

Sheth L.U.J. And Sir M.V. College

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

Output:-

The screenshot shows the RStudio interface with the following content:

```
> countries <- read.csv("countries.csv")
> str(countries)
'data.frame': 188 obs. of 21 variables:
 $ Country      : chr  "Afghanistan" "Albania" "Algeria" "Angol
a" ...
 $ Region       : chr  "Middle East/Central Asia" "Northern/East
ern Europe" "Africa" "Africa" ...
 $ Population..millions.: num  29.82 3.16 38.48 20.82 0.09 ...
 $ HDI          : num  0.46 0.73 0.73 0.52 0.78 0.83 0.73 NA 0.9
3 0.88 ...
 $ GDP.per.Capita : chr  "$614.66" "$4,534.37" "$5,430.57" "$4,66
5.91" ...
 $ Cropland.Footprint : num  0.3 0.78 0.6 0.33 NA 0.78 0.74 NA 2.68 0.
82 ...
 $ Grazing.Footprint : num  0.2 0.22 0.16 0.15 NA 0.79 0.18 NA 0.63
0.27 ...
 $ Forest.Footprint  : num  0.08 0.25 0.17 0.12 NA 0.29 0.34 NA 0.89
0.63 ...
 $ Carbon.Footprint  : num  0.18 0.87 1.14 0.2 NA 1.08 0.89 NA 4.85
4.14 ...
 $ Fish.Footprint    : num  0 0.02 0.01 0.09 NA 0.1 0.01 NA 0.11 0.06
...
 $ Total.Ecological.Footprint : num  0.79 2.21 2.12 0.93 5.38 ...
 $ Cropland          : num  0.24 0.55 0.24 0.2 NA 2.64 0.44 NA 5.42
0.71 ...
 $ Grazing.Land       : num  0.2 0.21 0.27 1.42 NA 1.86 0.26 NA 5.81
0.16 ...
 $ Forest.Land        : num  0.02 0.29 0.03 0.64 NA 0.66 0.1 NA 2.01
2.04 ...
 $ Fishing.Water      : num  0 0.07 0.01 0.26 NA 1.67 0.02 NA 3.19 0
...
 $ Urban.Land         : num  0.04 0.06 0.03 0.04 NA 0.1 0.07 NA 0.14
0.15 ...
 $ Total.Biocrapacity : num  0.5 1.18 0.59 2.55 0.94 ...
 $ Biocrapacity.Deficit.or.Reserve: num  -0.3 -1.03 -1.53 1.61 -4.44 ...
 $ Earths.Required    : num  0.46 1.27 1.22 0.54 3.11 1.82 1.29 6.86
```

The Environment pane on the right shows the following objects:

- chi_result: List of 9
- clean_exact: 55 obs. of 3 variables
- combined_data: 100 obs. of 2 variables
- countries: 188 obs. of 21 variables
- cyl_pivot: 32 obs. of 4 variables
- describe_vars: 18249 obs. of 6 variables
- df: 32 obs. of 4 variables
- duplicates_report: 13 obs. of 4 variables

