HOMEWORK SUMMARY - MAS291

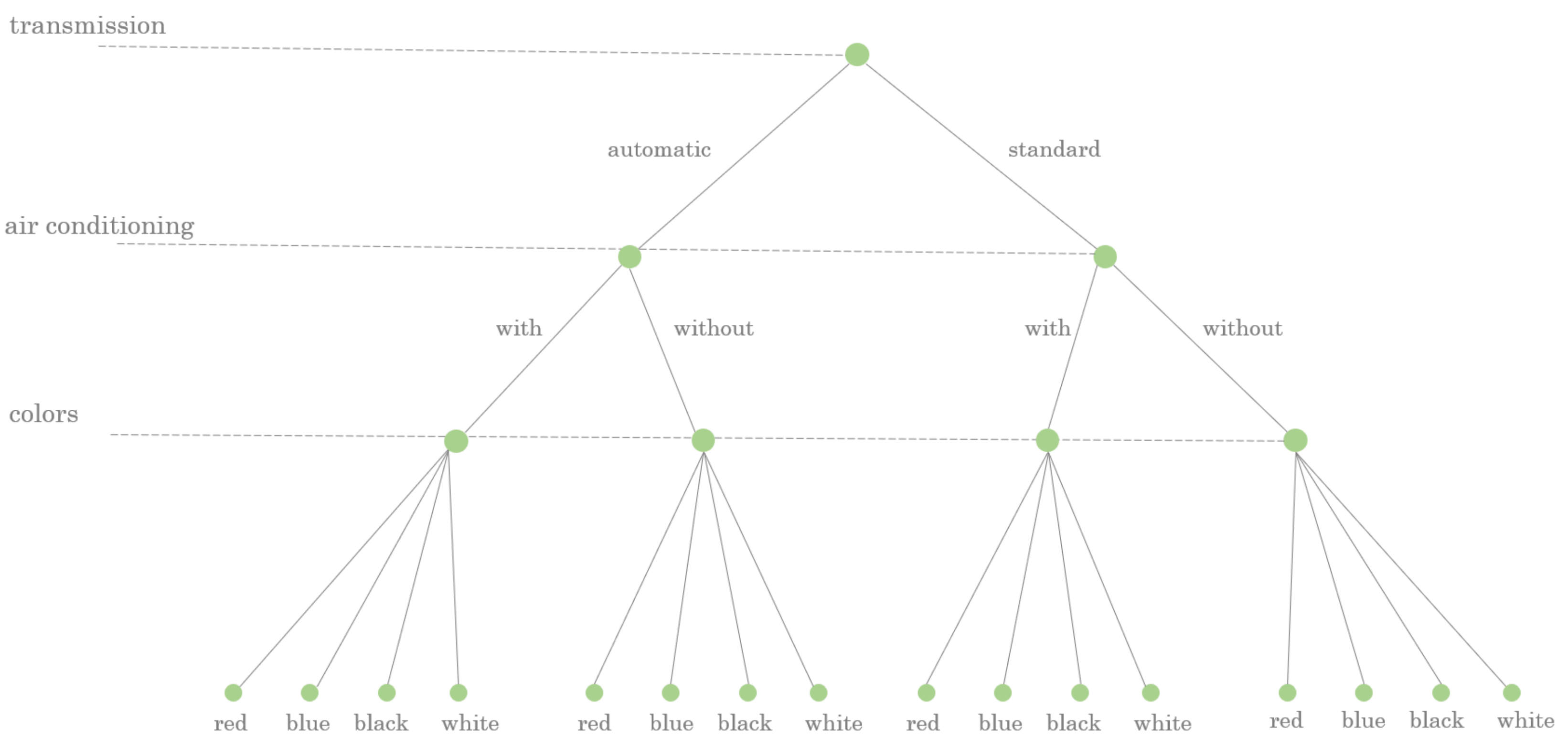
Nguyen Dang Loc - SE160199

Chap 2. Probability

x = 6:
+ (19x+39) mod 139 =14
+ (27x+57) mod 139 = 80
+ (33x+53) mod 139 = 112

2-14

An order for an automobile can specify either an automatic or a standard transmission, either with or without air conditioning, and with any one of the four colors red, blue, black, or white.
Describe the set of possible orders for this experiment.



A - automatic transmission	R - red
S - standard transmission	B - blue
W - with air conditioning	X - black
O - without air conditioning	F - white

Possible orders:

$$S = \{AWR, AWB, AWX, AWF, AOR, AOB, AOX, AOF, \\ SWR, SWB, SWX, SWF, SOR, SOB, SOX, SOF\}$$

2-80

Suppose that a patient is selected randomly from the those described in Exercise 2-57.

	Complete Response	Total
Ribavirin plus interferon alfa	16	21
Interferon alfa	6	19
Untreated controls	0	20

Let A denote the event that the patient is in the group treated with interferon alfa, and let B denote the event that the patient has a complete response. Determine the following probabilities.

$a) P(A)$ $\quad$ $d) P(A \cup B)$

$b) P(B)$ $\quad$ $e) P(A' \cup B)$

$c) P(A \cap B)$

Solve:

$|S| = 60$

a) $P(A) = 19/60 \approx 0.317$

b) $P(B) = 22/60 \approx 0.367$

c)
$A \cap B$: Patient who is treated with interferon alfa and has a complete response
$P(A \cap B) = 6/60 = 0.1$

d)
$A \cup B$: Patient who is treated with interferon alfa or has a complete response
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$= 19/60 + 22/60 - 6/60 \approx 0.583$

e)
$A' \cup B$: Patient who isn't treated with interferon alfa or has a complete response
$A' \cap B$: Patient who isn't treated with interferon alfa and has a complete response
According to the table: $|A' \cap B| = 16$

