Homework 1 - STAT 511
Amal Agarwal

Answer 1

(a) Y~ $\text{Binom}(n, p), y \in \{0, 1, 2, ...\}$
Probability mass function (pmf) of Y is
$P(Y = y) = \binom{n}{y} p^y (1 - p)^{(n-y)}, y \in \{0, 1, 2, ...\}$
Now the mean can be evaluated as

$$E[Y] = \sum_{y=0}^{\infty} y P(Y = y)$$

$$= \sum_{y=0}^{\infty} y \binom{n}{y} p^y (1 - p)^{(n-y)}$$

$$= \sum_{y=0}^{\infty} y \frac{n!}{y!(n - y)!} p^y (1 - p)^{(n-y)}$$

$$= \sum_{y=0}^{\infty} \frac{n(n - 1)!}{(y - 1)!(n - y)!} p p^{(y-1)} (1 - p)^{(n-y)}$$

$$= np \sum_{y=1}^{\infty} \frac{(n - 1)!}{(y - 1)!(n - y)!} p^{(y-1)} (1 - p)^{(n-y)}$$

$$= np \sum_{y=1}^{\infty} \binom{n - 1}{y - 1} p^{(y-1)} (1 - p)^{(n-y)}$$

$$= np \left(\because \sum_{y=1}^{\infty} \binom{n - 1}{y - 1} p^{(y-1)} (1 - p)^{(n-y)} = \sum_{x=0}^{\infty} \binom{n - 1}{x} p^{(x)} (1 - p)^{(n-1-x)} = 1 \right)$$
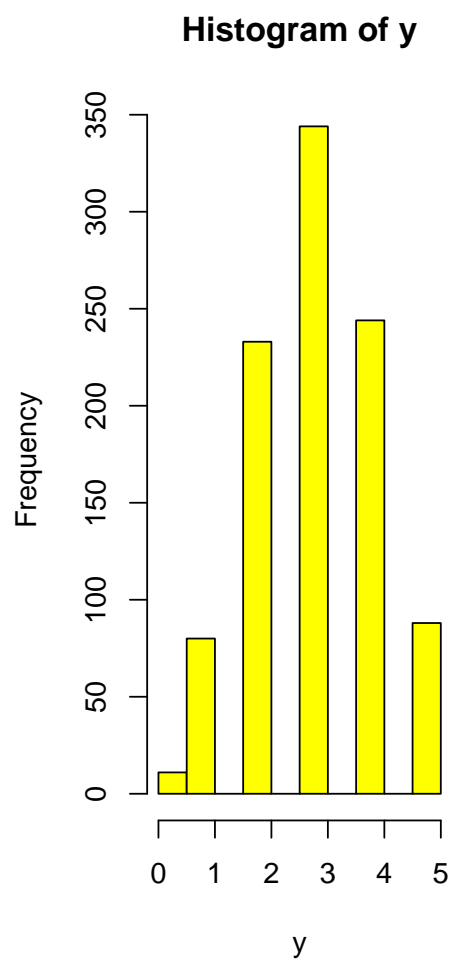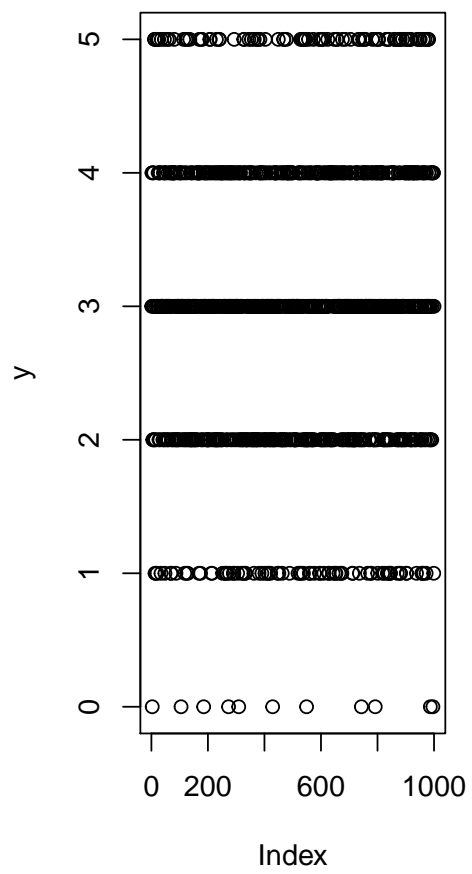
1

