

SHETH L.U.J. AND SIR M.V. COLLEGE

SUBJECT: Data Analysis with R

Aim: 7 Performing one-way ANOVA using aov() (R).

8 Performing two-way ANOVA using aov() (R).

9 Conducting Chi-square tests using chisq.test() (R)

7TH OUTPUT:

The screenshot displays the RStudio interface with the following components:

- Source Editor:** Contains R code for loading data, renaming columns, and performing an ANOVA test.
- Console:** Shows the execution of the R code, including the ANOVA results.
- Environment:** Lists the objects in the global environment, including 'anova_result' and 'students'.
- Files:** Shows the file explorer with 'R S103.Rproj' and 'StudentsPerformance.csv'.

R Code:

```
1 library(dplyr)
2 students <- read.csv("StudentsPerformance.csv")
3 # Rename column for easy usage
4 colnames(students)[7] <- "reading_score"
5
6 # Convert test preparation course to factor
7 students$test_preparation_course <- as.factor(students$test_preparation_course)
```

Console Output:

```
R - R 4.5.2 - D:\New folder\R S103/
> library(dplyr)
> students <- read.csv("StudentsPerformance.csv")
> # Rename column for easy usage
> colnames(students)[7] <- "reading_score"
> # Convert test preparation course to factor
> students$test_preparation_course <- as.factor(students$test_preparation_course)
> anova_result <- aov(reading_score ~ test_preparation_course, data = students)
> # ANOVA table
> summary(anova_result)

              Df Sum Sq Mean Sq F value    Pr(>F)
test_preparation_course 1 12449   12449    61.96 9.08e-15 ***
Residuals              998 200504      201
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> view(anova_result)
> view(students)
>
```

Environment:

- Global Environment
- anova_result: List of 13
- students: 1000 obs. of 8 vari...

Files:

- R S103.Rproj: 218 B
- StudentsPerformance.csv: 70.3 KB

Table 1: StudentsPerformance.csv (Sample Data)

gender	race_ethnicity	parental_level_of_education	lunch	test_preparation_course	math_score	reading_score	writing_score	
1	female	group B	bachelor's degree	standard	none	72	72	74
2	female	group C	some college	standard	completed	69	90	88
3	female	group B	master's degree	standard	none	90	95	93
4	male	group A	associate's degree	free/reduced	none	47	57	44
5	male	group C	some college	standard	none	76	78	75
6	female	group B	associate's degree	standard	none	71	83	78
7	female	group B	some college	standard	completed	88	95	92
8	male	group B	some college	free/reduced	none	40	43	39
9	male	group D	high school	free/reduced	completed	64	64	67
10	female	group B	high school	free/reduced	none	38	60	50
11	male	group C	associate's degree	standard	none	58	54	52
12	male	group D	associate's degree	standard	none	40	52	43
13	female	group B	high school	standard	none	65	81	73
14	male	group A	some college	standard	completed	78	72	70
15	female	group A	master's degree	standard	none	50	53	58
16	female	group C	some high school	standard	none	69	75	78
17	male	group C	high school	standard	none	88	89	86
18	female	group B	some high school	free/reduced	none	18	32	28

SHETH L.U.J. AND SIR M.V. COLLEGE

SUBJECT: Data Analysis with R

8th output:

The screenshot displays the RStudio interface with the following components:

- Console:** Shows the execution of R code to load the 'insurance' dataset, convert categorical variables to factors, and perform a two-way ANOVA. The ANOVA results are summarized below:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
sex	1	6.436e+08	6.436e+08	11.593	0.000682 ***
smoker	1	1.209e+11	1.209e+11	2177.284	< 2e-16 ***
sex:smoker	1	4.923e+08	4.923e+08	8.868	0.002954 **
Residuals	1334	7.406e+10	5.552e+07		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

- Environment:** Lists the loaded objects: 'anova_re...' (List of 13) and 'insurance' (1338 obs. of 7 vari...).
- Files:** Shows the project files: 'insurance.csv' (54.3 KB), 'R S103.Rproj' (218 B), and 'StudentsPerformance.csv' (70.3 KB).
- Data Viewer:** Displays the first 20 rows of the 'insurance' dataset. The columns are: age, sex, bmi, children, smoker, region, and charges.

SHETH L.U.J. AND SIR M.V. COLLEGE

SUBJECT: Data Analysis with R

9th output:

The image displays two screenshots of the RStudio interface, showing the execution of R code to analyze the Titanic dataset.

