

Simple Exposure Analysis

300958 Social Web Analysis

Week 4 Lab Solutions

1 Using R

- Use `read.csv` to read in the CSV files `keyMetrics.csv`, `LifetimeLikesByGenderAge.csv` and `WeeklyReachDemog.csv` into data frames. Make sure you give sensible names to the data frames you create. Use `as.is=TRUE` for `keyMetrics.csv` to avoid the mentioned problem.

```
> keyMetrics = read.csv("http://staff.scm.uws.edu.au/~lapark/300958/labs/keyMetrics.csv",  
as.is=TRUE)  
> LifetimeLikesByGenderAge =  
read.csv("http://staff.scm.uws.edu.au/~lapark/300958/labs/LifetimeLikesByGenderAge.csv", as.is=TRUE)  
> WeeklyReachDemog =  
read.csv("http://staff.scm.uws.edu.au/~lapark/300958/labs/WeeklyReachDemog.csv", as.is=TRUE)
```

- Identify and extract the "Weekly Total Impressions" variable from the `keyMetrics`.
HINT: the `names` function will give a list of column names.

```
> names(keyMetrics)
```

```
[1] "Date"  
[2] "Daily. People. Talking. About. This"  
[3] "Weekly. People. Talking. About. This"  
[4] "X28. days. People. Talking. About. This"  
[5] "Daily. Page. stories"  
[6] "Weekly. Page. stories"  
[7] "X28. days. Page. stories"  
[8] "Lifetime. Total. likes"  
[9] "Daily. New. likes"  
[10] "Daily. Unlikes"  
[11] "Daily. Friends. of. Fans"  
[12] "Daily. Page. engaged. users"  
[13] "Weekly. Page. engaged. users"  
[14] "X28. days. Page. engaged. users"  
[15] "Daily. Total. reach"  
[16] "Weekly. Total. reach"  
[17] "X28. days. Total. reach"
```

[18] "Daily. Organic. reach"
[19] "Weekly. Organic. reach"
[20] "X28. days. Organic. reach"
[21] "Daily. Paid. reach"
[22] "Weekly. Paid. reach"
[23] "X28. days. Paid. reach"
[24] "Daily. Viral. reach"
[25] "Weekly. Viral. reach"
[26] "X28. days. Viral. reach"
[27] "Daily. Total. impressions"
[28] "Weekly. Total. impressions"
[29] "X28. days. Total. impressions"
[30] "Daily. Organic. impressions"
[31] "Weekly. Organic. impressions"
[32] "X28. days. Organic. impressions"
[33] "Daily. Paid. impressions"
[34] "Weekly. Paid. impressions"
[35] "X28. days. Paid. impressions"
[36] "Daily. Viral. impressions"
[37] "Weekly. Viral. impressions"
[38] "X28. days. Viral. impressions"
[39] "Daily. Logged. in. page. views"
[40] "Weekly. Logged. in. page. views"
[41] "Daily. Logged. in. page. views. 1"
[42] "Weekly. Logged. in. page. views. 1"
[43] "Daily. Reach. of. page. posts"
[44] "Weekly. Reach. of. page. posts"
[45] "X28. days. Reach. of. page. posts"
[46] "Daily. Organic. reach. of. page. posts"
[47] "Weekly. Organic. reach. of. page. posts"
[48] "X28. days. Organic. reach. of. page. posts"
[49] "Daily. Paid. reach. of. page. posts"
[50] "Weekly. Paid. reach. of. page. posts"
[51] "X28. days. Paid. reach. of. page. posts"
[52] "Daily. Viral. reach. of. page. posts"
[53] "Weekly. Viral. reach. of. page. posts"
[54] "X28. days. Viral. reach. of. page. posts"
[55] "Daily. Total. Impressions. of. your. posts"
[56] "Weekly. Total. Impressions. of. your. posts"
[57] "X28. days. Total. Impressions. of. your. posts"
[58] "Daily. Organic. impressions. of. your. posts"
[59] "Weekly. Organic. impressions. of. your. posts"
[60] "X28. days. Organic. impressions. of. your. posts"
[61] "Daily. Paid. impressions. of. your. posts"
[62] "Weekly. Paid. impressions. of. your. posts"
[63] "X28. days. Paid. impressions. of. your. posts"
[64] "Daily. Viral. impressions. of. your. posts"

