Add on course - MID Term Examination

Biostatistics & DoE for Pharmaceutical Industry.

Dr B. C. Roy College of Pharmacy and Allied Health Sciences.

Ram Swarup Chattopadhyay.

18901920001

B. Pharm 4th year.

Full Marks - 30

Time - 1 hrs.

1) Write short notes on :- a) Hypothesis testing B) Central tendency of dispersion C) Multiple Linear Regression and its application in DoE.

2) A: Differentiate b/w F test, Z-test, & Student's t test. F test is ~~done~~ done prior to performing t-test.

3) With example differentiate b/w two way and one way ~~anova~~. Anova? Name a few non parametric test which ~~can~~ can be used in the place of ~~of~~ one way anova and two way anova.

4) A pharmaceutical company launched ~~a new~~ an ~~antidi~~ anti-diabetic ~~drug~~ drug ~~and~~ and perform a ~~st~~ survey. on a set of population. It is has been found depends on both dose & time of administration assuming 6 doses of the drug and 6 times of administration. Draw a latin square design. to the clinical trial. ~~Differentiate b/w the~~

Differentiate b/w and latin square design & 2^2 factorial, with example..

Prof. (Dr.) Samir Kumar Samanta
M. Pharm., Ph.D (J.U.)
Principal
Dr. B. C. Roy College of Pharmacy & AHS
Durgapur, West Bengal-713206

Dr. B. C. Roy College of Pharmacy & AHS * Durgapur

Mid-Term Examination

Add on Course :- Biostatistics and DoE for pharmaceutical Industry

Dr. B. C. Roy College of Pharmacy and Allied Health Science.

Name : Sinchan Kumar Roy

Roll no : 18901920101

Year : B. Pharm 4th year.

Full marks : 30 Time :- 1 hour

Question :- [3×5 = 15]

1. Write short notes on a. Hypothesis Testing. b. Central Tendency Dispersion. c. Multiple linear regression and its application in DoE

2. Differentiate between F Test, Z test & Student T test. F is done prior to performing T test why? [2.5×2 = 5]

3. With example differentiate between two way ANOVA & one way ANOVA. Name a few non-parametric test which can be used in place of one way ANOVA and two way ANOVA. [5]

4. A pharmaceutical company launched an anti-diabetic drug and perform Statistical survey over a set of population. It has been found that the drugs efficacy depends on both dosage and time of administration. Assuming 6 dosage of the drug & 6 time of administration in a day. Construct a latin square design to do this trial. Differentiate between latin square design & 2² factorial design with example. [3+2]

Dr. B. C. Roy College of Pharmacy & AHS * Durgapur

Prof. (Dr.) Samir Kumar Samanta
M. Pharm., Ph.D (J.U.)
Principal
Dr. B. C. Roy College of Pharmacy & AHS
Durgapur, West Bengal-713206

a. Hypothesis Testing :-

Hypothesis Testing is a method in statistical analysis of a data based on an design made on the bases on the theorical hypothesis. Hypothesis Testing can be done to find the result of a particular set of data or the on the data the researcher want to find out. It can help the researcher to determine the data collect whether match between the supposed result or alternate result.

Two types :- i. Null Hypothesis (H_0)
ii. Alternate Hypothesis (H_1/H_0)

i. Null Hypothesis :-

i. Denoted by H_0 (in general)
ii. The data result compared to the natural set of data.
iii. The data states that there is no difference.

ii. Alternate Hypothesis :-

i. Denoted by H_1/H_0.
ii. The Data result compared to the natural set of data and state that it is different from the natural data set.
iii. Mainly show the difference between the data.

Central Tendency Dispersion :—

- **Mean :-** Mean is referred the average value of the data set. Mean is ~~reffered to the~~ a Statistical parameter which is used to determine the average value of the data. It is determined by the ratio of the sum of the data and the number of data factor present.

$$\text{Mean} = \frac{\text{Total Sum of the Data}}{\text{Number of Data factor}}$$

- **Median :-** Median is referred to the ~~statistical~~ Statistical ~~word~~ mid data of the data set. The exact middle data in a data sheet is the value of the median in that particular data set.