$P(A' \cap B) = 16/60$

$P(A' \cup B) = P(A') + P(B) - P(A' \cap B)$

$= 1 - P(A) + P(B) - P(A' \cap B)$

$= (1 - 19/60) + 22/60 - 16/60$

≈ 0.783

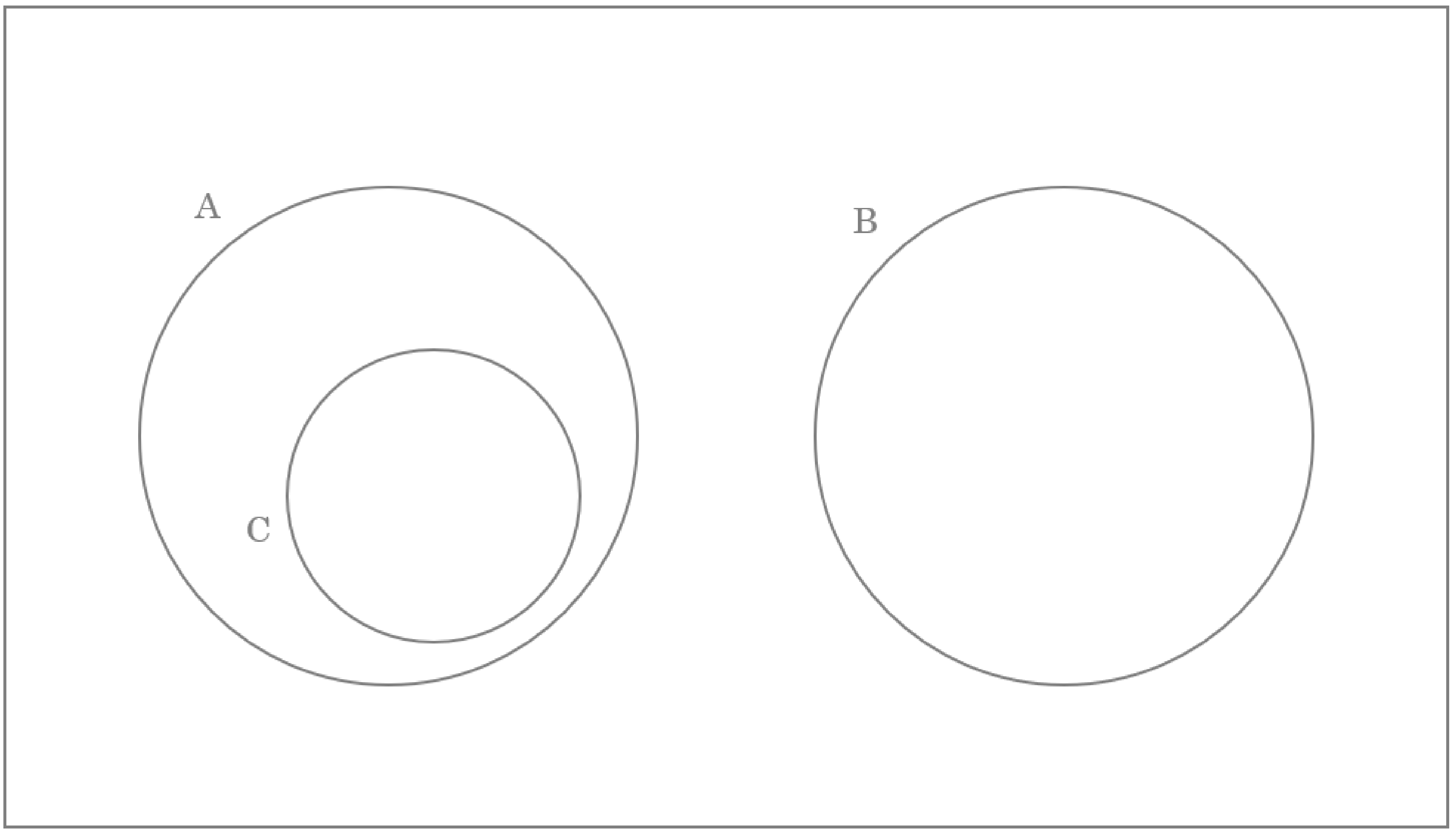
2-112

Suppose A and B are mutually exclusive events. Construct a Venn diagram that contains the three events A, B and C such that P(A | C) = 1 and P(B | C) = 0.

Solve:

$P(A|C) = P(A \cap C)/P(C) = 1$
$\Rightarrow P(A \cap C) = P(C)$
$\Rightarrow A \cap C = C$

$P(B|C) = P(B \cap C)/P(C) = 0$
$\Rightarrow B \cap C = \emptyset$



Chap 3. Discrete Random Variables and Probability Distributions

x = 6:
+ (13x+17) mod 160 = 95
+ (23x+27) mod 160 = 5
+ (33x+37) mod 160 = 75

3-5

Determine the range (possible values) of the random variable.
A batch of 500 machined parts contains 10 that do not conform to customer requirements. Parts are selected successively, without replacement, until a nonconforming part is obtained. The random variable is the number of parts selected.

Solution:

500 machined parts contains:
+ 10 non-conforming parts
+ 490 conforming parts

Let X denote the random variable → X is the number of parts selected

- $X = 1$ → The first part selected is non-conforming
- $X = 2$ → The second part selected is non-conforming
- ...
- $X = 491$ → The 491^{th} part selected is non-conforming
  (All 490 parts selected earlier are conforming)

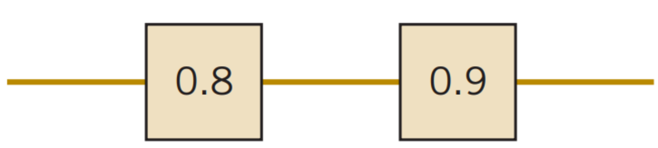
Hence, range of X is $\{1, 2, ..., 491\}$

3-75

Calculate the mean for the random variable in Exercise 3-37.

(3-37) Consider the circuit in Example 2-32. Assume that devices fail independently. What is the probability mass function of the number of failed devices?

(Example 2-32) The following circuit operates only if there is a path of functional devices from left to right. The probability that each device functions is shown on the graph. Assume that devices fail independently.



Solution:

The number of failed devices in total of 2 devices is the random varible.

Let X denote the random variable. Range of X is $\{0, 1, 2\}$

Calculate the PMF:

- $f(0) = P(X = 0) = P(x_1)P(x_2) = 0.8 \times 0.9 = 0.72$
- $f(2) = P(X = 2) = P(x_1{}')P(x_2{}') = (1 - P(x_1))(1 - P(x_2)) = 0.2 \times 0.1 = 0.02$
- $f(1) = P(X = 1) = 1 - P(X = 0) - P(X = 2) = 1 - 0.72 - 0.02 = 0.26$

Mean of the random variable:

$$\mu = E(X) = 0 \times f(0) + 1 \times f(1) + 2 \times f(2)$$
$$= 0 \times 0.72 + 1 \times 0.26 + 2 \times 0.02$$
$$= 0.3$$

3-95

The random variable X has a binomial distribution with n = 10 and p = 0.01. Determine the following probabilities.

(a) $P(X = 5)$
(b) $P(X \leqslant 2)$
(c) $P(X \geqslant 9)$
(d) $P(3 \leqslant X < 5)$

Solution:

$n = 10,\ p = 0.01$

(a) $P(X=5) = \binom{10}{5} 0.01^5 (1-0.01)^{10-5} \approx 2.39 \times 10^{-8}$

(b) $P(X \leqslant 2) = P(0) + P(1) + P(2)$

$$= \binom{10}{0} 0.01^0 (1-0.01)^{10} + \binom{10}{1} 0.01^1 (1-0.01)^9 + \binom{10}{2} 0.01^2 (1-0.01)^8$$

$$\approx 0.9999$$

(c) $P(X \geqslant 9) = P(9) + P(10)$

$$= \binom{10}{9} 0.01^9 (1-0.01)^1 + \binom{10}{10} 0.01^{10} (1-0.01)^0$$

$$= 9.91 \times 10^{-18}$$

(d) $P(3 \leqslant X < 5) = P(3) + P(4)$

$$= \binom{10}{3} 0.01^3 (1-0.01)^7 + \binom{10}{4} 0.01^4 (1-0.01)^6$$

$$\approx 1.138 \times 10^{-4}$$

Chap 4. Continuous Random Variables and Probability Distributions

x = 6:
+ (5x + 19) mod 109 = 49
+ (7x + 29) mod 109 = 71
+ (9x + 39) mod 109 = 93

4-49

Integration by parts is required. The probability density function for the diameter of a drilled hole in millimeters is $10e^{-10(x-5)}$ for $x > 5mm$. Although the target diameter is 5 millimeters, vibrations, tool wear, and other nuisances produce diameters greater than 5 millimeters.

(a) Determine the mean and variance of the diameter of the holes.

(b) Determine the probability that a diameter exceeds 5.1 millimeters.

Solution:

The diameter of the drilled hole is the random variable, denote as X.

The probability density function of random variable X is: $f(x) = 10e^{-10(x-5)}, x > 5$

a)

- $E(X) = \int_5^\infty x10e^{-10(x-5)}dx \ = e^{50} \int_5^\infty x10e^{-10x}dx$

Let $u = -e^{-10x} \Rightarrow du = 10e^{-10x}dx,$ then:

$$E(X) = e^{50} \int_5^\infty xdu$$
$$= e^{50}\Big(ux\Big|_5^\infty - \int_5^\infty udx\Big)$$
$$= e^{50}\Big(-e^{-10x}x\Big|_5^\infty - \int_5^\infty -e^{-10x}dx\Big)$$
$$= e^{50}\Big(0 + 5e^{-50}\Big) + e^{50}\int_5^\infty e^{-10x}dx$$
$$= 5 + e^{50}\tfrac{-1}{10}e^{-10x}\Big|_5^\infty$$
$$= 5 + \tfrac{1}{10}$$
$$= 5.1$$

- $E(X^2) = \int_5^\infty x^2 10e^{-10(x-5)}dx \ = e^{50}\int_5^\infty x^2 10e^{-10x}dx$

Let $u = -e^{-10x} \Rightarrow du = 10e^{-10x}dx,\ v = x^2 \Rightarrow dv = 2xdx$

$$
\begin{aligned}
E(X^2) &= e^{50}\int_5^\infty vdu \\
&= e^{50}\Big(uv\Big|_5^\infty - \int_5^\infty udv\Big) \\
&= e^{50}\Big(-e^{-10x}x^2\Big|_5^\infty - \int_5^\infty -e^{-10x}2xdx\Big) \\
&= e^{50}\Big(-e^{-10x}x^2\Big|_5^\infty\Big) + \tfrac{2}{10}e^{50}\int_5^\infty 10e^{-10x}xdx \\
&= e^{50}\Big(0 + 25e^{-50}\Big) + \tfrac{2}{10}E(X) \\
&= 25 + 0.2 \times 5.1 \\
&= 26.02
\end{aligned}
$$

- $V(X) = E(X^2) - E(X)^2 = 26.02 - 5.1^2 = 0.01$

b)

$$P(X > 5.1) = \int_{5.1}^\infty 10e^{-10(x-5)}dx \ = e^{50}\int_{5.1}^\infty 10e^{-10x}dx \ = e^{50}(-e^{-10x}\Big|_{5.1}^\infty) = e^{50} \times e^{-51} = e^{-1} \approx 0.36788$$

4-71

The compressive strength of samples of cement can be modeled by a **normal distribution** with a mean of 6000 kilograms per square centimeter and a standard deviation of 100 kilograms per square centimeter.
(a) What is the probability that a sample's strength is less than $6250 kg/cm^2$?
(b) What is the probability that a sample's strength is between 5800 and $5900 kg/cm^2$?
(c) What strength is exceeded by 95% of the samples?

