wants to investigate the absences among the staffs. The events are defined as $A = \{\text{employee is absent}\}$, and $B = \{\text{employee is sick}\}$. Which of the following statement is correct?
A. The probability of an employee being absent is greater than the probability that the employee is absent given that the employee is sick.
B. If the chance for an employee to be absent is greater that the chance for an employee to be sick, then A and B are dependent events.
C. If the manager knows that $P(A) = 0.1$ and $P(B) = 0.05$, then he/she can find $P(B|A)$.
D. None of above statement is correct.
10. Which of the following statement about "Statistic Power" is *incorrect*?
A. More stringent significance levels require larger samples to achieve the desired power level.
B. Power can be decreased by choosing a less stringent alpha level.
C. Smaller effect sizes always require larger sample sizes to achieve the desired power.
D. Any increase in power is most likely achieved by increased sample size.
11. Which of the following statement is correct?
A. The null hypothesis from a break-even analysis should imply that the two methods are equally profitable.
B. If the standard two-sample t-test rejects $H_0 : \mu_1 - \mu_2 \leq 10$, then we know for sure that μ_1 is more than 10 larger than μ_2 .
C. If the two-sample confidence interval for $(\mu_1 - \mu_2)$ included 0, then we should conclude that $\mu_1 - \mu_2$.
D. If the boxplots of the data for the two groups overlap, then the two means are not significantly different.

Please use the following information to answer Question 12 to 14

Suppose SBP(systolic blood pressure), birthweight (oz), and age (days) are measured for 16 infants. Let SBP be the

dependent variable, and birthweight and age be the independent variables. The outputs, *ANOVA table* and the *parameter estimates table*, of the regression analysis are shown as follows.

Source	SS	F	P-value
Model	591.0356	48.08063	<0.001
Error	79.90186		
Total	670.9375		

variable	Parameter	t value	P-value
Intercept	.53.450	11.794	<0.0001
birth weight	0.126	3.657	0.0029
age days	5.888	8.656	<0.0001

12. The degrees of freedom for the regression model should be
- A. 1
 - B. 2
 - C. 3
 - D. 4
13. The coefficient of determination (R^2), is in the interval
- A. [0.6, 0.7]
 - B. [0.7, 0.8]
 - C. [0.8, 0.9]
 - D. [0.9, 1.0]
14. If the level of significance is set at 0.05, which of following statement is incorrect? Let β_1 and β_2 be the partial-regression coefficients of the variable birthweight and age, respectively.
- A. $H_0: \beta_1 = \beta_2 = 0$
 - B. H_1 : both β_1 and β_2 are not equal to 0
 - C. Both birthweight and age are significant to SBP.
 - D. The degrees of freedom for Total is 15.

Please use the following information to answer Question 15 and 16

Two data sets, X and Y, are shown in Table.

variables	data					average	standard deviation
X	180	210	150	220	240	200	35.36
Y	16	14	3	22	20	15	7.42

15. The covariance of X and Y is between which following interval?
- A. [150, 200]
 - B. [200, 250]
 - C. [250, 300]
 - D. [300, 350]

16. The correlation coefficient, r , is between which following interval?

- A. [-0.9, 0]
- B. [0, 0.1]
- C. [0.7, 0.8]
- D. [0.8, 0.9]

Part B Partial Credit Questions (20 points)

1. [5 points]

A biotechnology company is testing a newly therapy. If the therapy lowers the blood pressure of a patient by more than 5 mm, it is viewed as effective. What are the H_0 (null hypothesis) and H_1 (alternative hypothesis) to test this new therapy?

2. [15 points]

Customers at a fast-food restaurant buy both sandwiches and drinks. The following joint distribution summarizes the numbers of sandwiches (X) and drinks (Y) purchased by customers.

		X	
		1 sandwich	2 sandwiches
Y	1 drink	0.40	0.20
	2 drinks	0.10	0.25
	3drinks	0	0.05

- (a) [4 pts] Find the expected value and variance of the number of sandwiches.
- (b) [4 pts] Find the expected value and variance of the number of drinks.
- (c) [7 pts] If the profit earned from selling a sandwich is \$1.50 and from a dink is \$1.00, what is the expected value and standard deviation of the profit made from each customer?

References:

1. Statistics for Business – Decision making and analysis, by Robert Stine & Dean Foster
2. Fundamentals of Biostatistics, by Bernard Rosner
3. Statistical Quality Control - A modern introduction, by Douglas C. Montgomery

