Statistical Methods for Data Science

Testing of Hypothesis

Time (min) for three drying formulas for curing glue are studied

Formula A	13	10	8	11	8	5		
Formula B	13	11	14	14	12	6	10	
Formula C	17	14	13	10	11	12	16	15

Test at 5% level of significance whether is any difference in the mean curing time of glue?

Find between which two formulas the mean difference has contributed significantly using least significant difference post-hoc test?

Analysis of Variance (ANOVA)

Statistical Methods for Data Science

Testing of Hypothesis

Statistical Methods for Data Science

Testing of Hypothesis

Statistical Methods for Data Science

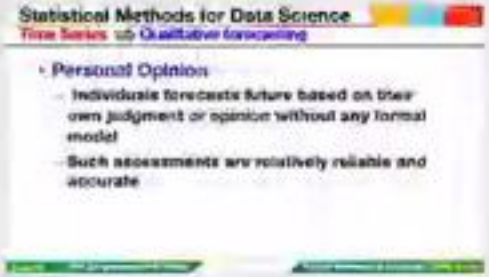
Statistical Methods for Data Science

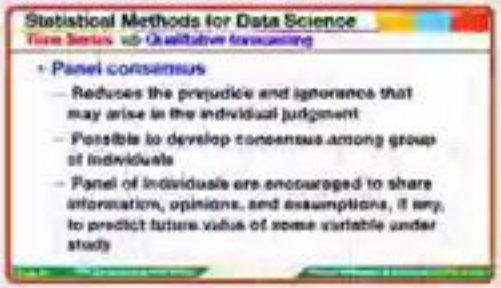
Time Series ➡ Qualitative forecasting

- **Panel consensus**
  - Reduces the prejudice and ignorance that may arise in the individual judgment
  - Possible to develop consensus among group of individuals
  - Panel of individuals are encouraged to share information, opinions, and assumptions, if any, to predict future value of some variable under study

Firefox
File Edit View History Bookmarks Tools Window Help
Prof Gangaboraiah B Anandaraiah PhD
Sat 26 Feb 6:21 PM
New Tab
Search with Google or enter address
Getting Started
Firefox
Amazon
mail.google
cran.r-project
accounts.go...
youtube
photos.google
docs.google
Recommended by Pocket
Webex
FEB
26

Firefox File Edit View History Bookmarks Tools Window Help Prof.Gangaboralah B Andanalah PhD Sat 26 Feb 6:21 PM





Statistical Methods for Data Science

Time Series ⇒ Qualitative forecasting

- **Panel consensus**
  - Reduces the prejudice and ignorance that may arise in the individual judgment
  - Possible to develop consensus among group of individuals
  - Panel of individuals are encouraged to share information, opinions, and assumptions, if any, to predict future value of some variable under study

Statistical Methods for Data Science

Time Series ➡ Qualitative forecasting

- **Panel consensus**
  - Reduces the prejudice and ignorance that may arise in the individual judgment
  - Possible to develop consensus among group of individuals
  - Panel of individuals are encouraged to share information, opinions, and assumptions, if any, to predict future value of some variable under study

Analysis of Variance (ANOVA)

Statistical Methods for Data Science

Testing of Hypothesis

Statistical Methods for Data Science

Testing of Hypothesis

Statistical Methods for Data Science

Chat
Participants (52)
Disable group chat
Rahul Sahu
Hi Professor

Statistical Methods for Data Science

Testing of Hypothesis

Time (min) for three drying formulas for curing glue are studied

Formula A	13	10	8	11	8	5		
Formula B	13	11	14	14	12	6	10	
Formula C	17	14	13	10	11	12	16	15

Test at 5% |
curing time

Virtual Classroom
Chat
Participants (52)
Disable group chat
Rahul Saha
Hi Professor

Statistical Methods for Data Science

Testing of Hypothesis

Time (min) for three drying formulas for curing glue are studied

Virtual Classroom
Chat
Participants (53)
Disable group chat
Rahul Saha
Hi Professor

Statistical Methods for Data Science

Testing of Hypothesis

Time (min) for three drying formulas for curing glue are studied

Formula A	13	10	8	11	8	5		
Formula B	13	11	14	14	12	6	10	
Formula C	17	14	13	10	11	12	16	15

Test at 5% level of significance whether is any difference in the mean curing time of glue?

Virtual Classroom
Chat
Participants (54)
Disable group chat
Good Evening

Statistical Methods for Data Science

Testing of Hypothesis

Time (min) for three drying formulas for curing glue are studied

Formula A	13	10	8	11	8	5		
Formula B	13	11	14	14	12	6	10	
Formula C	17	14	13	10	11	12	16	15

Test at 5% level of significance whether is any difference in the mean curing time of glue?

Impartus Prof.Gangaboraiah B Aridanaiah PhD Sat 26 Feb 6:23 PM

Virtual Classroom
Chat
Participants (54)
Disable group chat
Good Evening
Hi Sir
Good evening sir
Good Evenings Sir
Good evening Sir
GE All
Good evening Sir!
GE
ENANKO BHOWMICK JOINED
K MOHAMMED NAYAZUDDIN JOINED

Methods for Data Science

pothesis

ime (min) for three drying formulas for curing glue are tudied

nula A	13	10	8	11	8	5		
nula B	13	11	14	14	12	6	10	
nula C	17	14	13	10	11	12	16	15

el of significance whether is any difference in the mean f glue?

Slide B'lore

Impartus Prof.Gangaboraiah B Anbanaiah PhD Sat 26 Feb 6:23 PM

Virtual Classroom
Chat Participants (56)
Disable group chat
Hi Sir
Prakash Kumar
Good evening sir
S J Krishna Kaushikula
Good Evenings Sir.
Bhabhav Chakraborty
Good evening Sir
Rita
GE All !
Good evening Sir!
GE
ENANRO BHOWMICK JOINED
K MOHAMMED NAYAZUDDIN JOINED
ABHISHEK PANDEY JOINED

l Methods for Data Science

ypothesis

Time (min) for three drying formulas for curing glue are studied

rmula A	13	10	8	11	8	5		
rmula B	13	11	14	14	12	6	10	
rmula C	17	14	13	10	11	12	16	15

evel of significance whether is any difference in the mean of glue?

Slide B'lore

Impartus Prof.Gangaboraiah B Andanaiah PhD Sat 26 Feb 6:24 PM

Virtual Classroom
Chat Participants (55)
Disable group chat
Write Message

l Methods for Data Science

ypothesis

Time (min) for three drying formulas for curing glue are studied

rmula A	13	10	8	11	8	5		
rmula B	13	11	14	14	12	6	10	
rmula C	17	14	13	10	11	12	16	15

evel of significance whether is any difference in the mean of glue?

Slide B'lore

Impartus Prof.Gangaboraiah B Andanaiah PhD Sat 26 Feb 6:24 PM

Statistical Methods for Data Science

Testing of Hypothesis

Time (min) for three drying formulas for curing glue are studied

Formula A	13	10	8	11	8	5		
Formula B	13	11	14	14	12	6	10	
Formula C	17	14	13	10	11	12	16	15

Test at 5% level of significance whether is any difference in the mean curing time of glue?