The screenshot shows the RStudio interface with the following content:

```
> data <- read.csv("countries.csv")
> countries$continent <- as.factor(countries$continent)
Error in '$<-data.frame'(*tmp*, continent, value = integer(0)) :
replacement has 0 rows, data has 188

> countries$income_level <- as.factor(countries$income_level)
Error in '$<-data.frame'(*tmp*, income_level, value = integer(0)) :
replacement has 0 rows, data has 188

> contingency_table <- table(countries$continent, countries$income_level)
> print(contingency_table)
< table of extent 0 x 0 >
> chi_result <- chisq.test(contingency_table)
Error in chisq.test(contingency_table) :
at least one entry of 'x' must be positive

> print(chi_result)
Error: object 'chi_result' not found

> print(chi_result$expected)
Error: object 'chi_result' not found

> countries <- read.csv("countries.csv")
> str(countries)
'data.frame': 188 obs. of 21 variables:
 $ Country      : chr  "Afghanistan" "Albania" "Algeria" "Angol
a" ...
 $ Region       : chr  "Middle East/Central Asia" "Northern/East
ern Europe" "Africa" "Africa" ...
 $ Population..millions.: num  29.82 3.16 38.48 20.82 0.09 ...
 $ HDI          : num  0.46 0.73 0.73 0.52 0.78 0.83 0.73 NA 0.9
3 0.88 ...
 $ GDP.per.Capita : chr  "$614.66" "$4,534.37" "$5,430.57" "$4,66
5.91" ...
 $ Cropland.Footprint : num  0.3 0.78 0.6 0.33 NA 0.78 0.74 NA 2.68 0.
82 ...
 $ Grazing.Footprint : num  0.2 0.22 0.16 0.15 NA 0.79 0.18 NA 0.63
0.27 ...
 $ Forest.Footprint  : num  0.08 0.25 0.17 0.12 NA 0.29 0.34 NA 0.89
0.63 ...
 $ Carbon.Footprint  : num  0.18 0.87 1.14 0.2 NA 1.08 0.89 NA 4.85
4.14 ...
 $ Fish.Footprint    : num  0 0.02 0.01 0.09 NA 0.1 0.01 NA 0.11 0.06
...
 $ Total.Ecological.Footprint : num  0.79 2.21 2.12 0.93 5.38 ...
 $ Cropland          : num  0.24 0.55 0.24 0.2 NA 2.64 0.44 NA 5.42
0.71 ...
 $ Grazing.Land       : num  0.2 0.21 0.27 1.42 NA 1.86 0.26 NA 5.81
0.16 ...
 $ Forest.Land        : num  0.02 0.29 0.03 0.64 NA 0.66 0.1 NA 2.01
2.04 ...
 $ Fishing.Water      : num  0 0.07 0.01 0.26 NA 1.67 0.02 NA 3.19 0
...
 $ Urban.Land         : num  0.04 0.06 0.03 0.04 NA 0.1 0.07 NA 0.14
0.15 ...
 $ Total.Biocrapacity : num  0.5 1.18 0.59 2.55 0.94 ...
 $ Biocrapacity.Deficit.or.Reserve: num  -0.3 -1.03 -1.53 1.61 -4.44 ...
 $ Earths.Required    : num  0.46 1.27 1.22 0.54 3.11 1.82 1.29 6.86
```

The Environment pane on the right shows the following objects:

