

# **RxExam's Biostatistics** **Questions & Answers**

**2019-2020 Edition**

**MANAN SHROFF**

---

This book is not intended as a substitute for the advice of physicians. Students or readers must consult their physician about any existing problem. Do not use information in this book for any kind of self-treatment. Do not administer any dose of mentioned drugs in this book without consulting your physician.

The author is not responsible for any kind of misinterpreted, incorrect, or misleading information or any typographical errors in this book. Any doubtful or questionable answers should be checked in other available reference sources.

All rights reserved.

No part of this book may be reproduced or transmitted in any form or by any means, electronically Photocopying, recording, or otherwise, without prior written permission of the publisher.

RXEXAM® is a registered trademark of Pharmacy Exam of Krishna Publications Inc. Any unauthorized use of this trade mark will be considered a violation of law.

---

- 1.** In biostatistics, confounding is normally defined as:
- One chosen from a carefully defined population with the aid of a formal method to avoid bias.
  - A formal method to assign subjects by chance to one or the other treatment.
  - The effect of two or more variables that do not allow a conclusion about either one separately.
  - The systematic tendency of any factors associated with the design, conduct, analysis, and evaluation of the results of a trial to make the estimate of a treatment effect deviate from its true value.
- 2.** The data that categories patients as males or females are known as:
- Random data
  - Nominal data
  - Ordinal data
  - Interval data
- 3.** Which of the following data represents interval continuous data?
- A number of cigarettes smoked per day by a person.
  - A number of children in a household.
  - Height of children.
  - Number of languages a person speaks.
- 4.** Data can be transformed by using the logarithm, square root, or reciprocal. Which of the following is the most common data transformation used in medical research?
- data converted to log
  - data converted to root
  - data converted to reciprocal
  - data converted to ratio
- 5.** The scores of Naplex for 10 students are 75, 82, 90, 92, 67, 95, 110, 80, 82, 86. Find the mean for the above data.
- 83.6
  - 85.9
  - 836
  - 43
- 6.** The scores of Naplex for 10 students are 75, 82, 90, 92, 67, 95, 110, 80, 82, 86. Find the median for the above data.
- 110
  - 84
  - 82
  - 67
- 7.** The scores of Naplex for 10 students are 75, 82, 90, 92, 67, 95, 110, 80, 82, 86. Find the mode for the above data.
- 75
  - 90
  - 86
  - 82
- 8.** Classifying continuing educational experience into categories including "strongly agree," "agree," and "disagree," is an example of which type of variable or data?
- Nominal
  - Ordinal
  - Interval
  - Ratio
- 9.** Which of the following statements about the prevalence is/are TRUE?
- It is a measure of the rate of occurrence of a condition.
  - It is the number of new cases of a condition that develop during a specific period of time.
  - It is the proportion of a population found to have a condition at a single point in time.

- a. I only
- b. III only
- c. II and III only
- d. All

**10.** "One out of every 50 adults in the United States has depression." This statement is an example of:

- a. Incidence
- b. Prevalence

**11.** Which of the following statements best describe the relative risk (RR)?

- a. It is the risk of an event or outcome occurring in a group of interest in relation to a control group.
- b. It usually states how much the treatment reduced the risk of an outcome relative to the control group.
- c. It represents the absolute difference of the event rate between the treatment and control groups.
- d. It is a representation of the number of patients who need to be treated to prevent one additional event compared to treating the same number of patients with the control therapy.

**12.** Calculate the value of RR if the risk of heart failure associated with an invention drug is 5% versus 9% with placebo?

- a. 1.25
- b. 2.65
- c. 0.55
- d. 1.8

**13.** Which of the following statements best describe the interpretation of value of RR obtained in a previous question?

- a. there was no association between the study drug and heart failure.
- b. the risk of heart failure in patients taking the study drug was less than the risk of heart failure in patients taking placebo.

c. the risk of heart failure in patients taking the study drug was more than the risk of heart failure in patients taking placebo.