Solution: $\mu = 6000,\ \sigma = 100$

Let X denote the normal random variable (the compressive strength of samples of cement). Standardize: $Z = \dfrac{X - 6000}{100}$

a)

$$P(X < 6250) = P(Z < \frac{6250 - 6000}{100}) = P(Z < 2.5) = 0.993790$$

b)

$$
\begin{aligned}
P(5800 < X < 5900) &= P(\frac{5800 - 6000}{100} < Z < \frac{5900 - 6000}{100}) \\
&= P(-2 < Z < -1) \\
&= P(Z < -1) - P(Z < -2) \\
&= 0.158655 - 0.022750 \\
&= 0.135905
\end{aligned}
$$

c)

$$P(X > x) = 0.95 \Leftrightarrow P(Z > \frac{x - 6000}{100}) = 0.95$$
$$\Rightarrow P(Z \leqslant \frac{x - 6000}{100}) = 0.05 \Rightarrow \frac{x - 6000}{100} \approx -1.64 \Rightarrow x \approx 5836$$

4-93

An article in *International Journal of Electrical Power & Energy Systems* ["Stochastic Optimal Load Flow Using a Combined Quasi–Newton and Conjugate Gradient Technique" (1989, Vol.11(2), pp. 85–93)] considered the problem of optimal power low in electric power systems and included the effects of uncertain variables in the problem formulation. The method treats the system power demand as a normal random variable with 0 mean and unit variance.
(a) What is the power demand value exceeded with 95% probability?
(b) What is the probability that the power demand is positive?
(c) What is the probability that the power demand is more than – 1 and less than 1?

Solution:

Let X be power demand with mean = 0, variance = 1 $\Rightarrow$ X is standard normal distribution a)

$$P(X > x) = 0.95 \Rightarrow P(X \leqslant x) = 1 - P(X > x) = 1 - 0.95 = 0.05$$

$$\Rightarrow x \approx -1.64$$

b)

$$P(X > 0) = 1 - P(X \leqslant 0) = 1 - 0.5 = 0.5$$

c)

$$P(-1 < X < 1) = P(X < 1) - P(X < -1) = 0.841345 - 0.158655 = 0.68269$$

Chap 6. Descriptive Statistics

x = 6:

+ (13x + 17) mod 87 = 8
+ (11x + 21) mod 87 = 0 (50)
+ (9x + 29) mod 87 = 83

6-8

In Applied Life Data Analysis (Wiley, 1982), Wayne Nelson presents the breakdown time of an insulating fluid between electrodes at 34 kV. The times, in minutes, are as follows: 0.19, 0.78, 0.96, 1.31, 2.78, 3.16, 4.15, 4.67, 4.85, 6.50, 7.35, 8.01, 8.27, 12.06, 31.75, 32.52, 33.91, 36.71, and 72.89. Calculate the sample mean and sample standard deviation

Solution:

Sample mean:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum_{i=1}^{19} x_i}{19} = \frac{272.82}{19} \approx \boldsymbol{14.359\ min}$$

Sample variance:

$$\sum_{i=1}^{19} x_i = 272.82$$

$$\sum_{i=1}^{19} x_i^2 = 10333.9$$

$$s^2 = \frac{\sum_{i=1}^{19} x_i^2 - \frac{(\sum_{i=1}^{19} x_i)^2}{19}}{19-1} = \frac{10333.9 - \frac{272.82^2}{19}}{18} \approx 356.4718$$

Sample standard deviation:

$$s = \sqrt{356.471} \approx \boldsymbol{18.88\ min}$$

Use R:

```
x <- c(0.19, 0.78, 0.96, 1.31, 2.78, 3.16, 4.15, 4.67, 4.85, 6.50,
       7.35, 8.01, 8.27, 12.06, 31.75, 32.52, 33.91, 36.71, 72.89)
mean(x) // 14.35895
sd(x) // 18.88
```

6-50

Construct frequency distributions and histograms with 8 bins and 16 bins for the motor fuel octane data in Exercise 6-30. Compare the histograms. Do both histograms display similar information?

(6-30) An article in Technometrics (1977, Vol. 19, p. 425) presented the following data on the motor fuel octane ratings of several blends of gasoline:

88.5	98.8	89.6	92.2	92.7	88.4	87.5	90.9
94.7	88.3	90.4	83.4	87.9	92.6	87.8	89.9
84.3	90.4	91.6	91.0	93.0	93.7	88.3	91.8
90.1	91.2	90.7	88.2	94.4	96.5	89.2	89.7
89.0	90.6	88.6	88.5	90.4	84.3	92.3	92.2
89.8	92.2	88.3	93.3	91.2	93.2	88.9	
91.6	87.7	94.2	87.4	86.7	88.6	89.8	
90.3	91.1	85.3	91.1	94.2	88.7	92.7	

90.0	86.7	90.1	90.5	90.8	92.7	93.3	
91.5	93.4	89.3	100.3	90.1	89.3	86.7	
89.9	96.1	91.1	87.6	91.8	91.0	91.0	

Solution:

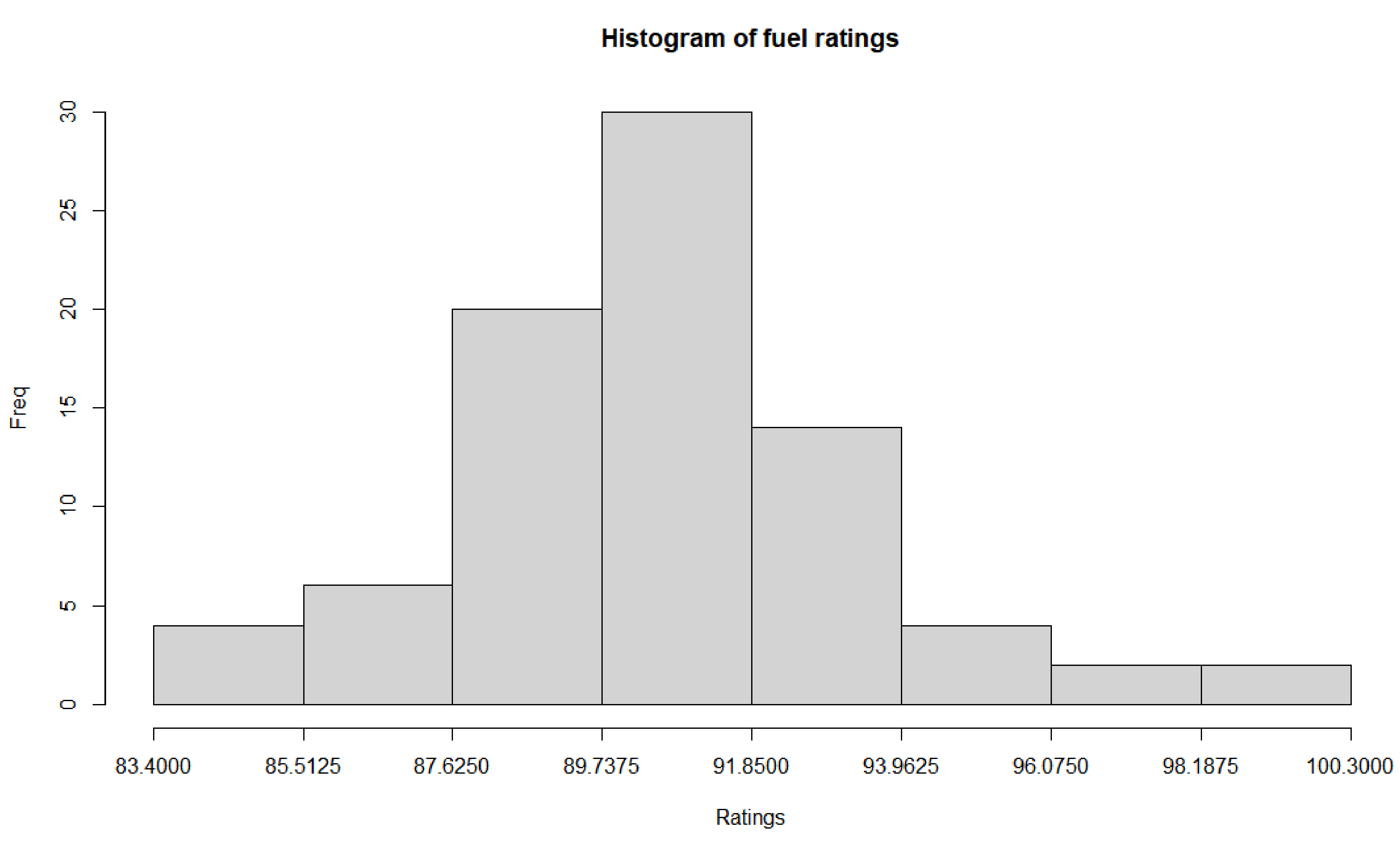
8 bins:

Construct frequency distributions: create 8 intervals of equal width

```
data <- read.table("D://MAS//650.txt", sep="", header=F,
                   na.strings="", stringsAsFactors=F) # read data
dt <- as.numeric(data) # convert to list of numeric variable
d <- (max(dt) - min(dt)) / 8
bins <- seq(min(dt), max(dt), by=d) # list of boundaries
fd <- cut(dt, bins) # group data into bins
transform(table(fd))
# Result
           fd Freq
1 (83.4,85.5]    3
2 (85.5,87.6]    6
3 (87.6,89.7]   20
4 (89.7,91.8]   30
5   (91.8,94]   14
6   (94,96.1]    4
7 (96.1,98.2]    2
8  (98.2,100]    2
```

Construct histogram

```
hist(dt, xlab = "Ratings", ylab = "Freq", breaks = seq(min(dt), max(dt),
     length.out = 9), xaxt='n', main="Histogram of fuel ratings")
axis(1, at = seq(min(dt), max(dt))
```



16 bins:

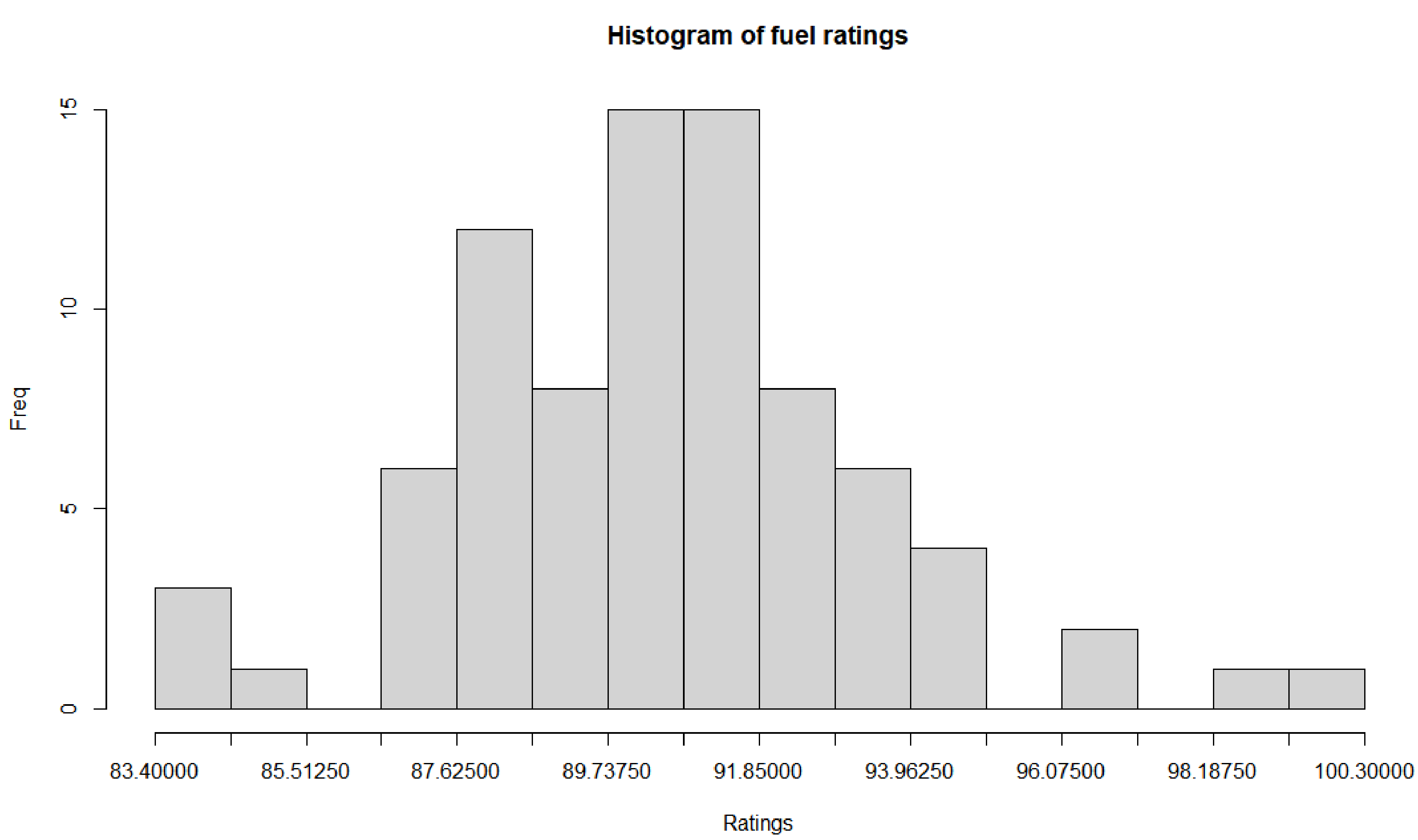
Construct frequency distributions: create 16 intervals of equal width

```
data <- read.table("D://MAS//650.txt", sep="", header=F,
                   na.strings="", stringsAsFactors=F) # read data
dt <- as.numeric(data) # convert to list of numeric variable
d <- (max(dt) - min(dt)) / 16
bins <- seq(min(dt), max(dt), by=d)
fd <- cut(dt, bins)
transform(table(fd))
# Result
```

```
            fd Freq
1  (83.4,84.5]    2
2  (84.5,85.5]    1
3  (85.5,86.6]    0
4  (86.6,87.6]    6
5  (87.6,88.7]   12
6  (88.7,89.7]    8
7  (89.7,90.8]   15
8  (90.8,91.8]   15
9  (91.8,92.9]    8
10   (92.9,94]    6
11     (94,95]    4
12   (95,96.1]    0
13 (96.1,97.1]    2
14 (97.1,98.2]    0
15 (98.2,99.2]    1
16  (99.2,100]    1
```

Construct histogram

```
hist(dt, xlab = "Ratings", ylab = "Freq", breaks = seq(min(dt), max(dt),
     length.out = 17), xaxt='n', main="Histogram of fuel ratings")
axis(1, at = seq(min(dt), max(dt), by=d))
```



The shapes of the 2 histograms are quite the same. The two histograms are likely to display similar information.

6-83

The pull-off force for a connector is measured in a laboratory test. Data for 40 test specimens follow (read down, then left to right). Construct and interpret either a digidot plot or a separate stem-and-leaf and time series plot of the data.