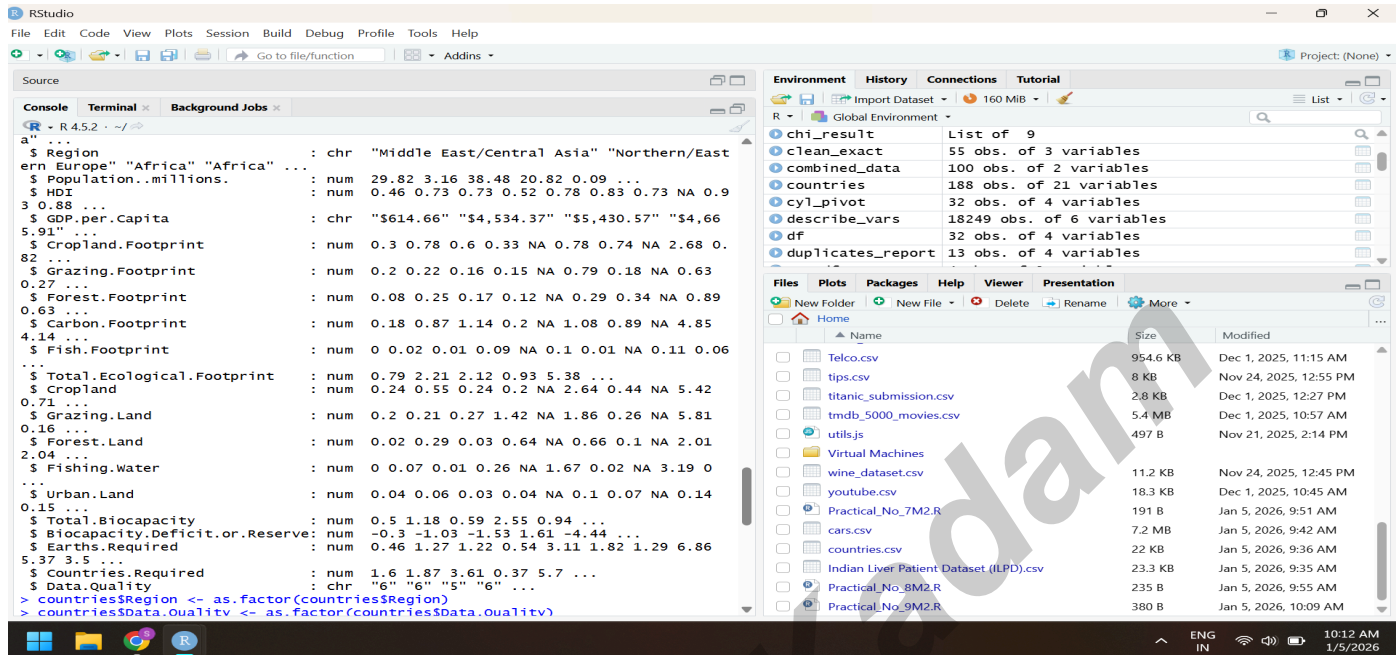
- chi_result: List of 9
- clean_exact: 55 obs. of 3 variables
- combined_data: 100 obs. of 2 variables
- countries: 188 obs. of 21 variables
- cyl_pivot: 32 obs. of 4 variables
- describe_vars: 18249 obs. of 6 variables
- df: 32 obs. of 4 variables
- duplicates_report: 13 obs. of 4 variables