- **Mode :-** Mode is reffered to the highest frequency of the data occured in a data set. The number of times a data repeat itself in a data set is known as Mode.

Example :

Class	~~Height~~ Weight (kg)
I	20
II	30
III	40
IV	20

- Mean = $\frac{20+30+40+20}{4}$

$= \frac{120}{4} = 30$ kg

- Mode =

- Median

20, 30, 40, 30

Multiple linear Regression :-

Multiple linear Regression is an statistical technique to shortening a large data set. We can delete that data set which is less than 0.05 and then the program that work analysis result be nearly same as the previous one. Minor change can be there.

Application :-

1. Factorial Design (2^K factorial Design)
2. It used to reduce the data set and optimize the data.

F Test	Z Test	T Test
1. Data set is <30.	1. Data set is ≥ 30.	1. Large Data Set.
2. Formula :- Used to determine whether the data value is equal variant or unequal variant.	2. Formula = $Z = \frac{\lvert\bar{x} - u\rvert}{\frac{\sigma}{\sqrt{n}}}$ $\bar{x}$ → Statistical mean of data u → Hypothesis mean σ → Standard Deviation. n → no. of data.	2. Formula = $T = \frac{\lvert\bar{x} - \text{Mean}\rvert}{\text{Standard Deviation}}$ $\bar{x}$ → Statistical Mean u → Hypothesis mean

. F test done prior to the T test because F test help us to forecast the data whether is unequal variant or equal variant.

value is less than 0.05, it is equal variant.

value is more than 0.05, it is unequal variant.

Dr. B. C. Roy College of Pharmacy & AHS * Durgapur

Prof. (Dr.) Samir Kumar Samanta
M. Pharm., Ph.D (J.U.)
Principal
Dr. B. C. Roy College of Pharmacy & AHS
Durgapur, West Bengal-713206

FACTORIAL CHART QUESTION 4

2^3 factorial.

A	B	C	AB	BC	CA	ABC
+1	-1	+1	+1	+1	+1	-1
-1	+1	+1	-1	+1	-1	-1
+1	-1	+1	-1	-1	+1	-1
-1	-1	+1	+1	-1	-1	+1
+1	+1	-1	+1	-1	-1	-1
-1	+1	-1	-1	-1	+1	+1
+1	-1	-1	-1	+1	-1	+1
-1	-1	-1	-1	+1	+1	-1

40:-

Table

Subject	Dosage	Time of Administration
1	2 = A	7 = B
2	3 = B	6 = C
3	5 = C	4 = D
4	4 = D	5 = A
5	5 = E	3 = F
6	6 = F	2 = E

Row	COLUMN	TREAT	RESULT
1	1	A	2
1	2	B	7
1	3	B	8
1	4	C	6
2	1	C	5
2	2	D	4
2	3	D	4
2	4	A	5
1	5	E	5
1	6	F	3
2	5	F	6
2	6	E	2

Latin square design is in a square form with specific number of row & column.

1	A	B
2	B	
3	C	
4	D	A

• 2^B factorial design is a design of data set there is several data / factor : 1

For Example :- $2^3 = 8$

A	B	C	AC	BC	CA	ABC
+1	+1	+1	+1	+1	+1	+1
−1	+1	+1	−1	+1	−1	−1

and so on.

Refer to page I.

Qs. One-way ANOVA :-

Anova stand for the analysis of varients. One way ANOVA means there is only one factor that effects the result of the data set.

Example :- The body temperature before and after administration of anti-pyretic drug.

Two way ANOVA :-

Two way ANOVA means there are two factors that effect the result of the data set.

Example :- The Anti-diabetic drug depends on the dose and the time of the administration.

Two non-parametric Test :-

i. Willcoxan Test
ii. Rank signed Test
iii. Mann-whitney Test.

. Mid Term Examination.

