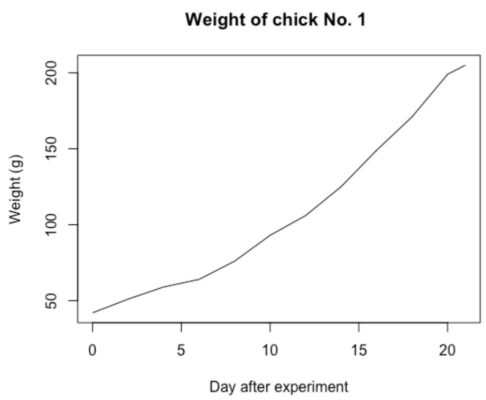
統計學與實習上
第六次作業

1. 資料集 ChickWeight 為 R 語言內建之資料集，紀錄編號 1 至 50 的小雞餵食 4 種不同配方之飼料，於第 0 至 21 天每隔兩天所測得的體重。檔案內共有四個欄位，分別為體重(weight, unit: g)、測重日(Time, unit: day)、小雞編號(Chick)、飼料編號(Diet)。請使用此資料回答以下問題。**(by R) (1.5 points)**

a. 請於 R 中使用函數或是使用讀取檔案方式讀入此資料，並使資料名稱為 ChickWeight。請以時間為 x 軸，體重為 y 軸，繪製出一張折線圖以呈現小雞一號於實驗中的體重成長。**(0.3 points)**

```
# a.
data(ChickWeight) # by function data()
ChickWeight <- read.csv('ChickWeight.csv', header = T, sep = ',') # or reading the .csv file
c1 <- ChickWeight[ChickWeight$Chick == 1,]
plot(c1$Time, c1$weight, type = 'l',
xlab = 'Day after experiment', ylab = 'Weight (g)',
main = 'Weight of chick No. 1 ')
```



b. 已知於多年前文獻中顯示食用飼料一號的小雞於第 21 天所測得的體重平均為 150g，本團隊想了解隨著雞隻品種改進，是否能在同樣飼料下成長更快速。請建立一資料名為 chick_21，其中僅包含第 21 天之資料，使用此資料檢驗食用飼料一號的 21 天小雞平均體重是否高於文獻值，即 $H_0: \mu_{Diet1,\ 21\,days} <= 150g$，並計算出其 95%信賴區間。**(0.5 points)**

```
# b.
chick_21 <- ChickWeight[ChickWeight$Time == 21,]
chick_21_d1 <- chick_21[chick_21$Diet == 1,]
t.test(chick_21_d1$weight, mu = 150, alternative = 'greater')
# H0: mu_(diet 1, 21 days) <= 150g; H1: mu_(diet 1, 21 days) > 150g
# The 95% CI: [152.0231, Inf) does not include 150, therefore, reject H0.
```

```r
# Conclude that the mean chick weight under 28 day is greater than before.
```

c. 雞隻飼料評估指標之一為過輕率，本實驗中設定小雞 21 天重小於 100g 即為過輕。於 chick_21 內新增一欄位名為 underweight，將過輕小雞紀錄為 1，非過輕者紀錄為 0。製作一比例列聯表顯示並討論食用四種飼料小雞的過輕率。**(0.7 points)**

```r
c.
chick_21$underweight <- ifelse(chick_21$weight < 100,1,0)
prop.table(table(underweight = chick_21$underweight, diets = chick_21$Diet),2)
# As shown in the table, chicks consuming diets 1 and 2 had the highest and second highest rates of underweight,
# and all the chicks with diets 3 and 4 have normal weight.
```

```
           feed
underweight     1     2     3     4
          0 0.875 0.900 1.000 1.000
          1 0.125 0.100 0.000 0.000
```

2.(Text, p.310) The Tennessee Tourism Institute (TTI) plans to sample information center visitors entering the state to learn the fraction of visitors who plan to camp in the state. Current estimates are that 35% of visitors are campers. How many visitors would you sample to estimate the population proportion of campers with a 95% confidence level and an allowable error of 2%? **(0.8 points)**