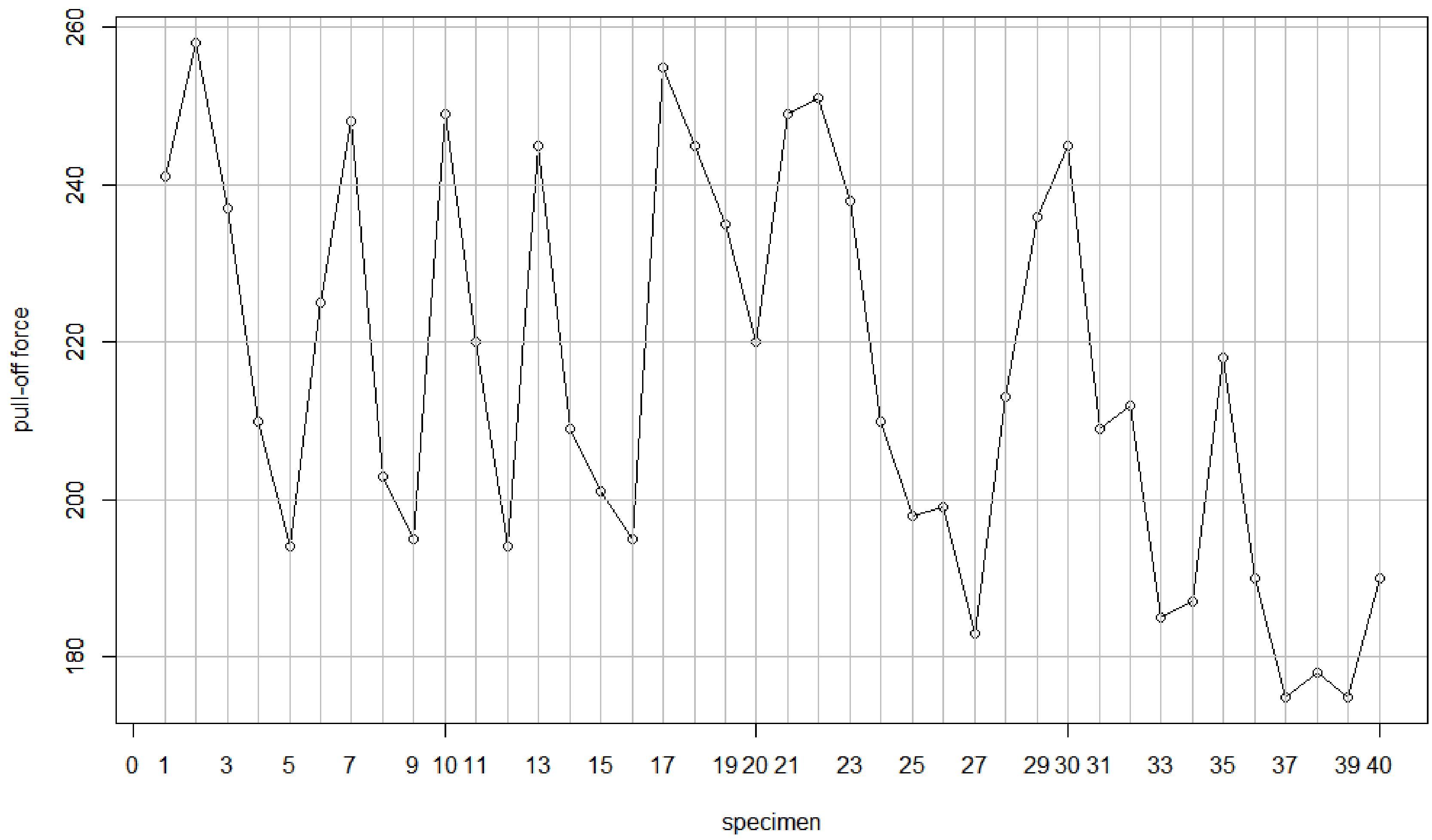
Solution:

Time series plot:

The specimen is on y-axis and the pull-off force is on the x-axis.

```
data <- c(241,258,237,210,194,225,248,203,195,249,220,194,245,
           209,201,195,255,245,235,220,249,251,238,210,198,199,
           183,213,236,245,209,212,185,187,218,190,175,178,175,190)

xValue <- 1:length(data)
yValue <- data
plot(data, type='o',xlab='specimen', ylab='pull-off force')
axis(1, at = seq(1, 40, by = 1), tck = 1, lty = 1, col = "gray")
axis(2, at = seq(min(yValue), max(yValue), by = 20), tck = 1, lty = 1, col = "gray")
```



Stem-and-leaf plot

```
stem(data)
  The decimal point is 1 digit(s) to the right of the |
  17 | 558
  18 | 357
  19 | 00445589
  20 | 1399
  21 | 00238
  22 | 005
  23 | 5678
  24 | 1555899
  25 | 158
```

In the time-series plot, it does not likely to present upwards trend or downwards trend

→ there is no obvious pattern or conclusion drawn from the plot.

Chap 7. Point Estimation of Parameters and Sampling Distributions

x = 6:

+ (3x + 13) mod 21 = 10
+ (5x + 15) mod 21 = 3
+ (7x + 17) mod 21 = 17

7-10

Suppose that the random variable X has the continuous uniform distribution

$$f(x) = \begin{cases} 1, & 0 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

Suppose that a random sample of n = 12 observations is selected from this distribution. What is the approximate probability distribution of $\overline{X} - 6$? Find the mean and variance of this quantity.

Solution:

Mean and variance of continuous unifrom distribution:

$$\mu = E(X) = \frac{0+1}{2} = 0.5,$$

$$\sigma^2 = V(X) = \frac{(1-0)^2}{12} = \frac{1}{12} \approx 0.0833$$

The sampling distribution of the sample mean $\overline{X}$ is approximately normal with

+ Mean: $\mu_{\overline{X}} = E(\overline{X}) = \mu$
+ Variance: $\sigma^2_{\overline{X}} = V(\overline{X}) = \dfrac{\sigma^2}{n}$

Mean and variance of the probability distribution $\overline{X} - 6$:

$$E(\overline{X} - 6) = E(\overline{X}) - 6 = 0.5 - 6 = -5.5$$

$$V(\overline{X} - 6) = V(\overline{X}) = \frac{\sigma^2}{n} = \frac{1/12}{12} \approx 0.00694$$

7-3

PVC pipe is manufactured with a mean diameter of 1.01 inch and a standard deviation of 0.003 inch. Find the probability that a random sample of n = 9 sections of pipe will have a sample mean diameter greater than 1.009 inch and less than 1.012 inch.

Solution:

Let $X_1, X_2, ..., X_9$ is a random sample of size $n = 9$ of pipe's diameter with
$\mu = 1.01,\ \sigma = 0.003$

The sampling distribution of $\overline{X}$ is normal with

+ Mean $\mu_{\overline{X}} = \mu = 1.009$,
+ Standard deviation $\sigma_{\overline{X}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{0.003}{\sqrt{9}} = 0.001$

Then:

$$\begin{aligned} P(1.009 < \overline{X} < 1.012) &= P(\frac{1.009 - 1.01}{0.003/\sqrt{9}} < Z < \frac{1.012 - 1.01}{0.003/\sqrt{9}}) \\ &= P(-1 < Z < 2) \\ &\approx 0.8185946 \end{aligned}$$

7-17

From the data in Exercise 6-21 on the pH of rain in Ingham County, Michigan:

5.47	5.37	5.38	4.63	5.37	3.74	3.71	4.96	4.64	5.11	5.65
5.39	4.16	5.62	4.57	4.64	5.48	4.57	4.57	4.51	4.86	4.56
4.61	4.32	3.98	5.70	4.15	3.98	5.65	3.10	5.04	4.62	4.51
4.34	4.16	4.64	5.12	3.71	4.64					

What proportion of the samples has pH below 5.0?

Solution:

A sample size $n = 39$, a subsample size $x = 26$ which contain values less than 5.

The proportion of the sample has pH below 5.0 is: $P(X < 5.0) = \dfrac{x}{n} = \dfrac{26}{39} \approx 0.667$

About 67% of the sample has pH below 5.0.