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

```

[65] "Weekly.Viral.impressions.of.your.posts"
[66] "X28.days.Viral.impressions.of.your.posts"
[67] "Daily.Total.Consumers"
[68] "Weekly.Total.Consumers"
[69] "X28.days.Total.Consumers"
[70] "Daily.Page.consumptions"
[71] "Weekly.Page.consumptions"
[72] "X28.days.Page.consumptions"
[73] "Daily.Negative.feedback"
[74] "Weekly.Negative.feedback"
[75] "X28.days.Negative.feedback"
[76] "Daily.Negative.feedback.from.users"
[77] "Weekly.Negative.feedback.from.users"
[78] "X28.days.Negative.feedback.from.users"
[79] "Daily.Total.check.ins"
[80] "Weekly.Total.check.ins"
[81] "X28.days.Total.check.ins"
[82] "Daily.Total.check.ins.1"
[83] "Weekly.Total.check.ins.1"
[84] "X28.days.Total.check.ins.1"
[85] "Daily.Total.check.ins.using.mobile.devices"
[86] "Weekly.Total.check.ins.using.mobile.devices"
[87] "X28.days.Total.check.ins.using.mobile.devices"
[88] "Daily.Total.check.ins.using.mobile.devices.1"
[89] "Weekly.Total.check.ins.using.mobile.devices.1"
[90] "X28.days.Total.check.ins.using.mobile.devices.1"

```

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

```

> impress <- as.numeric(keyMetrics[,-1,28])
> print(impress)

```

```

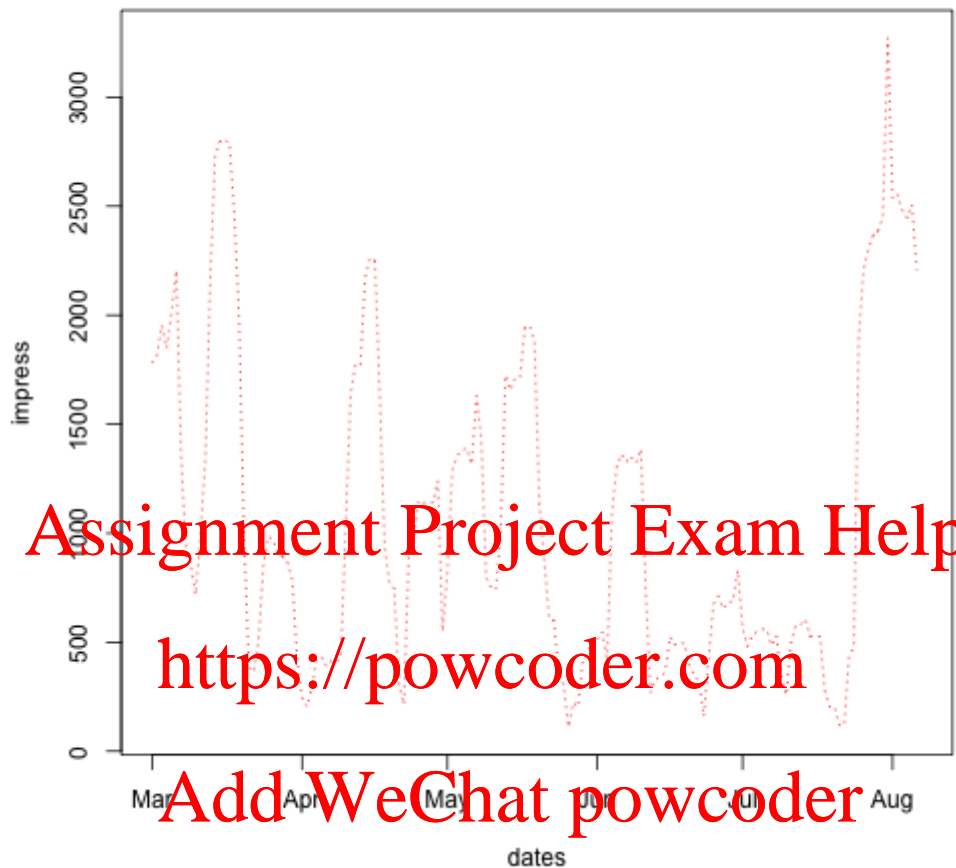
[1] 1783 1815 1955 1847 2012 2206 1353 921 856 717 1002 1337 2156 2731
[15] 2796 2808 2781 2461 1968 981 387 364 509 852 986 970 908 891
[29] 865 779 409 247 204 292 419 432 391 413 403 451 1049 1639
[43] 1779 1774 2174 2262 2248 1639 948 772 747 308 214 898 978 1148
[57] 1137 1140 1140 1240 549 824 1287 1355 1381 1390 1315 1630 1432 798
[71] 751 741 1002 1725 1656 1714 1715 1941 1952 1876 1132 881 612 599
[85] 350 308 110 223 203 417 444 512 514 549 412 1092 1307 1363
[99] 1329 1347 1322 1387 726 271 329 342 362 527 488 496 495 394
[113] 351 316 159 428 677 710 654 684 687 832 567 481 530 550
[127] 560 559 485 529 338 257 499 575 580 605 528 525 527 278
[141] 204 205 119 121 439 476 1894 2208 2305 2365 2387 2451 3272 2537
[155] 2556 2480 2447 2507 2207 NA

