

## Biostatistics

Archive Questions
Lectures 7 and 10-21


1. If you want to compare the presence of cancer risk factors between residents in Amman and Aqaba. The best way to collect data is:
a. Experiments
b. Surveys
c. Both of the above
d. None of the above

Answer: B. Surveys
2. The purposes of collecting random samples from a population include all of the following except:
a. To obtain a sample that will be representative for the population
b. To determine the sample size, for selecting a small one for example
c. To give every member in the population an equal chance to appear in the study
d. To give all members in the population the same probability of selection
e. To eliminate the selection bias

Answer: B. To determine the sample size, for selecting a small one for example
3. Methods of data collection include all of the following except:
a. Collecting data through comprehensive surveys
b. Collecting data through sample surveys
c. Collecting data through population census
d. Collecting data through hotel records
e. Collecting data through hospital records

Answer: D. Collecting data through hotel records
4. One of the following is not a source of secondary data:
a. Vital events registration
b. Hospital records
c. Online questionnaire
d. Disease registration
e. Environmental health

Answer: C. Online questionnaire
5. One of the following is not a method used to collect primary data:
a. Interviews
b. Questionnaires
C. Surveys
d. Experimentations
e. Census

## Answer: E. Census

1. When using alpha level of 0.05 , the test is considered to be statistically significant if:
a. $P=0.052$
b. $P=0.04$
c. $P=0.01$
d. $P=0.2$
e. A and D
f. B and C

## Answer: F. B and C

2. When we accept the null hypothesis at a level of significance equals 0.05 , this means:
a. $P>0.003$
b. $P<0.05$
c. $P>0.10$
d. $P<0.010$
e. $P<0.03$
3. When we accept the null hypothesis at a level of significance equals 0.05 . The widest range of possibilities for $p$ value will be when:
a. $P>0.003$
b. $P<0.05$
c. $\mathbf{P}>\mathbf{0 . 1 0}$
d. $P>0.05$
e. $P<0.03$

Answer: D. P > 0.05
4. For a specific statistical test, the $p$ value was equal to 0.04 . If the null hypothesis of that test was accepted, that is because:
a. Alpha was 0.05
b. Alpha was 0.01
c. Both $a$ and $b$
d. Neither of $\mathbf{a}$ or $\mathbf{b}$
5. To determine the critical region (area) in any statistical test, we have to have:
a. Alpha value
b. $P$ value
c. Degree of freedom
d. Both A and C
e. Both B and C

Answer: D. Both A and C
6. Consider having an alpha value of 0.01 , the test that is considered statistically insignificant will be when:
a. $P=0.04$
b. $P=0.005$
c. $P=0.003$
d. $\mathbf{P}=0.001$

Answer: $\mathbf{A} . \mathrm{P}=0.04$
7. When using a confidence level of 0.95 , the test is considered statistically significant if:
a. $P=0.21$
b. $P<0.04$
c. $\mathbf{P}>0.05$ but $<0.95$
d. All of the above

Answer: B. P < 0.04
8. When testing a hypothesis, it is important to know all of the following except:
a. Level of significance
b. Type of the test of significance
c. Degree of freedom
d. Type of the data we have
e. Type of the sample we have

Answer: E. Type of the sample we have
9. All of the following are true about $P$-value except:
a. It is a calculated probability of chance factor
b. All statistical significance tests should consider $P$ - values
c. It is reflecting the sampling error
d. It is the probability of the influencing factor
e. If it is small, conventionally less than $0.05, \mathrm{H}_{0}$ is rejected as it is implausible

Answer: D. It is the probability of the influencing factor
10. If $\boldsymbol{\alpha}=0.01$, the test is considered statistically insignificant when:
a. $\mathbf{P}=0.005$
b. $\mathbf{P}=0.007$
c. $P=0.001$
d. $P=0.000$
e. $P=0.013$
11. Obtaining a sound generalized information about population depending on the evidence of the sample is termed:
a. Presentation of data
b. Descriptive biostatistics
c. Confidence interval
d. Inferential biostatistics
e. Collection of data
12. If $\boldsymbol{\alpha}=0.001$, the test is considered statistically significant when:
a. $P=0.0100$
b. $P=0.0002$
c. $P=0.1000$
d. $P=0.0500$
e. $P=0.0040$

Answer: B. P= 0.0002

1. Concerning Chi square test all the following statements are true EXCEPT? Select one:
a. Applied when we have absolute frequencies.
b. It is used when we have two groups of population.
c. Cannot be applied when we have proportion rate alone
d. It is used when we have more than two groups of population
e. Degree of freedom is always equal to 1