Virtual Classroom
Chat Participants (57)
B'lore

Statistical Methods for Data Science

Testing of Hypothesis

Time (min) for three drying formulas for curing glue are studied

Formula A	13	10	8	11	8	5		
Formula B	13	11	14	14	12	6	10	
Formula C	17	14	13	10	11	12	16	15

Statistical Methods for Data Science

Testing of Hypothesis

Time (min) for three drying formulas for curing glue are studied

Formula A	13	10	8	11	8	5		
Formula B	13	11	14	14	12	6	10	
Formula C	17	14	13	10	11	12	16	15

Mean (Formula A) = 9.17 min; SD (Formula A) = 2.79 min

Mean (Formula A) = 11.43 min; SD (For in

Mean (Formula A) = 13.50 min; SD (For in

Virtual Classroom
Chat
Participants (57)
Disable group chat

Impartus Prof.Gangaboraiah B Andanaiah PhD Sat 26 Feb 6:25 PM

Statistical Methods for Data Science

Testing of Hypothesis

Time (min) for three drying formulas for curing glue are studied

Formula A	13	10	8	11	8	5		
Formula B	13	11	14	14	12	6	10	
Formula C	17	14	13	10	11	12	16	15

Mean (Formula A) = 9.17 min; SD (Formula A) = 2.79 min

Mean (Formula A) = 11.43 min; SD (Formula A) = 2.82 min

Mean (Formula A) = 13.50 min; SD (Formula A) = 2.45 min

Virtual Classroom
Chat
Participants
Disable group chat
Roudrav Chakraborty
T Test
Prakash Kumar
t test
Mukul Mathur
none of them
S L S N V V S Sudhakar
t test
Pritthi Mohan
anova
PRABHAT SWAIN JOINED
SUBHODEEP MITRA JOINED
Arijit Barai
yes t test, as we need to compar
Saurav Raj
t test in combination
AMIT MADHAV JOINED
Write Message
Slide
B'lore

ience

las for curing glue are

5		
6	10	
12	16	15

= 2.79 min

) = 2.82 min

) = 2.45 min

Formula A	Formula B
9.17	11.43
7.77	7.95
6	7
7.868	
0.000	
11.000	
1.449	
0.088	
1.796	
0.175	
2.201	

Former Professor of Statistics | KIMS, B'lore

Statistical Methods for Data Science

Testing of Hypothesis

	Formula A	Formula B
Mean	9.17	11.43
Variance	7.77	7.95
Observations	6	7
Pooled Variance	7.868	
Hypothesized Mean Difference	0.000	
df	11.000	
t Stat	1.449	
P(T<=t) one-tail	0.088	
t Critical one-tail	1.796	
P(T<=t) two-tail	0.175	
t Critical two-tail	2.201	

Chat Participants

Disable group chat

T Test

t test

none of them

t test

anova

PRABHAT SWAIN JOINED

SUBHODEEP MITRA JOINED

yes t test, as we need to compair

t test in combination

AMIT MADHAV JOINED

Statistical Methods for Data Science

Testing of Hypothesis

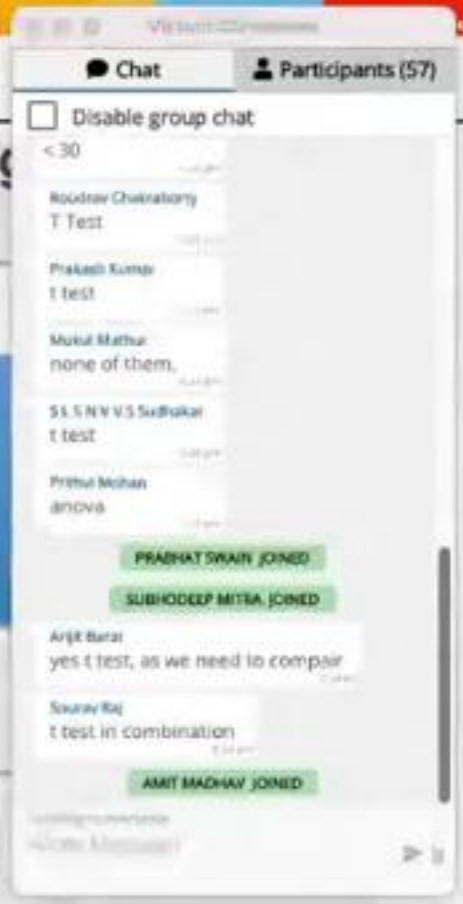
Time (min) for three drying formulas for curing studied

Formula A	13	10	8	11	8	5	
Formula B	13	11	14	14	12	6	10
Formula C	17	14	13	10	11	12	16

Mean (Formula A) = 9.17 min; SD (Formula A) = 2.79 min

Mean (Formula B) = 11.43 min; SD (Formula B) = 2.82 min

Mean (Formula C) = 13.50 min; SD (Formula C) = 2.45 min



e are

	Formula A	Formula B
	9.17	11.43
	7.77	7.95
s	6	7
nce	7.868	
d Mean Difference	0.000	
	11.000	
	1.449	
tail	0.088	
-tail	1.796	
tail	0.175	
-tail	2.201	

Statistical Methods for Data Science

Testing of Hypothesis

	Formula A	Formula B
Mean	9.17	11.43
Variance	7.77	7.95
Observations	6	7
Pooled Variance	7.868	
Hypothesized Mean Difference	0.000	
df	11.000	
t Stat	1.449	
P(T<=t) one-tail	0.088	
t Critical one-tail	1.796	
P(T<=t) two-tail	0.175	
t Critical two-tail	2.201	

Excel File Edit View Insert Format Tools Data Window Help Prof.Gangaboralah B Andanalah PhD Sat 26 Feb 6:31 PM

Book1

Home Insert Draw Page Layout Formulas Data Review View

Calibri (Body) 12 Wrap Text General Merge & Centre Conditional Formatting Format as Table Cell Styles Insert

Open recovered workbooks? Your recent changes were saved. Do you want to continue working where you left off?

Formula A

	A	B	C	D	E	F	G	H	I	J	K
	10	11	14								
	8	14	13		SUMMARY						
	11	14	10		*Groups*	*Average*	*Variance*		C - A	C - B	B - A
6	8	12	11		Formula A	9.17	7.77	2.79	4.33	2.07	2.
7	5	6	12		Formula B	11.43	7.95	2.82			
8		10	16		Formula C	13.50	6.00	2.45			
9			15								
10											

7R x 2C

Sheet2 Sheet3 Sheet4 Sheet5 Sheet1

Ready

Virtual Classroom

Chat Participants (61)

Disable group chat

SUBHODEEP MITRA JOINED

Arijit Barai
yes t test, as we need to compair

Sourav Raj
t test in combination

AMIT MADHAV JOINED

GURURAJA HEGDE V JOINED

ROUDRAV CHAKRABORTY JOINED

SHYAMALA GOWRI V G JOINED

ISHANT KUMAR JOINED

NEERAJ SABHNANI JOINED

PRESTI REKHA SAIKI JOINED

SAGI S JOINED

P V Saket
Sir, can you please show once how to calculate the p-value from t-score?

no sir

Sending to everyone
Write Message

Excel File Edit View Insert Format Tools Data Window Help

Prof.Gangaboralah B Andanalah PhD

Sat 26 Feb 6:31 PM

Book1

Home Data Review View

New
New from Template...
Open...
Open Recent
Close
Save
Save As...
Save as Template...
Browse Version History
Import
Reduce File Size...
Share
Always Open as Read-only
Restrict Permissions
Passwords...
Page Setup...
Print Area
Print...
Properties...

Anova: Single Factor

SUMMARY

Groups	Average	Variance		C - A	C - B	B - A
Formula A	9.17	7.77	2.79	4.33	2.07	2.26
Formula B	11.43	7.95	2.82			
Formula C	13.50	6.00	2.45			

12	11
6	12
10	16

Virtual Classroom

Chat | Participants (61)

Disable group chat

SUBHODEEP MITRA JOINED

Arijit Barat: yes t test, as we need to compair

Excel File Edit View Insert Format Tools Data Window Help

Prof.Gangaboralah B Andanalah PhD Sat 26 Feb 6:31 PM

Book1

Home Data Review View

New
New from Template...
Open...
Open Recent
Close
Save
Save As...
Save as Template...
Import
Reduce File Size...
Share
Always Open as Read-only
Restrict Permissions
Passwords...
Page Setup...
Print Area
Print...
Properties...