```

- Look at the help page for `plot` (see the examples) and try changing the axis labels (`xlab=` and `ylab=`), give a title (`main=`) and changing colours and line type (`col=` and

lty=)

```
> dates <- strptime(keyMetrics[-1,1], format="%m/%d/%y")
> plot(dates,impress, type="l", col=2, lty=3)
```



plot of chunk unnamed-chunk-3

2 χ^2 Test

2.1 Test for independence

Create the two column table.

```
> WeekReach = read.csv("http://staff.scm.uws.edu.au/~lapark/300958/labs/WeeklyReachDemog.csv",
as.is=TRUE)
> tab = matrix(as.numeric(WeekReach[158,3:16]), nrow=2, byrow=TRUE)
> colnames(tab) <- c("13-17", "18-24", "25-34", "35-44",
+ "45-54", "55-64", "65+")
> rownames(tab) <- c("Female", "Male")
>
> stretchTable = function(tab, variableNames) {
```

```

+   tabx = rep(rownames(tab), rowSums(tab))
+   l = ncol(tab)
+   m = nrow(tab)
+   cn = colnames(tab)
+   taby = c()
+   for (a in 1:m) {
+     for (b in 1:l) {
+       taby = c(taby, rep(cn[b], tab[a,b]))
+     }
+   }
+
+   d = data.frame(x = tabx, y = taby)
+   colnames(d) = variableNames
+   return(d)
+ }
>
> tab2 = stretchTable(tab, c("Gender", "Age"))
>
> # Verify that we the correct values
> table(tab2)

```

Assignment Project Exam Help

	Age					
Gender	13-17	18-24	25-34	35-44	45-54	55-64
Female	1	45	15	6	2	3
Male	2	121	31	7	4	2

