# MATH 3070 Lab Project 11

*Aini Liang*

*15 十一月, 2017*



*Remember: I expect to see commentary either in the text, in the code with comments created using* `#`*, or (preferably) both!* ***Failing to do so may result in lost points!***

*Because randomization is used in this assignment, I set the seed here, in addition to beginning each code block.* ***Do not change the seed!***

```
set.seed(6222016)
```

# Problem 1 (Verzani problem 7.1)

*Simulate 1000 rolls of a pair of dice, and compute the sum of each pair. Which is more common, a roll of 7 or 8?*

```
# Your code here
s = sample(1:6, 1000, T) + sample(1:6, 1000, T)
length(s[s == 7])
```

```
## [1] 164
```

```
length(s[s == 8])
```

```
## [1] 130
```

# Problem 2 (Verzani problem 7.2)

*For the* `rivers` *data set, take 1000 random samples of size 10. Compare the mean of the sample means computed from these samples, with the sample mean of the data in* `rivers`*.*

```
# Your code here
ss <- mean(replicate(1000, mean(sample(rivers, 10))))
ss
```

```
## [1] 591.3036
```

```
sampleMean = mean(rivers)
sampleMean
```

```
## [1] 591.1844
```

# Problem 3

*The data set* `Melanoma` *(**MASS**) includes data for 205 Danish patients with malignant melanoma. The variable* `time` *describes survival time in days, and* `sex` *describes the sex of the patient. Does survival time differ between the sexes?*

1. *Find \(\bar{x}_{\text{men}} - \bar{x}_{\text{women}}\), the mean difference in survival time (* `time` *) between men and women (* `sex` *) in* `Melanoma`*.*

```
# Your code here
library(MASS)
data <- Melanoma
men <- subset(Melanoma, subset = sex == 1, select = time, drop = TRUE)
women <- subset(Melanoma, subset = sex == 0, select = time, drop = TRUE)
mean(men) - mean(women)
```

```
## [1] -336.934
```

2. *Investigate whether the difference you observed in part 1 is significant, using procedures explored in the lecture. There are two groups in this investigation: male (* `Melanoma$sex == 0` *) and female (* `Melanoma$sex == 1` *). Randomly reassign the data in the* `time` *variable to the two groups, and compute the mean difference. Repeat 2000 times (this needs to be done relatively quickly; if it takes over a few minutes, I will dock points), and determine how frequently the difference in the mean survival time between men and women (that is, \(\bar{x}_{\text{men}} - \bar{x}_{\text{women}}\) in the simulated data is less than the same difference observed in the actual data. Does this analysis suggest the difference is due to "noise", or due to an actual difference in survival time between men and women?*

```
# Your code here
sss <- replicate(2000, {
    indicator <- sample(1:205, size = sum(Melanoma$sex == 0))
    with(Melanoma, mean(time[indicator]) - mean(time[-indicator]))
})
mean(sss < (mean(men) - mean(women)))
```

```
## [1] 0.0155
```