Similarily

$$E[Y^2] = \sum_{y=0}^{\infty} y^2 P(Y = y)$$

$$= \sum_{y=0}^{\infty} y^2 \binom{n}{y} p^y (1-p)^{(n-y)}$$

$$= \sum_{y=0}^{\infty} y^2 \frac{n!}{y!(n-y)!} p^y (1-p)^{(n-y)}$$

$$= \sum_{y=0}^{\infty} (y-1+1) \frac{n(n-1)!}{(y-1)!(n-y)!} p^y (1-p)^{(n-y)}$$

$$= \sum_{y=0}^{\infty} (y-1) \frac{n(n-1)(n-2)!}{(y-1)!(n-y)!} p^y (1-p)^{(n-y)} + \sum_{y=0}^{\infty} \frac{n(n-1)!}{(y-1)!(n-y)!} p^y (1-p)^{(n-y)}$$

$$= \sum_{y=0}^{\infty} \frac{n(n-1)(n-2)!}{(y-2)!(n-y)!} p^2 p^{(y-2)} (1-p)^{(n-y)} + \sum_{y=0}^{\infty} \frac{n(n-1)!}{(y-1)!(n-y)!} pp^{(y-1)} (1-p)^{(n-y)}$$

$$= n(n-1)p^2 \sum_{y=2}^{\infty} \frac{(n-2)!}{(y-2)!(n-y)!} p^{(y-2)} (1-p)^{(n-y)} + np \sum_{y=1}^{\infty} \frac{(n-1)!}{(y-1)!(n-y)!} p^{(y-1)} (1-p)^{(n-y)}$$
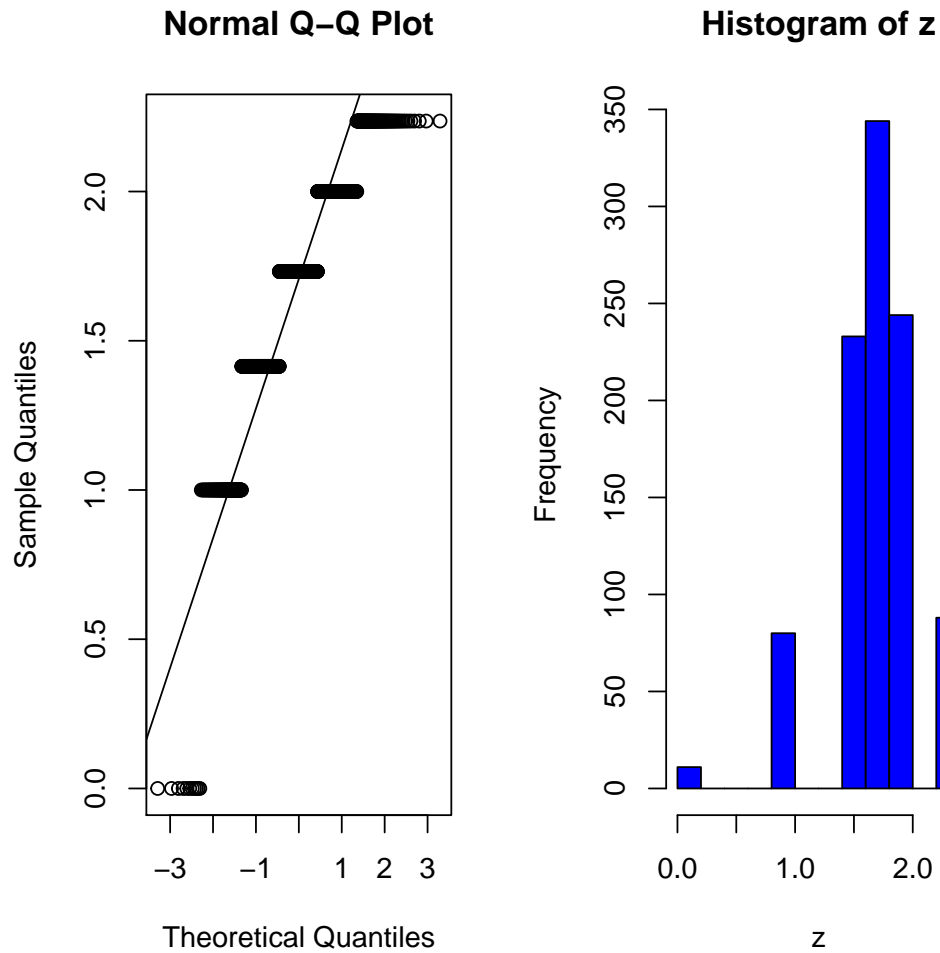
$$= n(n-1)p^2 + np$$
$$= n^2 p^2 - np^2 + np$$