Top Screenshot: The console shows the loading of the Titanic dataset and its structure.

```
# Load required library
library(dplyr)
# Import dataset
titanic <- read.csv("tested.csv")
# View structure and data
str(titanic)
# Convert variables to factors
```

The output of `str(titanic)` shows the structure of the data frame:

```
'data.frame': 418 obs. of 12 variables:
 $ PassengerId: int  892 893 894 895 896 897 898 899 900 901 ...
 $ Survived   : int  0 1 0 0 1 0 1 0 1 0 ...
 $ Pclass     : int  3 3 2 3 3 3 3 2 3 3 ...
 $ Name       : chr  "Kelly, Mr. James" "Wilkes, Mrs. James (Ellen Needs)" "Myles, Mr. Thomas Francis" "Wirz, Mr. Albert" ...
 $ Sex        : chr  "male" "female" "male" "male" ...
 $ Age        : num  34.5 47.62 27.22 14.30 26.18 21 ...
 $ Sibsp      : int  0 1 0 0 1 0 0 1 0 2 ...
 $ Parch      : int  0 0 0 0 1 0 0 1 0 0 ...
 $ Ticket     : chr  "330911" "363272" "240276" "315154" ...
 $ Fare       : num  7.83 7.9 9.69 8.66 12.29 ...
 $ Cabin      : chr  "" "" "" "" "" ...
 $ Embarked   : chr  "Q" "S" "Q" "S" ...
```

The output of `head(titanic)` shows the first six rows of the dataset:

PassengerId	Survived	Pclass	Name	Sex	Age	Sibsp	Parch	Ticket	Fare	Cabin
1	0	3	Kelly, Mr. James	male	34.5	0	0	330911	7.8292	
2	1	3	Wilkes, Mrs. James (Ellen Needs)	female	47.0	1	0	363272	7.0000	
3	0	2	Myles, Mr. Thomas Francis	male	62.0	0	0	240276	9.6875	
4	0	3	Wirz, Mr. Albert	male	27.0	0	0	315154	8.6625	
5	1	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	female	22.0	1	1	3101298	12.2875	
6	0	3	Svensson, Mr. Johan Cervin	male	14.0	0	0	7538	9.2250	

Bottom Screenshot: The console shows the conversion of variables to factors and the performance of a Chi-square test.

```
# Convert variables to factors
titanic$Sex <- as.factor(titanic$Sex)
titanic$Survived <- as.factor(titanic$Survived)
# create contingency table
table_data <- table(titanic$Sex, titanic$Survived)
# Display contingency table
table_data
```

The output of `table_data` shows the contingency table:

	0	1
female	0	152
male	266	0

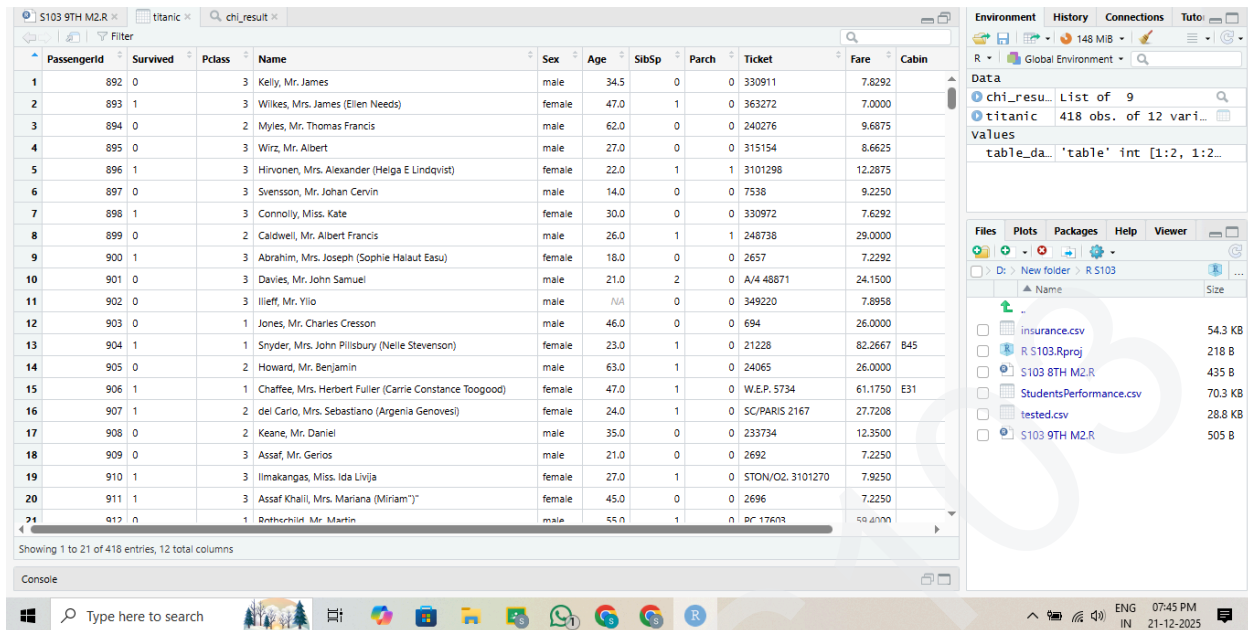
The output of `chi_result` shows the results of the Chi-square test:

```
Pearson's Chi-squared test with Yates' continuity correction

data: table_data
X-squared = 413.69, df = 1, p-value < 2.2e-16
```

