Question 1

(**a**) Mean = 292.3922, Median = 249
Since, the mean is greater than the median. The distribution is right skewed.

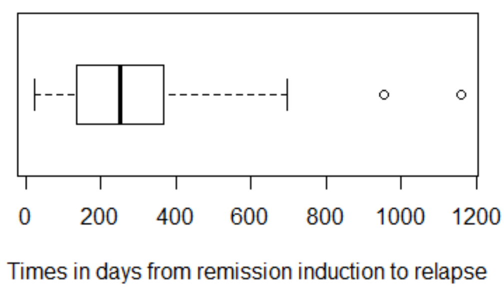

Figure 1: BoxPlot

Output 1: Q1(a) R Code Output

```
> boxplot(myData, outline = TRUE, horizontal = TRUE, xlab =
+         "Times in days from remission induction to relapse")
> boxplot.stats(myData)
$stats
[1]   24.0 135.5 249.0 367.0 697.0

$n
[1] 51

$conf
[1] 197.782 300.218

$out
[1]   955 1160
```

(**b**) There could be or could not be any outliers in the dataset. Sometimes uneven values are due to something significant and scientific and sometimes they are just error.
You can drop the outliers if you are very confident that these value are out of scope or if you can verify the results by doing the experiment again.

(**c**)Percent of patients in remission for less than one year: 74.50

Question 2

(**a**) Probability Distribution of Y

Y	PD
0	150000-1031/150000
50,000	1/150000
10,000	5/150000
1000	25/150000
10	1000/150000

(**b**) $\mathrm{E}[X] = \sum_{i=1}^{k} x_i\, p_i = x_1 p_1 + x_2 p_2 + \cdots + x_k p_k . \mathrm{E}[X] =$ 0.9

(**c**) Mean Expected Value =0.9 and the ticket cost is 1. So there will be a loss of $0.10.Hence, it is not worthwhile to purchase the ticket.

(**d**)

Standard Deviation = 142.00

Output 2: Q4(a) R Code Output

```
> vec <- c(50000,rep(10000, each = 5),
rep(1000,each = 25),rep(10,each = 1000),
+             rep(0,each = 150000-1031))
> sd(vec)
[1] 142.0094
```

Question 3

(Assuming sample is normally distributed.)

(**a**) The target population are the adults who have experienced feeling of depression or sadness.

(**b**) Probability that your sample will have 12 or more people with these feelings = 0.017

Output 3: Q3(b) R Code Output

```
> p_hat = 12/68
> p <- 0.10
> p
[1] 0.1
```

```
> n <- 68
> n
[1] 68
> SD <- sqrt(p*(1-p)/n)
> SD
[1] 0.03638034
> z_score = (p_hat - p)/SD
> ans_a = pnorm(z_score,lower.tail = FALSE)
> ans_a
[1] 0.01777771
```

(**c**) There is a 1.7% chance that in my sample of 68 people, more than 12 people will have such feelings.

(**d**) Probability that the sample percentage will differ from this by more than 5 percent = 0.16

Output 4: Q3(d) R Code Output

```

> z_score_1 = (0.15 - 0.10)/SD
> z_score_1
[1] 1.374369
> z_score_2 = (0.05-0.10)/SD
> z_score_2
[1] -1.374369
> ans_d = pnorm(z_score_1,lower.tail = FALSE) +
+         pnorm(z_score_2)
> ans_d
[1] 0.1693273
```

Question 4

(a)

$H0$: Mean = \$14200

$H1$: Mean $\neq$ \$14200

$Population\ Mean = 14200$

$Sample\ Mean = 15300$

$Standard\ Error = 2600/sqrt(75) = 300.22$

$t\ value = 3.66$

$-1.99 \leq Alpha\ Range \leq 1.99$

$Level\ of\ Significance = 0.05$

Result: Since, t value does not lie in the critical range, we reject the NULL hypothesis at 0.05 level of significance

Output 5: Q4(a) R Code Output

```
> pop_mean = 14200
> sample_mean = 15300
> pop_sd = 2600
> givenAlpha = 0.05
> mySampleSize = 75
> se <- pop_sd/sqrt(mySampleSize)
>
> t_value_stat = (sample_mean - pop_mean)/se
> t_value_stat
[1] 3.663954
> alpha_range = qt(1-givenAlpha/2,df = mySampleSize-1)
> critical_range = c(-alpha_range,alpha_range)
> critical_range
[1] -1.992543  1.992543
```

