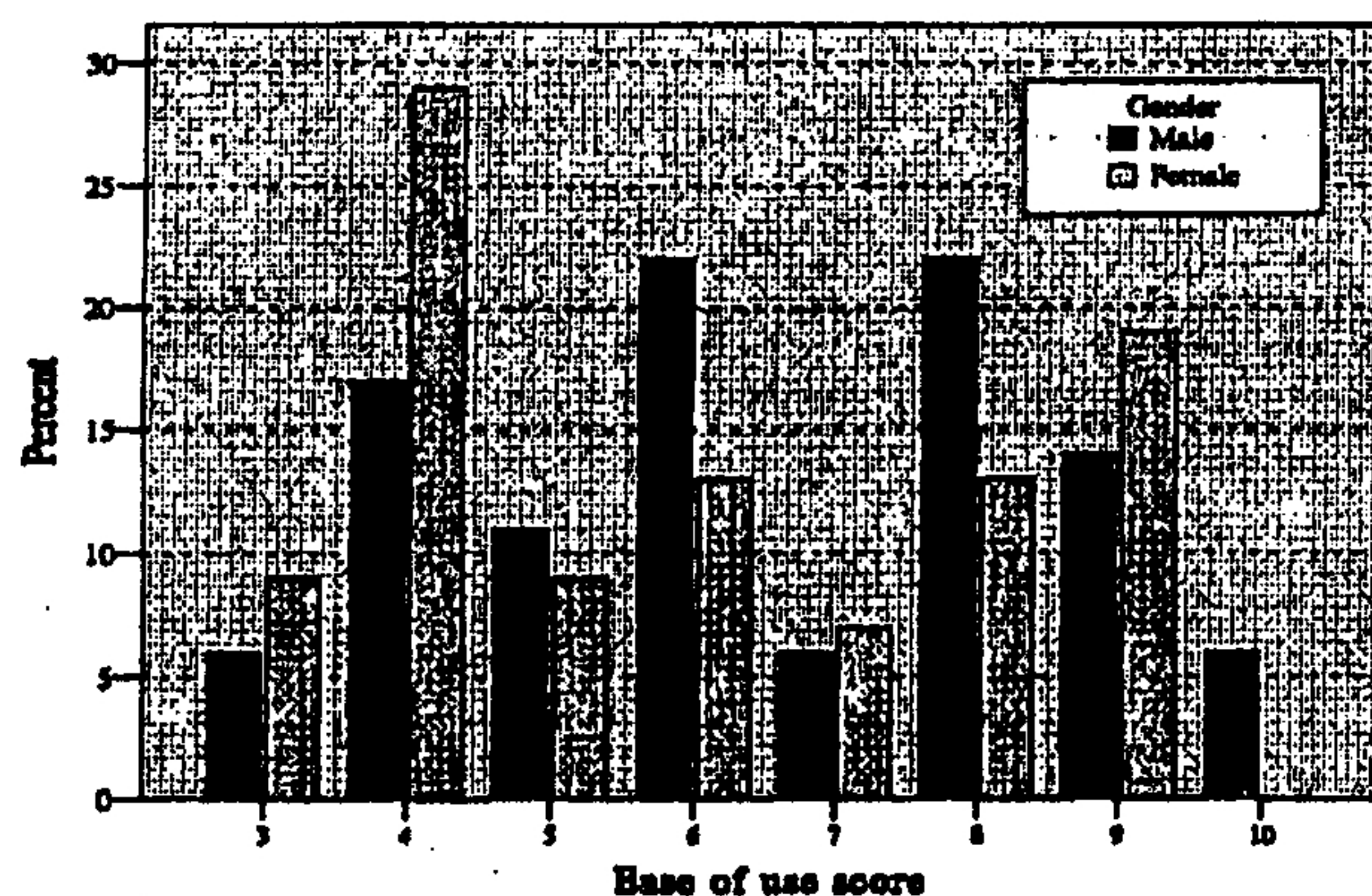


Part I. Multiple Choice Questions (54 points, 3 pts each)

There are five possible responses to each of the following multiple choice questions. There is only one "BEST" or "CORRECT" answer. Be sure to read all possible choices before selecting your answer.

Use the following information to answer Question 1 to Question 4

A distributor of appliances is doing a customer satisfaction survey for a manufacturer of DVD players. A sample of 68 clients is asked to rate a particular DVD player on appearance, functionality, ease of use, and price on a 1 to 10 scale, where 1 corresponds to the worst rating and 10 to the best possible rating. A graph of the ease of use ratings classified by gender is given below.



