

NAFACK SONKENG SONIA

ID: UD79629PU88846

BIO 614: BIostatISTICS

PHASE II COURSE

ATLANTIC INTERNATIONAL UNIVERSITY

HONOLULU, HAWAII

21th AUGUST 2023

INTRODUCTION

This course provides a broad overview of bio statistical methods of exploring organizing and presenting data and introduces fundamentals of probability with applications to 2X2 tables. Emphasizes the interpretation and conceptual foundations of statistical estimation, including concepts of population, sample parameter, and estimate; and approaches the inference using the likelihood function, confidence intervals, and hypothesis test. Introduces and employs the statistical computing packages (SPSS or STATA), to manipulate and present data. 25 multiple choice questions and answers (in bold) were developed as the main assignment as indicated below.

MULTIPLE CHOICE QUESTIONS & ANSWERS

1. For which variables the cumulative frequencies can be determined?

- a. Only for quantitative variables.
- b. For ordinal and quantitative variables.
- c. Only for discrete quantitative variables.
- d. For nominal, ordinal and quantitative variables.

Answer: b, for ordinal and quantitative variable we can add their frequency to have cumulative frequencies

2. In meteorology, the intensity of UV-B radiation is classified as follows: weak, moderate, strong, very strong, extreme. What is the type of this data?

- a. Categorical ordinal.
- b. Discrete numerical.
- c. Categorical nominal.
- d. Exponential

Answer: a, values of UV-B radiation are defined by an ordering relationship between the possible categories,

3. Celsius degree temperature scale

- a. is a ratio scale,
- b. is an interval scale,
- c. is an interval scale,
- d. is a ratio scale,

Answer: b, because it is possible to determine the order, the difference, and the sum of the values, and the zero point is determined because of a convention.

4. When you read scientific literature, do you know whether the statistical tests that were used were appropriate and why they were used?

- a. Always
- b. Mostly
- c. Rarely
- d. Never

Answer: b, mostly are acceptable.

5. Choose the best answer: definition of frequency in statistics.

- a. The frequency of an event is equal to the count of its occurrence in the population.
- b. The frequency of an event is equal to the count of observations.
- c. The frequency of an event is equal to the count of its occurrence per unit time.
- d. The frequency of an event is equal to the count of its occurrence in a series of observations.

Answer: d, the frequency represents a proportion or a fraction whose numerator is the occurrence of an event, and the denominator is the total number of observations.

6. Choose the correct answer on reference interval?

- a. Any interval that contains roughly 95% of the data.
- b. An interval determined from a (big) sample: the mean plus/minus the standard deviation.
- c. Any interval that contains roughly 68% of the data.
- d. An interval determined from a (big) sample: the mean plus/minus twice the standard deviation.

Answer: a, it is a specified interval of the distribution of values obtained from populations of healthy subjects. It is generally defined for an interval corresponding to 95% of the population, centered on the median.

7. Which parameter is used to measure a central tendency?

- a. Variance
- b. Arithmetic mean
- c. frequency
- d. Standard deviation
- e. Interval

Answer: b, the mean, is one of the main measures of central tendency of a statistical series. It is used to summarize the series studied around a central value.

8. Which of the following statements are true?

- a. The p-value is the probability of the sample data arising by chance.
- b. The p-value is an arbitrary value, designated as the significance level.
- c. The p-value is the chance of getting an observed effect if the null hypothesis was false.
- d. The p-value is the chance of getting an observed effect if the null hypothesis was true.
- e. A very small p-value allows us to say that there is enough evidence to accept the null hypothesis.

Answer: b and d, p-value is a statistical measure range between 0 and 1. it is used to give an indication that determines whether an observed result can be due to chance or not and if nulle hypotheses can be reject or not.

9. select the correct answer:The 95% confidence interval for the mean

- a. Contains the sample mean with 95% certainty.
- b. Is less likely to contain the population mean than the 99% confidence interval.
- c. Contains 95% of the observations in the population.

- d. Is approximately equal to the sample mean plus and minus two standard deviations.
- e. Can be used to give an indication of whether the sample mean is a precise estimate of the precise estimate mean.

Answers: b and e ; Narrow confidence intervals indicate the sample mean is a precise estimate.