**14.** What is the value of relative risk reduction if an RR is 0.33?

- a. 1.56
- b. 0.33
- c. 0.67
- d. 1.33

**15.** What would be the absolute risk reduction if the incidence of heart attack was 2% in the intervention group and 7% in the control group?

- a. 15%
- b. 5%
- c. 2%
- d. 7%

**16.** If the calculated absolute risk reduction (ARR) for heart attack event was 5.1%, what would be the number needed to treat?

- a. 11
- b. 2.65
- c. 3.2
- d. 20

**17.** Which of the following statements is **TRUE ABOUT** the standard deviation (SD)?

- a. It is used to calculate confidence intervals.
- b. It is used to determine variability of a sample around a sample mean.
- c. It is used to display bimodal or skewed data.
- d. A large SD shows that individual data points are clustered closer to the mean.

**18.** If the sample size in a study is 100 subjects and the SD for blood glucose is 10 mg/dl, what is the standard error of the mean?

- a. 65.9

d. the regression model explained 30% of the total variance is not a good fit.

**146.** In a study examining the relationship between the drinking a coffee in the late evening and the likelihood of insomnia, the  $r^2$  value for the regression line is 1.0. What does this indicate?

a. For each additional cup of coffee, chances of occurring insomnia increase by 1.0%.

b. Since  $r^2$  equals 1.0, this indicates there is no relationship between drinking coffee and developing insomnia.

c. Since  $r^2$  equals 1.0, the regression line perfectly fits the data.

d. For each additional cup of coffee, there is no effect on sleep.

**147.** Which of the following information is/are TRUE ABOUT an odds ratio (OR)?

I. If the OR = 1 indicates there is no difference between the two arms of the study.

II. If the OR is > 1 the control is better than the intervention.

III. If the OR is < 1 the intervention is better than the control.

a. I only

b. I and II only

c. II and III only

d. All

**148.** Suppose a study examined the relationship between workers exposed to an asbestos for more than 10 years and the risk of developing asthma. If the study found an increased risk of asthma in the group who exposed to asbestos (OR: 1.16, 95% CI: 0.9-1.5). What does this mean?

a. Employees who exposed to asbestos for more than 10 years were 1.16 times less likely to develop asthma than those who did not.

b. Employees who exposed to asbestos for more than 10 years had 16% decreased odds of developing asthma than those who did not.

c. The results are not significant because the confidence interval includes 1.0.

d. The results are not significant because the confidence interval is greater than 1.0.

**149.** What does "OR 0.4 95%CI 0.4-0.6 p <0.05" mean?

a. The odds of death in the intervention groups are 60% less than the odds of death in the control groups with the true population effect between 60% and 40%. This result was statistically significant.

b. The odds of death in the intervention groups are 60% more than the odds of death in the control groups with the true population effect between 60% and 40%. This result was statistically significant.

c. The odds of death in the intervention groups are 60% less than the odds of death in the control groups with the true population effect between 40% and 60%. This result was statistically NOT significant.

d. The odds of death in the intervention groups are 60% more than the odds of death in the control groups with the true population effect between 40% and 60%. This result was statistically NOT significant.

**150.** A drug company-funded double blind randomised controlled trial evaluated the efficacy of an adenosine receptor antagonist Cangrelor vs Clopidogrel in patients undergoing urgent or elective Percutaneous Coronary Intervention (PCI) who were followed up for specific complications for 48 hrs (Bhatt et al. 2009). The results section reported "OR 0.65 95% confidence interval [CI], 0.55 to 0.83; P=0.005" What does this mean?

a. The odds of death, myocardial infarction, ischemia-driven revascularization, or stent thrombosis at 48 hours after randomization in the Cangrelor arm were 35% less than in the Clopidogrel arm with the true population effect between 34% and 7%. This result was not statistically significant.

b. The odds of death, myocardial infarction, ischemia-driven revascularization, or stent thrombosis at 48 hours after randomization in the Cangrelor arm were 35% less than in the Clopidogrel arm with the true population effect between 45% and 17%. This result was statistically significant.

c. The odds of death, myocardial infarction, ischemia-driven revascularization, or stent thrombosis at 48 hours after randomization in the Cangrelor arm were 35% less than in the Clopidogrel arm with the true population effect between 45% and 17%. This result was statistically NOT significant.

d. The odds of death, myocardial infarction, ischemia-driven revascularization, or stent thrombosis at 48 hours after randomization in the Cangrelor arm were 35% less than in the Clopidogrel arm with the true population effect between 34% and 7%. This result was statistically significant.