<https://powcoder.com>

Add WeChat powcoder

Compute the χ^2 randomisation distribution.

```

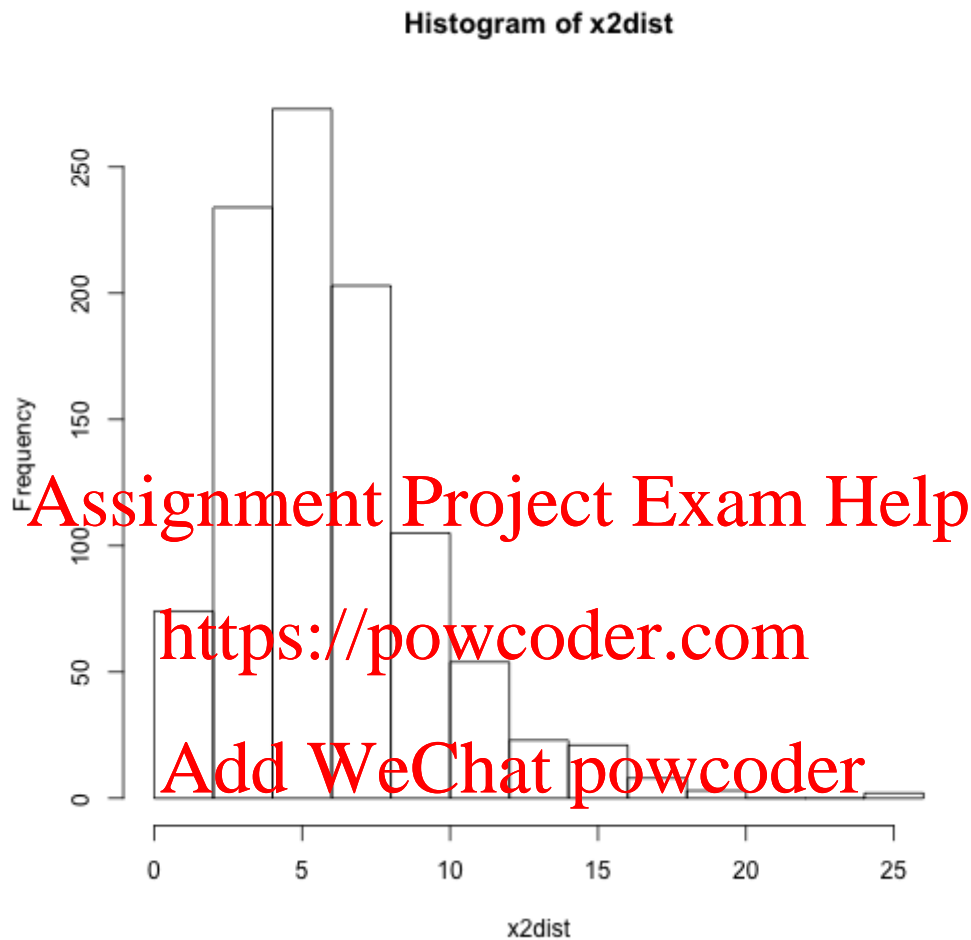
> ## define the functions to use
> expectedIndependent = function(X) {
+   n = sum(X)
+   p = rowSums(X)/sum(X)
+   q = colSums(X)/sum(X)
+   return(p %o% q * n) # outer product creates table
+ }
>
> chiSquaredStatistic = function(X, E) {
+   return(sum((X - E)^2/E))
+ }
>
> ## compute the expected table
> E = expectedIndependent(table(tab2)) # compute expected counts if independent
>
> ## compute the randomisation distribution
> x2dist = replicate(1000, { # compute 1000 randomised chi-squared statistics

```

```

+ ageShuffle = sample(tab2$Age)
+ genderShuffle = sample(tab2$Gender)
+ Xindep = table(genderShuffle, ageShuffle)
+ chiSquaredStatistic(Xindep, E)
+ })
>
> hist(x2dist)

```



plot of chunk unnamed-chunk-5

Compute the χ^2 statistic for `tab`.

```
> x2 = chiSquaredStatistic(table(tab2), E)
```

Compute the p value of the test.

```

> ## pval is the proportion of x2dist that is greater than x2
> pval = mean(x2dist > x2)
> print(pval)

```

```
[1] 0.586
```

The p -value is large, so we cannot reject H_0 . Our conclusion is that we don't have enough evidence to show that age and gender are dependent.

2.2 Test for preference

```
> age = colSums(tab)
> n = sum(age)
> k = length(age)
```

Given the sample size n and the age table, compute the expected age group frequencies if all ages have the same proportion.

```
> ## If all have the same proportion and the proportions must sum to 1:
> ep = rep(1/k, k)
> ## We want the expected frequencies
> E = ep*n
```