Answer: E. Degree of freedom is always equal to 1
2. One of the following statements concerning Chi square test is correct:
a. The tabulated value depends on sample size
b. It depends on both the observed and the expected values of each cell
c. It is used for qualitative data
d. The expected value of each cell must be greater than the observed value
e. Degree of freedom in its case is $\mathbf{N}$-1

Answer: B. It depends on both the observed and the expected values of each cell
3. Degree of freedom in applying chi square test of $\mathbf{4} \times \mathbf{4}$ contingency table is:
a. 16
b. 1
c. 6
d. 4
e. 9
4. $\mathbf{C} \times \mathbf{R}$ table is well known as:
a. Table of an association
b. Contingency table
c. Cross- tabulation table
d. All of the above

Answer: d. All of the above
5. All of the following are correct statements concerning the Chi square test except:
a. Chi square is usually upper one- sided test
b. Degree of freedom depends on numbers of both columns and rows
c. The tabulated value depends on the degree of freedom
d. Chi square depends on both the observed and the expected values
e. None of the above is incorrect

Answer: A. Chi square is usually upper one- sided test
6. All of the following are correct statements concerning the Chi square test except:
a. Chi square is always upper one- sided test
b. Degree of freedom depends on numbers of both columns and rows
c. The tabulated value depends on the degree of freedom
d. Chi square depends on both the observed and the expected values
e. Chi square test is used for continuous data

## Answer: E. Chi square test is used for continuous data

7. Concerning Chi square test all the following statements are true EXCEPT? Select one:
a. The degree of freedom is $(\mathbf{c}-1) \times(r-1)$
b. Applied in qualitative data
c. Can be applied when we have proportion rate alone
d. It is used when we have more than two groups of population
e. It is used when we have two groups of population

Answer: C. Can be applied when we have proportion rate alone
8. In Chi square test, the tabulated Chi square value depends on:
a. The product of $(\mathbf{c}-1) \times(r-1)$
b. Level of significance and degree of freedom
c. The designated level of $\alpha$
d. The difference between $O$ and $E$
e. The expected value for each cell

## Answer: B. Level of significance and degree of freedom

9. In Chi square test, the null hypothesis could be rejected if:
a. The expected value for each column is equal to the expected value of the corresponding row
b. The difference between the observed and the expected values for each cell is large
c. The observed value for each cell is equal to zero
d. The difference between the observed and the expected values for each cell is small
e. The expected values in all cells are equal

Answer: B. The difference between the observed and the expected values for each cell is large
10. One of the following statements is correct in Ci square test:
a. One not rejects the alternative hypothesis when there is no significant difference
b. Null hypothesis uses proportion one equals proportion two
c. Alternative hypothesis uses proportion one equals proportion two
d. Null hypothesis uses proportion one does not equal proportion two
e. Null hypothesis uses mean one equals mean two

Answer: B. Null hypothesis uses proportion one equals proportion two
11. The following table presents the distribution of women according to the care received during pregnancy and the complications experienced during delivery.

| Care Received | Complications |  | Total |
| :---: | :---: | :---: | :---: |
|  | Present | Absent |  |
| No | 50 | 90 | 140 |
| Yes | 20 | 140 | 160 |
| Total | 70 | 230 | 300 |

The expected value for women who did not receive care and experienced complications during delivery is equal to:
a. $(\mathbf{7 0} \times 230) \div \mathbf{3 0 0}$
b. $(\mathbf{7 0} \times 140) \div \mathbf{3 0 0}$
c. $(50 \times 20) \div \mathbf{7 0}$
d. $(70 \times 50) \div \mathbf{3 0 0}$
e. $(70 \times 50) \div \mathbf{2 4 0}$

Answer: B. $(\mathbf{7 0 \times 1 4 0 )} \div \mathbf{3 0 0}$
12. The following table presents the distribution of 1000 women suffering from cystitis according to the prescribed antibiotic therapy as well as the treatment outcome

| Treatment <br> Outcome | Prescribed Antibiotic |  |  | Total |
| :---: | :---: | :---: | :---: | :---: |
|  | TIMP- SMX | Amoxicillin | Cyclacillin |  |
| Cured | 110 | 60 | 130 | 300 |
| Improved | 105 | 150 | 210 | 465 |
| Not cured | 35 | 90 | 110 | 235 |
| Total | 250 | 300 | 450 | 1000 |