## Biostatistics Answers

**1. (c)** The effect of two or more variables that do not allow a conclusion about either one separately is defined as confounding.

**Random sample:** one chosen from a carefully defined population with the aid of a formal method to avoid bias and confounding.

**Randomization:** in comparative trials, a formal method to assign subjects by chance to one or the other treatment.

**Random errors:** errors that follow the principle of indifference.

**Bias:** The systematic tendency of any factors associated with the design, conduct, analysis, and evaluation of the results of a trial to make the estimate of a treatment effect deviate from its true value.

**Confounding:** The effect of two or more variables that do not allow a conclusion about either one separately.

**Validity:** (a). In logic, an argument is valid if the conclusions follow from the premises

(b). In pharmacological sciences, a method is valid if it measures what it should, is reproducible and responsive to change, e.g., by a treatment.

**2. (b)** Nominal data.

Observations recorded during research constitute data. There are three types of data i.e. nominal, ordinal, and interval data.

Statistical methods for analysis mainly depend on type of data. Generally, data show picture of the variability and central tendency. Therefore, it is very important to understand the types of data.

**1. Nominal data:** This is synonymous with categorical data where data is simply assigned

“names” or categories based on the presence or absence of certain attributes/characteristics without any ranking between the categories. For example, patients are categorized by gender as males or females; by religion as Hindu, Muslim, or Christian. It also includes binominal data, which refers to two possible outcomes. For example, outcome of cancer may be death or survival, drug therapy with drug ‘X’ will show improvement or no improvement at all.

**2. Ordinal data:** It is also called as ordered, categorical, or graded data. Generally, this type of data is expressed as scores or ranks. There is a natural order among categories, and they can be ranked or arranged in order. For example, pain associated with cancer may be classified as mild, moderate, and severe. Since there is an order between the three grades of pain, this type of data is called as ordinal. To indicate the intensity of pain, it may also be expressed as scores (mild = 1, moderate = 2, severe = 3). Hence, data can be arranged in an order and rank.

**3. Interval data:** This type of data is characterized by an equal and definite interval between two measurements. For example, weight is expressed as 100, 110, 120, 130, 140 lbs. The interval between 100 and 110 is same as that between 130 and 140.

**3. (c)** Height of children.

This type of data is characterized by an equal and definite interval between two measurements. For example, weight is expressed as 100, 110, 120, 130, 140 lbs. The interval between 100 and 110 is same as that between 130 and 140.

Interval type of data can be either continuous or discrete. A continuous variable can take any value within a given range.

For example: hemoglobin (Hb) level may be taken as 11.3, 12.6, 13.4 gm % while a discrete variable is usually assigned integer values i.e. does not have

fractional values. For example, blood pressure values are generally discrete variables or number of cigarettes smoked per day by a person.

**Examples of Interval Continuous Data:**

1. Height of children
2. Weight of cars
3. Time to wake up in the morning

**Examples of Interval Discrete Data:**

1. A number of cigarettes smoked per day by a person.
2. A number of children in a household.
3. A number of languages a person speaks.

**4. (a)** data converted to log.

Sometimes, certain data may be converted from one form to another form to reduce skewness and make it to follow the normal distribution. For example, drug doses are converted to their log values and plotted in dose response curve to obtain a straight line so that analysis becomes easy. Data can be transformed by taking the logarithm, square root, or reciprocal. Logarithmic conversion is the most common data transformation used in medical research.