Do you want to continue working where you left off?

G2

1	Formul											
2	13			Anova: Single Factor								
3	10											
4				SUMMARY								
5				*Groups*	*Average*	*Variance*		C - A	C - B	B - A		
6	12	11		Formula A	9.17	7.77	2.79	4.33	2.07	2.26		
7	6	12		Formula B	11.43	7.95	2.82					
8	10	16		Formula C	13.50	6.00	2.45					

Sheet2 Sheet3 Sheet4 Sheet5 Sheet1

Ready

Virtual Classroom

Chat Participants (61)

Disable group chat

SUBHODEEP MITRA JOINED

Arijit Barat
yes t test, as we need to compair

Formula A	Formula B	Formula C
13	13	17
10	11	14
	14	13
	14	10
	12	11
	6	12
	10	16

Anova: Single Factor

SUMMARY

Groups	*Average*	*Variance*		C - A	C - B	B - A
Formula A	9.17	7.77	2.79	4.33	2.07	2.26
Formula B	11.43	7.95	2.82			
Formula C	13.50	6.00	2.45			

Virtual Classroom

Chat | Participants (61)

Disable group chat

yes t test, as we need to compair

Sourav Raj

t test in combination

Excel File Edit View Insert Format Tools Data Window Help

Prof Gangabiralah B Andanalah PhD Sat 26 Feb 6:32 PM

Sort...
Auto-filter
Advanced Filter...
Subtotals...
Validation...
Table...
Text to Columns...
Consolidate...
Group and Outline
Summarise with Pivot Table
Table Tools
Get External Data

Formula A	Formula B	Formula C
13	13	17
10	11	14
	14	13
	14	10
	12	11
	6	12
	10	16

Groups	*Average*	*Variance*		C - A	C - B	B - A
Formula A	9.17	7.77	2.79	4.33	2.07	2.26
Formula B	11.43	7.95	2.82			
Formula C	13.50	6.00	2.45			

Virtual Classroom

Chat | Participants (61)

Disable group chat

t test in combination

AMIT MADHAV JOINED

GURUBAJA HEGDE Y JOINED

Excel File Edit View Insert Format Tools Data Window Help Prof.Gangaboralah B Andanalah PhD Sat 26 Feb 6:33 PM

Book1

Home Insert Draw Page Layout Formulas Data Review View

Open recovered workbooks? Your recent changes were saved. Do you want to continue working where you left off? Yes No

D3

Formula A	Formula B	Formula C
13	13	17
10	11	14
	14	13
	14	10
	12	11
	6	12
	10	16

Data Analysis

Analysis Tools

- Regression
- Sampling
- t-Test: Paired Two Sample for Means
- t-Test: Two-Sample Assuming Equal Variances
- t-Test: Two-Sample Assuming Unequal Variances
- z-Test: Two Sample for Means

OK

Cancel

SUMMARY

Groups	*Average*	*Variance*		C - A	C - B	B - A
Formula A	9.17	7.77	2.79	4.33	2.07	2.26
Formula B	11.43	7.95	2.82			
Formula C	13.50	6.00	2.45			

Sheet2 Sheet3 Sheet4 Sheet5 Sheet1

Ready

Virtual Classroom

Chat Participants (61)

Disable group chat

t test in combination

AMIT MADHAV JOINED

GURURAJA HEGDE V JOINED

Epic Pen

Prof.Gangaboraiah B Andanaiah PhD Sat 26 Feb 6:33 PM

Book1

Home | Insert | Draw | Page Layout | Formulas | Data | Review | View

Open recovered workbooks? Your recent changes were saved. Do you want to continue working where you left off?

Data Analysis

Analysis Tools

- Regression
- Sampling
- t-Test: Paired Two Sample for Means
- t-Test: Two-Sample Assuming Equal Variances
- t-Test: Two-Sample Assuming Unequal Variances
- z-Test: Two Sample for Means

OK

Cancel

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Formula A	Formula B	Formula C										
2	13	13	17										
3	10	11	14										
4	8	14	13		SUMMARY								
5	11	14	10		*Groups*	*Average*	*Variance*		C - A	C - B	B - A		
6	8	12	11		Formula A	9.17	7.77	2.79	4.33	2.07	2.26		
7	5	6	12		Formula B	11.43	7.95	2.82					
8		10	16		Formula C	13.50	6.00	2.45					

Sheet2 | Sheet3 | Sheet4 | Sheet5 | Sheet1

Ready

Virtual Classroom

Chat | Participants (61)

Disable group chat

t test in combination

AMIT MADHAV JOINED

GURURAJA HEGDE V JOINED

Epic Pen Prof.Gangaboralah B Andanalah PhD Sat 26 Feb 6:34 PM

Book1

Home | Insert | Draw | Page Layout | Formulas | Data | Review | View

Open recovered workbooks? Your recent changes were saved. Do you

B1

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Formula A	Formula B	Formula C										
2	13	13	17										
3	10	11	14										
4	8	14	13										
5	11	14	10		*Groups*	*Average*	*Variance*		C - A	C - B	B - A		
6	8	12	11		Formula A	9.17	7.77	2.79	4.33	2.07	2.26		
7	5	6	12		Formula B	11.43	7.95	2.82					
8		10	16		Formula C	13.50	6.00	2.45					

Sheet2 | Sheet3 | Sheet4 | Sheet5 | Sheet1

Ready

t-Test: Two-Sample Assuming Equal Variances

Input

Variable 1 Range: $A$1:$A$7

Variable 2 Range: $B$1:$B$8

Hypothesized Mean Difference: 0

✓ Labels

Alpha: 0.05

Output options

Output Range:

• New Worksheet Ply:

New Workbook

OK

Cancel

Virtual Classroom

Chat | Participants (61)

Disable group chat

ROUDRAV CHAKRABORTY JOINED

SHYAMALA GOWRI V G JOINED

ISHANT KUMAR JOINED

Excel File Edit View Insert Format Tools Data Window Help Prof.Gangaboralah B Andanalah PhD Sat 26 Feb 6:35 PM

Book1

Home Insert Draw Page Layout Formulas Data Review View

Get External Data | Refresh All | Queries & Connections | Properties | Edit Links | Sort | Filter | Clear | Reapply | Advanced | Text to Columns | Flash Fill | Remove Duplicates | Data Validation | Consolidate | What-If Analysis | Group | Ungroup | Subtotal | Show Detail | Hide Detail | Analysis Tools | Data Analysis

Open recovered workbooks? Your recent changes were saved. Do you want to continue working where you left off? Yes

A1 t-Test: Two-Sample Assuming Equal Variances

t-Test: Two-Sample Assuming Equal Variances

	Variable 1	Variable 2
Mean	[illegible]	[illegible]
Variance	[illegible]	[illegible]
Observations	6	7
Pooled Variance	[illegible]	
Hypothesized Mean Difference	0	
df	11	
t Stat	[illegible]	
P(T<=t) one-tail	[illegible]	
t Critical one-tail	[illegible]	
P(T<=t) two-tail	[illegible]	
t Critical two-tail	[illegible]	

Sheet2 Sheet3 Sheet4 Sheet5 Sheet6 Sheet1

Virtual Classroom

Chat | Participants (60)