10. A study was conducted into the influence of spaying of bitches on their subsequent development of urinary incontinence. Young adult bitches presenting for spaying were randomly allocated to immediate ovariohysterectomy or to a deferred operation 6 months later. The bitches were followed over the 6 months. What type of variable is 'development of urinary incontinence'?

- a. Qualitative variable
- b. Quantitative variable
- c. Categorical variable
- d. Binary variable
- e. Continuous variable

Answer: a, c, and d.

11. What statistical analysis would you use to answer the question in the above study?

- a. Two-sample t-test?
- b. Correlation?
- c. Chi-squared test?
- d. Paired t-test?
- e. Linear regression?

Answer: c, Chi-square test of independence is used to evaluate whether two categorical or nominal variables are likely to be related or not. In this case urinary continence is a categorical variable

13. In a one-sample t-test the calculated t value is 1.897 and the t value that belongs to the significance level is 2.013.

What should be your decision?

- a. I accept the null hypothesis.
- b. I repeat my calculation because this situation cannot happen in one-sample t-test.
- c. I cannot say anything without knowing the probabilities.
- d. I reject the null hypothesis.

Answer: a, 1.897 is less than 2.013 so the probability to make an error by reject the null hypothesis is high. So, we accept the null hypotheses.

14. The sample p-value calculated during a Wilcoxon test is 0.035. The critical p-value is 5%. Choose the correct statement.

- a. The null hypothesis is rejected, a significant difference can be assumed between the values measured before and after the treatment.
- b. The difference between the medians is significant in 97% of all cases.

- c. The null hypothesis is accepted; there is no significant difference between the values measured before and after the treatment.
- d. There is no significant difference between the medians in 97% of all cases.

Answer: a, p value (0.035) is less than 5% so the probability of making error by reject the null hypotheses is reduce.

15. choose the correct answers: The paired t-test:

- a. Tests the null hypothesis that the two populations mean are equal.
- b. Must have equally sized numbers of observations in each group.
- c. Assumes that the data in each group are normally distributed.
- d. Is appropriate for comparing the means of independent groups of observations.
- e. When appropriately used, is more powerful when the sample size is large.

Answers: b and e the paired t test is required to calculate the difference for each pair of observations

16. Answer true or false for the following statements:

A correlation coefficient:

- a. Should not be calculated when there is an underlying relationship between the two variables, but it is not linear.
- b. Does not provide evidence of a causal relationship between two variables.
- c. Should not be used to judge the biological importance of the relationship between two variables.
- d. Should be performed only when certain assumptions are satisfied (e.g., variables measured on a random sample of individuals, both the variables are quantitative and at least one of the two variables need to be normally distributed).

Answers: all the answers are true. It is used to give the strength of relationship between variables

17. A study was conducted to investigate the relationship between a sheep's live weight (kg) and its chest girth (cm). A random sample of 66 sheep was weighed and simultaneously had their chest girth measured. Answer true or false for the following statements:

- a. Analysis of the data from this study could be performed using a two-sample t-test.
- b. A scatter diagram should be used to present the data.
- c. Chest girth measurement is a categorical variable.
- d. Simple linear regression could be used to describe the straight-line relationship between sheep live weight and chest girth.
- e. Sheep live weight can be predicted by measuring chest girth as long as the relationship between these variables is linear and chest girth is measured without error.

Answers:

- a. False. Simple linear regression or correlation analysis should be used to analyse the data.**

- b. True.
- c. False. This variable is a continuous variable.
- d. True.
- e. True.

16. What is the aim of a test for independence?

- a. It tests whether the means are independent from the choice of group.
- b. Tests the effect of risk factors.
- c. Tests whether the probabilities of possible outcomes of a quality are independent of the presence of another effect.
- d. Tests whether two random variables are independent.

Answers: c, this test determine whether two questions are independent or not, or in other words, whether the answers to one condition the answers to the other.

17. Choose the right statement. The y axis intercept of the regression line...

- a. cannot be zero.
- b. cannot be more than one.
- c. cannot be less than negative one.
- d. can be any real number.

Answer: d, the slope of the regression line can be any real number.

18. What is the reason of using an ANOVA instead of several t-test on the same sample?

- a. With ANOVA we can reduce the multiplicity (the increase in the first type error).
- b. Variance analysis has higher power than t-tests.
- c. Comparing variances reduces higher the second type error than comparing means.
- d. Normality could not be interpreted for multiple Comparisons.