(**b**)
Confidence Level = 99 %
Critical Value = 2.37
Margin Error = Standard Error * Critical Value = 713.86
Confidence Interval = Sample Mean +- Margin Error
Confidence Interval : $14586 \leq SampleMean \leq 16014$

Result : We are 99% confident that the sample mean will lie between 14586 and 16014

Output 6: Q4(b) R Code Output

```
> val = qt(0.99,df = 74)
> val
[1] 2.377802
> marginError <- val * se
> marginError
[1] 713.8688
```

```
> confidenceInterval = c(sample_mean - marginError , sample_mean
+ marginError)
> confidenceInterval
[1] 14586.13 16013.87
```

(**c**)

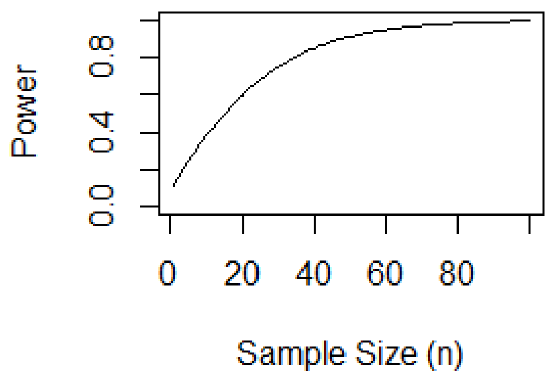

Figure 2: Power Curve

Output 7: Q4(c) R Code Output

```
> power_curve <- function(n = 75,alpha = 0.05,H0 = 14200,H1 = 15300,
+                             sig = 2600)
+ {
+   zee <- qnorm(p = alpha,mean = 0,sd = 1,lower.tail = FALSE)
+   cr <- zee * (sig/sqrt(n)) + H0;
+   power <- pnorm(q = cr,mean = H1,sd = sig/sqrt(n),
lower.tail = FALSE)
+   power
+ }
> power_curve()
[1] 0.9782616
> mycurve1 <- lapply(X = 1:100, FUN = power_curve,alpha = .05
+                        ,H0 = 14200,H1 = 15300,sig = 2600)
> plot(x = 1:100,y = mycurve1,type = "line",
xlab = "Sample Size (n)", ylim = c(0,1),
```

```
+         ylab = "Power")
```

Question 5

(a)
$H0:$ Mean = 2700
$H1$: Mean $\neq$ 2700
$Population\ Mean = 2600$
$Sample = 2620$
$Standard\ Error = 450/sqrt(36) = 75$
$t\ value = -1.06$
$-2.03 \leq Alpha\ Range \leq 2.03$
$Level\ of\ Significance = 0.05$

Result: Since, t value is in range of critical Values, we do not reject the NULL hypothesis.

Output 8: Q5(a) R Code Output

```
> given_pop_mean = 2700
> given_sample_mean = 2620
> given_pop_sd = 450
> alpha = 0.05
> size = 36
> standard_e = given_pop_sd/sqrt(size)
> t_value = (given_sample_mean - given_pop_mean)/standard_e
> t_value
[1] -1.066667
> d_range = qt(1-alpha/2,df = size-1)
> f_range = c(-d_range,d_range)
> f_range
[1] -2.030108  2.030108
```

(**b**)(Assuming sample is normally distributed.)
Power = 0.26

Output 9: Q5(b) R Code Output

```
> lower_bound = qnorm(alpha/2,mean = 2700,sd = standard_e)
> lower_bound
```

```
[1] 2553.003
> upper_bound = qnorm(alpha/2,mean = 2700,
sd = standard_e, lower.tail = FALSE)
> upper_bound
[1] 2846.997
> given_mean = 2600
> z_lower = (lower_bound - given_mean)/standard_e
> z_lower
[1] -0.6266307
> z_upper = (upper_bound - given_mean)/standard_e
> z_upper
[1] 3.293297
> p_lower = pnorm(z_lower)
> p_lower
[1] 0.2654507
> p_upper = pnorm(z_upper,lower.tail = FALSE)
> p_upper
[1] 0.0004950985
> power = p_lower + p_upper
> power
[1] 0.2659458
```