Disable group chat

ROUDRAV CHAKRABORTY JOINED

SHYAMALA GOWRI V G JOINED

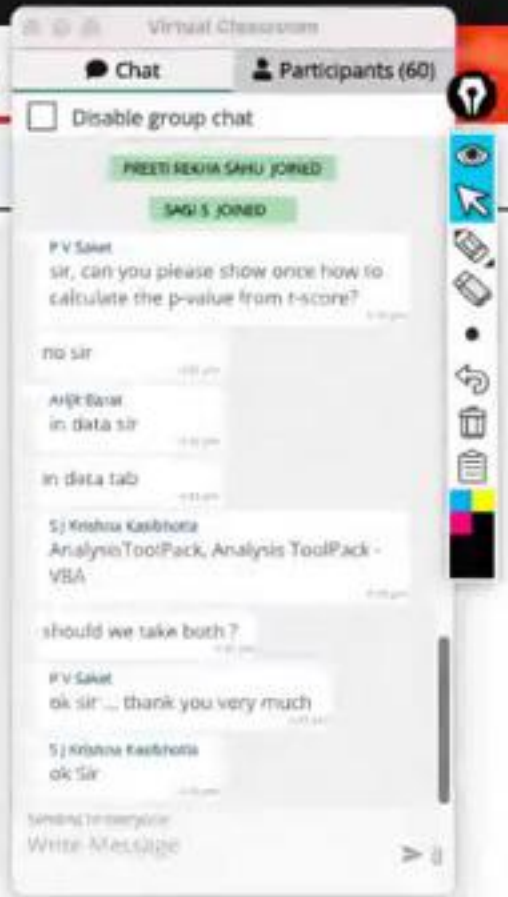
ISHANT KUMAR JOINED

Statistical Methods for Data Science

Testing of Hypothesis

t-Test: Two-Sample Assuming Equal Variances

	Formula A	Formula C
Mean	9.17	13.50
Variance	7.77	6
Observations	6	8
Pooled Variance	6.736	
Hypothesized Mean Difference	0	
df	12	
t Stat	3.092	
P(T<=t) one-tail	0.005	
t Critical one-tail	1.782	
P(T<=t) two-tail	0.009	
t Critical two-tail	2.179	



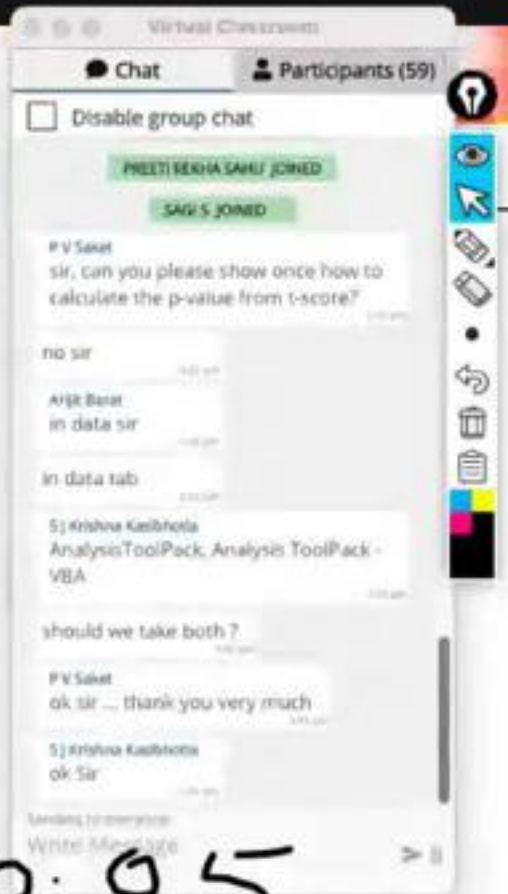
nce

esis

t-Test: Two-Sample Assuming Equal Variances

	Formula B	Formula C
Mean	11.43	13.50
Variance	7.95	6
Observations	7	8
Pooled Variance	6.901	
Hypothesized Mean Difference	0	
df	13	
t Stat	1.524	
P(T<=t) one-tail	0.076	
t Critical one-tail	1.771	
P(T<=t) two-tail	0.152	
t Critical two-tail	2.160	

$p \leq \alpha = 0.05$



Statistical Methods for Data Science

Testing of Hypothesis ➡ Analysis of Variance (ANOVA)

Types of ANOVA

- One-way
- Two-way
- Three-way (LSD)
- Repeated measures of ANOVA (Rm ANOVA)

Virtual Classroom
Chat Participants (58)
Disable group chat

Epic Pen Prof.Gangaboraiah B Andanaiah PhD Sat 26 Feb 6:50 PM

Statistical Methods for Data Science

Testing of Hypothesis ➡ Analysis of Variance (ANOVA)

$r =$

Types of ANOVA

- One-way
- Two-way
- Three-way (LSD)
- Repeated measures of ANOVA (Rm ANOVA)

Trg

Obs 1 Obs 2 — Im Obs 3

1
2
3

t-Test: Two-Sample Assuming Equal Variances

	Formula B	Formula C
Mean	11.43	13.50
Variance	7.95	6
Observations	7	8
Pooled Variance	6.901	
Hypothesized Mean Difference	0	
df	13	
t Stat	1.524	
P(T<=t) one-tail	0.076	
t Critical one-tail	1.771	
P(T<=t) two-tail	0.152	
t Critical two-tail	2.160	

Chat
Participants (59)
Disable group chat

Session 7

Testing of hypothes

Virtual Classroom
Chat
Participants (59)
Disable group chat
Arijit Barat

• ve hypothesis

Virtual Classroom
Chat
Participants (59)
Disable group chat
Arijit Barat

Testing of Hypothesis ➡ One sample Variance (χ^2 – test)

A manufacturer of car batteries claim that the life of his batteries is approximately normally distributed with a standard deviation equal to 0.9 year. If a random sample of 10 batteries has a standard deviation of 1.2 years, do you think that $\sigma > 0.9$ year? Use a 0.05 level of significance.

Virtual Classroom
Chat
Participants (59)
Disable group chat
Arijit Barat

Fisher's F – test for ratio of variances

Data Science

pothesis

Ratio of two Var

Virtual Classroom

Chat | Participants (59)

Disable group chat

Arijit Barat

Statistical Methods for Data Science

Testing of Hypothesis ➡ Ratio of two Variance (F – test)

1. State null and alternative hypothesis ➡ $H_0 : \sigma_1^2 = \sigma_2^2$ vs $H_1 : \sigma_1^2 < \sigma_2^2$ or $H_1 : \sigma_1^2 > \sigma_2^2$ or $H_1 : \sigma_1^2 \neq \sigma_2^2$
2. Specify the level of significance 'α'
3. Fisher's F - Distribution
4. the test statistic ➡

Virtual Classroom
Chat Participants (59)
Disable group chat
Arijit Barai

Epic Pen Prof.Gangaboraiah B Andanaiah PhD Sat 26 Feb 6:58 PM

Statistical Methods for Data Science

Testing of Hypothesis ➡ Ratio of two Variance (F – test)

1. State null and alternative hypothesis ➡ $H_0 : \sigma_1^2 = \sigma_2^2$ vs $H_1 : \sigma_1^2 < \sigma_2^2$ or $H_1 : \sigma_1^2 > \sigma_2^2$ or $H_1 : \sigma_1^2 \neq \sigma_2^2$
2. Specify the level of significance 'α'
3. Fisher's F - Distribution
4. Compute the test statistic ➡ $$F = \frac{S_1^2}{S_2^2} \cong F_{(\alpha, n_1\text{-}1, n_2\text{-}1)}$$

Virtual Classroom
Chat Participants (59)
Disable group chat
Arijit Barat

t-Test: Two-Sample Assuming Equal Variances

	Formula A	Formula C
Mean	9.17	13.50
Variance	7.77	6
Observations	6	8
Pooled Variance	6.736	
Hypothesized Mean Difference	0	
df	12	
t Stat	3.092	
P(T<=t) one-tail	0.005	
t Critical one-tail	1.782	
P(T<=t) two-tail	0.009	
t Critical two-tail	2.179	

Chat
Participants (59)
Disable group chat
in data lab

Statistical Methods for Data Science

Testing of Hypothesis ➡ One-way Analysis of Variance

ANOVA ➡ Testing equality of k group means against not equal

- Samples are drawn from normal population
- The population variances should be equal
- The sample size should be less than 30 (i.e., n < 30)

ANOVA

- Groups should be independent
- Subjects should be allocated randomly to both groups
- However even if sample size more than 30 (i.e., n > 30) ANOVA should be continue to apply, because of central limit theorem it approaches normal.

