

# Exercise 9

0316213 Yu-Wen Pu

2018-05-22

```
knitr::opts_chunk$set(results = "hold")
set.seed(1830)
```

```
ttl <- function(x) {
  paste("#### ", x, "\n", sep = "")
}
# e.g. ttl("19.11") outputs "#### 19.11\n"
```

```
q <- strsplit("10.1、10.3、10.4、10.5、10.6、10.8、10.9、10.10、10.12、10.13", ", ")
q <- sapply(q, ttl)
q <- paste(q, collapse = "")
cat(q)
```

## 10.1

```
d <- read.table("health.dat", header = TRUE, sep = ",")
reg <- lm(d$Y ~ d$X1)
a <- reg$coefficients[["(Intercept)"]]
b <- reg$coefficients[["d$X1"]]
cat("Y' =", round(b, 2), "* X +", round(a, 2), fill = TRUE)
```

```
## Y' = 0.07 * X + 3.53
```

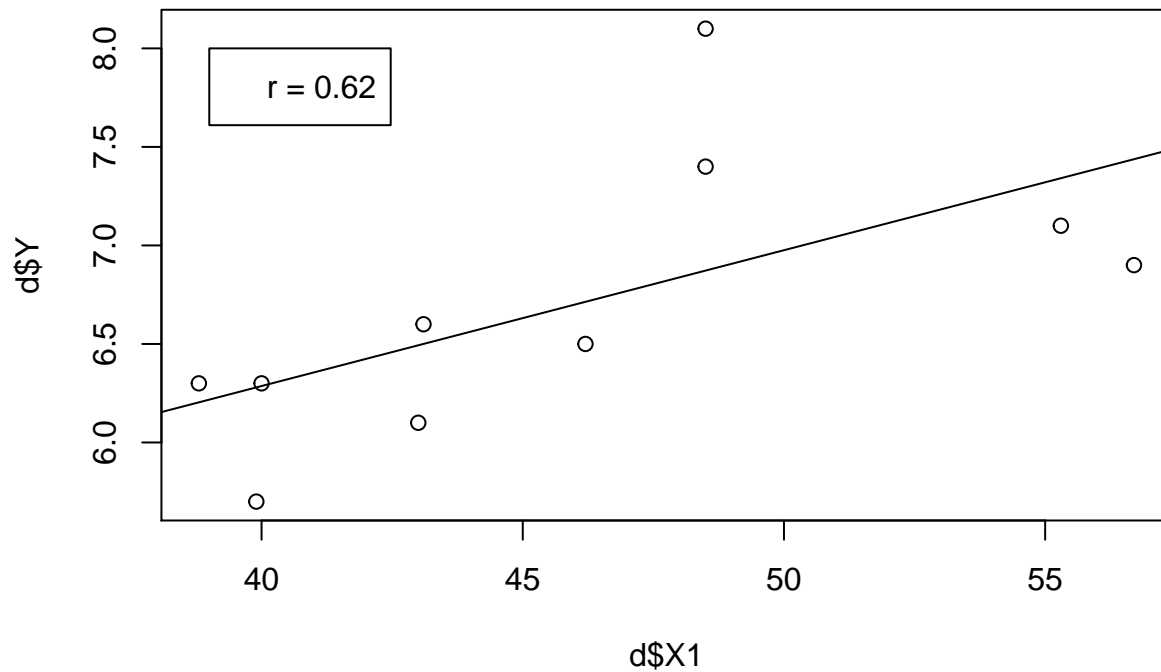
## 10.3

```
round(b * 70 + a, 2)
```

```
## [1] 8.35
```