**5. (b)** 85.90

Mean is the common measure of central tendency, most widely used in calculations of averages. It is the average of a data set. So, in the above example, the mean of the data can be calculated by adding up the individual values and dividing the sum by number of students (n).

The sample mean  $\bar{x}$  is calculated as

$$\bar{x} = \frac{x_1 + x_2 + x_3 + x_4 + \dots + x_n}{n} = \sum \frac{x_i}{n}$$

$$\bar{x} = \text{Mean} = \frac{75+82+90+92+67+95+110+80+82+86}{10}$$

Mean = 85.90

**6. (b)** 84.

Median is an average, which is obtained by getting middle values of a set of data arranged or ordered from lowest to the highest (or vice versa). In this process, 50% of the population has the value smaller than median and 50% of samples have the value larger than median.

Median is a better indicator of central value when one or more of the lowest or the highest observations are wide apart or are not evenly distributed.

Median in case of even number of observations is taken arbitrary as an average of two middle values and in case of odd number, the central value forms the median.

To find median in above data, we need to arrange the data either ascending or descending. Therefore:

67, 75, 80, 82, 82, 86, 90, 92, 95, 110

In case of even numbers, the median should be the average of two middle values. Therefore, in above case, it should be:

$$= \frac{(82+86)}{2}$$

= 84

If we remove the last data set 110 from the calculation, the median in case of odd number (9) should be the central value of the data set:

67, 75, 80, 82, 82, 86, 90, 92, 95

= 82

**7. (d)** 82

The mode is the number that occurs most often within a set of numbers or is the most common number in a data set. In above case the mode should be 82 since it appears twice in data set. To get better idea, look the following data set:



75, 82, 75, 92, 67, 95, 75, 80, 82, 86.

In this data set, the mode will be 75 and not 82 since it appears three times.

A set of numbers can have more than one mode (this is known as bimodal) if there are multiple numbers that occur with equal frequency, and more times than the others in the set.

75, 82, 75, 92, 67, 95, 75, 80, 82, 86, 82.

In this example, both the number 75 and the number 82 are modes; since they both appear three times.

If no number in a set of numbers occurs more than once, that set has no mode:

For example: 75, 82, 85, 90, 92

**8. (b)** Classifying continuing educational experience into categories is a categorical variable as the value functions as a label rather than a numeric value. The type of categorical variable is ordinal since categories can be ranked in a specific order.

**9. (b)** III only.

Incidence is a measure of the rate of occurrence of a condition. In other words, incidence is the number of new cases of a condition that develop during a specific period of time. For example, the incidence of hypertension may be documented as the number of new cases of hypertension per year within a given sample or population.

Incidence can be measured using methods such as randomized controlled trials (e.g., to determine the incidence of adverse effects of a medication) or cohort studies, which follow groups of patients over time.

Prevalence, on the other hand, is the proportion of a population found to have a condition at a single point in time. For example, the prevalence of hypertension may be documented as the percentage of the US population with hypertension.

Prevalence is measured using cross-sectional study designs, such as a survey or a census, or by using administrative data such as medical, hospital, or prescription drug claims.

**10. (b)** This statement is an example of prevalence since the statement is an expression of a proportion of a population found to have a condition at a single point in time.

**11. (a)** The relative risk (RR) is the risk of an event or outcome occurring in a group of interest in relation to a control group. The group of interest can be an intervention (e.g., a medication) or a pre-specified characteristic e.g., those with hypertension).

An example would be the risk of heart attack in patients taking a study drug versus a control group of patients taking a placebo.

The relative risk (RR) can be calculated by dividing the event rate (i.e., heart attacks) in the intervention group by the event rate (i.e., heart attacks) in the control group.

**12. (c)** 0.55.

The relative risk (RR) can be calculated by using the following formula:

$$RR = \frac{\text{event rate in intervention group}}{\text{event rate in control group}}$$

$$RR = \frac{5}{9} = 0.55$$

**13. (b)** The relative risk (RR) is the risk of an event or outcome occurring in a group of interest in relation to a control group. The group of interest can be an intervention (e.g., a medication) or a pre-specified characteristic (e.g., those with hypertension).

An example would be the risk of heart failure in patients taking a study drug versus patients taking placebo. The RR can be calculated by dividing the event rate (i.e., heart failure) in the intervention group by the event rate in the control group.

1. If the RR is 1, there is no difference between the two groups; or, in other words, there is no association between exposure to a factor and the outcome of interest (e.g., there was no association between the study drug and heart failure).

2. An RR less than 1 indicates a negative association (e.g., an RR of 0.55 - we found in question 11 - means the risk of heart failure in patients taking the study drug was less than the risk of heart failure in patients taking placebo).

3. While an RR greater than 1 indicates a positive association (e.g., an RR of 2 means the risk of heart failure in patients taking the study drug was greater than the risk of heart failure in patients taking placebo).

**14. (c) 0.67**

The relative risk reduction (RRR) is another measure of effect size, which reports how much the treatment reduced the risk of an outcome relative to the control group. The RRR can be calculated by subtracting the relative risk (RR) from 1.

The Relative Risk Reduction (RRR) can be calculated by the following formula:

$$RRR = 1 - \text{Relative Risk (RR)}$$

$$RRR = 1 - 0.33 = 0.67$$

**15. (b) 5%.**

ARR represents the absolute difference of the event rate between the treatment and control groups. The ARR can be easily calculated by subtracting the event rate in the intervention group from the event rate in the control group.

ARR = the event rate in the control group - the event rate in the intervention group

$$ARR = 7\% - 2\% = 5\%$$

**16. (d) 20**

The number needed to treat (NNT) is a representation of the number of patients who need to be treated to prevent one additional event compared to treating the same number of patients with the control therapy.

It can be easily calculated as it is the inverse of the ARR. The NNT should always be rounded up to the nearest integer, as you cannot have a fraction of a patient (e.g., an NNT of 1.6 would be rounded up to an NNT of 2).

$$NNT = \frac{1}{ARR}$$

$$NNT = \frac{1}{0.051} = 19.60 = 20$$

The NNT in this example demonstrates that for every 20 patients treated, 1 additional heart attack can be avoided compared to treating those patients with the control therapy.

**17. (b)** It is used to determine variability of a sample around a sample mean.

1. Standard deviations (SD) determine the approximate distribution of individual data points in regard to a sample mean; therefore, the SD is a measure of the variability in the data.

2. Typically, a study will find a range of individual values or observations, where some values fall either below the mean or above the mean. Thus, individual deviations from the mean can be either positive or negative.

3. Variance is calculated as the mean of the deviations squared, and the SD is calculated as the square root of the variance.

4. The SD provides value to readers when the sample is either normally or nearly normally distributed and is used for continuous variables.

5. If the SD is reported for bimodal or skewed data, it may be falsely larger than expected. Therefore, it would not be an appropriate tool to use to measure the variability for those distributions.

margin been utilized, the treatment would have been found to be “non-inferior” if the event rates were as high as 8.75%.

On the other hand, overly stringent NI margins can prevent an effective therapy from being deemed non-inferior.

**113. (b)** After a non-inferiority clinical trial, a new therapy may be accepted as effective, even if its treatment effect is slightly lesser than the current standard. It is therefore possible that, after a series of trials where the new therapy is slightly worse than the preceding drugs, an ineffective or harmful therapy might be incorrectly declared efficacious; this is known as 'bio-creep'.

To provide an example, suppose that a theoretical Drug A is deemed to be superior to placebo. Several years later, Drug B is found to be non-inferior to drug A with a certain NI margin. Then, Drug C is compared to Drug B via a NI study at some point later. During each of these studies, the new agent is found to be “not acceptably” worse than the previous agent. It may then be unknown if Drug C is more effective than placebo.

**114. (d)** The Chi-Squared Test is used to test for statistical significance of a dependent categorical variable between two or more independent groups. It is used to determine whether observed sample frequencies differ significantly from expected frequencies specified in the null hypothesis.

**115. (c)** The only assumption for the Chi-Squared Test is that all cells must have an expected value of at least five.

The chi-square goodness of fit test is appropriate when the following conditions are met:

1. The sampling method is simple random sampling.
2. The variable under study is categorical.
3. The expected value of the number of sample observations in each level of the variable is at least 5.

**116. (a)** A categorical variable (sometimes called a nominal variable) is one that has two or more categories, but there is no intrinsic ordering to the categories. For example, gender is a categorical variable having two categories (male and female) and there is no intrinsic ordering to the categories. Hair color is also a categorical variable having a number of categories (blonde, brown, brunette, red, etc.) and again, there is no agreed way to order these from highest to lowest. A purely categorical variable is one that simply allows you to assign categories, but you cannot clearly order the variables. If the variable has a clear ordering, then that variable would be an ordinal variable, as described below.

An ordinal variable is similar to a categorical variable. The difference between the two is that there is a clear ordering of the variables. For example, suppose you have a variable, economic status, with three categories (low, medium and high). In addition to being able to classify people into these three categories, you can order the categories as low, medium and high. Now consider a variable like educational experience (with values such as elementary school graduate, high school graduate, some college and college graduate). These also can be ordered as elementary school, high school, some college, and college graduate. Even though we can order these from lowest to highest, the spacing between the values may not be the same across the levels of the variables. If these categories were equally spaced, then the variable would be an interval variable.

An interval variable is similar to an ordinal variable, except that the intervals between the values of the interval variable are equally spaced. For example, suppose you have a variable such as annual income that is measured in dollars, and we have three people who make \$50,000, \$60,000 and \$70,000. The second person makes \$10,000 more than the first person and \$10,000 less than the third person, and the size of these intervals is the same. If there were two other people who make \$100,000 and \$110,000, the size of that interval between these two people is also the same (\$10,000).

**117.(c)** Stroke is a categorical dependent variable and therefore Chi Squared Test or Fischer’s Exact Test can be used. The Chi Squared Test can only be used when all cells must have an expected value of at least five. In cases where the expected value is not at least five for all cells, a Fischer’s Exact Test should be used.

The formula for calculating expected cell value in Chi Squared Test is:

$$\text{Expected Cell} = \frac{\text{Row Total} \times \text{Column Total}}{\text{Total number subjects}}$$

In above example, we can calculate it by the following way:

	Stroke	No Stroke	Total
Intervention Group	2	63	65
Expected Cell Value	?	?	
Control Group	3	52	55
Expected Cell Value	?	?	
Total	5	115	120

As shown in the above table, we need to find values for each expected cell denoted with “?”. Using the formula, we can calculate expected cell value in intervention and control groups.

A. Intervention Group With Stroke

$$\text{Expected Cell} = \frac{65 \times 5}{120} = 2.70$$

B. Intervention Group Without Stroke

$$\text{Expected Cell} = \frac{65 \times 115}{120} = 62.29$$

C. Control Group With Stroke

$$\text{Expected Cell} = \frac{55 \times 5}{120} = 2.29$$

D. Control Group Without Stroke

$$\text{Expected Cell} = \frac{55 \times 115}{120} = 52.70$$

So, placing all values in above table, it should look like:

	Stroke	No Stroke	Total
Intervention Group	2	63	65
Expected Cell Value	2.70	62.29	
Control Group	3	52	55
Expected Cell Value	2.29	52.70	
Total	5	115	120

There were two cells (e.g. 2.70 and 2.29) in our calculation have expected cell value less than 5. Therefore, this data does not meet the chi-squared assumption for at least five expected in each cell. So, the Fischer’s Exact Test shall be used instead.

**118. (c)** The degrees of freedom for a Chi-squared Test can be calculated by using the following formula:

$$\text{Degrees of Freedom} = (\text{Rows} - 1) \times (\text{Columns} - 1)$$

Therefore, the degrees of freedom in 2 x 3 contingency chi squared test should be:

$$DF = (R - 1) \times (C - 1)$$

$$DF = (2-1) \times (3-1)$$

$$DF = 1 \times 2 = 2$$

**119. (a)** HbA1c is a continuous dependent variable. As we are testing two independent groups and it meets the general rule of thumb having at least 30

subjects in each group, the t-test is the best option.

**Continuous variable:** A continuous variable has numeric values, and the relative magnitude of the values is significant. This means that data can be ranked low to high or high to low. An example of a continuous variable is value of HbA1c in a diabetic patient.

**Discrete variable:** Suppose we flip a coin and count the number of tails. The number of tails could be any integer value between 0 and plus infinity. We could not, for example, get 4.6 tails.

**120. (a)** Variables are classified as either independent or dependent variables. In research studies, the independent variable is the variable that is varied or manipulated by the researcher.

For example, patients in a clinical trial may be assigned to take a placebo or an anti-diabetic medication. The type of study medication that the patient is taking is an independent variable. The dependent variable is the outcome or response that is measured, such as blood glucose level, in this example. Researchers are typically interested in understanding the effect of the independent variable (or treatment) on the dependent variable (or outcome); for example, what is the effect of an anti-diabetic medication on blood glucose level?

**121. (b)** To do a Chi Square statistic calculation, we first want to find out values for each expected cell; it must be greater than 5.

Observed Value				
	Biology	Physics	Chemistry	Total
Male	12	16	21	49
Female	20	11	20	51
Total	32	27	41	100

Expected Value				
	Biology	Physics	Chemistry	Total
Male	?	?	?	?
Female	?	?	?	?
Total	?	?	?	?

To calculate expected value for each cell, we need to use the following formula:

$$\text{Expected Cell} = \frac{\text{Row Total} \times \text{Column Total}}{\text{Total number subjects}}$$

$$\text{Expected Cell} = \frac{49 \times 32}{100} = 15.68$$

$$\text{Expected Cell} = \frac{49 \times 27}{100} = 13.23$$

$$\text{Expected Cell} = \frac{49 \times 41}{100} = 20.09$$

$$\text{Expected Cell} = \frac{51 \times 32}{100} = 16.32$$

$$\text{Expected Cell} = \frac{51 \times 27}{100} = 13.77$$

$$\text{Expected Cell} = \frac{51 \times 41}{100} = 20.91$$

Expected Value				
	Biology	Physics	Chemistry	Total
Male	15.68	13.23	20.09	49
Female	16.32	13.77	20.91	51
Total	32	27	41	100

Now, we can construct the null hypothesis.

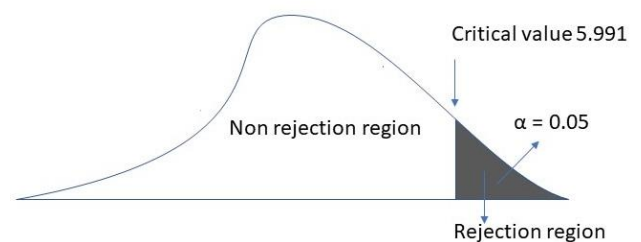
Ho: Gender has NO influence on a subject selected.

Ha: Gender has influence on a subject selected.

Degrees of freedom = (Rows-1)x(Columns-1)

Degrees of freedom = (2-1)x(3-1) = 2

The Chi Square value at 0.05 level of significance with a degree of freedom equals 2 would be 5.991. Please refer to Biostatistics Table file.



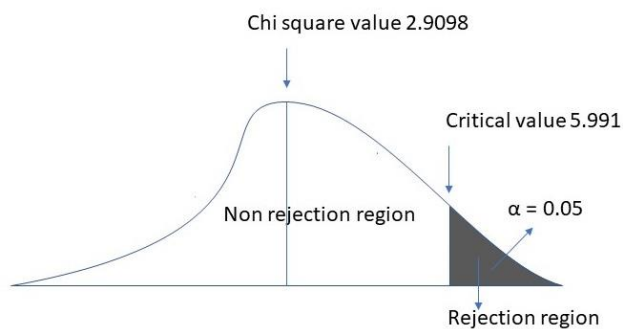
The next step is to find out the Chi square test value to compare it with our previously obtained critical value. The formula to find the Chi Square test value is:

$$\chi^2 = \sum \frac{(\text{observed value} - \text{expected value})^2}{\text{Expected value}}$$

Therefore:

$$\begin{aligned} \chi^2 &= \frac{(12-15.68)^2}{15.68} + \frac{(20-16.32)^2}{16.32} + \frac{(16-13.23)^2}{13.23} + \\ &\frac{(11-13.77)^2}{13.77} + \frac{(21-20.09)^2}{20.09} + \frac{(20-20.91)^2}{20.91} \\ &= \frac{(-3.68)^2}{15.68} + \frac{(3.68)^2}{16.32} + \frac{(2.77)^2}{13.23} + \frac{(-2.77)^2}{13.77} \\ &+ \frac{(0.91)^2}{20.09} + \frac{(-0.91)^2}{20.91} \\ &= \frac{13.54}{15.68} + \frac{13.54}{16.32} + \frac{7.67}{13.23} + \frac{7.67}{13.77} + \frac{0.82}{20.09} \\ &+ \frac{0.82}{20.91} \\ &= 0.8635 + 0.8296 + 0.5797 + 0.5570 + 0.0408 + \\ &0.0392 \\ &= 2.9098 \end{aligned}$$

Since Chi Squared value 2.9098 is less than our critical value (5.991); we are failed to reject the null hypothesis. In other words; a gender does NOT have any influence on the subject selected.



**122. (d)** The pain scores are an ordinal variable, so a non-parametric test must be used. If the data is part of one single group, Wilcoxon signed rank test shall be used. If the data is part of two independent groups, Mann Whitney U test shall be used. For three or more independent groups, Kruskal Wallace test should be used.

In our example data is a pre/post comparison of paired data of the same single group (same 100 patients are used before and after administering Oxycodone); therefore, the Wilcoxon Signed-Rank test is the best option.

**123. (c)** For the continuous dependent variables, t-test, paired t-test or ANOVA can be used depend on the presence of independent groups. If there is a single independent group, paired t-test shall be used. If there are two independent groups, then, independent groups t-test shall be performed. For three or more independent groups, ANOVA should be used.

**124. (d)** All. Analysis of Variance or ANOVA is used to test for statistical significance of a continuous dependent variable between three or more independent groups assuming the data are normally distributed.

When a statistically significant p-value is found using ANOVA, it means that there is a statistically significant difference between two or more of the groups. A student's t-test can then be used to tease out exactly where the differences are.

ANOVA can be used for a comparison between two groups, but the student's t-test will give the same answer and is much easier to calculate. In its simplest form, ANOVA provides a statistical test of whether the population means of several groups are equal, and therefore generalizes the t-test to more than two groups. ANOVA is useful for comparing (testing) three or more group means for statistical significance. It is conceptually similar to multiple two-sample t-tests, but is more conservative, resulting in fewer type I errors, and is therefore suited to a wide range of practical problems.

Clopidogrel in patients undergoing urgent or elective Percutaneous Coronary Intervention (PCI) who were followed up for specific complications for 48 hrs (Bhatt et al. 2009). The results section reported "OR 0.65 95% confidence interval [CI], 0.55 to 0.83; P=0.005" This means:

The odds of death, myocardial infarction, ischemia-driven revascularization, or stent thrombosis at 48 hours after randomization in the Cangrelor arm were 35% ( $1-0.65=0.35$ ) less than in the Clopidogrel arm with the true population effect between ( $1-0.55=0.45$ ) 45% and ( $1-0.83=0.17$ ) 17%. This result was statistically significant ( $p(0.005) < 0.05$ ) since  $p < 0.05$ .