★ **Multiple Choice Questions (80%)**

1. Suppose that the probability of parents having a child with blond hair is 1/4. If there are 6 children in the family, what is the probability that half of them (3 children) will have blond hair?

   A) 0.13　　B) 0.23　　C) 0.33　　D) 0.43

2. A statistics professor classifies his students according to their grade point average (GPA, random variable Y) and their gender (random variable X). The accompanying table gives the proportion of students falling into the various categories. One student is selected at random.

	Under 2.0 ($Y = 0$)	2.0-3.0 ($Y = 1$)	3.0-4.0 ($Y = 2$)
Male ($X = 0$)	0.05	0.2	0.20
Female ($X = 1$)	0.15	0.25	0.15

   What is the approximated value of the conditional variance of the random variable Y conditional on $X = 1$, $Var(Y|X = 1)$?

   A) 0.11　　B) 0.33　　C) 0.55　　D) 0.77

3. Following the information provided in the question 2, what is the approximated value of the correlation coefficient between two random variables X and Y, $Corr(X,Y)$?

   A) -0.13　　B) -0.23　　C) -0.33　　D) -0.43

4. In a survey of 400 likely voters, 215 responded that they would vote for the incumbent and 185 responded that they would vote for the challenger. Let p denote the fraction of all likely voters who preferred the incumbent at the time of the survey, and let $\hat{p}$ be the fraction of survey respondents who preferred the incumbent. What is the standard error of $\hat{p}$?

   A) 0.085　　B) 0.065　　C) 0.045　　D) 0.025

5. On a single toss of a die, we can get only one of six possible outcomes: 1, 2, 3, 4, 5, or 6. These are mutually exclusive events. If the die is fair, the probability of getting a 2 or 3 on a single toss of the die is

   A) 1/3　　B) 1/6　　C) 2/3　　D) 3/4

6. A laboratory blood test is 95% effective in detecting a certain disease when it is, in fact, present. However, the test also yields a "false positive" result for 1% of the healthy people tested. If 0.1% of the population actually has the disease, what is the probability that a person has the disease given that her test result is positive?

   A) 0.27　　B) 0.47　　C) 0.67　　D) 0.87

7. Which one of the following statement about the characteristics of the normal distribution is correct?

   A) It has zero mean　　　　　　　B) It has a fat tail

   C) It is a symmetric distribution　　D) All of the above

8. Which one of the following statement about the central limit theorem is correct?

A) It states that under the general conditions, when the sample size is large, sample average (sample estimator) will be close to the population mean (estimator) with very high probability.

B) The theorem is applied any sample size.

C) It states that under the general conditions, when the sample size is large, the sampling distribution of the standardized sampling average is approximately standard normal.

D) None of the above.

9. The p-value is the probability that

A) The null hypothesis is true.

B) The alternative hypothesis is true.

C) Falsely accepting the null hypothesis given that the alternative hypothesis is true.

D) Observing a more extreme test statistic than the actual one under the null hypothesis.

10. Which of the following statements is true?

A) If we reject the one-sided null hypothesis at level α, we will also reject the two-sided null hypothesis at level α.

B) If we reject the two-sided null hypothesis at level α, we will also reject the one-sided null hypothesis at level $1 - \alpha$.

C) If we reject the two-sided null hypothesis at level α, it is possible that we do not reject the one-sided null hypothesis at level α.

D) If we do not reject the one-sided null hypothesis at level α, we will also not reject the two-sided null hypothesis at level $1 - \alpha$.

11. Which of the following statements is true?

A) In a linear regression model, each explanatory variable is associated with an error term.

B) It is ok to use OLS to estimate the model $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + u$ even though the model contains a nonlinear term.

C) The sample values of the explanatory variable can be the same in the simple regression model, however they must be different in the multiple regression model.

D) When comparing the performance between regression models, one should always choose the one with higher R^2.

12. Which of the following quantities will not be affected by changing the units of measurement of the dependent and explanatory variables?

A) R^2 B) Regression coefficients   C) Standard errors   D) Confidence intervals

13. Which of the following statements is true?

A) Sum of OLS residuals is zero.

B) The explained sum of squares is always greater than the residual sum of squares.

C) R^2 will be higher if the regression model contains a constant term.

D) R^2 is negative if all estimated coefficients are negative.

2

14. In the regression analysis, what will increase when more explanatory variables are added?

   A) Total sum of squares   B) Explained sum of squares

   C) Residual sum of squares   D) None of the above

15. Which of the following models suffers from perfect multicollinearity?

   A) $Y = \beta_0 + \beta_1 X + \beta_2(1/X) + u$

   B) $Y = \beta_0 + \beta_1 X + \beta_2 \log X + u$

   C) $Y = \beta_0 + \beta_1 \log X + \beta_2 \log X^2 + u$

   D) None of the above

16. Consider the estimated equation: $\log(\widehat{wages}) = \hat{\beta}_0 + \hat{\beta}_1 exp + \hat{\beta}_2 exp^2$, where exp stands for the length of working experience. Which of the following statements is true?

   A) When experience is increased by one unit, wages will be increased by $\hat{\beta}_1$ unit.

   B) When experience is increased by one unit, wages will be increased by $\hat{\beta}_1 + \hat{\beta}_2$ unit.

   C) When experience is increased by one unit, wages will be increased by $\hat{\beta}_1 + 2\hat{\beta}_2 exp$ unit.

   D) None of the above.

★ **Essay Questions (20%)**

1. Could you please explain in detail what is "Gauss-Markov Theorem" with its implication in a regression analysis?

2. Could you please explain in detail the "Ordinary Least Squares (OLS)" approach in a regression analysis?

3. Explain and give examples to illustrate the difference between cross-sectional, time-series, and panel data.

4. Explain the bias-variance tradeoff in the regression analysis. (Hint: the objective function of OLS is the mean squared error (MSE): $E(Y - X\beta)^2 = [E(Y - X\beta)]^2 + \text{Var}(Y - X\beta)$. Is it possible to have a lower MSE than the one obtained by OLS by lowering the variance part?)