NATIONAL UNIVERSITY OF SINGAPORE
Department of Statistics and Applied Probability

2018/19 Semester 2 **ST2137 Computer Aided Data Analysis** **Tutorial 5**

1. During election campaigns, there have been television debates between two candidates. A researcher was interested in determining whether or not a particular debate between two candidates was effective in changing viewer's preferences for the candidates. The researcher randomly selected a sample of voters before the debate and asked them to indicate their preferences for the two candidates. The same voters were asked for their preferences for the two candidates after the television debate. The results of the survey are given as follows.

Gender	Preference before TV debate	Preference after TV debate	
		Candidate A	Candidate B
Male	Candidate A	67	28
	Candidate B	46	54
Female	Candidate A	58	42
	Candidate B	37	61

   (a) For each gender group, use SAS to test if the debate between two candidates was effective in changing viewers' preferences for the candidates.
   (b) Repeat part (a) using R
   (c) Repeat part (a) using SPSS.

2. A large corporation is interested in determining whether an association exists between the commuting time of its employees and the level of stress-related problems observed on the job. A study of 116 assembly line workers was conducted. The file "stress.txt" contains the results of the study. Variables and their corresponding descriptions are given as follows.

Variable	Description
stress	Stress level (H = High, M = Moderate, L = Low)
time	Commuting Time (S = Short, Under 15 minutes, M = Moderate, 15-45 minutes, L = Long, Over 45 minutes)

   Use SAS, R and SPSS to do parts (a) to (b)
   (a) Test at the 0.05 level of significance if there is evidence of a significant relationship between commuting time and stress?
   (b) Construct a side by side bar chart to compare the stress levels over various commuting time levels.

3. A manufacturer of flashlight batteries took a sample of 13 batteries from a day's production and used them continuously until they failed to work. The life as measured by the number of hours until failure was given in the data file "battery.txt".
   (a) Use SAS to find out if there is evidence that the mean life of batteries is more than 400 hours. Use a 5% significance level.
   (b) Repeat (a) using R.
   (c) Repeat (a) using SPSS.

4. (Refer to Question 1 in Tutorial 4) One of the major measures of the quality of service provided by any organization is the speed with which the organization responds to customer complaints. During a recent year a company got 50 complaints. The file "furniture.txt" consists of the number of days between the receipt of the complaint and the resolution of the complaint.
   (a) Using a nonparametric method in SAS to determine if the mean number of days between receipt of the complaint and the resolution of the complaint is 20 days or more. Use a 5% significance level.
   (b) Repeat (a) using R
   (c) Repeat (a) using SPSS.

Partial answers to selected questions

1. (a) SAS code (Refer to p7.53-7.54)

```
data t5q1m; set t5q1; where(gender="M");
proc freq data=t5q1m; tables before*after/agree; weight count;
```

Male

	After debate	
Before debate	Candidate A	Candidate B
Candidate A	67	28
Candidate B	46	54

H_0: Debate was not effective in changing male viewers' preference. McNemar's test: Observed Chi-square = 4.38 > $\chi^2_{0.05}(1) = 3.84$ or p-value = 0.0364. Reject H_0.

Female

	After debate	
Before debate	Candidate A	Candidate B
Candidate A	58	42
Candidate B	37	61

H_0: Debate was not effective in changing female viewers' preference. McNemar's test: Observed Chi-square = 0.317 > $\chi^2_{0.05}(1) = 3.84$ or p-value = 0.5737. Do not reject H_0.

(b) R code (Refer to p7.55-7.56)

```
t5q1m <- matrix(c(67,28,46,54), nr=2, byrow=T, dimnames = list("Before
TV debate"=c("Candidate A", "Candidate B"), "After TV
debate"=c("Candidate A", "Candidate B")))
mcnemar.test(t5q1m)
```

(c) SPSS (Refer to p7.57-7.58)

"Data" → "Weight Cases.." → Move "Count" to Frequency variable
"Analyze → "Descriptive Statistic" → "Crosstabs.." → Move "before" to Rows, Move after to "after" and "gender to Level 1 of 1 → "Statistics.." → "McNemar"

2. (a) H_0: Stress level and commuting time are independent. Observed Chi-square = 9.8311 > $\chi^2_{0.05}(4) = 9.48$ or p-value = 0.0434. Reject H_0.

(b) SAS code (Refer to p7.60)

```
proc sgpanel data=t5q2; panelby time; vbar stress/group=stress;
```

R code (refer to p7.62)

```
stressgp <- ifelse(stress=="L","(3) Low",ifelse(stress=="M", "(2)
Moderate", "(1) High"))
timename <- c("(1) Under 15 min", "(2) Between 15 to 45 min", "(3)
Over 45 min")
stressname <- c("(1) High", "(2) Moderate", "(3) Low")
barplot(matrix(table(stressgp, timegp),3,3,byrow=F,
  dimnames=list(stressname,timename)), beside=TRUE,col=c(1:3))
legend(1,28, stressname, fill=c(1:3))
```

SPSS (Refer to p7.65-66)

"Graphs" → "Legacy Dialogs..." → "Bar…"
Choose "Cluster" → "Define"
Move "time" to Category Axis and "stress" to "Define Clusters by" panel

3. (a) SAS code (Refer to p8.6)

```
proc ttest data= t5q3 h0=400 side=U; var life;
```

H_0: $\mu = 400$ against H_1: $\mu > 400$. $t_{obs} = 1.244 < t_{0.05}(12) = 1.78$ or p-value $= 0.1186$. Do not reject H_0.

(b) R code (Refer to p8.22)

```
t.test(life, mu=400, conf.level=0.95, alternative="greater")
```

(c) SPSS code (Refer to p8.9 to 8.10)

"Analyze" → "Compare Means" → "One-Sample T-test.." → Move "life" to Test Variable(s) → Change the value in Test Value box to 400

4. (a) SAS code (Refer to p8.4)

```
proc univariate data=t5q4 mu0=20; var days;
```

H_0: $\mu = 20$ against H_1: $\mu > 20$. Sign test: p-value $= 0.0038/2 = 0.0019^*$, Signed rank test: p-value $= 0.0012/2 = 0.0006^*$. Reject H_0.

(b) R code (Refer to p8.7-8.8)

```
daysnon0 <- days[days != 20]
binom.test(sum(daysnon0>20),length(daysnon0),alternative="greater")
wilcox.test(daysnon0,mu=20,alternative="greater")
```

H_0: $\mu = 20$ against H_1: $\mu > 20$. Sign test: p-value $= 0.0019$, Signed rank test: p-value $= 0.0009$

(c) SPSS code (Refer to p8.11-8.13

Create a new variable "mu" with 50 values of μ_0 (i.e. 20)
"Analyze" → "Nonparametric Tests" → "Legacy Dialogs" → "2 Related Samples…"
Move "days" and "mu" to the Test Pair(s) List → Choose Test Type: Wilcoxon and Sign