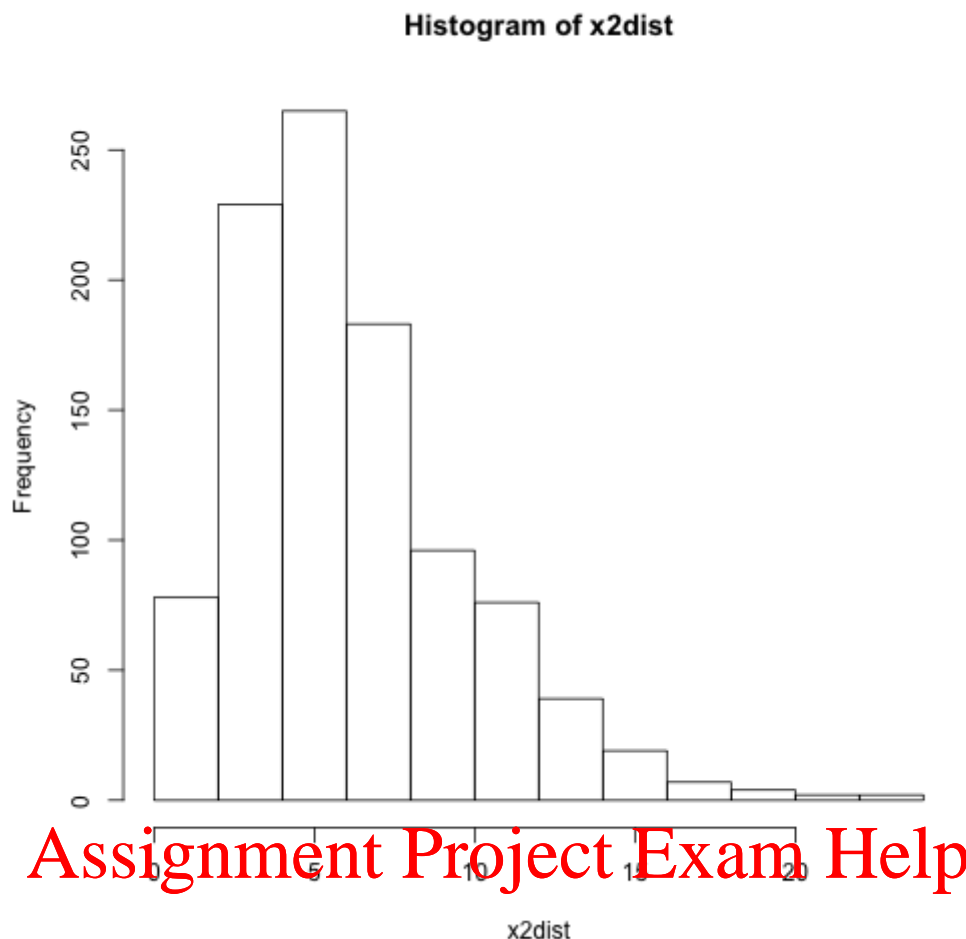
Compute the χ^2 value using the observed and expected frequencies.

```
> x2 = chiSquaredStatistic(age, E)
> print(x2)
```

```
[1] 614.8197
```

- Sample from the set of age categories with replacement using the given proportions and sample size (using `sample`).
- Compute the χ^2 value of the sample.
- Repeat at least 1000 times to obtain a distribution of χ^2 values given H_0 .

```
> x2dist = replicate(1000, {
+   r = table(sample(k, size = n, prob = ep, replace = TRUE))
+   chiSquaredStatistic(r, E)
+ })
> hist(x2dist)
```



Assignment Project Exam Help

<https://powcoder.com>

plot of chunk unnamed-chunk-11

Compute the p value.

Add WeChat powcoder

```
> pval = mean(x2dist > x2)
> print(pval)
```

```
[1] 0
```

The p -value is very small, so we reject H_0 , meaning that we have evidence that the ages are not of equal proportion.

3 More R - a challenge

- Construct the 2×2 table containing the number of males and females (independent of age), versus the months of April and May. Does a χ^2 test show that the reach for gender is independent of the two months?

```
> dates = strptime(WeeklyReachDemog$Date, format="%m/%d/%y")
> months = format(dates, "%b")
> female = rowSums(WeeklyReachDemog[, 3:9], na.rm=TRUE)
```



```

> male = rowSums(WeeklyReachDemog[, 10:16], na.rm=TRUE)
> X = matrix(c(sum(female[months == "Apr"]),
+             sum(female[months == "May"]),
+             sum(male[months == "Apr"]),
+             sum(male[months == "May"])), 2, 2)
> colnames(X) = c("Female", "Male")
> rownames(X) = c("April", "May")
> print(X)

```

	Female	Male
April	985	2545
May	1099	3048

Create the two column single entry table

```

> X2 = stretchTable(X, c("Months", "Gender"))
>
> # Verify that we the correct values
> table(X2)