Add on course → Biostatistics & DOE for Pharmaceutical Industry.

Dr. B.C. Roy College of Pharmacy & A.H.S.

Name → Pritam Jana.

Univ. Roll No. → 18901920073.

Year → B. Pharm. 4th year.

Time → 1hr.

F.M. → 30.

(A) Question 1:- Write some notes on; A) Hypothesis testing. B) Central tendency of Dispersion. C) Multiple linear regression & its application in DOE. (3×5 = 15)

(B) Differentiate between F test, Z test & student t test. F test is done prior to Performing t test why? (2.5+2.5 = 5)

(C) With example. Differentiate b/w two way & one way Anova? Name a few Parametric test which can be used in place of one way & two way ANOVA (5)

(D) A Pharmaceutical company launched an Antidiabetic drug & Perform a Statistical survey over a set of population; It has been found that the drugs efficacy depends on both dose & time of administration assuming 6 doses of the drug & 6 times of administration in a day, Construct a Latin square design to do this clinical trial. (3+2)

Differentiate b/w LSD & 2^3 factorial design. with example?

Answer:-

A) a) Hypothesis testing:- Hypothesis testing is determine/measure the difference between Hypothesized or standardard mean value and sample mean value; which we get after calculation.

There are several Hypothesis testing → like → null hypothesis and alternate hypothesis.

For null hypothesis → there is no significant difference between Hypothesized* and Calculated mean.

Null hypothesis is represented by H_0; defined as μ = Hypothesized mean;
μ_a = Calculated mean;
For a tablet avg. wt is 325 mg; is Represent as H_0:

• If null hypothesis not follow then alternate hypothesis says that Hypothesized mean is not equals to calculated mean;

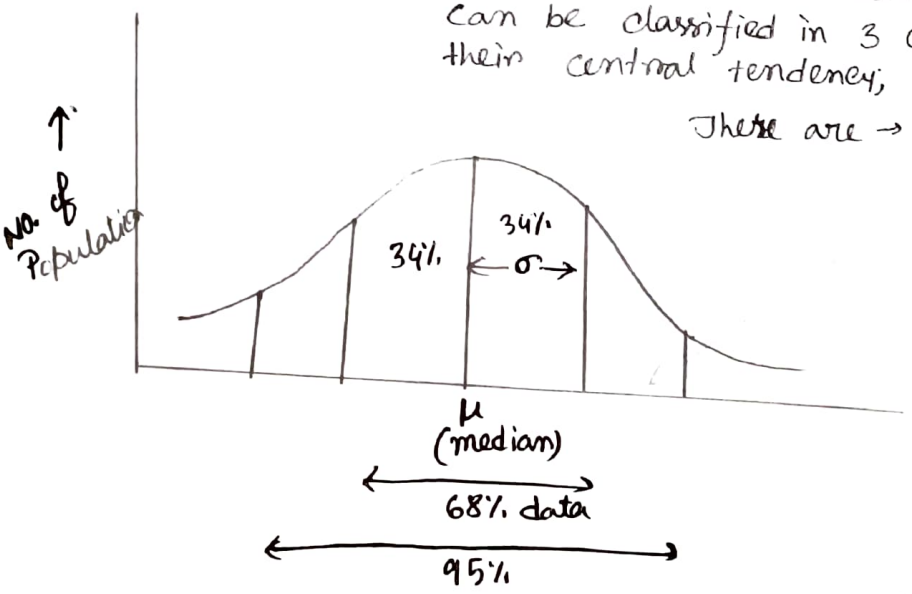
$H_0: \mu \neq \mu_a$;

For the above same example for alternate Hypothesis;

$\mu_a \neq 325$ mg;
It might be greater than 325 mg or less than 325mg

Central tendency of Dispersion:- For a set of data there is several dissimilarities which can be classified in 3 classes; which is their central tendency;

There are → Mean.
Median.
Standard deviation



• Mean:- Mean is nothing but the arithmetic average or mean of a set of data; which is lies in the centre of the data on histogram Plot;
represent as → $\overline{X} = \frac{X}{N}$ N = No. of operation.