Virtual Classroom
Chat Participants (58)
Disable group chat

Epic Pen
Prof.Gangaboraiah B Andanaiah PhD
Sat 26 Feb 7:00 PM

Statistical Methods for Data Science

Testing of Hypothesis ➡ One-way Analysis of Variance

Epic Pen Prof.Gangaboraiah B Andanaiah PhD Sat 26 Feb 7:02 PM

Statistical Methods for Data Science

Testing of Hypothesis ➡ One-way Analysis of Variance

G_1	G_2	G_3	. . .	G_k
X_{11}	X_{21}	X_{31}	. . .	X_{k1}
X_{12}	X_{22}	X_{32}	. . .	X_{k2}
X_{13}	X_{23}	X_{33}	. . .	X_{k3}
. . .	. . .	. . .	. . .	. . .
X_{1n_1}	X_{2n_2}	X_{3n_3}	. . .	X_{kn_k}
C_1	C_2	C_3	. . .	C_k

$$n_1 + n_2 + n_3 + \ldots + n_k = n$$

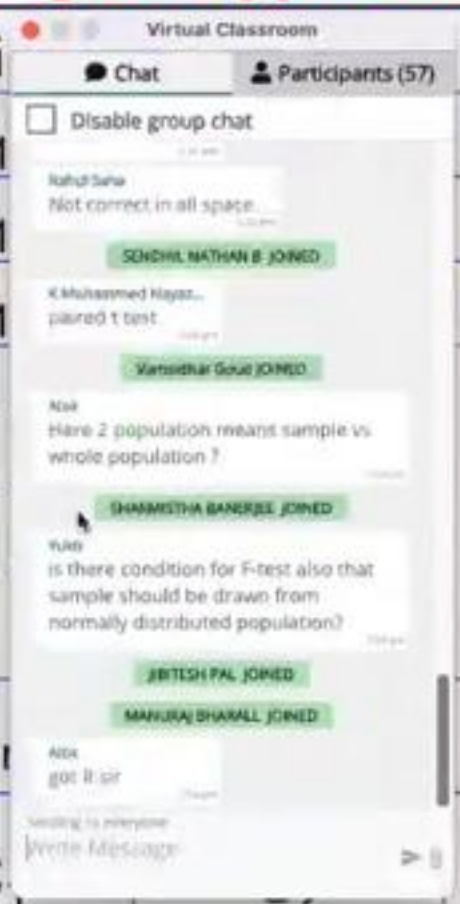
Virtual Classroom

Impartus Prof.Gangaboraiah B Andanaiah PhD Sat 26 Feb 7:04 PM

Statistical Methods for Data Science

Testing of Hypothesis ➡ One-way Analysis of Variance

G		G_3	. . .	G_k
X_1		X_{31}	. . .	X_{k1}
X_1		X_{32}	. . .	X_{k2}
X_1		X_{33}	. . .	X_{k3}
.		.	.	.
.		.	.	.
.		.	.	.
X_{1n}		X_{1n_3}	. . .	X_{1n_k}
C		C_3	. . .	C_k



$$n_1 + n_2 + n_3 + \ldots + n_k = n$$

$$C_1 + C_2 + C_3 + \ldots + C_k = G$$

The hypotheses to be tested are

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$$

vs

$$H_1 : \mu_1 \neq \mu_2 \neq \ldots \neq \mu_k$$

Statistical Methods for Data Science

Testing of Hypothesis ➡ One-way Analysis of Variance

G_1	G_2	G_3	. . .	G_k
X_{11}	X_{21}	X_{31}	. . .	X_{k1}
X_{12}	X_{22}	X_{32}	. . .	X_{k2}
X_{13}	X_{23}	X_{33}	. . .	X_{k3}
. . .	. . .	. . .	. . .	. . .
X_{1n_1}	X_{1n_2}	X_{1n_3}	. . .	X_{1n_k}
C_1	C_2	C_3	. . .	C_k

$$n_1 + n_2 + n_3 + \ldots + n_k = n$$

$$C_1 + C_2 + C_3 + \ldots + C_k = G$$

The hypotheses to be tested are

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$$

vs

$$H_1 : \mu_1 \neq \mu_2 \neq \ldots \neq \mu_k$$

Virtual Classroom
Chat Participants (57)
Disable group chat

Epic Pen Prof.Gangaboraiah B Andanaiah PhD Sat 26 Feb 7:05 PM

Statistical Methods for Data Science

innovate achieve lead

Testing of Hypothesis ➡ One-way Analysis of Variance

G_1	G_2	G_3	. . .	G_k
X_{11}	X_{21}	X_{31}	. . .	X_{k1}
X_{12}	X_{22}	X_{32}	. . .	X_{k2}
X_{13}	X_{23}	X_{33}	. . .	X_{k3}
. . .	. . .	. . .	. . .	. . .
X_{1n_1}	X_{1n_2}	X_{1n_3}	. . .	X_{1n_k}
C_1	C_2	C_3	. . .	C_k

$$n_1 + n_2 + n_3 + \ldots + n_k = n$$

$$C_1 + C_2 + C_3 + \ldots + C_k = G$$

The hypotheses to be tested are

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$$

vs

$$H_1 : \mu_1 \neq \mu_2 \neq \ldots \neq \mu_k$$

Virtual Classroom
Chat Participants (57)
Disable group chat

Statistical Methods for Data Science

Testing of Hypothesis ➡ One-way Analysis of Variance

1 State null and alternative hypothesis ➡

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$$

vs

$$H_1 : \mu_1 \neq \mu_2 \neq \ldots \neq \mu_k$$

Virtual Classroom
Chat
Participants (57)
Disable group chat

Statistical Methods for Data Science

Testing of Hypothesis ➡ One-way Analysis of Variance

1. State null and alternative hypothesis
2. Specify the level of significance ‘α’
3. Student’s t-distribution

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$$

vs

$$H_1 : \mu_1 \neq \mu_2 \neq \ldots \neq \mu_k$$

Statistical Methods for Data Science

Testing of Hypothesis ➡ One-way Analysis of Variance

1. State null and alternative hypothesis

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k$$

vs

$$H_1 : \mu_1 \neq \mu_2 \neq \ldots \neq \mu_k$$

2. Specify the level of significance 'α'

3. Fisher's F- ~~Student's t~~-distribution

4. Compute the test statistic

$$F = \frac{MGSS}{MWSS} \approx F_{(k-1, n-k)}$$

5. Define the critical region/ rejection criteria

Virtual Classroom
Chat Participants (57)
Disable group chat

Statistical Methods for Data Science

Testing of Hypothesis ➡ One-way Analysis of Variance

G_1	G_2	G_3	. . .	G_k
X_{11}	X_{21}	X_{31}	. . .	X_{k1}
X_{12}	X_{22}	X_{32}	. . .	X_{k2}
X_{13}	X_{23}	X_{33}	. . .	X_{k3}
. . .	. . .	. . .	. . .	. . .
X_{1n_1}	X_{1n_2}	X_{1n_3}	. . .	X_{1n_k}
C_1	C_2	C_3	. . .	C_k