Thus

$$Var[Y] = E[Y^2] - (E[Y])^2$$
$$= (n^2 p^2 - np^2 + np) - (np)^2$$
$$= np - np^2$$
$$= np(1-p)$$

(b) If $E[Y] = np = 3, Var[Y] = np(1-p) = 3(1-p) = 1 \Rightarrow p = \frac{2}{3}$ and $n = \frac{9}{2}$. But n has to be a natural number and thus here the variance can't be made equal to 1

(c) Here the variance can't be made equal to 1. Suppose we make $n = 5$ and $p = 3/5$. With these parameters, the mean is again 3 but the variance modifies to 6/5. Taking these parameters, the plot and histogram of simulated random variables are:
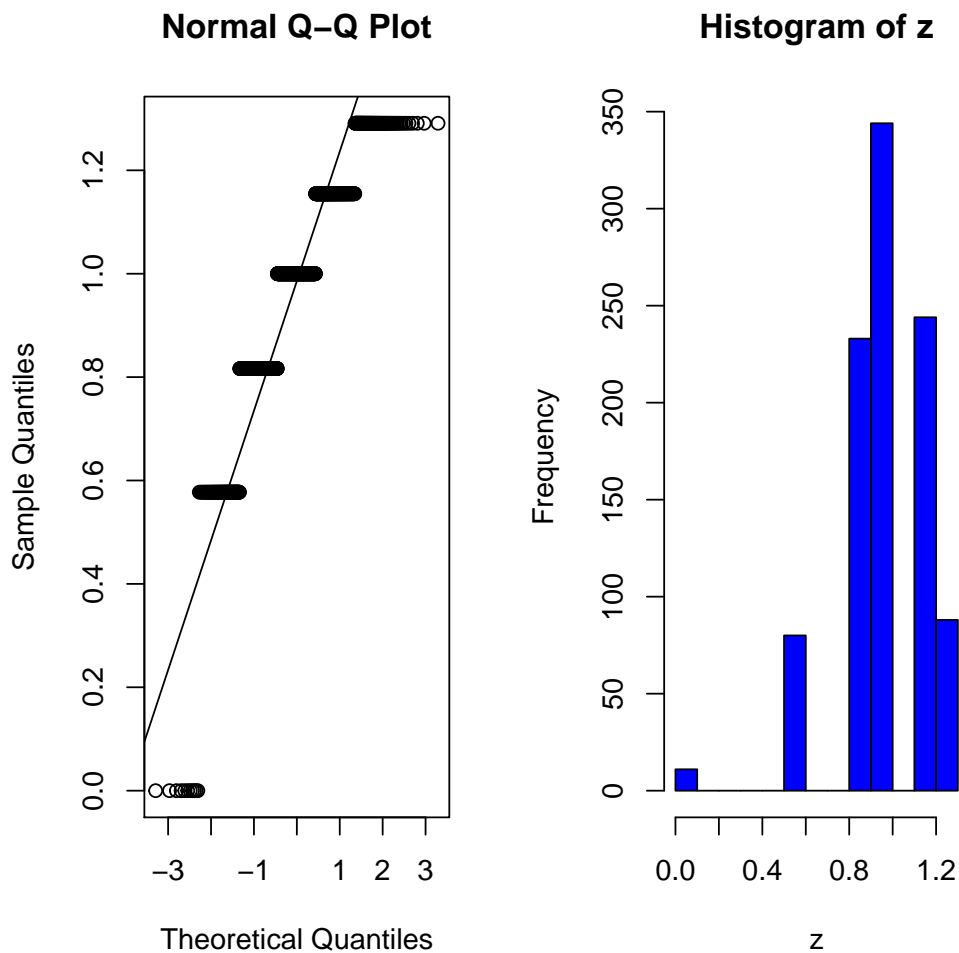
2

Histogram of y

(d)   The square root power transformation gives the following plots:

Normal Q–Q Plot

Sample Quantiles

Theoretical Quantiles

Histogram of z

Frequency

z

The cube root power transformation gives the following plots:

