Statistics 101 Elementary Statistics
Reviewer for Final Examination S101-FE-001

PART I. TRUE OR FALSE.

1. Conclusions taken from descriptive statistics only apply to the data at hand.
2. In the experiment of tossing two fair coins, the probability that the outcomes are identical is 0.50.
3. In probability sampling, all the elements of the population are given equal chances of inclusion in the sample.
4. Confidence interval estimate for the difference of the mean of paired data requires us to compute for the absolute value of the difference of each pair of observations.
5. As the sample size increases, the probability of rejecting the null hypothesis increases.
6. Systematic sampling requires information on the arrangement of the elements in the sampling frame to determine the reliability of the estimates.
7. Sturge's formula is used to give an approximation of the class size.
8. If A and B are mutually exclusive events with P(A)=0.3 and P(B)=0.5, then $P(A \cap B)=0.15$.
9. A coefficient of correlation for two variables X and Y equal to zero implies that X and Y are not correlated.
10. If a and b are constants and X and Y are independent random variables, then Var(aX – bY) is equal to $a^2Var(X) + b^2Var(Y)$.
11. In cluster sampling, it is ideal to form clusters so that the elements are heterogeneous with respect to the characteristic being studied within each cluster.
12. If the set of observations is negatively skewed, the mode is less than the mean.
13. If the observations in the sample of size 100 are all distinct then the 7th decile is greater than the 3rd decile.
14. If the results of an investigation show that one sleeping tablet works better than another at the 5% level of significance, the conclusion would be similar if tested at the 10% level of significance.
15. In the simple linear regression model, the random error terms are independent from one another and are all normally distributed with mean and variance equal to 0 and 1, respectively.
16. In stratified random sampling, we divide the population into groups called strata and randomly select the strata to which every element of the sampled stratum will be a part of the sample for the study.

PART II. MUTIPLE CHOICE.

1. A rubbing alcohol company has to make their products within a standard of 70% isopropyl solution so that its germ-killing qualities are optimal and that there are no side-effects in the skin of their buyers. The quality control department conducted a survey of their products in the production line and from a sample of 30 units, the average concentration was 72%. At 0.05 level of significance, there was no sufficient evidence to conclude that the concentration of the alcohol they produce was not different from 70%. The value 70% is a

A. Population B. Parameter C. Sample D. Statistic

2. A nongovernment organization compares the household expenditures of two districts in Quezon City. What method of data collection is most appropriate for this particular case?

A. Experimental Method
B. Survey Method
C. Observational Method
D. Use of Existing Studies

3. Suppose that $Z \sim N(0,1)$ and c is a real number. Which of the following is **NOT** correct?

A. $P(Z \geq c) = 1 - P(Z < c)$
B. $P(Z \geq c) = P(Z > c)$
C. $P(Z \geq c) = P(Z \leq -c)$
D. $P(-c < Z < c) = P(Z < c) - P(Z \leq -c)$

4. Suppose that a dataset contains n distinct values (n is an even number) and its distribution is skewed to the right. What percentage of the observations will be greater than the median?

A. More than 50% B. Less than 50% C. Exactly than 50% D. Indeterminate

5. A researcher wishes to study the relationship between two variables X and Y. Five respondents of the same sex, age, and income bracket were selected and the researcher made the appropriate measurements for variables X and Y. The following measurements were obtained:

X	**5**	**4**	**3**	**2**	**1**
Y	**1**	**2**	**3**	**4**	**5**

The value of the Pearson's product moment correlation coefficient is:

A. -1 B. +1 C. -1/2 D. +1/2

6. If a sample has 500 observations, how many observations are in between the 20th and 95th percentiles?

A. 100 B. 475 C. 375 D. 250

7. An article in the *Philippine Consumer Reports* gives the price for 14 different brands of water-packed light tuna based on prices paid in a sample of supermarkets

0.99	1.92	1.23	0.85	0.65	0.53	1.41
1.12	0.63	0.67	0.69	0.60	0.60	0.66

With only the level of measurement as the primary consideration, which central tendency is best and is most applicable to these data?

A. Mean B. Median C. Mode D. None of the above

8. Assuming a linear relationship between X and Y, which of the following is true if the coefficient of correlation r equals -0.30?

A. There is no correlation.
B. Variable X is larger than Variable Y.
C. Variable X is smaller than Variable Y.
D. The slope is negative.

9. An automobile manufacturer claims that a particular make of car averages at least 20 miles per gallon for highway driving. The statistician for a consumer magazine believes this statement is erroneous and that the average mileage per gallon is less than 20. State the hypotheses to use in a statistical test for the statistician to verify her belief.

A. Ho: $\mu = 20$ vs. Ha: $\mu < 20$
B. Ho: $\mu = 20$ vs. Ha: $\mu \neq 20$
C. Ho: $\mu = 20$ vs. Ha: $\mu > 20$
D. Ho: $\mu = 20$ vs. Ha: $\mu \approx 20$

10. We have a negative relationship between number of drinks consumed and number of marks in a driving test. One individual scores 3 on number of drinks consumed, another individual scores 5 on the number of drinks consumed. What will be their respective scores on the driving test if the intercept is 18 and the slope is 3?

A. 51 and 87 B. 27 and 33 C. 9 and 3 D. Impossible to predict

11. A scientist is weighing each of the 30 fishes. She obtains a mean of 30 g and a standard deviation of 2 g. After completing the weighing, she finds that the scale was misaligned, and always under reported every weight by 2 g (i.e., a fish that really weighed 26 g was reported to weigh 24 g). What is the mean and standard deviation after correcting for the error in the scale?

A. 28 g, 2 g B. 30 g, 4 g C. 32 g, 2 g D. 32 g, 4 g

12. Consider rolling a six-sided die once. Let A be the set of outcomes where an odd number comes up. Let B be the set of outcomes where a 1 or a 2 comes up. In terms of A and B, what is the event getting 2 dots?