The expected value for those who have been cured by amoxicillin is:
a. 18
b. 60
c. 90
d. 70.5
e. 139.5
13. The following table presents the distribution of 1000 women suffering from cystitis according to the prescribed antibiotic therapy as well as the treatment outcome

| Treatment <br> Outcome | Prescribed Antibiotic |  |  | Total |
| :---: | :---: | :---: | :---: | :---: |
|  | TIMP- SMX | Amoxicillin | Cyclacillin |  |
| Cured | 110 | 60 | 130 | 300 |
| Improved | 105 | 150 | 210 | 465 |
| Not cured | 35 | 90 | 110 | 235 |
| Total | 250 | 300 | 450 | 1000 |

The degree of freedom of the test statistics is:
a. 997
b. 1
c. 4
d. $\alpha$
e. 6

Answer: C. 4
14. One statement is correct regarding Chi square validity:
g. Chi square can apply when the total number is less than 20
$h$. When the overall total is less than 20 , all expected values should be at least 1
i. More than $\mathbf{2 0 \%}$ of the expected numbers should be less than 5
j. None of the expected numbers should be less than 1
k. Applied to the table showing just proportions or percentiles

Answer: D. None of the expected numbers should be less than 1
15. One statement is not correct regarding Chi square validity:
a. None of the expected numbers should be less than 1
b. If the overall total is less than 40 , all the expected values should be at least 5
c. There should be no cell zero
d. Less than $20 \%$ of the expected numbers should be less than 1
e. The total number should be more than 40

Answer: D. Less than 20\% of the expected numbers should be less than 1

Lectures \#14, \#15 and \#16 (T- Test):
** Use the following table for $t$ - test values whenever needed:


1. One statement is INCORRECT for the assumption of $\mathbf{t}$ test:
a. Normal distribution of the population of the sample
b. Randomization of the sample
c. The data are categorical
d. Dependency of the sample
e. The data are continuous

Answer: C. The data are categorical
2. If we want to know whether Indian women are taller than Jordanian women according to the height (cm) and we know the following information for Indian women: sample size $=60$, mean height $=180 \mathrm{~cm}$, standard deviation $=5$ and for Jordanian women: sample size=50; mean height 170 cm , standard deviation= 3
(Assuming that level of significance or alpha=0.05, and two-sided test). The calculated value will be:
a. 10.5
b. 6.3
c. 3.8
d. 7.2
e. 12.4
3. Assuming that level of significance or alpha=0.05, and two-sided test The Calculated Value $(t)=1.78$, and the sample size $(n)=78$. We conclude that:
a. $\mathbf{P}<0.001$
b. $P=0.010$
c. $P<0.05$ and $>0.010$
d. $\mathbf{P}<0.05$ and $>0.020$
e. $P<0.100$ and $>0.050$

Answer: $\mathrm{E} . \mathrm{P}<0.100$ and $>0.050 ?$
4. If it is known that the mean blood sugar of adults in Jordan is: $\mathbf{1 2 0} \mathbf{~ m m o l} / \mathbf{l}$ and we want to test whether mean blood sugar of adults in Al- Karak governorate is the same or different from the Jordanian population. The sample size $=81$ adults, their arithmetic mean of blood sugar= $124 \mathrm{mmol} / 1$ and standard deviation=18.
(Assuming that level of significance or alpha=-0.05, and two-sided test). The Calculated Value(t) will be:
a. 1.55
b. 2.00
c. 3.22
d. 2.50
e. 4.15

Answer: b. 2.00
5. In two tailed $\mathbf{t}$ - test at alpha $=\mathbf{0 . 0 1}$ and total subjects $=29$. The critical $\mathbf{t}$ value is:
a. 2.16
b. 2.46
c. 1.3
d. 1.96
e. 3.46

Answer: A. 2.76
6. A first-year medical student student was given the following information: mean of the sample $=85$, mean of the population $=90, S D=4.14$, sample size $=31$ and $\alpha=$ 0.05 . The degree of freedom they must use will be:
a. 30
b. 31
c. 32
d. 5.57

Answer: A. 30
7. A first-year medical student student was given the following information: mean of the sample $=85$, mean of the population $=90, S D=4.14$, sample size $=31$ and $\alpha=$ 0.05 . The critical value of the statistical test of choice will be:
a. 2.04
b. 1.7
c. 6.72
d. 1.13
e. 6.62
8. The temperature of 10 subjects suffering from tonsillitis before ( $40,40,37,38,39$, 39, 38, 38, 39,38 ) and after 4 hours of Panadol therapy became (37.38.38,38.37.37.38.38.37,37), Respectively (Assuming that level of significance or alpha=0.05, and two-sided test) The Calculated value (t) is:
a. 5.1
b. 2.1
c. 3.2
d. 6.6
e. 2.7