Normal Q–Q Plot   **Histogram of z**

These qq plots show that even after the square root and cube root transformations, the data is far from being normally distributed. This is expected since binomial is a discrete distribution. However if we increase n to 50 and plot the graphs again, we get

Normal Q–Q Plot

Sample Quantiles / Theoretical Quantiles

Histogram of z

Frequency / z
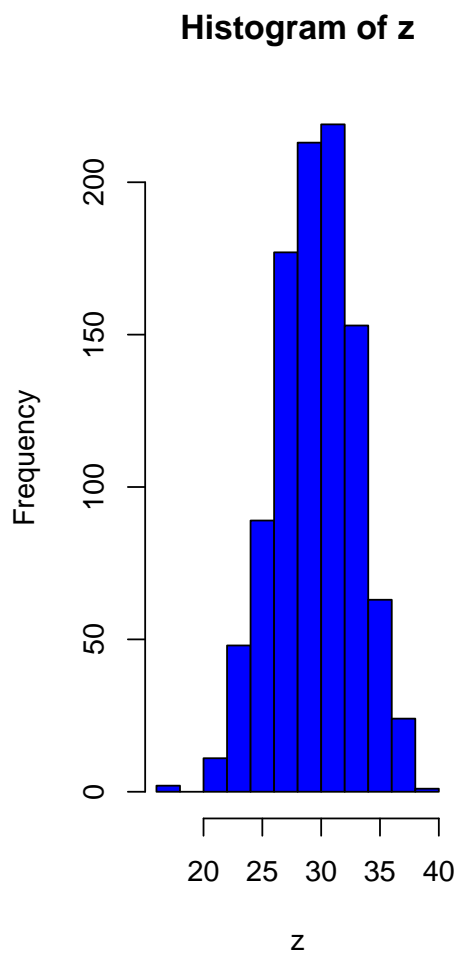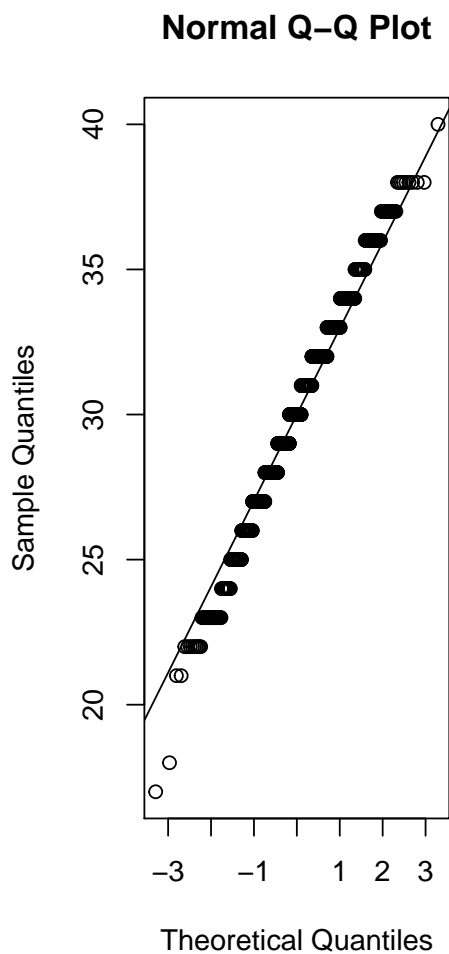
This is significantly better which demonstrates the well known result that if $n \to \infty$, binomial distribution becomes close to normal.

Answer 2

(a) Y$\sim$ Normal$(\mu, \sigma^2), y \in R$
Probability density function (pdf) of Y is

$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(y-\mu)^2}{2\sigma^2}}, y \in R$

Now the mean can be evaluated as

$$E[Y] = \int_{-\infty}^{+\infty} y f(y) \ dy$$

$$= \int_{-\infty}^{+\infty} y \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(y-\mu)^2}{2\sigma^2}} \ dy$$

Assume $\frac{y-\mu}{\sqrt{2\sigma^2}} = x$ Thus we get

$$E[Y] = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} \sqrt{2\sigma^2}(\mu + \sqrt{2\sigma^2}x)e^{-x^2} \, dx$$

$$= \frac{1}{\sqrt{\pi}} \left[\int_{-\infty}^{+\infty} \mu e^{-x^2} \, dx + \int_{-\infty}^{+\infty} \sqrt{2\sigma^2}x e^{-x^2} \, dx \right]$$

