

Exercise 7

0316213 Yu-Wen Pu

2018-05-06

```
knitr::opts_chunk$set(results = "hold", fig.retina = 2)
set.seed(1830)
```

8.6

- 我只關心我是否有被多收錢，不關心我是否有被少收錢。
- 我應該以 upper tail 來判斷。
- 若以 lower tail 來判斷，則很可能只有被少收錢時才會 reject null hypothesis。

8.9

- a)
 - null hypothesis: 兩次猜測的平均誤差 \geq 第一次猜測的誤差。
 - alternative hypothesis: 兩次猜測的平均誤差 $<$ 第一次猜測的誤差。
- b)
 - Type I error: 我判定「兩次猜測的平均誤差 $<$ 第一次猜測的誤差」，但事實上「兩次猜測的平均誤差 \geq 第一次猜測的誤差」。
 - Type II error: 我判定「兩次猜測的平均誤差 \geq 第一次猜測的誤差」，但事實上「兩次猜測的平均誤差 $<$ 第一次猜測的誤差」。
- c) one-tailed test，因為依題意，我們只關心兩次猜測的平均誤差會不會小於第一次猜測的誤差。要是我們也想確認兩次猜測的平均誤差會不會大於第一次猜測的誤差，那就得用 two-tailed test 了。

8.20

- a) 政府專賣的價格與企業私營的價格相同。
- b) H_1 的分佈的平均值； H_0 的分佈的平均值。
- c) 因為這樣的差異可能不夠顯著，只是巧合。
- d) 我們已知企業私營的價格的分佈，並抽樣獲得政府專賣的價格的樣本。

19.1

```
chisq.test(c(25, 32, 10))
```

```
##
## Chi-squared test for given probabilities
##
## data:  c(25, 32, 10)
## X-squared = 11.313, df = 2, p-value = 0.003494
```

p 值小於 5%，因此我們可以相信，這個分佈不是隨機的，或許有些老師真的特別受歡迎。

19.2

- 這三門課除了授課教師不同外，上課時間也不同。因此 Prof. Klansky 的課最多人選，也可能是因為 11:00 a.m. 最多人有空。
- 如何改進這個實驗：選擇上課時間相同，只有授課教師不同的課來檢驗。在這種情況下，我們才能假設如果老師都一樣受歡迎，學生的分佈會是隨機的。

19.3

```
chisq.test(c(8, 10, 20, 8, 4), p = c(.1, .2, .4, .2, .1))
```

```
##  
## Chi-squared test for given probabilities  
##  
## data: c(8, 10, 20, 8, 4)  
## X-squared = 2.4, df = 4, p-value = 0.6626
```

Yes. p 值遠大於 5%。

19.4

population: 作者的小孩。

19.5

```
chisq.test(c(169, 83))
```

```
##  
## Chi-squared test for given probabilities  
##  
## data: c(169, 83)  
## X-squared = 29.349, df = 1, p-value = 6.044e-08
```

p 值遠小於 5%，因此我們可以相信，這個分佈不是隨機的，黑人小孩受到了社會上的種族偏見的影响，較偏好白種的娃娃。

19.6

```
chisq.test(c(28, 61))
```

```
##  
## Chi-squared test for given probabilities  
##  
## data: c(28, 61)  
## X-squared = 12.236, df = 1, p-value = 0.0004688
```

p 值小於 5%，因此我們可以相信，這個分佈不是隨機的，或許是因為在 1970 年黑人的權利已逐漸受到重視，偏好白種娃娃的黑人小孩的比例相較於 1939 年也顯著減少了。

19.7

```
x <- matrix(c(169, 83, 28, 61), byrow = TRUE, ncol = 2)  
x <- cbind(x, rowSums(x))  
x <- rbind(x, colSums(x))  
colnames(x) <- c("White Doll", "Black Doll", "Total")  
rownames(x) <- c("1939", "1970", "Total")  
knitr::kable(x)
```

	White Doll	Black Doll	Total
1939	169	83	252
1970	28	61	89
Total	197	144	341

```
x <- matrix(c(169, 83, 28, 61), byrow = TRUE, ncol = 2)
chisq.test(x, correct = FALSE)
```

```
##
##  Pearson's Chi-squared test
##
## data:  x
## X-squared = 34.173, df = 1, p-value = 5.042e-09
```

p 值遠小於 5%，因此我們可以相信，「娃娃的選擇」與「調查年份」之間不是獨立的，不同時間做調查會得到不同的結果。

19.7 探討的是「娃娃的選擇」與「調查年份」之間是否獨立；**19.5** 和 **19.6** 探討的是「娃娃的選擇」與否是隨機的。