$$n_1 + n_2 + n_3 + \ldots + n_k = n$$

$$C_1 + C_2 + C_3 + \ldots + C_k = G$$

Virtual Classroom
Chat Participants (56)
Disable group chat

Testing of Hypothesis ➡ One-way Analysis of Variance

Calculation of sum of squares

1. Correction factor (CF) : $\frac{G^2}{n}$

2. Total sum of squares (TSS) : $\sum_i \sum_j x_{ij}^2 - CF$

3. Between

Calculation of sum of squares

1. Correction factor (CF) : $\frac{G^2}{n}$

2. Total sum of squares (TSS) : $\sum_i \sum_j x_{ij}^2 - CF$

3. Between group sum of squares (GSS) : $\sum_{i=1}^{k} \frac{C_i^2}{n_i} - CF$

Testing of Hypothesis ➡ One-way Analysis of Variance

Calculation of sum of squares

1. Correction factor (CF) : $\frac{G^2}{n}$

$$\left(\frac{C_1^2}{n_1}+\frac{C_2^2}{n_2}+\cdots+\frac{C_k^2}{n_k}\right)-CF$$

$$TV = BGV + \underline{WGV}$$

2. Total sum of squares (TSS) : $\sum_i \sum_j x_{ij}^2 - CF$

3. Between group sum of squares (GSS) : $\sum_{i=1}^{k} \frac{C_i^2}{n_i} - CF$

4. Within group sum of squares (WSS) : TSS - GSS

Virtual Classroom
Chat Participants (56)
Disable group chat

Statistical Methods for Data Science

Testing of Hypothesis ➡ One-way Analysis of Variance

Calculation of sum of squares

1. Correction factor (CF) : $\frac{G^2}{n}$

$$\left(\frac{C_1^2}{n_1}+\frac{C_2^2}{n_2}+\cdots+\frac{C_k^2}{n_k}\right)-CF$$

$$TV = BGV + \underline{WGV}$$

2. Total sum of squares (TSS) : $\sum_i \sum_j x_{ij}^2 - CF$

3. Between group sum of squares (GSS) : $\sum_{i=1}^{k} \frac{C_i^2}{n_i} - CF$

4. Within group sum of squares (WSS) : TSS - GSS

Epic Pen Prof.Gangaboraiah B Andanalah PhD Sat 26 Feb 7:14 PM

Statistical Methods for Data Science

Testing of Hypothesis ➡ One-way Analysis of Variance

Calculation of sum of squares

1. Correction factor (CF) : $\frac{G^2}{n}$

2. Total sum of squares (TSS) : $\sum_i \sum_j x_{ij}^2 - CF$

3. Between group sum of squares (GSS) : $\sum_{i=1}^{k} \frac{C_i^2}{n_i} - CF$

4. Within group sum of squares (WSS) : TSS - GSS

Virtual Classroom

Epic Pen

Prof.Gangaboralah B Andanalah PhD

Sat 26 Feb 7:14 PM

Statistical Methods for Data Science

Testing of Hypothesis ➡ One-way Analysis of Variance

Calculation of sum of squares

1. Correction factor (CF) : $\frac{G^2}{n}$

2. Total sum of squares (TSS) : $\sum_i \sum_j x_{ij}^2 - CF$

3. Between group sum of squares (GSS) : $\sum_{i=1}^{k} \frac{C_i^2}{n_i} - CF$

Virtual Classroom
Chat
Participants (55)
Disable group chat

Statistical Methods for Data Science

Testing of Hypothesis ➡ One-way Analysis of Variance

Calculation of sum of squares

1. Correction factor (CF) : $\frac{G^2}{n}$

2. Total sum of squares (TSS) : $\sum_i \sum_j x_{ij}^2 - CF$

3. Between group sum of squares (GSS) : $\sum_{i=1}^{k} \frac{C_i^2}{n_i} - CF$

4. Within group sum of squares (WSS) : TSS - GSS

Epic Pen Prof.Gangaboraiah B Andanaiah PhD Sat 26 Feb 7:18 PM

Statistical Methods for Data Science

Testing of Hypothesis ➡ One-way Analysis of Variance

One-way ANOVA table

Source of variation	df	Sum of squares	Mean sum of squares	F-ratio
Between groups	k - 1	GSS	$MGSS = \frac{GSS}{k-1}$	$F = \frac{MGSS}{MWSS}$
Within groups	n - k	WSS	$MWSS = \frac{WSS}{n-k}$	
Total	n - 1	TSS	$F \approx F$ - distribution with $k-1$ and $n-k$ df	

Epic Pen
Prof.Gangaboraiah B Andanaiah PhD
Sat 26 Feb 7:19 PM

Statistical Methods for Data Science

Testing of Hypothesis

Σ

Virtual Classroom
FEB 26

Virtual Classroom

Chat | Participants (56)

Disable group chat

paired t test

Statistical Methods for Data Science

Testing of Hypothesis ➡ One-way Analysis of Variance

One-way ANOVA table

Source of variation	df	Sum of squares	Mean sum of squares	F-ratio
Between groups	k - 1	GSS	$MGSS = \frac{GSS}{k-1}$	$F = \frac{MGSS}{MWSS}$
Within groups	n - k	WSS	$MWSS = \frac{WSS}{n-k}$	
Total	n - 1	TSS	$F \approx F$ - distribution with $k-1$ and $n-k$ df	

Statistical Methods for Data Science

Testing of Hypothesis

Three drying formulas for curing glue are studied

Formula A	13	10	8	11	8	5		
Formula B	13	11	14	14	12	6	10	
Formula C	17	14	13	10	11	12	16	15

Test at 5% level of significance whether is any difference in the mean curing time of glue?

Find between which two formulas the mean difference has contributed significantly using least significant difference post-hoc test?

Virtual Classroom
Chat
Participants (57)
Disable group chat
paired t test

Statistical Methods for Data Science

Testing of Hypothesis ➡ One-way Analysis of Variance

One-way ANOVA table

Source of variation	df	Sum of squares	Mean sum of squares	F-ratio
Between groups	2	64.595	32.298	F=4.523
Within groups	12	128.548	7.142	
Total	14	193.143	$F(4.523) > F_{(0.05;\ 2,\ 12)} = 3.555$	

H_0 may be rejected and **H_1 may be not rejected (no need for post-hoc test)**

P – value = 0.026

-Test: Two-Sample Assuming Equal Variances

	Formula A	Formula B
Mean	9.17	11.43
Variance	7.77	7.95
Observations	6	7
Pooled Variance	7.868	
Hypothesized Mean Difference	0	
df	11	
t Stat	-1.449	
P(T<=t) one-tail	0.088	
t Critical one-tail	1.796	
P(T<=t) two-tail	0.175	
t Critical two-tail	2.201	

Formula A	Formula B	Formula C
13	13	17
10	11	14
8	14	13
11	14	10
8	12	11
5	6	12
	10	16

Anova: Single Factor

SUMMARY

Groups	*Average*	*Variance*	
Formula A	9.17	7.77	2.79
Formula B	11.43	7.95	2.82
Formula C	13.50	6.00	2.45

C - A	C - B	B - A
4.33	2.07	2.26

Epic Pen Prof.Gangaboralah B Andanalah PhD Sat 26 Feb 7:23 PM

Statistical Methods for Data Science

Testing of Hypothesis ➡ One-way Analysis of Variance

One-way ANOVA table

Source of variation	df	Sum of squares	Mean sum of squares	F-ratio
Between groups	2	64.595	32.298	F=4.523
Within groups	12	128.548	7.142	
Total	14	193.143	$F(4.523) > F_{(0.05;\,2,\,12)} = 3.555$	

H_0 may be rejected and **H_1 may be not rejected (no need for post-hoc test)**

P – value = 0.026

Virtual Classroom

Testing of Hypothesis ➡ One-way Analysis of Variance

One-way ANOVA table

Source of variation	df	Sum of squares	Mean sum of squares	F-ratio
Between groups	2	64.595	32.298	F=4.523
Within groups	12	128.548	7.142	
Total	14	193.143	$F(4.523) > F_{(0.05;\ 2,\ 12)} = 3.555$	

H_0 may be rejected and **H_1 may be not rejected (no need for post-hoc test)**

P – value = 0.026

Virtual Classroom
Chat
Participants (54)
Disable group chat

Testing of Hypothesis ➡ One-way Analysis of Variance

- Reject H_0, if $F > F_{(\alpha, k-1, n-k)}$
- If H_0 is rejected, then further mean differences between any two groups should be tested using Post-hoc tests. Following are different Post-hoc tests.