The second integrand is an odd function and thus the integral from $-\infty$ to $+\infty$ goes to zero. Thus we have

$$E[Y] = \frac{1}{\sqrt{\pi}} \left[\int_{-\infty}^{+\infty} \mu e^{-x^2} \, dy \right]$$

$$= \frac{1}{\sqrt{\pi}} \mu\sqrt{\pi}$$

$$= \mu$$

Similarily

$$E[Y^2] = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} y^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dy$$

Now substituting $y - \mu = t$

$$E[Y^2] = \frac{1}{\sqrt{2\pi\sigma^2}} \left[\int_{-\infty}^{+\infty} t^2 e^{-\frac{t^2}{2\sigma^2}} \, dt + 2\mu \int_{-\infty}^{+\infty} t e^{-\frac{t^2}{2\sigma^2}} \, dt + \mu^2 \int_{-\infty}^{+\infty} e^{-\frac{t^2}{2\sigma^2}} \, dt \right]$$
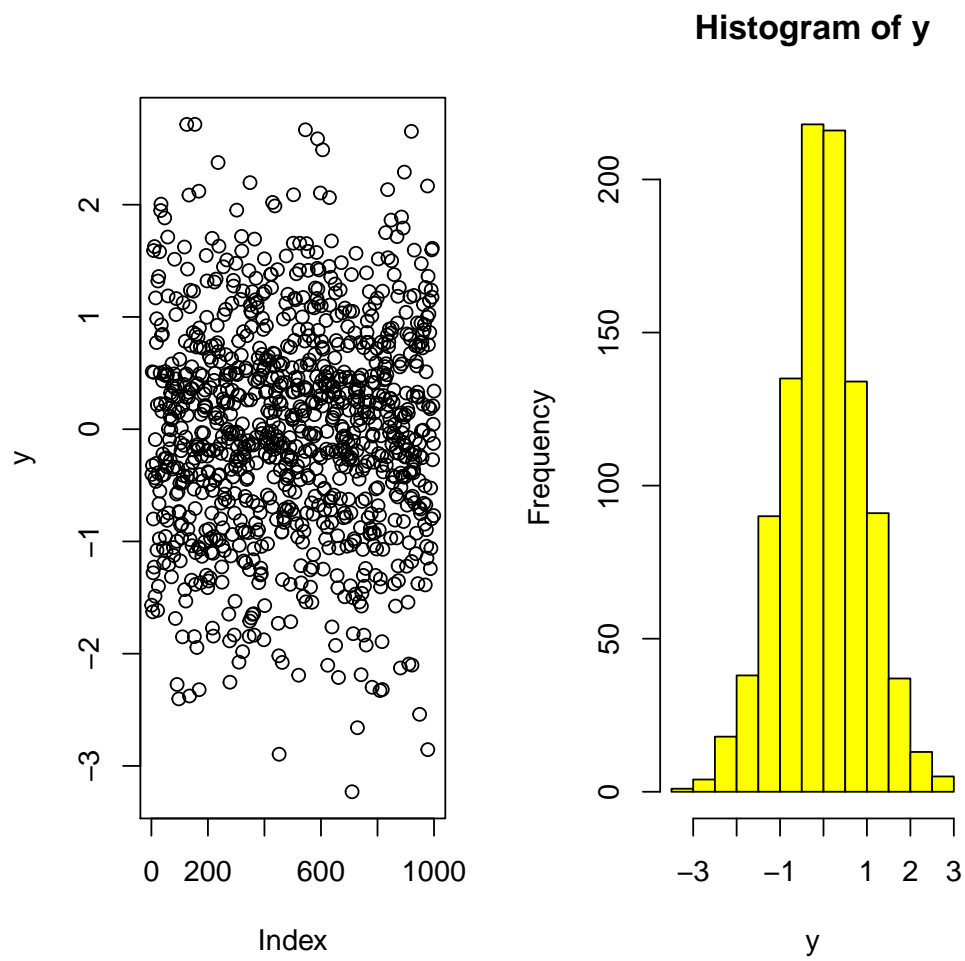
The second integral again goes to zero. Thus we obtain

$$E[Y^2] = \frac{1}{\sqrt{2\pi\sigma^2}} \left[\frac{\sqrt{\pi}}{2}(2\sigma^2)^{3/2} + \mu^2\sqrt{2\pi\sigma^2} \right] = \sigma^2 + \mu^2$$