Use R:

```
data <- read.table("D://MAS//717.txt", sep="", header=F,
                   na.strings="", stringsAsFactors=F) # read data
dt <- as.numeric(data)
res <- length(dt[dt<5])/length(dt) # 0.6666667
```

Chap 8. Statistical Intervals for a Single Sample

x = 6:

+ (3x + 17) mod 26 + 46 = 55
+ (5x + 17) mod 26 + 46 = 67
+ (7x + 17) mod 26 + 46 = 53

8-55

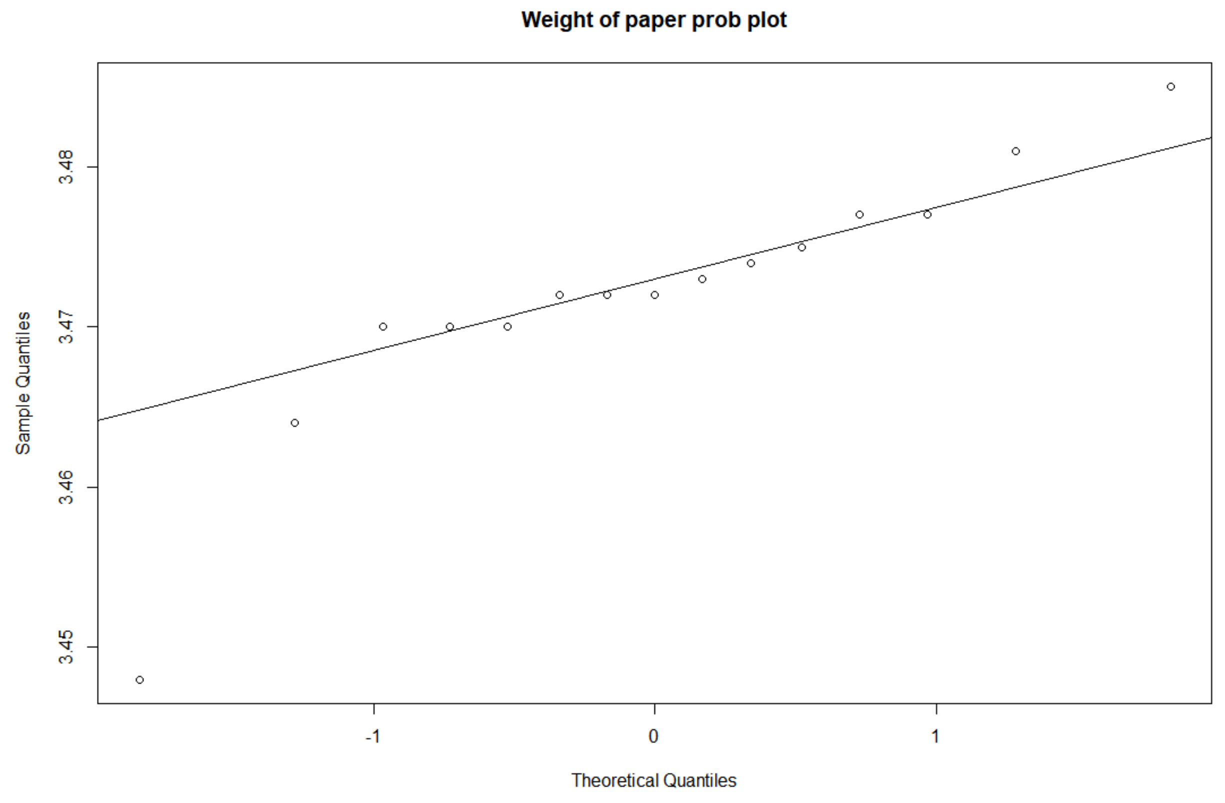
An article in Technometrics (1999, Vol. 41, pp. 202–211) studied the capability of a gauge by measuring the weight of paper. The data for repeated measurements of one sheet of paper are in the following table. Construct a 95% one-sided upper confidence interval for the standard deviation of these measurements. Check the assumption of normality of the data and comment on the assumptions for the confidence interval.

Observations
3.481, 3.448, 3.485, 3.475, 3.472,
3.477, 3.472, 3.464, 3.472, 3.470,
3.470, 3.470, 3.477, 3.473, 3.474

Solution:

Code in R

```
# plot
dt <- c(3.481, 3.448, 3.485, 3.475, 3.472,
        3.477, 3.472, 3.464, 3.472, 3.470,
        3.470, 3.470, 3.477, 3.473, 3.474)
qqnorm(dt, main = 'Weight of paper prob plot')
qqline(dt)
```



The pattern of normal probability plot is roughly linear and close to the line, we can say that the the distribution of observed values is approximately normal.

Sample size $n = 15$, the sample mean and sample standard deviation:

$$\overline{x} = \frac{\sum_{i=1}^{15} x_i}{15} = 3.472$$

$$s = \frac{\sum_{i=1}^{1} 5(x_i - \overline{x})}{14} = 0.0083$$

The CI is 95% $\Rightarrow 1 - \alpha = 0.95 \Rightarrow \chi^2_{1-\alpha,n-1} = \chi^2_{0.975,7} = 6.57$

A 95% one-sided upper confidence interval for σ^2 is:

$$\sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{1-\alpha,n-1}} \Rightarrow \sigma^2 \leq \frac{14 \times 0.0083^2}{6.57} \approx 0.000147$$

$$\Rightarrow \sigma \leq 0.01212$$

8-67

The U.S. Postal Service (USPS) has used optical character recognition (OCR) since the mid-1960s. In 1983, USPS began deploying the technology to major post ofices throughout the country (www.britannica.com). Suppose that in a random sample of 500 handwritten zip code digits, 466 were read correctly.

(a) Construct a 95% confidence interval for the true proportion of correct digits that can be automatically read.

(b) What sample size is needed to reduce the margin of error to 1%?

(c) How would the answer to part (b) change if you had to assume that the machine read only one-half of the digits correctly?

Solution:

A point estimate of p: $\hat{p} = \frac{x}{n} = \frac{466}{500} = 0.932$

$np = 500 \times 0.932 = 466 > 5$

$n(1-p) = 500 \times (1 - 0.932) = 34 > 5$

a)

$\frac{\alpha}{2} = \frac{1 - 0.95}{2} = 0.025 \Rightarrow z_{0.025} = 1.96$

A 95% confidence interval of the true proportion:

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$\Leftrightarrow 0.932 - 1.96\sqrt{\frac{0.932(1-0.932)}{500}} \leq p \leq 0.932 + 1.96\sqrt{\frac{0.932(1-0.932)}{500}}$$

$$\Leftrightarrow 0.9099 \leq p \leq 0.9541$$

b)

Error $E = 0.01$

$$n \geq \left(\frac{z_{\alpha/2}}{E}\right)^2 p(1-p) = \left(\frac{1.96}{0.01}\right)^2 \times 0.932 \times (1 - 0.932) = 2434.65$$

Sample size required $n = 2435$

c)

$p = 0.5 \Rightarrow p(1-p) = 0.25$

$$n \geq \left(\frac{z_{\alpha/2}}{E}\right)^2 p(1-p) = \left(\frac{1.96}{0.01}\right)^2 \times 0.25 = 9604$$