## 10.4

```
plot(d$Y ~ d$X1)
abline(reg)
legend(39, 8, paste("r =", round(cor(d$X1, d$Y), 2)))
```



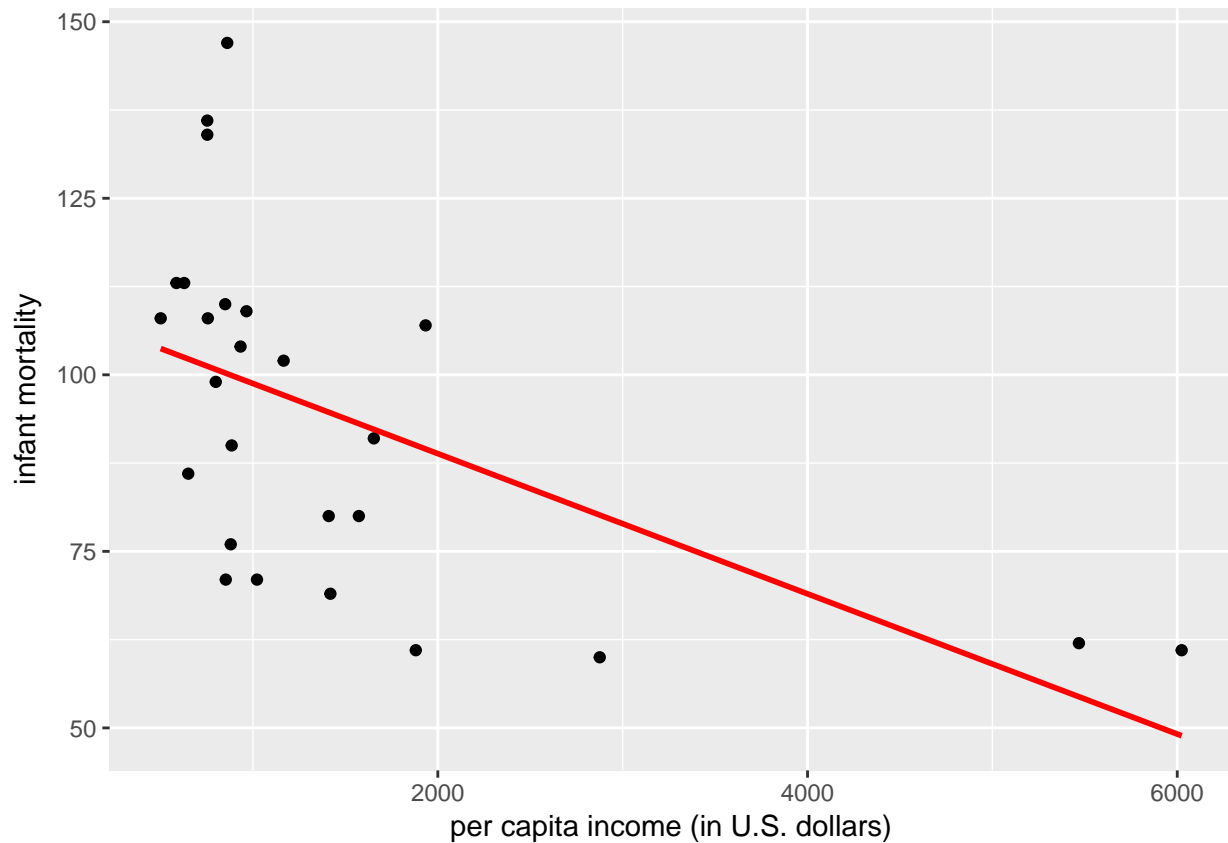
資料量過少，而且 70 離給定資料中 X1 的最大值 (56.7) 太遠，預測的準確度可能不高。

## 10.5

因為 Senegal 的收入和表中大部份的國家相去不遠，Ethiopia 或 Namibia 的收入就有些極端。

## 10.6

```
d <- read.table("../week11b/SubSaharanInfMort.dat", header = TRUE, sep = ",",
               na.strings = "NA", fileEncoding = "ISO-8859-1")
library(ggplot2)
ggplot(d, aes(x = income, y = infmort)) +
  geom_point() +
  geom_smooth(method = "lm", color = "red", se = FALSE) +
  labs(x = "per capita income (in U.S. dollars)", y = "infant mortality")
```



從以上 regression line 可以看出，收入愈高，嬰兒死亡率愈低。

## 10.8

```
d <- read.table("../week12a/Tab10-1.dat", header = TRUE, sep = "")
reg <- lm(d$Symptoms ~ d$Stress)
a <- reg$coefficients[["(Intercept)"]]
b <- reg$coefficients[["d$Stress"]]
cat("predicted:", b * 21.467 + a, "(using 3 digits)", fill = TRUE)
cat("predicted:", b * mean(d$Stress) + a, "(using all digits)", fill = TRUE)
cat("mean(Symptoms):", mean(d$Symptoms), fill = TRUE)
```

```
## predicted: 90.70071 (using 3 digits)
## predicted: 90.70093 (using all digits)
## mean(Symptoms): 90.70093
```