```

Assignment Project Exam Help

	Gender	
Months	Female	Male
April	985	2545
May	1099	3048

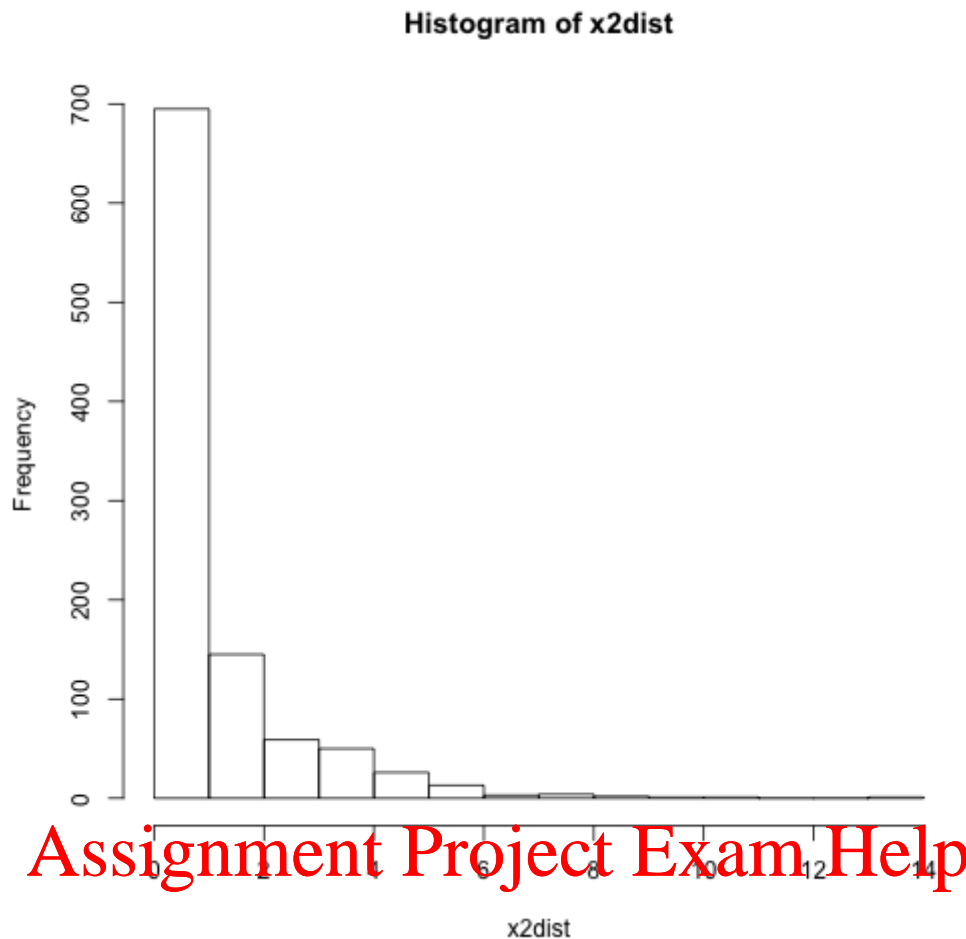
<https://powcoder.com>

Add WeChat powcoder

```

> ## compute the expected table
> E = expectedIndependent(table(X2)) # compute expected counts if independent
>
> ## compute the randomisation distribution
> x2dist = replicate(1000, { # compute 1000 randomised chi-squared statistics
+   genderShuffle = sample(X2$Gender)
+   monthShuffle = sample(X2$Month)
+   Xindep = table(monthShuffle, genderShuffle)
+   chiSquaredStatistic(Xindep, E)
+ })
>
> hist(x2dist)

```



Assignment Project Exam Help

https://powcoder.com

plot of chunk unnamed-chunk-15

Compute the χ^2 statistic of our original sample.

Add WeChat powcoder

```
> Xstat = chiSquaredStatistic(X, E)
```

Compute p -value.

```
> pval = mean(x2dist > Xstat)
> print(pval)
```

```
[1] 0.163
```

The p -value is large, so we don't reject H_0 . We don't have enough evidence to say that month and gender are not independent.

Use the R χ^2 test and compare the p -values to ours.

```
> chisq.test(X)
```

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

data: X

X-squared = 1.8266, df = 1, p-value = 0.1765

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder