

# Tabulating some count data from a Workshop

Miranda

4 December 2015

**Motivational example inspired by R.Muenchen R for SAS and SPSS**

**Categorical data**

**Steps:**

**Enter some data**

**summarise some data**

**showing some data missing**

**converting variables to factors**

**labelling and using summary**

**dropping levels of a factor**

**using a package naming the Questions in the data frame**

**The data is from a training workshop**

**The data is added as vectors and then saved as a csv file.**

**Training Question: Which workshop did you do?**

**for this example we are entering code in the Source pane**

**as part of the syntax. You should be able to bring**

**your own larger file into R and work with it.**

**Entering the data as two vectors**

```
training <- c(1,2,1,2,1,2,1,2)
gender <- c("F", "F", "F", NA, "M", "M", "M", "M")
```

## Entering the answers to the questions

```
Task1 <- c(1,2,2,3,4,5,5,4)
Task2 <- c(1,1,2,1,5,4,3,5)
Task3 <- c(5,4,4,NA,2,5,4,5)
Task4 <- c(1,1,3,3,4,5,4,5)
```

## To get a simple table of frequencies of a categorical variable

```
table(training)
## training
## 1 2
## 4 4

table(gender)
## gender
## F M
## 3 4
```

## The mean values of the answers for the 4 tasks are computed

```
mean(Task1)
## [1] 3.25

mean(Task2)
## [1] 2.75

mean(Task3)
## [1] NA

mean(Task4)
## [1] 3.25
```

## We notice there are missing

### dealing with missing

```
mean(Task3, na.rm=TRUE)
## [1] 4.142857
```

## making a labelled factor of the Training

```
training <- factor (  
  training,  
  levels=c(1,2,3,4),  
  labels = c("Workshop1", "Workshop2", "Workshop3", "Workshop4")  
)
```

Note in this data there are just two types of training.

By adding in the extra levels of this factor (ie 3 and 4) variable,

it can be useful in the future.

This could also be entered as just two levels.

```
training  
## [1] Workshop1 Workshop2 Workshop1 Workshop2 Workshop1 Workshop2 Workshop1  
## [8] Workshop2  
## Levels: Workshop1 Workshop2 Workshop3 Workshop4  
  
table(training)  
  
## training  
## Workshop1 Workshop2 Workshop3 Workshop4  
##          4          4          0          0
```

## making the variable gender into a factor

```
gender <- factor(gender)  
  
gender  
## [1] F    F    F    <NA> M    M    M    M  
## Levels: F M
```

## Save as a data.frame

```
Tdata <- data.frame(training,gender,Task1,Task2,  
                   Task3,Task4)  
  
Tdata
```

```
##   training gender Task1 Task2 Task3 Task4
## 1 Workshop1     F     1     1     5     1
## 2 Workshop2     F     2     1     4     1
## 3 Workshop1     F     2     2     4     3
## 4 Workshop2  <NA>     3     1    NA     3
## 5 Workshop1     M     4     5     2     4
## 6 Workshop2     M     5     4     5     5
## 7 Workshop1     M     5     3     4     4
## 8 Workshop2     M     4     5     5     5
```

## Save as a csv file

```
write.csv(Tdata, file="Tdata.csv")
```

## Check the csv file is written in the folder

## Converting Task responses to factors and labelling

```
tasklevels <- c(1,2,3,4,5)
tasklabels <- c("Strongly Disagree",
               "Disagree",
               "Neutral",
               "Agree",
               "Strongly Agree")

tasklevels
## [1] 1 2 3 4 5

tasklabels
## [1] "Strongly Disagree" "Disagree"          "Neutral"
## [4] "Agree"             "Strongly Agree"
```

## Since there is an order to the variable

## we can label using the ordered factor

```
Tdata$Task1 <- ordered(Tdata$Task1,tasklevels,tasklabels)
Tdata$Task2 <- ordered(Tdata$Task2,tasklevels,tasklabels)
Tdata$Task3 <- ordered(Tdata$Task3,tasklevels,tasklabels)
Tdata$Task4 <- ordered(Tdata$Task4,tasklevels,tasklabels)
```