SHETH L.U.J. AND SIR M.V. COLLEGE

SUBJECT: Data Analysis with R



The screenshot displays the R Studio environment. The main window shows a data frame with 21 rows and 12 columns. The columns are: Passengerid, Survived, Pclass, Name, Sex, Age, SibSp, Parch, Ticket, Fare, Cabin, and Embarked. The data represents passengers on the Titanic. The Environment pane on the right shows the 'titanic' data frame with 418 observations and 12 variables. The Files pane shows the project files, including 'insurance.csv', 'R S103.Rproj', 'S103 8TH M2.R', 'StudentsPerformance.csv', 'tested.csv', and 'S103 9TH M2.R'.

Passengerid	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
1	0	3	Kelly, Mr. James	male	34.5	0	0	330911	7.8292		
2	1	3	Wilkes, Mrs. James (Ellen Needs)	female	47.0	1	0	363272	7.0000		
3	0	2	Myles, Mr. Thomas Francis	male	62.0	0	0	240276	9.6875		
4	0	3	Wirz, Mr. Albert	male	27.0	0	0	315154	8.6625		
5	1	3	Hirvonen, Mrs. Alexander (Heiga E Lindqvist)	female	22.0	1	1	3101298	12.2875		
6	0	3	Svensson, Mr. Johan Cervin	male	14.0	0	0	7538	9.2250		
7	1	3	Connolly, Miss. Kate	female	30.0	0	0	330972	7.6292		
8	0	2	Caldwell, Mr. Albert Francis	male	26.0	1	1	248738	29.0000		
9	1	3	Abraham, Mrs. Joseph (Sophie Halaut Easu)	female	18.0	0	0	2657	7.2292		
10	0	3	Davies, Mr. John Samuel	male	21.0	2	0	A/4 48671	24.1500		
11	0	3	Ilieff, Mr. Ylio	male	NA	0	0	349220	7.8958		
12	0	1	Jones, Mr. Charles Cresson	male	46.0	0	0	694	26.0000		
13	1	1	Snyder, Mrs. John Pillsbury (Nelle Stevenson)	female	23.0	1	0	21228	82.2667	B45	
14	0	2	Howard, Mr. Benjamin	male	63.0	1	0	24065	26.0000		
15	1	1	Chaffee, Mrs. Herbert Fuller (Carrie Constance Toogood)	female	47.0	1	0	W.E.P. 5734	61.1750	E31	
16	1	2	del Carlo, Mrs. Sebastiano (Argenia Genovesi)	female	24.0	1	0	SC/PAIRS 2167	27.7208		
17	0	2	Keane, Mr. Daniel	male	35.0	0	0	233734	12.3500		
18	0	3	Assaf, Mr. Gerios	male	21.0	0	0	2692	7.2250		
19	1	3	Ilmakangas, Miss. Ida Livija	female	27.0	1	0	STON/OZ. 3101270	7.9250		
20	1	3	Assaf Khalil, Mrs. Mariana (Miriam)"	female	45.0	0	0	2696	7.2250		
21	0	1	Bonthechilid, Mr. Martin	male	55.0	1	0	PC 17613	59.4000		