Problem 1

A simulation of a Taco Bell restaurant was performed to compare two different queuing systems. Each system was replicated 12 times. As output from the first system, the average simulated customer time in system, or sojourn time (in minutes) for each replication was:

$$D_1 = \{5, 8, 7, 4, 5, 12, 14, 3, 10, 10, 2, 4\}$$

(Integers are used for simplicity.) A second run of the simulation model was made using an independent stream of random numbers. Use hypothesis testing to determine whether there is a statistically significant difference between the average customer sojourn times for the two systems. The average sojourn times for the 12 replications of the second system are:

$$D_2 = \{10, 6, 10, 13, 10, 4, 2, 13, 6, 5, 13, 12\}$$

For simplicity, assume that the average sojourn times are drawn from normal populations with equal population variances. Please show all your calculations and report and interpret the test's $p-$value, both its magnitude and the conclusion of the test.

Hypothesis:

$$H_0 : \mu_0 = \mu_1$$
$$H_a : \mu_0 \neq \mu_1$$

Calculations by hand:

$$\bar{y} = \frac{\sum f(y)}{n} : \ \ \bar{y_1} = 7; \ \ \bar{y_2} = 8.667$$

$$s^2 = \frac{\sum_{i=1}^{n_1} (y_i - \bar{y})^2 + \sum_{i=1}^{n_2} (y_j - \bar{y})^2}{n_1 + n_2 - 2} : \ \ s^2 = 14.54545$$

$$z = \frac{\bar{y_1} - \bar{y_2}}{\sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{n}}} : z = \frac{7 - 8.667}{\sqrt{\frac{14.54545 + 15.15152}{12}}} = -1.0595$$

Then tests statistics $z = -1.0595$ and $-z_{\alpha/2} = -1.96$. Since $|z| < z_{\alpha/2}$, we fail to reject H_0.

R Code and Output:

```
D1 <- c(5,8,7,4,5,12,14,3,10,10,2,4)
D2 <- c(10,6,10,13,10,4,2,13,6,5,13,12)

# Perform Parametric Two-sample T-test
t.test(D1,D2)

# Output
Welch Two Sample t-test
data:  D1 and D2
```

```
t = -1.0595, df = 21.991, p-value = 0.3009
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -4.929224  1.595890
sample estimates:
mean of x mean of y
 7.000000  8.666667
```

From the R code output we can see that the $p-$value $= 0.3009$. This large $p-$value indicates that there is insufficient evidence to suggest that $\mu_0 \neq \mu_1$, as the $p-$value > 0.05 and we fail to reject the H_0. Here we achieve the same result as previously when we compared z test statistic and the critical value $z_{\alpha/2}$.

Problem 2

The thief of Baghdad is trying to escape from a circular dungeon. There are three identical doors that he may choose from. The first door leads to a tunnel that takes 5 minutes to traverse and returns him to the dungeon. The second door leads to a tunnel that takes 12 minutes to traverse and also returns him to the dungeon. The third door leads to a tunnel that takes 8 minutes to traverse and leads to freedom (so that he can save the princess). The thief is a Markov thief in the sense that he can't mark a door previously chosen. A Markov custodian will remove any marks while he is traversing tunnels.

Let:
X = time for the thief to escape;
Y = initial door choice.

(a) Find the expected time for the thief to escape ($E(X)$). *Hint:* use conditional expectation.
$P(Y = 1) = P(Y = 2) = P(Y = 3) = 1/3$

$$\begin{aligned}
E[X] &= E[X|Y=1]P(Y=1) + E[X|Y=2]P(Y=2) + E[X|Y=3]P(Y=3) = \\
&= \frac{5+E[X]}{3} + \frac{12+E[X]}{3} + \frac{8}{3} = \frac{25+2E[X]}{3} \\
E[X] &= \frac{25+2E[X]}{3} \\
3E[X] &= 25 + 2E[X] \\
E[X] &= 25
\end{aligned}$$

Then, the expected time for the thief to escape $\boxed{E(X) = 25}$ minutes.

(b) Write an R program that estimates the expected time to escape for the thief of Baghdad in order to verify your solution to part (a). Estimate the expected time to escape by averaging the escape times for 1000 thieves.

R Code and Output:

```
thief <- function(t1, t2, t3){
  doors <- c(1,2,3)              # door choices
  esc_t <- 0                     # the escape time variable
  status <- 0                    # the escape status binary variable
  while (status == 0){
    d <- sample(doors, 1)        # rand. pick a door while still in dungeon
      if (d == 3){               # if door 3 is picked, escape
        esc_t = t3 + esc_t       # update time variable
        status = 1
      }
      if (d == 1){               # if door 1 & 2 are picked, continue
          esc_t = t1 + esc_t     # update time variable
        }
      if(d == 2){
          esc_t = t2 + esc_t
        }
    }
  return(esc_t)
}

n <- 1000                        # number of thieves to simulate
thief_times <- replicate(n, thief(5,12,8))
mean(thief_times)                # get the expected escape times

# Output:
[1] 25.074999999999999
```

Based on the output of the R code, the estimated expected time to escape is approximately equal to the solution in part (a).