Answer: E. 2.7
9. A first-year medical student student was given the following information: mean of the sample $=85$, mean of the population $=90, S D=4.14$, sample size $=31$ and $\alpha=$ 0.05 . The calculated value of the statistical test of choice will be:
a. 2.04
b. 1.7
c. 6.72
d. 1.13
e. 6.62

Answer: c. 6.72
10. In two tailed $\mathbf{t}$ - test at $\alpha=0.001$, and total subjects $=29$. The critical $\mathbf{t}$ value is:
a. 2.76
b. 1.70
c. 3.68
d. 2.48
e. 1.96

Answer: c. 3.68
11. One statement is incorrect to assume in one sample $t$ - test:
a. Normal distribution of the population of the chosen sample
b. Independency of the sample
c. Randomization of the sample
d. Dependency of the sample
e. None of the above

## Answer: D. dependency of the sample

12. You would like to see whether your colleagues' weights differ from general population. The colleagues' weights are normally distributed; the average population weight is 70 kg . The sample size=100, the sample mean= 75:20. (2sided, Set alpha=0.05). So, the calculated value of $t$ test is:
a. 1.40
b. 0.15
c. 3.05
d. 2.75
e. 2.50

Answer: E. 2.50
13. You would like to see whether your colleagues' weights differ from general population. The colleagues' weights are normally distributed; the average population weight is 70 kgs . The sample size $=100$, the sample mean= 75:20 $\mathbf{k g s}$. (2-sided, Set alpha=0.05). So, the decision to be taken according to your calculated value of $\mathbf{t}$ test is:
a. Accept the null hypothesis
b. Reject the alternative hypothesis
c. Fail to reject the null hypothesis
d. Accept the alternative hypothesis
e. Cannot be determined and more information are needed.

Answer: d. Accept the alternative hypothesis
14. The critical value for two tailed $t$ - test at alpha= 0.05 , and the total subjects in the study $=31$, is:
a. 2.15
b. 3.82
c. 1.72
d. 2.04
e. 2.83

Answer: d. 2.04
15. The critical value for two tailed $t$ - test at alpha= 0.001 , and the total subjects in the study $=13$, is:
a. 4.22
b. 3.85
c. 2.65
d. 4.32
e. 3.93

## Lecture \#17 (ANOVA Test):

No questions regarding ANOVA test nor Post Hoc Analysis were found in the archives.


1. One of the following observational studies are related to the grouped data:
a. Ecological study
b. Cross-sectional study
c. Cohort study
d. Case control study
e. Clinical trial

Answer: A. Ecological study
** Ecological study concept was not covered in our lectures. However, it can be solved by exclusion: (E) clinical trial is an experimental study. (B), (C) and (D) are types of observational studies that generally involve collecting data at the individual level and do not typically involve grouping the data **
2. One of the following statements is correct:
a. A case control study is more expensive, in comparison to cohort study.
b. A case control study needs a long period to yield the results, whereas cohort study yielding relatively quick results.
c. A case control study is more appropriate when the disease under investigation is rare in comparison to cohort study.
d. A case control study needs a large sample size in comparison to cohort study.
e. A case control study starts with exposure to risk factor or suspected cause, while cohort study starts with a disease

Answer: C. A case control study is more appropriate when the disease under investigation is rare in comparison to cohort study.
3. One of the following is false about cohort study:
a. Proceeds from cause to effect
b. More than one disease can be detected.
c. Relative risk can be determined
d. Helpful for evaluation of rare disease
e. Cases have to be followed for a long time.

Answer: D. Helpful for evaluation of rare disease
4. A case control study was conducted, to investigate relation of tobacco viewed with oral cancer it's observed that $\mathbf{3 5}$ out of 50 patients were tobacco chewers, as compared to $\mathbf{2 0}$ tobacco chewers out of $\mathbf{5 0}$ control subjects. The odds ratio of oral cancer associated with tobacco chewing was found to be 3.5, this means:
a. There is a $\mathbf{3 . 5 - f o l d}$ increase in the risk of oral cancer among individuals who chew tobacco compared to those who do not.
b. There is a $3.5 \%$ increase in the risk of oral cancer among individuals who chew tobacco compared to those who do not.
c. There is a 3.5 -fold increase in the risk of oral cancer among individuals who do not chew tobacco compared to those who do.
d. There is a $3.5 \%$ increase in the risk of oral cancer among individuals who do not chew tobacco compared to those who do.