Name:- Mithil Kadam

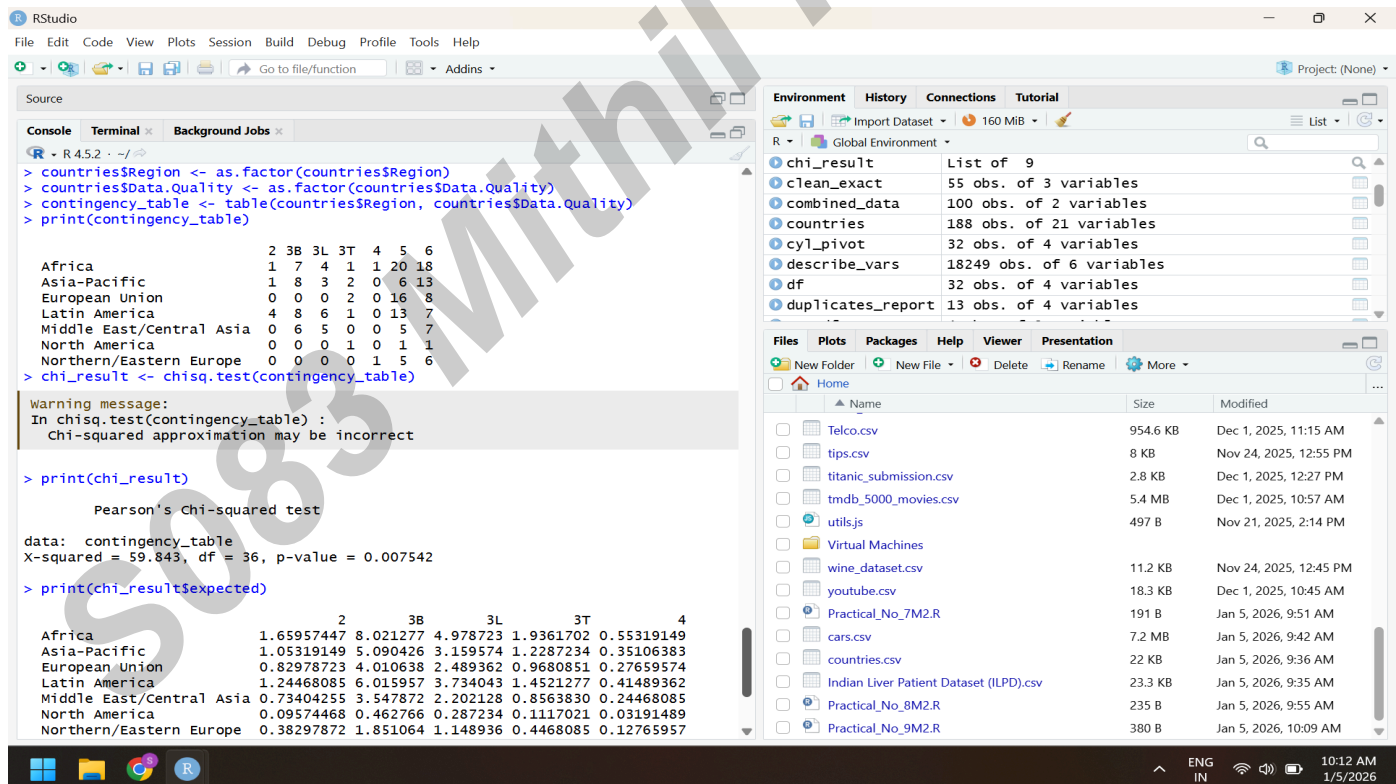
Roll No. S083

Subject:- Data Analysis With SAS / SPSS / R

Sheth L.U.J. And Sir M.V. College



```
a" ...  
$ Region : chr "Middle East/Central Asia" "Northern/East  
ern Europe" "Africa" "Africa" ...  
$ Population..millions. : num 29.82 3.16 38.48 20.82 0.09 ...  
$ HDI : num 0.46 0.73 0.73 0.52 0.78 0.83 0.73 NA 0.9  
3 0.88 ...  
$ GDP.per.Capita : chr "$614.66" "$4,534.37" "$5,430.57" "$4,66  
5.91" ...  
$ Cropland.Footprint : num 0.3 0.78 0.6 0.33 NA 0.78 0.74 NA 2.68 0.  
82 ...  
$ Grazing.Footprint : num 0.2 0.22 0.16 0.15 NA 0.79 0.18 NA 0.63  
0.27 ...  
$ Forest.Footprint : num 0.08 0.25 0.17 0.12 NA 0.29 0.34 NA 0.89  
0.63 ...  
$ Carbon.Footprint : num 0.18 0.87 1.14 0.2 NA 1.08 0.89 NA 4.85  
4.14 ...  
$ Fish.Footprint : num 0 0.02 0.01 0.09 NA 0.1 0.01 NA 0.11 0.06  
...  
$ Total.Ecological.Footprint : num 0.79 2.21 2.12 0.93 5.38 ...  
$ Cropland : num 0.24 0.55 0.24 0.2 NA 2.64 0.44 NA 5.42  
0.71 ...  
$ Grazing.Land : num 0.2 0.21 0.27 1.42 NA 1.86 0.26 NA 5.81  
0.16 ...  
$ Forest.Land : num 0.02 0.29 0.03 0.64 NA 0.66 0.1 NA 2.01  
2.04 ...  
$ Fishing.Water : num 0 0.07 0.01 0.26 NA 1.67 0.02 NA 3.19 0  
...  
$ Urban.Land : num 0.04 0.06 0.03 0.04 NA 0.1 0.07 NA 0.14  
0.15 ...  
$ Total.Biocapacity : num 0.5 1.18 0.59 2.55 0.94 ...  
$ Biocapacity.Deficit.or.Reserve : num -0.3 -1.03 -1.53 1.61 -4.44 ...  
$ Earths.Required : num 0.46 1.27 1.22 0.54 3.11 1.82 1.29 6.86  
5.37 3.5 ...  
$ Countries.Required : num 1.6 1.87 3.61 0.37 5.7 ...  
$ Data.Quality : chr "6" "6" "5" "6" ...  
> countries$Region <- as.factor(countries$Region)  
> countries$Data.Quality <- as.factor(countries$Data.Quality)  
> countries$Data.Quality <- as.factor(countries$Data.Quality)
```



```
> countries$Region <- as.factor(countries$Region)  
> countries$Data.Quality <- as.factor(countries$Data.Quality)  
> contingency_table <- table(countries$Region, countries$Data.Quality)  
> print(contingency_table)  
  
      2 3B 3L 3T 4 5 6  
Africa      1 7 4 1 1 20 18  
Asia-Pacific 1 8 3 2 0 6 13  
European Union 0 0 2 0 0 16 8  
Latin America 4 8 6 1 0 13 7  
Middle East/Central Asia 0 6 5 0 0 5 7  
North America 0 0 0 0 1 0 1  
Northern/Eastern Europe 0 0 0 0 1 5 6  
> chi_result <- chisq.test(contingency_table)  
  
Warning message:  
In chisq.test(contingency_table) :  
Chi-squared approximation may be incorrect  
  
> print(chi_result)  
  
Pearson's Chi-squared test  
data: contingency_table  
X-squared = 59.843, df = 36, p-value = 0.007542  
> print(chi_result$expected)  
  
      2      3B      3L      3T      4  
Africa      1.65957447 8.021277 4.978723 1.9361702 0.55319149  
Asia-Pacific 1.05319149 5.090426 3.159574 1.2287234 0.35106383  
European Union 0.82978723 4.010638 2.489362 0.9680851 0.27659574  
Latin America 1.24468085 6.015957 3.734043 1.4521277 0.41489362  
Middle East/Central Asia 0.73404255 3.547872 2.202128 0.8563830 0.24468085  
North America 0.09574468 0.462766 0.287234 0.1117021 0.03191489  
Northern/Eastern Europe 0.38297872 1.851064 1.148936 0.4468085 0.12765957
```