• Median:- In statistical evaluation median separates the higher half of value to lower half of the values.

Median is represent by (M) and defines the 50% value; means at 50% of the value/middle value, upon which the 50% greater data lies & below which 50% lesser data lies.

For a data set; for odd no. variables; $M = \left(\frac{n+1}{2}\right)^{th}$ Position.

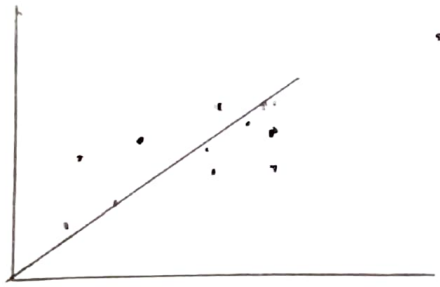
· Standard deviation (σ) :- It is represent by σ; which defines the deviation of data/value from standard one.
How much difference is there between standard value & calculated value,

Formula is — $\sigma = \sqrt{\frac{\Sigma (x - \bar{x})^2}{n-1}}$; where. $\bar{x} = \frac{sum}{n}$

MLR or Multiple Linear Regression is a statistical Evaluation Tool.

Multiple Linear Regression. is a statistical technique that uses several explanatory variables to predict the outcome of response variables.

It allow us to obtain predicted values for specific variables under certain conditions.



→ By applying MLR; we nullify the unessential Process or Parameter in DOE; and Perform the Process optimization.

· Those value which is far beyond the straight line are neglected; generally those P value is > 0.05; are not ~~calculate~~ taken for DOE and Process Perform.

· So, By applying Multiple Linear Regression; we optimize Process by rejecte the unnecessary Parameters; which not influence the result or outcome.

Dr. B. C. Roy College of Pharmacy & AHS * Durgapur

Prof. (Dr.) Samir Kumar Samanta
M. Pharm., Ph.D (J.U.)
Principal
Dr. B. C. Roy College of Pharmacy & AHS
Durgapur, West Bengal-713206

(B)

Z test

A one sample Z-test is used to check if there is a difference b/w the sample mean and Population mean when standard deviation is known

calculate by—

$$Z\ test = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$\bar{x}$ = mean of sample
μ = " " Population
σ = Std. deviation

→ when Population is >30 then z test Perform

student t-test

student-t test is used to determine or perform to see wheather the difference between response of two group is statistically significant or not.

calculated by—

$$t = \frac{\bar{m} - \mu}{s/\sqrt{n}}$$

→ If Population is <30 then t-test Perform.

F-te

F-test also known as Fischer is Performed before t-test to evaluate the data is come from same population, or source or not, It perform to measure the P value and observe if there is any significant difference or not

(C) <SD

most

•• F test is done prior to performing t test, bez →

F test or Fischer test is performed before t-test to evaluate the data is come from same population or source or not;

To analyse that the data should be proper for performing t-test;

By running F-test we have a P value;

If the P value is <0.05 that defines that there some significant difference b/w the values;

The P value must be >0.05 for further performing t-test

F-test abr. Fischen -med

(4)

LSD

latin sq. design is use of most experiment dosage experiment limit are heterogeneous and the heterogeneity in two direction

(5)

2^k factorial design.

F.D. is an experiment whose design consist of two or more factor each with different possible values.

•• Latin square Design construction :-

Time of Administration →

No. of doses ↓	A	B	C	D	E	F
1	A	B	C	D	E	F
2	B	C	D	E	F	A
3	C	D	E	F	A	B
4	D	E	F	A	B	C
5	E	F	A	B	C	D
6	F	A	B	C	D	E

A, B, C, D, E, F defines 6 times a day

Main difference between One way ANOVA & two way ANOVA is — Number of independent variables

A - One way anova has one independent variables; where as two way anova has 2 or more than two variables; which are independent and one can be dependent variable.

Ex. of Non parametric test are → Wilcoxon signed test.

→ signed [illegible]

→ Wilcoxon [illegible] test.

→ Kushkal [illegible]

→ Mann whitney Test etc

Biostat & DOE
Mid-term examination

Siddhanta Mishra
B. Pharm 1st year

Date: 24/01/21

Biostatistics & DOE for Pharmaceutical Industry (Add-on course)
Dr. B.C. Roy College of Pharmacy & Allied Health Sciences

Siddhanta Mishra
Univ. Roll- 18901922105.
B. Pharm 4th year.

Time - 1 hr.

Full Marks - 30

Q.1) Write short notes on — 3x5 = 15

a) Hypothesis testing
b) Central Tendency of dispersion.
c) Multiple Linear regression and its application in DOE.
d)

Q.2) Differentiate b/w F-test, Z-test and student t-test. 2.5 x 2 = 5

Z-test is done prior to performing t-test. why? (5)

Q.3) With example differentiate b/w Two way Anova and One way Anova.
Name a few non parametric test which can be used in place of One Anova and Two way Anova.

Q.4) A Pharmaceutical company launched an antidiabetic drug and performed the statistical subway over a set of population. It has been found that the drugs activity efficacy depends on both dose and Time of administration. assuming 6 doses of the drug and 6 time of administration in a day construct a LSD to do this clinical trial. (3+2).

Differentiate b/w LSD and 2^3 factorial design with example.

Dr. B. C. Roy College of Pharmacy & AHS * Durgapur

Prof. (Dr.) Samir Kumar Samanta
M. Pharm., Ph.D (J.U.)
Principal
Dr. B. C. Roy College of Pharmacy & AHS
Durgapur, West Bengal-713206

Q.1) **Hypothesis testing** :- Hypothesis testing is used to assess the output of data sets (Paired / non-paired) is same/or different — Ultimately it used to check the result that the result is valid or not.

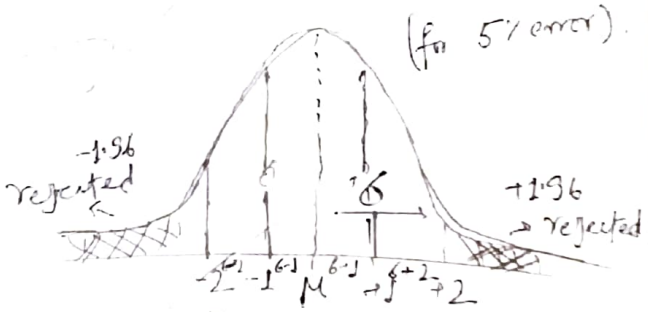
It uses null hypothesis and Alternate hypothesis.

If there is any significant differences b/w two data sets then null hypothesis will be rejected otherwise it will be accepted.

It depends on p-value (Probability of error)

If p-value < 0.05 then null hypothesis is accepted

p value > 0.05 " " Rejected



Types - a) F test b) T-test c) Z test d) One way Anova e) Two way Anova.

Hypothesis testing also done by some non parametric test like Wilcoxon signed rank test, Wilcoxon ranked sum test, Mann-whitney test.

Prof. (Dr.) Sujit Kumar Samanta
M. Pharm., Ph.D (J.U.)
Principal
Dr. B. C. Roy College of Pharmacy & AHS
Durgapur, West Bengal-713206

b) Central Tendency of dispersion (Mean, Mode, Median).

Suppose data set is

$f(x) = 2, 4, 3, 6, 2, 8, 3, 2, 5$

Mean: $(\mu) = \left(\frac{\Sigma x_i}{n}\right)$

Mean $(\mu) = \frac{2+4+3+6+2+8+3+2+5}{9} = \frac{35}{9}$

$= 3.88$

Median :-

To calculate mean value we have to assigned numeric in ascending / discending order.

for $f(x)$ 2, 2, 2, 3, (3), 4, 5, 6, 8,

for odd number the position of median will be $\left(\frac{n+1}{2}\right)^{th}$ position

for even " " $\left(\frac{n}{2} + \frac{n+1}{2}\right)$ position

For that data set, Median will be $\frac{9+1}{2} = 5^{th}$ number numerical. that is $= 3$

Mode :- To assess mode of a data sets, we have to prepare frequency table of given data sets.

$f(x) = 2, 2, 2, 3, 3, 4, 5, 6, 8$

Numeric	Frequency (f)
(2)	3
3	2
4	1
5	1
6	1
8	1

The data having maximum frequency will Mode of the data sets.

Here, in this data set Mode is 2

Q) Multiple linear regression :-

Q.2

F-test	Z-test	t-test
F test is done prior to the t-test	Z test is done for more than 30 popul. (i.e $n \geq 30$) $z = \frac{\bar{x} - M}{\sigma}$	t-test is done for less than 30 population. $t = \frac{\bar{x} - M}{S}$

Why F test is done prior to performing t-test.

=> In statistical data analysis, F test is done prior to t-test because t-test is depend on the result or outcome of the F-test.

Generally T-test is done with equal variants and unequal variants.

To confirm the the variants of test is done before t-test.

Interpretation of F-test result and its app^n in the t-test =>

By performing f test we get a p value.

If The p value greater than 0.05 implies that there is no significant difference and data sets are homogeneous hence, t-test is done with equal variants

p value is less than 0.05, then unequal variance t-test is done.

One way Anova

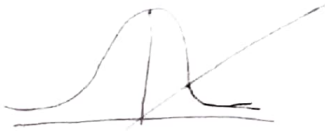
One way Anova is done for data set that is depend on single factor.

Ex- Hardness testing of diff^{n} tablets by one same testes.

Two way Anova

Two way is Anova is conducted for data set that is depend on single factor.

Ex -> Hardness testing of diff^{n} tabs by different tester.



▷ Non-parametric tests which are used in place of— one way & two way Anova.

⇌ Will-coxon signed-rank test (In place of one way Anova)
Will-coxon ranked sum test (in place of one way Anova)
Friedman test (In place of Two-way Anova).

Dose	Time of administration (min)					
	A	B	C	D	E	F
1	A	B	C	D	E	F
2	C	A	D	B	F	A
3	D	A	F	C	B	E
4	F	C	E	A	D	B
5	E	F	B			
6	B					

A = 6 am
B = 10 am
C = 2 Pm
D = 6 Pm
E = 10 Pm
F = 2 am

Dose (D)	Time of Administration (min)					
	A	B	C	D	E	F
1	A	B	C	D	E	F
2	B	C	D	E	F	A
3	C	D	E	F	A	B
4	D	E	F	A	B	C
5	E	F	A	B	C	D
6	F	A	B	C	D	E

In LSD, rows and the number of row and column are. In this design, same data is not overlapped in same row and columns.

- But in 2^3 factorial design same data may overlapped in same row / column.
- 2^3 factorial design ignore intermediate level and Highest and lowest level but in LSD does not ignore intermediate level.

Q.3

Q) Multiple linear Regression

MLR means (Multiple linear regression) is a statistical technique that uses several explanatory variable to predict the outcome of a responsible variable.

Application →

(i) To Predict values for specific variable under certain condition.

(ii) Processing of predicting result.

(iii) Used to investigating relationship

(iv) Analyze and predicting curves.

Q.4 Diff^n 2^(3) & 2^3 factorial design

LSD

A	B	C	D
B	C	D	A
C	D	A	B
D	A	B	C

2^3 FD

A	B	C	AB	BC	CA	ABC
+1	+1	+1	1	1	1	1
1	1	−1	1	−1	−1	−1
1	−1	+1	−1	−1	+1	−1
1	−1	−1	−1	1	−1	+1
−1	−1	−1	1	1	1	−1
−1	−1	+1	1	−1	−1	+1
−1	+1	−1	−1	−1	+1	+1
−1	+1	+1	−1	+1	−1	−1

Dr. B. C. Roy College of Pharmacy & AHS * Durgapur