- This graph is referred to as
 - stem-leaf plot
 - histogram
 - box plot
 - bar chart
 - none of the above
- Which one of the following statements is correct?
 - Gender and score are nominal scales
 - Score is an interval scale
 - Gender and brands of DVD player are categorical variables
 - Price and score are continuous variables
 - None of the above
- Let \hat{p} denote the percentage of the sampled male clients rated the DVD player not easy to use (rating of 4 or less).
 - $\hat{p} = 28\%$
 - $\hat{p} = 17\%$
 - The standard error of \hat{p} can be estimated by $\frac{\hat{p}(1-\hat{p})}{68}$
 - The standard error of \hat{p} is estimated by $\sqrt{\hat{p}(1-\hat{p})/68}$
 - The standard error of \hat{p} cannot be estimated based on the information given here
- The 80th percentile of ease of use score for the sampled female clients is
 - 0.8
 - 4
 - 9
 - 0.2
 - None of the above

(背面仍有題目,請繼續作答)

5. Which one of the following statistics can not measure the variability or dispersion of a data set?

- A) interquartile range B) variance C) covarince
D) standard deviation E) coefficient of variation

Use the following information to answer Question 6 to Question 7

An investigator carries out a large sample two-sided test of the null hypothesis that $\mu = 100$. He reports the value of test statistic of 2.12 with a p-value of 0.034.

6. Which of the following statement is correct?

- A) The 99% confidence interval for μ covers 100.
B) The null hypothesis is rejected at $\alpha = 0.01$.
C) The 95% confidence interval for μ covers 100.
D) The null hypothesis can not be rejected at $\alpha = 0.10$.
E) None of the above

7. Let $z = 2.12$. Which one of the following formulae is used to compute the p-value?

- A) $P(Z > z)$ B) $P(Z > 2z)$ C) $P(Z > z)/2$
D) $2P(Z > z)$ E) $P(Z > z) - P(Z < -z)$

Use the following information to answer Question 8 to Question 12

A kitchen utensil manufacturer selected 18 similar stores to try out three different promotional displays for a new-energy cooking pot. The display that generates the highest sales in this study is to be used in the manufacturer's national promotion program. Each display was assigned at random to six stores. Sales (in dollars) for the stores during the two-week observation period are given as follow and some descriptive statistics are given too.

Display	Store						mean	std
	1	2	3	4	5	6		
X	2161	1769	2748	1782	2830	3183	2412.2	592.5
Y	2379	1913	1119	1208	1962	1689	1711.7	480.4
Z	1479	1024	1598	963	1913	2251	1538.0	500.2

$$\sum(x - \bar{x})(y - \bar{y}) = -91621, \sum(x - \bar{x})(z - \bar{z}) = 1434194$$

8. The data set collected here is based on

- A) simple random sampling B) clustered sampling C) stratified sampling
D) completely randomized design E) randomized complete block design

9. Based on the descriptive statistics given above, the correlation of the sales between display X and display Z is

- A) 0.4513 B) 0.8065 C) 0
D) 4.8392 E) 0.9678

10. How can we interpret this correlation?

- A) No meaning for this study, since this is not a paired data set.
B) A mistake, since the correlation must be less than one.
C) A mistake, since the correlation should be negative.
D) If sales of display X increase then sales of display Z increase too.
E) There is a strong interaction between display X and display Z.

11. What is the best reason for randomly assigning treatment levels to the experimental units?
- A) Randomization makes the experiment easier to conduct.
 - B) Randomization will tend to average out all other uncontrolled factors so that they are not confounded with the treatment effects.
 - C) Randomization is required by statistical consultants before they will help you analyze the experiment.
 - D) Randomization makes the analysis easier because the data can be collected and entered into the computer in any order.
 - E) Randomization is used to remove the effects of another factor from the comparison.
12. If sales are normally distributed, which testing procedure would you use to compare the average sales from these displays?
- A) F test in a one-way ANOVA table for completely randomized design.
 - B) F test for "displays" in a two-way ANOVA table for block design.
 - C) Chi-square test for a 3 by 6 contingency table.
 - D) Paired t -test for a block design.
 - E) Two-sample t test for independent random samples.