Answer: A. There is a 3.5-fold increase in the risk of oral cancer among individuals who chew tobacco compared to those who do not.
** OR (Odds ratio) is used for the analysis of case control studies. Calculating OR is not required as per Prof. Waqar. However, its interpretation is **
5. If you want to compare the presence of cancer risk factors between residents in Amman and Aqaba, the best study to conduct is:
a. Cross-sectional study
b. Clinical trial
c. Case control study
d. Cohort study
e. Survey

Answer: E. Survey
** "Case control study can be useful for studying the association between risk factors and disease outcomes but may not be the most appropriate for comparing risk factors between different populations. Cross-sectional study could be used to compare risk factors, but a survey would be a more appropriate and efficient method for this specific research question." **
6. All the following are true about Case Series studies, EXCEPT:
a. Experience of a group of patients with a similar diagnosis
b. Cases may be identified from a single or multiple source
c. Assesses incident disease
d. Generally, report on new or unique condition
e. May be only realistic design for rare disorders
** Case Series were only mentioned as a type of descriptive observational studies**
7. The relation of lung cancer and cigarette smoking was studied, a group of patients with lung cancer was collected and compared with lung cancer free group the number and duration of cigarette smoking was then compared in the ho groups of the following statements, one is correct:
a. This study is a prospective cohort.
b. Incidence of disease can be calculated
c. Prevalence of cigarette smoking can be calculated
d. Relative risk can be measured
e. Prevalence disease can be calculated

Answer: C. Prevalence of cigarette smoking can be calculated ** This is a case- control study. In case control study we measure exposure (causative or risk factors) only **
8. One of the following statements is NOT true about cohort study:
a. Provides incidence of disease
b. Indicated when there is good evidence of association between exposure and disease
c. Done when ample funds are available
d. Study starts with exposure to risk factor
e. Done when disease is very is rare

Answer: E. Done when disease is very is rare
9. The value of an odd's ratio was studied for disease and cause and found as 1 . It indicates:
a. the presence of the cause does not increase or decrease the likelihood of developing the disease.
b. No association
c. Positive association
d. Unsatisfactory data
e. Incorrect calculation
f. Both a and b

Answer: F. Both a and b

10. The Temporal Sequence of exposure and effect may be difficult to determine in:
a. Case series studies
b. Clinical trial
c. Case control study
d. Cross-sectional study
e. Cohort study

## Answer: D. Cross- sectional study

11. Recall bias is most associated with one of the following study designs:
a. Case control study
b. Cohort study
c. Case report
d. Cross-sectional study
e. Case series study
12. All the following are advantages of Case Series studies, EXCEPT:
a. Useful for hypothesis generation
b. Informative for very rare disease with few established risk factors
c. Characterizes averages for disorder
d. Can study cause and effect relationships
e. Cannot assess disease frequency

Answer: D. Can study cause and effect relationships
13. Prevalence rate of a disease is calculated from:
a. Prospective cohort study
b. Case series study
c. Case report study
d. Cross Sectional study
e. Case-control study

Answer: D. Cross Sectional study
14. Odds ratio for bladder cancer and smoking is 2.4, this means:
a. odds exposure of controls is $\mathbf{2 . 4}$ times more than odds of exposure of cases
b. people with bladder cancer are 2.4 times more likely to be smokers
c. smokers have $2.4 \%$ more risk of bladder cancer than non-smokers
d. smokers have 24 times more risk than non-smokers
e. smoking is a protective factor for bladder cancer

Answer: B. people with bladder cancer are 2.4 times more likely to be smokers
15. The false statement about case control study is:
a. can be used for rare outcome
b. can be used for multiple exposures with one outcome
c. exposure is retrospective
d. usually used for rare exposure
e. the sample must contain all cases in population

Answer: D. Usually used for rare exposures ** (E) is not necessarily true too, as some exclusion criteria may be applied on the cases **
16. One of the following is false about cross-sectional study:
A. disease etiology cannot be studied
B. no follow up
C. gives prevalence of disease
D. is taken in one-time frame
E. unit of analysis is the group

Answer: E. Unit of analysis is the group
** As per (A), "Cross sectional studies can provide valuable information about disease prevalence and the distribution of exposures and outcomes in a population, but they are limited in their ability to establish a causal relationship. This is because cross-sectional studies capture data at a single point in time, making it challenging to determine the temporal sequence of exposure and outcome". **

## Lecture \#21 (Correlation and Linear Regression):

- No questions regarding correlation and Linear Regression were found in the archives.
- Prof. Waqar won't ask us to calculate Pearson's Correlation Coefficient (r) not the linear regression coefficient (b), but will definitely ask about their interpretation.