Sample size required $n = 9604$

8-53

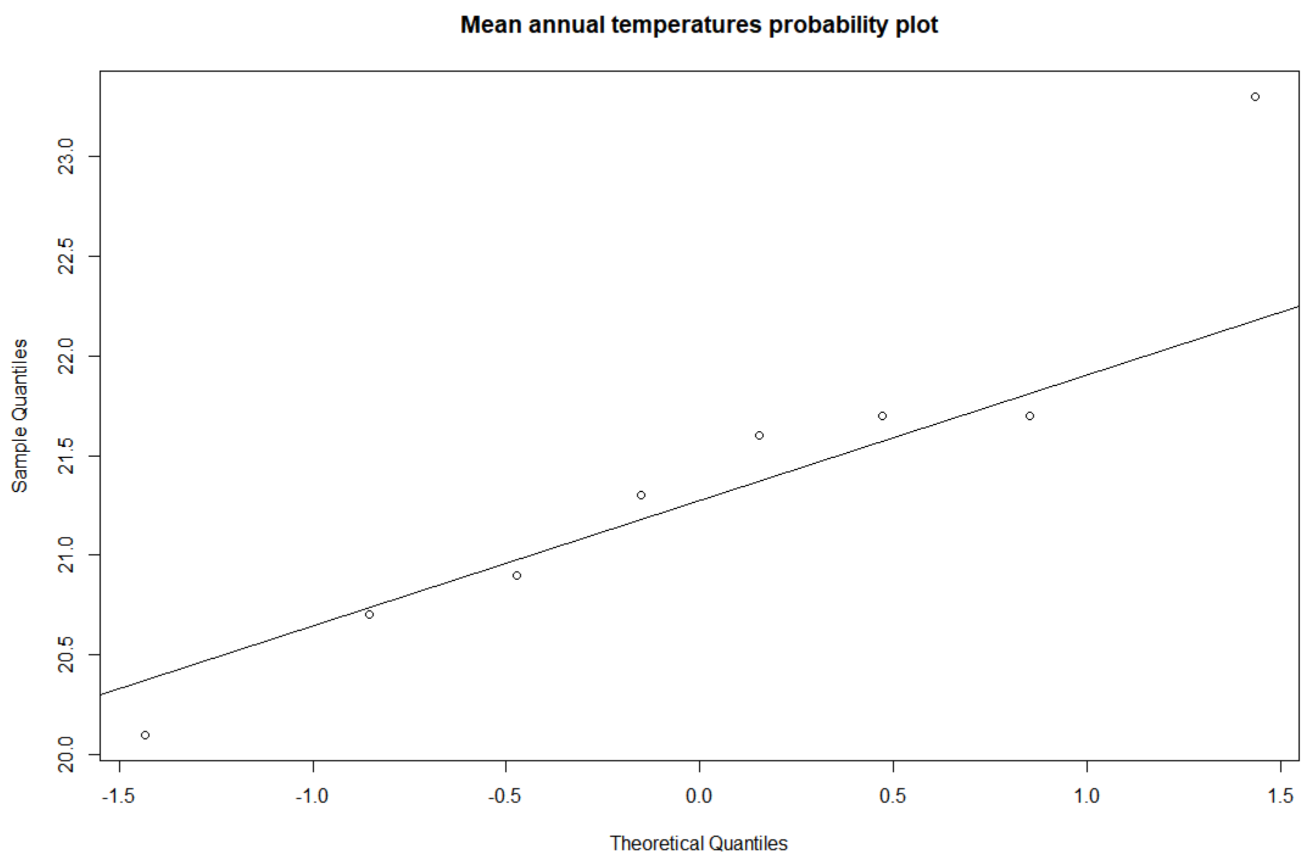
An article in Urban Ecosystems, “Urbanization and Warming of Phoenix (Arizona, USA): Impacts, Feedbacks and Mitigation” (2002, Vol. 6, pp. 183–203), mentions that Phoenix is ideal to study the effects of an urban heat island because it has grown from a population of 300,000 to approximately 3 million over the last 50 years, which is a period with a continuous, detailed climate record. The 50-year averages of the mean annual temperatures at eight sites in Phoenix follow. Check the assumption of normality in the population with a probability plot. Construct a 95% confidence interval for the standard deviation over the sites of the mean annual temperatures.

Site	Average Mean Temperature (°C)
Sky Harbor Airport	23.3
Phoenix Greenway	21.7
Phoenix Encanto	21.6
Waddell	21.7
Litchi eld	21.3
Laveen	20.7
Maricopa	20.9
Harlquahala	20.1

Solution:

Code in R

```
# plot
dt <- c(23.3, 21.7, 21.6, 21.7, 21.3, 20.7, 20.9, 20.1)
qqnorm(dt, main = 'Mean annual temperatures probability plot')
qqline(dt)
```



The pattern of normal probability plot is roughly linear and close to the line, we can say that the the distribution of observed values is approximately normal.

Sample size $n = 8$, the sample mean and sample standard deviation:

$$\overline{x} = \frac{\sum_{i=1}^{8} x_i}{8} = 21.4125$$

$$s = \frac{\sum_{i=1}^{8}(x_i - \overline{x})}{7} = 0.8955$$

The CI is 95% ⇒ $1 - \alpha = 0.95 \Leftrightarrow \alpha = 0.05$

$$\frac{\alpha}{2} = 0.025 \Rightarrow \chi^2_{\alpha/2,n-1} = \chi^2_{0.025,7} = 16.01$$

$$1 - \frac{\alpha}{2} = 0.975 \Rightarrow \chi^2_{1-\frac{\alpha}{2},n-1} = \chi^2_{0.975,7} = 1.69$$

A 95% confidence interval for σ^2 is

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}} \Leftrightarrow \frac{7 \times 0.8955}{16.01} \le \sigma^2 \le \frac{7 \times 0.8955}{1.69}$$

$$\Leftrightarrow 0.392 \le \sigma^2 \le 3.709$$

$$\Leftrightarrow 0.626 \le \sigma \le 1.926$$

Chap 9. Tests of Hypotheses for a Single Sample

x = 6:

+ (3x + 7) mod 103 = 25
+ (5x + 9) mod 103 = 39
+ (7x + 11) mod 103= 53

9-25

The proportion of adults living in Tempe, Arizona, who are college graduates is estimated to be p = 0.4. To test this hypothesis, a random sample of 15 Tempe adults is selected. If the number of college graduates is between 4 and 8, the hypothesis will be accepted; otherwise, you will conclude that p ≠ 0.4.

(a) Find the type I error probability for this procedure, assuming that p = 0.4.

(b) Find the probability of committing a type II error if the true proportion is really p = 0.2.

Solution:

a)

$$\alpha = P(\overline{x} < 4) + P(\overline{x} > 8),\ p = 0.4$$

$$= P\left(\frac{\overline{x} - np}{\sqrt{np(1-p)}} < \frac{4 - 15 \times 0.4}{\sqrt{15 \times 0.4 \times (1-0.4)}}\right) + P\left(\frac{\overline{x} - np}{\sqrt{np(1-p)}} > \frac{8 - 15 \times 0.4}{\sqrt{15 \times 0.4 \times (1-0.4)}}\right)$$
$$= P(z < -1.054) + P(z > 1.054)$$
$$= 2 \times P(z < -1.054)$$
$$= 0.292$$

b)

$$\alpha = P(4 \leq \overline{x} \leq 8),\ p = 0.2$$
$$= P\left(\frac{4 - 15 \times 0.4}{\sqrt{15 \times 0.2 \times (1-0.2)}} \leq \frac{\overline{x} - np}{\sqrt{np(1-p)}} \leq \frac{8 - 15 \times 0.2}{\sqrt{15 \times 0.2 \times (1-0.2)}}\right)$$
$$= P(0.64 \leq z \leq 3.23)$$
$$= P(z < 3.23) - P(z < 0.64)$$
$$= 0.257$$

9-39

Output from a software package follows:

One-Sample Z:
Test of mu = 20 vs mu > 20
The assumed standard deviation = 0.75

Variable	N	Mean	StDev	SE Mean	Z	P
x	10	19.889	?	0.237	?	?

(a) Fill in the missing items. What conclusions would you draw?
(b) Is this a one-sided or a two-sided test?
(c) Use the normal table and the preceding data to construct a 95% two-sided CI on the mean.
(d) What would the P-value be if the alternative hypothesis is $H_1 : \mu \neq 20$?

Solution:

a)

$$\sigma = SE\ Mean \times \sqrt{n} = 0.237 \times \sqrt{10} = 0.7495$$
$$Z = \frac{\overline{x} - \mu}{\sigma/\sqrt{n}} = \frac{19.889 - 20}{0.75/\sqrt{10}} = -0.468$$
$$P = 1 - \Phi(z_0) = 1 - \Phi(-0.468) = 1 - 0.32 = 0.68$$

b)

This is one-sided test because $H_1 : \mu > 20$

c)

A 95% two-sided CI on the mean:

$$\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$
$$\Leftrightarrow 19.889 - 1.96\frac{0.7495}{\sqrt{10}} \leq \mu \leq 19.889 + 1.96\frac{0.7495}{\sqrt{10}}$$
$$\Leftrightarrow 19.4245 \leq \mu \leq 20.354$$

d)

With $H_1 : \mu \neq 20$, P-value is

$$P = 2 \times (1 - \Phi(z_0)) = 2 \times (1 - \Phi(-0.468)) = 2 \times (1 - 0.32) = 1.36$$

9-53

A hypothesis will be used to test that a population mean equals 10 against the alternative that the population mean is greater than 10 with unknown variance. What is the critical value for the test statistic T_0 for the following significance levels?
(a) $\alpha = 0.01$ and $n = 20$
(b) $\alpha = 0.05$ and $n = 12$
(c) $\alpha = 0.10$ and $n = 15$

Solution:

$H_0 : \mu = 10$
$H_1 : \mu > 10$

a) $t_{0.01,19} = 2.539$

a) $t_{0.05,11} = 1.796$

a) $t_{0.10,15} = 1.345$

Chap 10. Statistical Inference for Two Samples

x = 6:
+ (3x + 13) mod 39 = 31
+ (5x + 15) mod 39 = 45 (Wilcoxon rank-sum test...) → 54
+ (x mod 10) + 82 = 88

10-31

An article in Radio Engineering and Electronic Physics [1984, Vol. 29 No. (3), pp. 63–66] investigated the behavior of a stochastic generator in the presence of external noise. The number of periods was measured in a sample of 100 trains for each of two different levels of noise voltage, 100 and 150 mV. For 100 mV, the mean number of periods in a train was 7.9 with
s = 2.6 For 150 mV, the mean was 6.9 with s = 2.4.
(a) It was originally suspected that raising noise voltage would reduce the mean number of periods. Do the data support this claim? Use α= 0.01 and assume that each population is normally distributed and the two population variances are equal. What is the P-value for this test?
(b) Calculate a confidence interval to answer the question in part (a).