Chat Participants (54)
Disable group chat

Impartus Prof.Gangaboraiah B Andanaiah PhD Sat 26 Feb 7:27 PM

Statistical Methods for Data Science

Test of Hypothesis ➡ One-way Analysis of Variance

Virtual Classroom
Chat | Participants (52)
Disable group chat
whole population ?
SHARMISTHA BANERJEE JOINED
Yuki
is there condition for F-test also that sample should be drawn from normally distributed population?
JIBITESH PAL JOINED
MANURAJ BHARALI JOINED
Alex
got it sir
Yuki
ok sir
K.SREEKANTH JOINED
AMOGH SINGHAL JOINED
Rahul Saha
Sir, please share the excel also
Write Message

One-way ANOVA table

	df	Sum of squares	Mean sum of squares	F-ratio
Bet	3	349.52	116.51	F=4.455
Wit	36	922.95	25.64	
	39	1272.47	$F(4.46) > F_{(0.05;\ 3,\ 36)} = 4.38$	

H_0 may be rejected and **H_1 may be accepted, and continue with post-hoc test.**

PowerPoint File Edit View Insert Format Arrange Tools Slide Show Window Help Prof.Gangaboraiah B Andanaiah PhD Sat 26 Feb 7:28 PM

Statistical Methods for Data Science

Testing of Hypothesis ⇒ One-way Analysis of Variance

S^2 = ... an sum of squares

k = ... ations in each Group = 10 df of

with ...

... responding to $\alpha = 0.05$ for 36 df is

The LSD is $\left\{ t_{0.05} \sqrt{S^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)} \right\}$

Want to save your changes to Session 10 - ANOVA and Chi-square test.pptx?

Save

Don't Save

Cancel

Statistical Methods for Data Science

Testing of Hypothesis ⇒ One-way Analysis of Variance

Statistical Methods for Data Science

Testing of Hypothesis ⇒ One-way Analysis of Variance

One-way ANOVA table

Source of variation	df	Sum of squares	Mean sum of squares	F-ratio
Between groups	3	349.52	116.51	7.54
Within groups	36	922.95	25.64	
Total	39	1272.47		

Statistical Methods for Data Science

Testing of Hypothesis ⇒ One-way Analysis of Variance

Statistical Methods for Data Science

Virtual Classroom

Statistical Methods for Data Science

Time Series ➡ Qualitative forecasting

- **Panel consensus**
  - Reduces the prejudice and ignorance that may arise in the individual judgment
  - Possible to develop consensus among group of individuals
  - Panel of individuals are encouraged to share information, opinions, and assumptions, if any, to predict future value of some variable under study

Forecasting and Time series

Virtual Classroom
Chat
Participants (52)
Disable group chat
whole population ?

Forecasting and Time series

Chat | Participants (53)

Disable group chat

is there condition for F-test also that sample should be drawn from

aboraiah PhD (Stats)

Former Professor of Statistics | KIMS, B'lore

Time Series ➡ Definition of Forecasting

Forecasting is the process of making **predictions of the future** based on **past and present data** and most commonly by analysis of trends. A commonplace example might be estimation of some variable of interest at some specified future date.

Chat
Participants (53)
Disable group chat

Epic Pen Prof.Gangaboralah B Andanalah PhD Sat 26 Feb 7:36 PM

Statistical Methods for Data Science

Time Series ➡ Definition of Forecasting

Forecasting is the process of making **predictions of the future** based on **past and present data** and most commonly by analysis of trends. A commonplace example might be estimation of some variable of interest at some specified future date.

Virtual Classroom

Time Series ➡ Definition of Forecasting

- Forecasting is required in many situations:

  (a) deciding whether to build another power

Virtual Classroom
Chat Participants (53)
Disable group chat
JIBITESH PAL JOINED

Time Series ➡ Definition of Forecasting

Forecasts can be required several years in advance (for the case of capital investments), or only a few minutes beforehand (for telecommunication routing).

Whatever the circumstances or time horizons

Virtual Classroom
Chat Participants (53)
Disable group chat
JIBITESH PAL JOINED

Time Series ➡ Definition of Forecasting

- For example, forecasts of electricity demand can be highly accurate because all three conditions are usually satisfied. We have a good idea of the contributing factors: electricity demand is driven largely by temperatures, with smaller effects for calendar variation such as holidays, and economic conditions.

Virtual Classroom
Chat
Participants (53)
Disable group chat
JBITESH PAL JOINED

Time Series ➡ Definition of Forecasting

- how well we understand the factors that contribute to it;
- how much data is available;
- whether the forecasts can affect the thing we are trying to forecast.

Virtual Classroom
Chat
Participants (53)
Disable group chat
JIBITESH PAL JOINED

Time Series ➡ Definition of Forecasting

- For example, forecasts of electricity demand can be highly accurate because all three conditions are usually satisfied. We have a good idea of the contributing factors: electricity demand is driven largely by temperatures, with smaller effects for calendar variation such as holidays, and economic conditions.

Virtual Classroom
Chat
Participants (53)
Disable group chat
JIBITESH PAL JOINED

Time Series ➡ Definition of Forecasting

- On the other hand, when forecasting currency exchange rates, only one of the conditions is satisfied: there is plenty of available data. However, we have a limited understanding of the factors that affect exchange rates, and forecasts of the exchange rate have a direct effect on the rates themselves.

Virtual Classroom
Chat
Participants (53)
Disable group chat
JIBITESH PAL JOINED

Statistical Methods for Data Science

Time Series ➡ Definition of Forecasting

- If there are well-publicised forecasts that the exchange rate will increase, then people will immediately adjust the price they are willing to pay and so the forecasts are self-fulfilling. In a sense, the exchange rates become their own forecasts. This is an example of the "efficient market hypothesis".

Virtual Classroom
Chat
Participants (53)
Disable group chat
JIBITESH PAL JOINED

Time Series ➡ Definition of Forecasting

- Consequently, forecasting whether the exchange rate will rise or fall tomorrow is about as predictable as forecasting whether a tossed coin will come down as a head or a tail. In both situations, you will be correct about 50% of the time, whatever you forecast. In situations like this, forecasters need to be aware of their own limitations, and not claim more than is possible.

Virtual Classroom
Chat Participants (52)
Disable group chat
JIBITESH PAL JOINED

Forecasts are rarely perfect.

Forecasting the future involves uncertainty. Therefore, it is almost impossible to make a perfect prediction. Forecasters know that they have to live with a certain amount of error, which is the difference between what is forecast and what actually happens. The goal of forecasting is to generate good forecasts *on the average* over time and to keep forecast errors as low as

Virtual Classroom
Chat
Participants (52)
Disable group chat
Alex
got it sir

Statistical Methods for Data Science

Time Series ➡ Forecasting frame

Forecasting should be an integral part of the decision-making activities of management, as it can play an important role in many areas of a company. Modern organisations require short-term, medium-term and long-term forecasts, depending on the specific application.

Short-term Forecasts

Short-term forecasts generally involve some form of scheduling which may include for example the seasons of the year for planning purposes.

The cyclical and seasonal factors are more important in these situations.