Use the following information to answer Question 13 to Question 15

The manufacturer of Anthony Big's exercise equipment is interested in the relationship between the number of months (X) and the number of hours (Y) the customer used the equipment last week. The result was the linear regression $\hat{Y} = 10 - 0.5X$, the number 0.5 in the equation means

13. A) the average customer used the equipment for 30 minutes last week.
B) who has owned the equipment an extra month used the equipment 30 minutes less last week than the average customer who has owned it one month less.
C) who just bought the equipment used it 30 minutes last week.
D) who bought the equipment one-half month ago.
E) None of the above
14. How can we interpret the number 10 in the equation?
- A) The average hours that the customer used the equipment last week is 10.
 - B) The maximum hours that the customer used the equipment last week is 10.
 - C) The median hours that the customer used the equipment last week is 10.
 - D) It is meaningless for this study.
 - E) None of the above
15. If one customer has owned the equipment for 2 years, how many hours he or she used this equipment last week?
- A) -2 hours
 - B) We should not use this regression to make such prediction since the value of explanatory variable (X) is outside the range of X when this regression line is constructed.
 - C) The customer will not use the equipment any more.
 - D) 2 hours
 - E) None of the above.

(背面仍有題目,請繼續作答)

Use the following information to answer Question 16 to Question 18

A company that supplies quotation machines to stockbroker firms has had increased sales volume every year for the past 10 years. Given in thousands of dollars, the sales figures are

year	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
X	1	2	3	4	5	6	7	8	9	10
Sales	60	80	110	130	160	170	190	230	240	270

16. Let the estimated regression line based on (sales, year) be $\hat{Y} = a + b(\text{year})$ with R-square of 0.99 and the estimated regression line based on (sales, X) be $\hat{Y}^* = c + d(X)$, then
- A) $a \neq c$ $b = d$ $\hat{Y}^* = \hat{Y}$ B) $a \neq c$ $b = d$ $\hat{Y}^* \neq \hat{Y}$
 C) $a = c$ $b = d$ $\hat{Y}^* = \hat{Y}$ D) $a \neq c$ $b \neq d$ $\hat{Y}^* \neq \hat{Y}$
 E) $a = c$ $b \neq d$ $\hat{Y}^* \neq \hat{Y}$
17. Which one of the following is not the assumption of the error terms in a simple linear regression?
- A) The error terms are independent.
 B) The error terms have constant variances.
 C) The error terms are normally distributed.
 D) The error terms have zero expected value.
 E) None of the above
18. Any reason(s) why we should question the use of a simple linear regression for this data set for prediction?
- A) No, the assumptions required by the analysis of a simple linear regression are all satisfied.
 B) No, the R-square is very high.
 C) Yes, the sales may be serially correlated.
 D) Yes, the relationship between years and sales is not linear.
 E) None of the above

Part II. Show all work clearly to receive full credit. (46 points)

19. A survey shows that 30% of the fashions that were found to be unprofitable were marketed by the major fashion clothes stores; 60% of the fashions found to be profitable were marketed by the major fashion clothes stores. If 70% of all fashions are profitable to market,
- (a) What is the probability that the major fashion clothes stores market a particular fashion? (3pts)
 (b) Find the probability that a fashion will be profitable if the major fashion clothes stores market it. (5pts)
20. The lifetime of a printer costing 200 is exponentially distributed with mean 2 years.
- (a) The exponential distribution has the form $f(t) = \lambda e^{-\lambda t}$ for $t > 0$, what is λ ? (2pts)
 (b) What is the probability that the printer fails during the first year? (5pts)
 (c) The manufacturer agrees to pay a full refund to a buyer if the printer fails during the first year following its purchase, and a one-half refund if it fails during the second year. If the manufacturer sells 100 printers, how much should it expect to pay in refunds?
 (Hint: $e^{-1} = 0.36788$, $e^{-2} = 0.13534$, $e^{-0.5} = 0.6065$) (10 pts)

21. A random poll of 100 lunch customers is taken, resulting in the following contingency table of observed values.

		Coffee	Tea	Soft Drink	others (Milk, etc.)
age	21-34	6(9.9)	3	30(9.45)	6(iii)
	35-55	10(i)	4	13(7.04)	5(6.72)
	> 55	6(ii)	3(2.30)	4	10(iv)

The chi-square test is proposed to determine whether or not the type of beverage ordered with lunch at a restaurant is independent of the age of the consumer.

- (a) Some expected frequencies are given in the parentheses in the above table; please find expected frequencies in (i) to (iv)? (4pts)
- (b) What is the degree of freedom for the chi-square test based on this data set? (3pts)
- (c) The p-value from chi-square test is 0.0036, what's the conclusion at $\alpha = 0.01$? (5pts)
- (d) Any reason(s) why we should question the use of chi-square test for this data set. (9pts)
- If the answer is yes, please make a suggestion.