## Use summary to look at the data

```
summary(Tdata[ c("Task1", "Task2", "Task3", "Task4")])

##           Task1           Task2           Task3
## Strongly Disagree:1 Strongly Disagree:3 Strongly Disagree:0
## Disagree           :2 Disagree           :1 Disagree           :1
## Neutral             :1 Neutral           :1 Neutral           :0
## Agree              :2 Agree             :1 Agree             :3
## Strongly Agree     :2 Strongly Agree :2 Strongly Agree     :3
##                   NA's                 :1
##           Task4
## Strongly Disagree:2
## Disagree           :0
## Neutral             :2
## Agree              :2
## Strongly Agree     :2
##
```

## Dropping some levels of a factor

### showing this functionality

```
summary(training [,drop=TRUE])

## Workshop1 Workshop2
##           4         4
```

## Plotting this data as a summary

### Using the library(Hmisc) we can label the tasks:

```
library(Hmisc)

## Warning: package 'Hmisc' was built under R version 3.2.2

## Loading required package: grid
## Loading required package: lattice
## Loading required package: survival
## Loading required package: Formula
## Loading required package: ggplot2

## Warning: package 'ggplot2' was built under R version 3.2.1

##
## Attaching package: 'Hmisc'
##
## The following objects are masked from 'package:base':
```

```
##
##      format.pval, round.POSIXt, trunc.POSIXt, units

label (Tdata$Task1) <- "I am able to calculate seed rate"
label (Tdata$Task1) <- "I am able to calculate fertilizer rate"
label (Tdata$Task1) <- "I can randomise my trial"
label (Tdata$Task1) <- "I can analyse my results in RStudio"
```

## To get a nice summary of the categorical variables

```
describe(Tdata)

## Tdata
##
## 6 Variables      8 Observations
## -----
-
## training
##      n missing  unique
##      8      0      2
##
## Workshop1 (4, 50%), Workshop2 (4, 50%)
## -----
-
## gender
##      n missing  unique
##      7      1      2
##
## F (3, 43%), M (4, 57%)
## -----
-
## Task1 : I can analyse my results in RStudio
##      n missing  unique
##      8      0      5
##
## Strongly Disagree (1, 12%), Disagree (2, 25%)
## Neutral (1, 12%), Agree (2, 25%), Strongly Agree (2, 25%)
## -----
-
## Task2
##      n missing  unique
##      8      0      5
##
## Strongly Disagree (3, 38%), Disagree (1, 12%)
## Neutral (1, 12%), Agree (1, 12%), Strongly Agree (2, 25%)
## -----
-
## Task3
##      n missing  unique
```

```

##          7          1          3
##
## Disagree (1, 14%), Agree (3, 43%)
## Strongly Agree (3, 43%)
## -----
-
## Task4
##      n missing  unique
##      8         0       4
##
## Strongly Disagree (2, 25%), Neutral (2, 25%)
## Agree (2, 25%), Strongly Agree (2, 25%)
## -----
-

str(Tdata)

## 'data.frame':   8 obs. of  6 variables:
## $ training: Factor w/ 4 levels "Workshop1","Workshop2",...: 1 2 1 2 1 2 1
2
## $ gender  : Factor w/ 2 levels "F","M": 1 1 1 NA 2 2 2 2
## $ Task1   : Ord.factor w/ 5 levels "Strongly Disagree"<...: 1 2 2 3 4 5 5
4
## ..- attr(*, "label")= chr "I can analyse my results in RStudio"
## $ Task2   : Ord.factor w/ 5 levels "Strongly Disagree"<...: 1 1 2 1 5 4 3
5
## $ Task3   : Ord.factor w/ 5 levels "Strongly Disagree"<...: 5 4 4 NA 2 5 4
5
## $ Task4   : Ord.factor w/ 5 levels "Strongly Disagree"<...: 1 1 3 3 4 5 4
5

names(Tdata)

## [1] "training" "gender"    "Task1"    "Task2"    "Task3"    "Task4"

```