Problem 3

Emma must pass 40 classes before she can graduate. She may take as many classes as many times as desired until the required 40 classes are passed. Any time Emma takes a course, there is a 90% chance of passing. Let X be the number of courses Emma must take until forty are passed.

Given:
X = the number of courses Emma must take until forty are passed;

$P(\text{passing a class}) = 0.9$;
$r = 40$ number of classes needed to pass (successes);
Need to use Negative Binomial Distribution: consider events A and B, where
$A = \{\text{the first } (y-1) \text{ trials contain } (r-1) \text{ successes}\}$
$B = \{\text{trial } y \text{ results in a success}\}$.
$E[X] = \mu = \frac{r}{p}$

(a) What is the probability mass function of X?

$$P(X) = \binom{x-1}{r-1} p^{r-1} q^{y-r}, \quad x = r, r+1, r+2, ..., 0 \le p \le 1$$
$$P(X) = \binom{x-1}{40-1} (0.9)^{40-1} (0.1)^{x-40} = \binom{x-1}{40-1} (0.1)^{x} \cdot 1.64 \cdot 10^{38}$$

(b) What is the expected number of courses that Emma must take before passing 40?

$$E[X] = \mu = \frac{r}{p} = \frac{40}{0.9} = 44.44$$

Then, the expected number of courses Emma must take before passing 40 is $\boxed{E(X) = 44.44}$ classes.

(c) Write an R program that simulates 10,000 students like Emma. Estimate the mean number of courses a student must take to pass 40 and compare with the theoretical result obtained in (b).

R Code and Output:

```
n <- 10000
size <- 400     # 10*(number of successes) = 10*40
courses <- rnbinom(n, r, .9)

mean(courses)

# Output:
[1] 44.366100000000003
```

Problem 4

Consider the Weibull distribution as parameterized in the SIMAN textbook with probability distribution function:

$$f(x) = \alpha\beta^{-\alpha} x^{\alpha-1} e^{-(x/\beta)^{\alpha}} \qquad x > 0;\ \alpha > 0,\ \beta > 0.$$

(a) Find the cumulative distribution function.

$$F(X) = \int_0^x \alpha\beta^{-\alpha}t^{\alpha-1}e^{-(t/\beta)^\alpha}dt =$$
$$= \int_0^{\frac{x^\alpha}{\beta^\alpha}} \frac{\alpha\beta^{-\alpha}t^{\alpha-1}e^{-u}}{\alpha\beta^{-\alpha}t^{\alpha-1}}du =$$
$$= \int_0^{\frac{x^\alpha}{\beta^\alpha}} e^{-u}du = -e^{-u}\Big|_0^{\frac{x^\alpha}{\beta^\alpha}} =$$
$$= -e^{-\frac{x^\alpha}{\beta^\alpha}} - (-e^0) =$$
$$= 1 - e^{-\frac{x^\alpha}{\beta^\alpha}},\ x > 0;\ \alpha > 0,\ \beta > 0.$$

Substitution work:

$$u = \frac{t^\alpha}{\beta^\alpha}$$
$$du = \frac{\alpha t^{\alpha-1}}{\beta^\alpha}dt$$
$$dt = \frac{du}{\alpha\beta^{-\alpha}t^{\alpha-1}}$$
$$t = x => u = \frac{x^\alpha}{\beta^\alpha}$$
$$t = 0 => u = 0$$

Then, Weibull distribution's cdf is

$$F(X) = \begin{cases} 1 - e^{-x^\alpha/\beta^\alpha}, & \text{for } x > 0;\ \alpha > 0,\ \beta > 0 \\ 0, & \text{for } x \le 0. \end{cases}$$

(b) Write an R program that generates the random variable X using the inverse-cdf technique and estimates $E[X]$ and $V[X]$ using 1000 variates. (Use $\alpha = 3$ and $\beta = 2$.)
Inverse-cdf technique:

$$Y = 1 - e^{-X^\alpha/\beta^\alpha}$$
$$e^{-X^\alpha/\beta^\alpha} = 1 - Y$$
$$-X^\alpha/\beta^\alpha = \ln(1-y)$$
$$X^\alpha = -\beta^\alpha \ln(1-y)$$
$$X = -\beta\sqrt[\alpha]{\ln(1-y)},\ \ \alpha = 3,\ \beta = 2$$
$$X = -2\sqrt[3]{\ln(1-y)}$$

R Code and Output:

```
a = 3 # shape
b = 2 # scale
options(digits = 22)
u <- runif(1001)

x <- abs(-2*(abs(log(1-u)))^p)
mean(x)
var(x)

# Outputs:
1.7883    # E[X]
0.4327    # V[X]
```