兩者幾乎一致。

## 10.9

不會對 correlation 造成任何影響。

## 10.10

slope 不變，intercept 會跟著被減 10。

## 10.12

令  $y = 1 \times x + 0$ ，即  $y = x$ ，得：

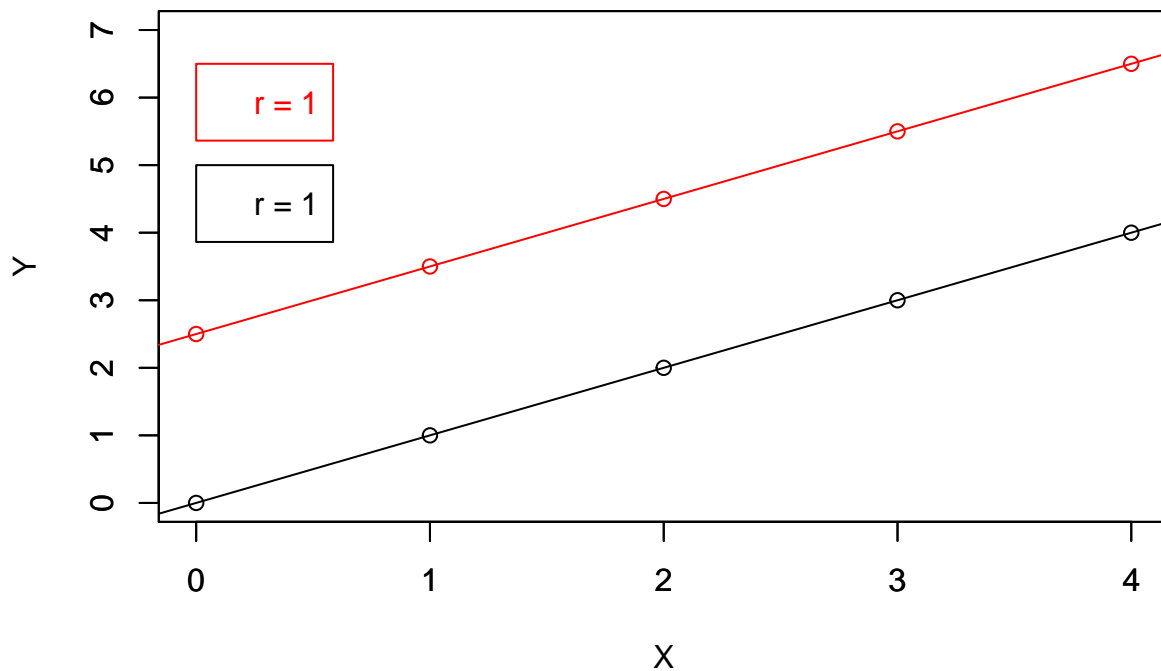
(0, 0)  
(1, 1)  
(2, 2)  
(3, 3)  
(4, 4)

## 10.13

```
d1 <- data.frame(X = 0:4, Y = 0:4)
d2 <- data.frame(X = d1$X, Y = d1$Y + 2.5)
d2
```

```
##   X   Y
## 1 0 2.5
## 2 1 3.5
## 3 2 4.5
## 4 3 5.5
## 5 4 6.5
```

```
plot(d1$Y ~ d1$X, ylim = c(0, 7), xlab = "X", ylab = "Y")
abline(lm(d1$Y ~ d1$X))
legend(0, 5, paste("r =", cor(d1$X, d1$Y)))
par(new = TRUE)
plot(d2$Y ~ d2$X, ylim = c(0, 7), xlab = "", ylab = "", col = "red")
abline(lm(d2$Y ~ d2$X), col = "red")
legend(0, 6.5, paste("r =", cor(d2$X, d2$Y)), text.col = "red", box.col = "red")
```



- a) slope 不變，intercept 會跟著被加 2.5。
- b) 不會對 correlation 造成任何影響。