Short-term Forecasts

Such forecasts are usually prepared every 6 months or on a more frequent basis.

These forecasts are needed for the scheduling of personnel, production and transportation. As part of the scheduling process, forecasts of demand are often also required.

Short-term Forecasts

Some airport operators undertake 'ultra short term' forecasts for (e.g.) the next month in order to provide for specific requirement such as adequate staffing in the peaks.

Virtual Classroom
Chat
Participants (52)
Disable group chat
Alok
got it sir

Medium-term Forecasts

Medium-term forecasts are generally prepared for planning, scheduling, budgeting and resource requirements purposes.

Medium-term forecasts are needed to determine future resource requirements, in order to purchase raw materials, hire personnel, or buy machinery and equipment.

Virtual Classroom
Chat
Participants (52)
Disable group chat
Alex
got it sir

Medium-term Forecasts

Medium-term forecasts are generally prepared for planning, scheduling, budgeting and resource requirements purposes.

Medium-term forecasts are needed to determine future resource requirements, in order to purchase raw materials, hire personnel, or buy machinery and equipment.

Virtual Classroom
Chat
Participants (52)
Disable group chat
Alok
got it sir

Time Series ➡ Forecasting frame

Long-term Forecasts

Long-term forecasts are used mostly in connection with strategic planning to determine the level and direction of capital expenditures and to decide on ways in which goals can be accomplished.

The trend element generally dominates long term situations and must be considered in the determination of any long-run decisions.

Virtual Classroom
Chat
Participants (52)
Disable group chat
Alok
got it sir

Step 1: Problem definition

Often this is the most difficult part of forecasting. Defining the problem carefully requires an understanding of the way the forecasts will be used, who requires the forecasts, and how the forecasting function fits within the organisation requiring the forecasts. A forecaster needs to spend time talking to everyone who will be involved in collecting data, maintaining databases, and using the forecasts for ning

Chat Participants (51)
Disable group chat
Alex
got it sir

Statistical Methods for Data Science

Time Series ➡ The basic steps in a forecasting task

Step 1: Problem definition

Often this is the most difficult part of forecasting. Defining the problem carefully requires an understanding of the way the forecasts will be used, who requires the forecasts, and how the forecasting function fits within the organisation requiring the forecasts. A forecaster needs to spend time talking to everyone who will be involved in collecting data, maintaining databases, and using the forecasts for

Time Series ➡ The basic steps in a forecasting task

Step 3: Preliminary (exploratory) analysis

Always start by graphing the data. Are there consistent patterns? Is there a significant trend? Is seasonality important? Is there evidence of the presence of business cycles? Are there any outliers in the data that need to be explained by those with expert knowledge? How strong are the relationships among the variables available for analysis? Various tools have been developed to help with this analysis.

Virtual Classroom
Participants (51)
group chat

Step 3: Preliminary (exploratory) analysis

Always start by graphing the data. Are there consistent patterns? Is there a significant trend? Is seasonality important? Is there evidence of the presence of business cycles? Are there any outliers in the data that need to be explained by those with expert knowledge? How strong are the relationships among the variables available for analysis? Various tools have been developed to help with this analysis.

Virtual Classroom
Participants (52)
group chat

Step 4: Choosing and fitting models

The best model to use depends on the availability of historical data, the strength of relationships between the forecast variable and any explanatory variables, and the way in which the forecasts are to be used. It is common to compare two or three potential models. Each model is itself an artificial construct that is based on a set of assumptions (explicit and implicit) and usually involves one or more parameters which must be estimated using ... own historical data.

Virtual Classroom
Participants (51)
group chat

Step 4: Choosing and fitting models

The best model to use depends on the availability of historical data, the strength of relationships between the forecast variable and any explanatory variables, and the way in [illegible] forecasts are to be used. It is common to cor[illegible] three potential models. Each model is itse[illegible] construct that is based on a set of assun[illegible]licit and implicit) and usually involves one o[illegible]meters which must be estimated using the kr[illegible]cal data.

Impartus Prof.Gangaboraiah B Andanaiah PhD Sat 26 Feb 7:57 PM

Statistical Methods for Data Science

Time Series ➡ The basic steps in a forecasting task

Step 4: Choosing and fitting models

The best model to use depends on the availability of historical data, the strength of relationships between the forecast variable and any explanatory variables, and the way in which the forecasts are to be used. It is common to compare two or three potential models. Each model is itself an artificial construct that is based on a set of assumptions (explicit and implicit) and usually involves one or more parameters which must be estimated using historical data.

Virtual Classroom
Chat
Disable gro
26

Step 5: Using and evaluating a forecasting model

There are also organisational issues in using and acting on the forecasts. When using a forecasting model in practice, numerous practical issues arise such as how to handle missing values and outliers, or how to deal with short time series.

Chat
Participants (49)
Disable group chat

Statistical Methods for Data Science

Time Series ➡ Types of forecasting methods

Forecasting Methods

Qualitative

- Personal Opinion or Judgment
- Panel consensus
- Delphi Method
- Market Research

Quantitative

Time Series

- Smoothing methods
- Exponential smoothing
- Trend projection method

Causal

- Regression Trend Analysis

Virtual Classroom
Chat
Participants (49)
Disable group chat

Virtual Classroom

Chat | Participants (49)

Disable group chat

got it sir

Yukti
ok sir

K SREEKANTH JOINED

AMOGH SINGHAL JOINED

Rahul Saha
Sir, please share the excel also.

Alok
quick break sir

ARIJIT BARAT JOINED

S KRISHNAMOORTHY JOINED

MUKUL MATHUR JOINED

K Mohammed Nayaz...
pattern

Amogh Singhal
Sir, is there a difference conceptually between prediction and forecasting ?

YUKTI JOINED

Sending to everyone

Write Message

Impartus Prof Gangaboraiah B Arabanalah PhD Sat 26 Feb 8:06 PM

Virtual Classroom

Chat | Participants (49)

Disable group chat

got it sir

Yukti
ok sir

K SREEKANTH JOINED

AMOGH SINGHAL JOINED

Rahul Saha
Sir, please share the excel also.

Alok
quick break sir ...

ARJIT BARAT JOINED

S KRISHNAMOORTHY JOINED

MUKUL MATHUR JOINED

K Mohammed Nayaz...
pattern

Amogh Singhal
Sir, is there a difference conceptually between prediction and forecasting ?

YUKTI JOINED

Sending to everyone

Write Message

Impartus Prof.Gangaboralah B Andanalah PhD Sat 26 Feb 8:06 PM

Virtual Classroom

Chat | Participants (48)

Disable group chat

OK Sir

K SREEKANTH JOINED

AMOGH SINGHAL JOINED

Rahul Saha
Sir, please share the excel also.

Alok
quick break sir

ARJIT BARAT JOINED

S KRISHNAMOORTHY JOINED

MUKUL MATHUR JOINED

K Mohammed Nayar...
pattern

Amogh Singhal
Sir, is there a difference conceptually between prediction and forecasting ?

YUKTI JOINED

Abhishek Pandey
diff in forecast & Prediction

Sending to everyone

Write Message

Statistical Methods for Data Science

Time Series ➡ Qualitative forecasting

- **Personal Opinion**
  - Individuals forecasts future based on their own judgment or opinion without any formal model
  - Such assessments are relatively reliable and accurate

Statistical Methods for Data Science

Time Series ➡ Qualitative forecasting

- **Personal Opinion**
  - Individuals forecasts future based on their own judgment or opinion without any formal model
  - Such assessments are relatively reliable and accurate

Statistical Methods for Data Science
Time Series ➡ Qualitative forecasting
• Personal Opinion
– Individuals forecasts future based on their own judgment or opinion without any formal model
– Such assessments are relatively reliable and accurate