Name:- Mithil Kadam

Roll No. S083

Subject:- Data Analysis With SAS / SPSS / R

Sheth L.U.J. And Sir M.V. College

The screenshot displays the RStudio interface. The console on the left shows the execution of a Chi-squared test on a contingency table. The output includes a warning message, the test statistics (X-squared = 59.843, df = 36, p-value = 0.007542), and the expected frequencies for each cell in the table.

Warning message:
In `chisq.test(contingency_table)` :
Chi-squared approximation may be incorrect

Pearson's Chi-squared test
data: contingency_table
X-squared = 59.843, df = 36, p-value = 0.007542

Expected frequencies:

	2	3B	3L	3T	4
Africa	1.65957447	8.021277	4.978723	1.9361702	0.55319149
Asia-Pacific	1.05319149	5.090426	3.159574	1.2287234	0.35106383
European Union	0.82978723	4.010638	2.489362	0.9680851	0.27659574
Latin America	1.24468085	6.015957	3.734043	1.4521277	0.41489362
Middle East/Central Asia	0.73404255	3.547872	2.202128	0.8563830	0.24468085
North America	0.09574468	0.462766	0.287234	0.1117021	0.03191489
Northern/Eastern Europe	0.38297872	1.851064	1.148936	0.4468085	0.12765957

	5	6
Africa	18.255319	16.5957447
Asia-Pacific	11.585106	10.5319149
European Union	9.127660	8.2978723
Latin America	13.691489	12.4468085
Middle East/Central Asia	8.074468	7.3404255
North America	1.053191	0.9574468
Northern/Eastern Europe	4.212766	3.8297872

The file explorer on the right shows a list of files in the 'Home' directory, including 'Telco.csv', 'tips.csv', 'titanic_submission.csv', 'tmdb_5000_movies.csv', 'utils.js', 'Virtual Machines', 'wine_dataset.csv', 'youtube.csv', 'Practical_No_7M2.R', 'cars.csv', 'countries.csv', 'Indian Liver Patient Dataset (ILPD).csv', 'Practical_No_8M2.R', and 'Practical_No_9M2.R'.

Name:- Mithil Kadam

Roll No. S083

Subject:- Data Analysis With SAS / SPSS / R