Biostatistics (BIO545)

Quiz 4
Total marks : 60
Correct answers are bold

Ques 1: The Shapiro-Wilk test is performed on residuals of an ANOVA model and yields a p-value of 0.002. What is the most appropriate interpretation?
A. The data is normally distributed, and ANOVA results are valid.
B. The assumption of normality is violated, and a non-parametric alternative should be considered.
C. There is no need to check normality for ANOVA since it assumes equal variances.
D. The p-value suggests equal variance between groups, making ANOVA results more robust.

Ques 2: When interpreting a Kaplan-Meier survival plot, how can you estimate the median survival time?
A. The time at which the survival probability first reaches 0.5.
B. The time at which the number of censored observations exceeds uncensored observations. C. The time at which the hazard function reaches its peak.
D. The time when the survival probability curve becomes flat.

Ques 3: Which of the following is the primary outcome measure obtained from a Cox proportional hazards regression model?
A. Odds Ratio (OR) B. Relative Risk (RR) **C. Hazard Ratio (HR)** D. Incidence Rate Ratio (IRR)

Ques 4: Holding all other factors constant (alpha, effect size), increasing the sample size (n) will generally: a) Decrease statistical power. **b) Increase statistical power.**
c) Have no effect on statistical power. d) Increase the probability of a Type I error.

Ques 5: What is the primary purpose of the Log-Rank Test.
A. To estimate hazard ratios in Cox regression.
B. To compare the survival distributions between groups.
C. To determine the median survival time.
D. To calculate the probability of survival at a specific time point.

Ques 6: Which of the following best defines the critical value in hypothesis testing?
A. The probability of making a Type I error.
B. The threshold beyond which the test statistic leads to rejecting the null hypothesis. C. The p-value used to determine statistical significance.
D. The mean of the sample data

Ques 7: Which factor does NOT influence sample size calculations?
A. Effect size B. Significance level (α) **C. Number of missing values in the dataset** D. Desired power

Ques 8: A public health researcher wants to examine whether smoking status (smoker vs. non-smoker) is associated with the presence of lung disease (yes or no) in a population sample. Which statistical test should be used? A. Independent t-test B. Paired t-test C. One-way ANOVA **D. Chi-square test**

Ques 9: Which test should be used to examine whether a new biomarker (continuous variable) is significantly correlated with age (continuous variable)?
A. Chi-square test **B. Pearson's correlation** C. ANOVA D. Log-Rank Test

Ques 10: Type I error occurs when:
A. The null hypothesis is true but is incorrectly rejected.
B. The null hypothesis is false but is not rejected.
C. The alternative hypothesis is accepted when it is false.
D. The test statistic is greater than the critical value.

Ques 11: What does the hazard function measure in survival analysis?
A. The chance of surviving after time t. **B. Hazard variation over time**
C. The total number of events in the study. D. The average survival time.

Ques 12: Which of the following factors does NOT influence the required sample size for a study? A. Effect size B. Variability in the data **C. The color of the data points in the plot** D. Significance level (α)

Ques 13: A nutritionist conducts a study on three diet plans (low-carb, Mediterranean, and vegan) and measures the cholesterol levels of 150 participants after six months. Which statistical test should be used? **A. One-way ANOVA** B. Two-way ANOVA C. Paired t-test D. Linear Regression

Ques 14: A hospital is conducting a clinical trial to compare the survival rates of two cancer treatments. They track patients over five years and record their survival times. Which test should be used to compare the survival distributions?
A. Cox Regression B. Kaplan-Meier Curve **C. Log-Rank Test** D. Chi-Square Test

Ques 15: A group of plant biologists is studying a species of plant where height is influenced by a single gene with two alleles: A (dominant, tall) and a (recessive, dwarf). They grow genetically known plants AA, Aa, and aa—in different environmental conditions (e.g., varying light, water, and nutrient availability). After several growth cycles, they observe that: Some AA and Aa plants have similar intermediate heights. A few aa plants under ideal conditions are nearly as tall as some Aa plants grown in poor conditions. The overall distribution of heights forms overlapping bell-shaped curves for each genotype. Based on this case study, which of the following conclusions is most consistent with the observations?
A. The A and a alleles do not affect plant height under any environmental condition.
B. Environmental variation can significantly mask the effects of genotype on phenotype.
C. Only heterozygous (Aa) plants are affected by environmental conditions.
D. Phenotypic traits are fixed and independent of external influences once the genotype is known.

Ques 16: A researcher is studying the effect of three different diets (Diet A, Diet B, and Diet C) on blood sugar levels in diabetic patients. The sample sizes for each group are small, and initial testing shows that the data are not normally distributed. Which of the following is the most appropriate statistical test to determine if there is a significant difference in blood sugar levels across the three diets?
A. One-way ANOVA B. Paired t-test **C. Kruskal-Wallis H test** D. Chi-square test

Ques 17: A researcher is studying the epidemiology of hypertension in a community of 10,000 individuals. At the beginning of the year, 800 individuals are already diagnosed with hypertension. Over the course of the year, 200 new cases are identified. Which of the following statements is correct?
A. The prevalence of hypertension at the start of the year is 2%.
B. The incidence of hypertension over the year is 10%.
C. The prevalence of hypertension at the start of the year is 8%.
D. The incidence of hypertension over the year is 20%.

Ques 18: A new blood test claims to be 95% accurate at detecting a rare disease (meaning there's a 5% chance of a false positive). The disease itself only affects 1% of the population. If someone tests positive on this blood test, what is the chance they actually have the disease?
A. ~5% **B. ~16.1%** C. ~ 50% D. ~95%

Given:
- Test accuracy for true positive: 95% ⇒ P(Positive | Disease) = 0.95
- False positive rate: 5% ⇒ P(Positive | No Disease) = 0.05
- Prevalence of the disease: 1% ⇒ P(Disease) = 0.01
- P(No Disease) = 0.99
Step 1: Find P(Positive)
P(Positive) = P(Positive | Disease) * P(Disease) + P(Positive | No Disease) * P(No Disease)
= 0.95 * 0.01 + 0.05 * 0.99 = 0.0095 + 0.0495 = 0.059

Step 2: Compute the Posterior Probability
P(Disease | Positive) = (0.95 * 0.01) / 0.059 = 0.0095 / 0.059 ≈ 0.161 x 100=16.1%

Ques 19: What does an AUC score of 0.5 indicate?
A. The model has perfect performance. B. The model performs worse than random guessing. **C. The model lacks discriminatory power, equivalent to random guessing.** D. The model has a high false positive rate

Ques 20: A teacher is studying the relationship between rank in class and stress level (measured on an ordinal scale: low, medium, high) among students. The data is not normally distributed and involves ranked/ordinal values. Which method is most appropriate for this analysis?
A. Pearson correlation B. Chi-Square Test **C. Spearman correlation** D. Linear regression