2185, found by $n=0.35(1-0.35)\left(\frac{1.96}{0.02}\right)^2 = 2184.9$

3. (Text, p.351) An insurance company, based on past experience, estimates the mean damage for a natural disaster in its area is $5,000. After introducing several plans to prevent loss, it randomly samples 200 policyholders and finds the mean amount per claim was $4,800 with a standard deviation of $1,300. Does it appear the prevention plans were effective in reducing the mean amount of a claim? Use the 0.05 significance level. **(0.7 points)**

Step 1: State the Hypothesis

H_0: $\mu \geq 5000$ H_1: $\mu < 5000$

Step 2: Select a Level of Significance

$\alpha = 0.5$

Step3: Select the test statistic.

We'll use T test.

Step4: Formulate the decision rule

Reject H_o if $t < -1.652$

Step5: Take sample, compute the test statistic, make decision.
the sample statistic :
t -2.176, found by $\frac{4800-5000}{1300/\sqrt{200}}$
C.I. for populations means:
$4800+1.652\ 1300/\sqrt{200}$=(-∞, 4951.86)
Step6: Interpret the result.
Reject H_o and conclude that the plans were effective.

4. (Text, p.345) The owners of the Westfield Mall wished to study customer shopping habits. From earlier studies, the owners were under the impression that a typical shopper spends 0.75 hour at the mall, with a standard deviation of 0.10 hour. Recently the mall owners added some specialty restaurants designed to keep shoppers in the mall longer. The consulting firm, Brunner and Swanson Marketing Enterprises, was hired to evaluate the effects of the restaurants. A sample of 45 shoppers by Brunner and Swanson revealed that the mean time spent in the mall had increased to 0.80 hour. Develop a hypothesis test to determine if the mean time spent in the mall changed. Use the 0.10 significance level. **(0.8 points)**

H_o: $\mu = 0.75$ H_1: $\mu \neq 0.75$
Reject H_o if $z > 1.645$ or $z<-1.645$
$z=\frac{0.8-0.75}{0.1/\sqrt{45}}=3.35$
Reject H_o. The mean time spent in the mall is more than 0.75 hours.

5.五年前的一次農業普查中，某一村里有 20%的家庭屬於農戶。今欲了解此比例是否已改變，故隨機抽取 400 戶為樣本，發現其中有 70 戶為有實際從事農業的農家。根據此項調查，請以兩種方法：(1)檢定統計量法、(2)信賴區間法檢定並**解釋**兩種統計數值的意義，說明目前此村里屬於農戶的百分比是否已不同於五年前（顯著水準 α=0.05）？ **(1.2 points)**

Step 1: State the Hypothesis
H_0: $\pi = 0.2$ H_1: $\pi \neq 0.2$
Step 2: Select a Level of Significance
$\alpha = 0.05$
Step3: Select the test statistic.
Use the z-statistic as the binomial distribution can be approximated by the normal distribution as $n\pi = 80 > 5$ and $n(1-\pi) = 320 > 5$.
Step4: Formulate the decision rule
Reject H_0 if $z > 1.96$, $z < -1.96$
Step5: Take sample, compute the test statistic, make decision.
the sample statistic : -1.25 > -1.96→統計值未落入拒絕域
found by $z=\frac{0.175-0.2}{\sqrt{\frac{(0.8)(0.2)}{400}}}$
C.I. for populations means: (0.1358, 0.2142)→在信賴區間中包含母體平均（0.2）
found by
$$\left(\frac{70}{400}\right) \pm 1.96\sqrt{\frac{(0.8)(0.2)}{400}}$$
Step6: Interpret the result.
在 0.05 的顯著水準下，無法拒絕虛無假設，此村里屬於農戶的百分比可能和五年前相同