A. $A^c \cup B^c$ B. $A^c \cap B^c$ C. $A^c \cup B$ D. $A^c \cap B$

13. Which of the following is **TRUE** about the level of significance?

A. It is the probability of rejecting null hypothesis when in fact it is true.
B. It is the probability of rejecting null hypothesis when in fact it is false.
C. It is the probability of accepting null hypothesis when in fact it is true.
D. It is the probability of accepting null hypothesis when in fact it is false.

14. You are allowed to choose four whole numbers from 1 to 10 (inclusive, without repetition). Which of the following is **FALSE**?

A. The numbers 4, 5, 6, 7 have the smallest possible standard deviation.
B. The numbers 1, 2, 3, 4 have the smallest possible standard deviation.
C. The numbers 1, 5, 6, 10 have the largest possible standard deviation.
D. The numbers 1, 2, 9, 10 have the largest possible standard deviation.

15. Which of the following datasets is skewed to the left?

A. 1, 2, 3, 4, 8 B. 10, 12, 14, 18, 20 C. 1, 4, 6, 8, 10 D. 3, 9, 9, 9, 15

16. Suppose we have defined a random variable $T = \dfrac{\overline{X} - \mu}{S/\sqrt{n}}$ which follows a t-distribution with $n - 1$ degrees of freedom. Which of the following statements is an assumption about the random sample from which $\bar{X}$ and S have been computed?

A. The random sample should come from a normal population.
B. The random sample should come from a normal population with an unknown mean and a known variance but the sample size is small.
C. The random sample should come from any continuous population distribution with an unknown mean and variance.
D. The random sample should come from a continuous symmetric population distribution with an unknown mean and variance.

17. Suppose 95% confidence interval was constructed. Which of the following is (are) the cause(s) of the increase in the width of the confidence interval?

I. The level of confidence was set to a value higher than 95% when the other factors are held constant.
II. The Type I error was set to a value higher than 5% when the other factors are held constant.
III. The sample size was reduced given that the other factors are held constant.

A. I only B. I and II only C. I and III only D. I, II and III

18. Assume that in your hand you hold an ordinary six-sided die and a one-peso coin. You toss both the die and the coin on a table. The probability that a tail appears on the coin and any number more than 3 on the die is

A. 1/12 B. 3/12 C. 6/12 D. 8/12

19. Let A be a random variable with $E(A) = 13$ and $Var(A) = 6$. Which among the following is **FALSE**?

A. $E(3A + 10) = 49$
B. $E(A^2) = 175$
C. $Var(10) = 10$
D. $Var(2A) = 24$

20. If the p-value for your test statistic satisfies p-value > 0.25 then:

A. You would not reject Ho for $\alpha = 0.05$.
B. You would reject Ho for $\alpha = 0.05$.
C. You would reject Ho for $\alpha = 0.10$.
D. Your acceptance region has a lower limit of 0.25

21. What type of decision is reached when the calculated value of the test statistic falls in the critical region when a false null hypothesis is being tested?

A. Correct Decision
B. Type I error
C. Type II error
D. Cannot be determined

22. A pretest-posttest designed experiment is interested in estimating the difference in the true pretest and posttest means using a fixed sample of 37 units. The confidence level was supposed to be 0.99, but instead a 95% CI for the difference in means was computed, which of the following statements best describes the miscalculation?

A. The resulting CI will be narrower than the CI based on paired samples.
B. The resulting CI will be wider than the CI based on paired samples.
C. The resulting CI will not differ from the CI based on paired samples.
D. There is no sufficient information to describe the result of miscalculation.

23. High concentration of trace metals in drinking water affects the flavor and poses a health hazard on drinkers. A health officer wanted to compare the zinc concentration found in bottom water and surface water for six randomly selected brands of bottled water. The data are as follows

Location	Brand					
	1	2	3	4	5	6
Bottom water	0.43	0.27	0.39	0.71	0.60	0.59
Surface water	0.42	0.24	0.57	0.62	0.64	0.65

Assuming that the differences in zinc concentration found in bottom water and surface water are normally distributed, test at 5% level of significance whether the data suggest that the true average zinc concentration of the bottom water (μ_1) exceeds that of surface water (μ_2). The appropriate test procedure to be used is

A. z-test on two population means using independent samples
B. t-test on two population means using independent samples
C. t-test on two population means using paired samples
D. Chi-Square Test of Independence

24. The director of a certain company is interested in seeing whether employment application status (accepted, waitlisted, denied acceptance) to his company is related to or independent from the type of community in which an applicant comes from. Here are the results:

Observed Frequencies	Acceptance Status			
Type of Community	*Accepted*	*Waitlisted*	*Denied*	Total
Urban	6	9	9	24
Rural	9	19	23	51
Suburban	7	8	10	25
Total	22	36	42	100

What is the expected number of accepted applicants from the urban community?

A. 5.28
B. 8.64
C. 11.22
D. 27.28

25. The Chi- Square Test of Independence should not be used if more than ___ % of the expected frequencies are less than 5 or when any expected frequency is less than 1.

A. 80
B. 20
C. 70
D. 30

26. If the correlation coefficient r=0.5 then the coefficient of determination is:
(The coefficient of determination is just r^2.)

A. 0.10
B. 0.25
C. 1.00
D. 2.50

27. We have a negative linear relationship between number of drinks consumed and driving test score. One individual scores 3 on number of drinks consumed, while another individual scores 5 on the number of drinks consumed. What will be their respective scores on the driving test if the intercept is 18 and the slope is 3?

A. 51 and 87
B. 9 and 3
C. 27 and 33
D. Impossible to predict