Solution:

$n_1 = 100,\ \overline{x}_1 = 7.9,\ s_1 = 2.6$
$n_2 = 100,\ \overline{x}_2 = 6.9,\ s_2 = 2.4$

assumed equal variances

a)

Parameter of interest: the difference in mean number of periods, $\mu_1 - \mu_2$ and $\Delta_0 = 0$

Null hypothesis: $H_0 : \mu_1 = \mu_2$

Alt. hypothesis: $H_1 : \mu_1 > \mu_2$

Test statistic: $t_0 = \dfrac{\overline{x_1} - \overline{x_2} - \Delta_0}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \dfrac{\overline{x_1} - \overline{x_2}}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

Computations:

+ The pooled estimator of σ^2:
$$S_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{(100 - 1)2.6^2 + (100 - 1)2.4^2}{100 + 100 - 2}} = 2.5$$
+ Test statistic: $t_0 = \dfrac{\overline{x_1} - \overline{x_2} - \Delta_0}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \dfrac{7.9 - 6.9}{2.5\sqrt{\frac{1}{100} + \frac{1}{100}}} = 2.828$

+ Degree of freedom: $v = n_1 + n_2 - 2 = 100 + 100 - 2 = 198$

+ $t_\alpha = t_{0.01,198} = 2.345$

Conclusion:

$t_0 = 2.828 > t_\alpha = 2.345 \Rightarrow$ reject H_0

Hence, the data support the claim that raising noise voltage would reduce the mean number of periods.

P-value: $P = P(T > |t_0|) = P(T > 2.828) \approx 0.01$

b)

$t_{\alpha/2,v} = t_{0.01/2,198} = 2.601$
$$\overline{x_1} - \overline{x_2} - t_{\alpha/2,n_1+n_2-2}s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \le \mu_1 - \mu_2 \le \overline{x_1} - \overline{x_2} + t_{\alpha/2,n_1+n_2-2}s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$7.9-6.9-2.601\times 2.5\sqrt{\frac{1}{100}+\frac{1}{100}}\le\mu_1-\mu_2\le 7.9-6.9+2.601\times 2.5\sqrt{\frac{1}{100}+\frac{1}{100}}$$

$$0.0804\le\mu_1-\mu_2\le 1.9196$$

10-54

An article in the Journal of Aircraft (1986, Vol. 23, pp. 859–864) described a new equivalent plate analysis method formulation that is capable of modeling aircraft structures such as cranked wing boxes and that produces results similar to the more computationally intensive finite element analysis method. Natural vibration frequencies for the cranked wing box structure are calculated using both methods, and results for the first seven natural frequencies follow:

Freq.	Finite Element Cycle/s	Equivalent Plate, Cycle/s
1	14.58	14.76
2	48.52	49.10
3	97.22	99.99
4	113.99	117.53
5	174.73	181.22
6	212.72	220.14
7	277.38	294.80

(a) Do the data suggest that the two methods provide the same mean value for natural vibration frequency? Use α = 0.05. Find the P-value.
(b) Find a 95% confidence interval on the mean difference between the two methods.

Solution:

```
> x <- c(14.58, 48.52, 97.22, 113.99, 174.73, 212.72, 277.38)
> y <- c(14.76, 49.10, 99.99, 117.53, 181.22, 220.14, 294.80)
> d <- x-y
> df <- data.frame(x, y, d)
> df
       x      y      d
1  14.58  14.76  -0.18
2  48.52  49.10  -0.58
3  97.22  99.99  -2.77
4 113.99 117.53  -3.54
5 174.73 181.22  -6.49
6 212.72 220.14  -7.42
7 277.38 294.80 -17.42
```

$n=7,\ \overline{d}=-5.486, s_d=5.929$

a)

Parameter of interest: the difference in mean value for natural vibration frequency of the two methods, $\mu_D=\mu_1-\mu_2=0$

Null hypothesis: $H_0:\mu_D=0$

Alt. hypothesis: $H_1:\mu_D\neq 0$

Test statistic: $t_0=\dfrac{\overline{d}}{s_d/\sqrt{n}}=\dfrac{-5.486}{5.924/\sqrt{7}}\approx -2.449$

Conclusion:

$t_{\alpha,n-1}=t_{0.05,6}=-1.943>t_0=-2.449\Rightarrow$ reject H_0

Hence, the data **doesn't support** the claim that the two methods provide the same mean value for natural vibration frequency.

P-value: $P=2P(T<-|t_0|)=2P(T<-2.449)\approx 0.03$

b)

A 95% confidence interval on the mean difference between the two methods

$\overline{d}-t_{\alpha/2,n-1}S_D/\sqrt{n}\le\mu_D\le\overline{d}+t_{\alpha/2,n-1}S_D/\sqrt{n}$

$\Leftrightarrow -5.486-t_{0.025,6}\times 5.924/\sqrt{7}\le\mu_D\le -5.486+t_{0.025,6}\times 5.924/\sqrt{7}$

$\Leftrightarrow -5.486-2.447\times 5.924/\sqrt{7}\le\mu_D\le -5.486+2.447\times 5.924/\sqrt{7}$

$\Leftrightarrow -10.965\le\mu_D\le -0.007$

10-88

A random sample of 500 adult residents of Maricopa County indicated that 385 were in favor of increasing the highway speed limit to 75 mph, and another sample of 400 adult residents of Pima County indicated that 267 were in favor of the increased speed limit.
(a) Do these data indicate that there is a difference in the support for increasing the speed limit for the residents of the two counties? Use α = 0.05. What is the P-value for this test?
(b) Construct a 95% conidence interval on the difference in the two proportions. Provide a practical interpretation of this interval

Solution:

$n_1 = 500,\ x_1 = 385$
$n_2 = 400, x_2 = 267$

$$\widehat{p_1} = \frac{x_1}{n_1} = \frac{385}{500} = 0.77,\ \widehat{p_2} = \frac{x_2}{n_2} = \frac{267}{400} = 0.6675$$

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{385 + 267}{500 + 400} = 0.724$$

a)

Parameter of interest: the difference in 2 proportions, p_1 and p_2

Null hypothesis: $H_0 : p_1 = p_2$

Alt. hypothesis: $H_1 : p_1 \neq p_2$

Test statistic: $z_0 = \dfrac{\widehat{p_1} - \widehat{p_2}}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

Computation:

+ $z_0 = \dfrac{\widehat{p_1} - \widehat{p_2}}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \dfrac{0.77 - 0.6675}{\sqrt{0.724(1-0.724)\left(\frac{1}{500} + \frac{1}{400}\right)}} = 3.4182$
+ $z_{\alpha/2} = z_{0.025} = 1.96$

Conclusion:

$z_0 = 3.4182 > 1.96 = z_{\alpha/2} \Rightarrow$ Reject H_0

P-value: $P = 2(1 - \Phi(3.4182)) \approx 0.001$

b)

A two-sided 95% CI on the difference of the 2 proportions:

$$\hat{p}_1 - \hat{p}_2 - z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \leq p_1 - p_2 \leq \hat{p}_1 - \hat{p}_2 + z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

$\Leftrightarrow 0.77 - 0.6675 - 1.96\sqrt{\frac{0.77(1-0.77)}{500} + \frac{0.6675(1-0.6675)}{400}} \leq p_1 - p_2 \leq 0.77 - 0.6675 - 1.96\sqrt{\frac{0.77(1-0.77)}{500} + \frac{0.6675(1-0.6675)}{400}}$

$\Leftrightarrow 0.0434 \leq p_1 - p_2 \leq 0.1616$