Answer: a, ANOVA is a special type of analysis of variance which is used when we only have two population means to compare. Hence, to avoid an increase in error while using a t-test to compare more than two populations, ANOVA is the best to used.

20. How can you decrease the chance to do a second type error?

- a. Increase the significance level.
- b. Decrease the first type error.
- c. Increase the sample size.
- d. We can't decrease the second type error, only the first type error.

Answer: c, when the sample size is increase, the chance to do an error is reduced.

21. What does the odds ratio show? (Regarding illness and risk factor.)

- a. It shows how many times higher the probability of an illness in the presence of the risk factor than in the absence.
- b. It shows how many times higher the probability of the occurrence of the illness than the nonoccurrence of the illness.
- c. It shows how many times higher the odds of a disease in the presence of the risk factor than its absence.
- d. It shows how many times the sampling error is more likely than the correct sampling.

Answer: c,

22. What test can we use if we have 1 numerical, continuous variable in 2 not paired groups and the groups are not normally distributed?

- a. ANOVA
- b. Wilcoxon rank
- c. Mann-Whitney U
- d. Kolmogorov-Smirnov

Answer: c,

23. Choose the condition of a chi-square test for independency.

- a. Each value in the expected contingency table must be greater than $5n$ (where n is the count of the dataset.)
- b. Each value in the observed contingency table must greater than 2 and at least 50% of the values must be greater than 5.
- c. Each value in the observed frequency table must be great than $5n$ (where n is the number of cells in the table.)
- d. At least four-fifth of the values in the expected frequency table must be larger than 5.

Answer: b, the chi-square test is used from the distribution when no more than 20% of the expected counts are less than 5 and all individual expected counts are 1 or greater. All four expected counts in a 2×2 table should be 5 or greater.

24. We are studying the applicability of a diagnostic test. What is the name of the parameter given by the ratio of true positive tests and all positive tests?

- a. Positive predictive value
- b. Sensitivity
- c. Specificity
- d. Negative predictive value

Answer: a, positive predictive value is the ratio of patients truly diagnosed as positive to all those who had positive test results (including healthy subjects who were incorrectly diagnosed as patient).

25. When is (parametric one way) ANOVA applicable?

- a. If the samples to be compared are independent and normally distributed.
- b. If the variance of the samples is equal and all samples are normally distributed.
- c. When it yields less alpha and beta error than a series of paired t-tests.
- d. If more than half of samples to be compared are normally distributed.

Answer: b, ANOVA test is used when variance want to be compared.

CONCLUSION

This course had as main objective to describe the role biostatistics serve in public health, differentiate among different sampling methods and discuss their strength and limitations, describe the strengths and limitations of designed experiments and observational studies, distinguish among different measurement scales, choose the appropriate descriptive and inferential statistical methods based on these distinctions and interpret the results, to be able to determine preferred methodological alternatives to commonly used statistical methods when assumptions are not met, to be able to apply basic concepts of probability, random variation, and commonly used statistical probability distributions, and to be able to develop presentations based on statistical analysis for both public health professionals and educated lay audiences. Books and articles were read even videos on U-tubes were also explored for better understanding of the course. This could be seen in the 25 multiple choice questions and answers. To me the objective of the course has been made.

References

- 1.FFPH, Antony Stewart. *Basic Statistics and Epidemiology A Practical Guide FOURTH EDITION*. Boca Raton London New York: CRC Press, 2016.
- 2.Indranil Saha, Bobby Paul. *ESSENTIALS OF BIOSTATISTICS AND RESEARCH METHODOLOGY*. 5A Bhawani Dutta Lane, Kolkata-700073: ACADEMIC PUBLISHERS, OCTOBER 2020.
- 3.Surveillance, • Monitoring the Health of Populations : Statistical Principles and Methods for Public Health. *Monitoring the Health of Populations: Statistical Principles and Methods for Public Health Surveillance*. Oxford University Press, Oct,2003.
- 4.Torre, Giuseppe La. *Applied Epidemiology and Biostatistics*. edizioniseed, n.d.
- 5.Wassertheil-Smoller, Sylvia. *Biostatistics and Epidemiology: A Primer for Health Professionals*. Springer science + Business Media,LLC , 2013.