Hence
$$Var[Y] = E[Y^2] - (E[Y])^2$$
$$= (\sigma^2 + \mu^2) - (\mu)^2 = \sigma^2$$
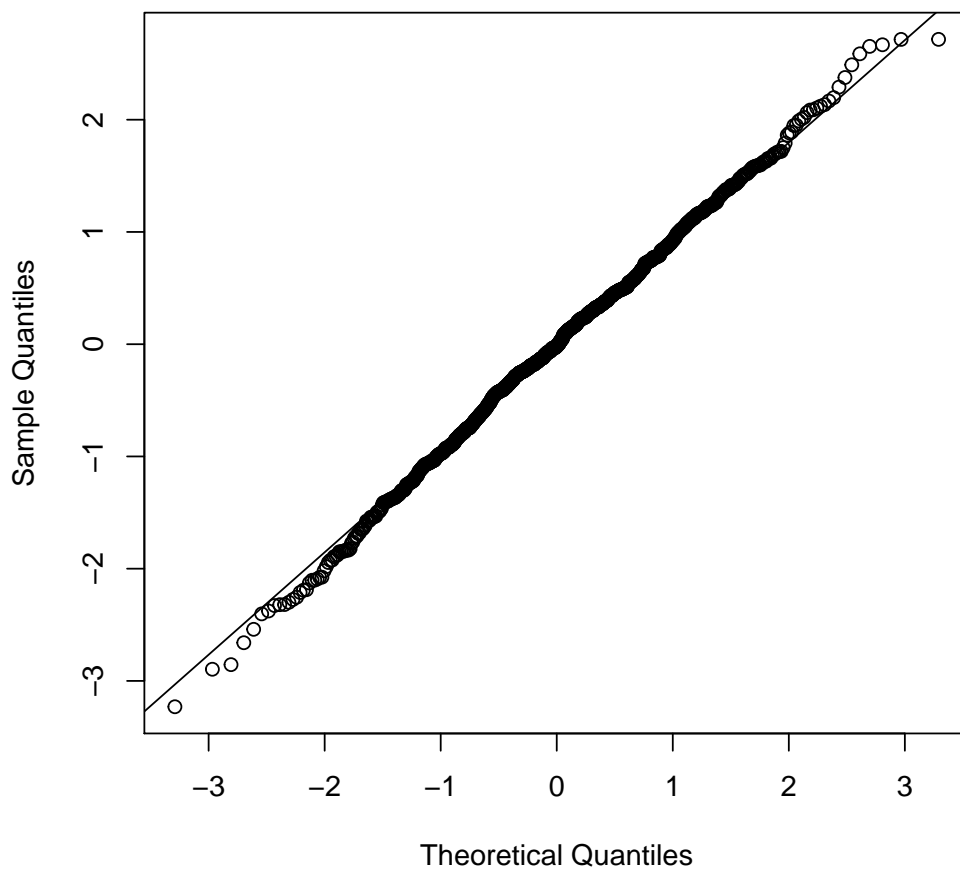
(b) The set of parameters (μ, σ^2) that make $E[Y] = 3$ and $Var[Y] = 1$ are $(3, 1)$.

(c) If $E[Y] = \mu = 3, Var[Y] = \sigma^2 = 1$, the plots of the simulated random variables and the histogram are given as:

Histogram of y

(d) Since this simulation is done for the normal distribution itself, no transformation is required to make it normal. The qq plot is given as

Normal Q−Q Plot

Answer 3

(a) Y$\sim$ Poisson$(\lambda), y \in \{0, 1, 2, ...\}$
Probability mass function (pmf) of Y is
$P(Y = k|\lambda) = \frac{\lambda^k}{k!}e^{-\lambda}, y \in \{0, 1, 2, ...\}$
Now the mean can be evaluated as

$$E[Y] = \sum_{k=0}^{\infty} kP(Y = k)$$

$$= \sum_{k=0}^{\infty} k\frac{\lambda^k}{k!}e^{-\lambda}$$

$$= \sum_{k=0}^{\infty} \frac{\lambda\lambda^{(k-1)}}{(k-1)!}e^{-\lambda}$$

$$= \lambda\sum_{k=1}^{\infty} \frac{\lambda^{(k-1)}}{(k-1)!}e^{-\lambda}$$
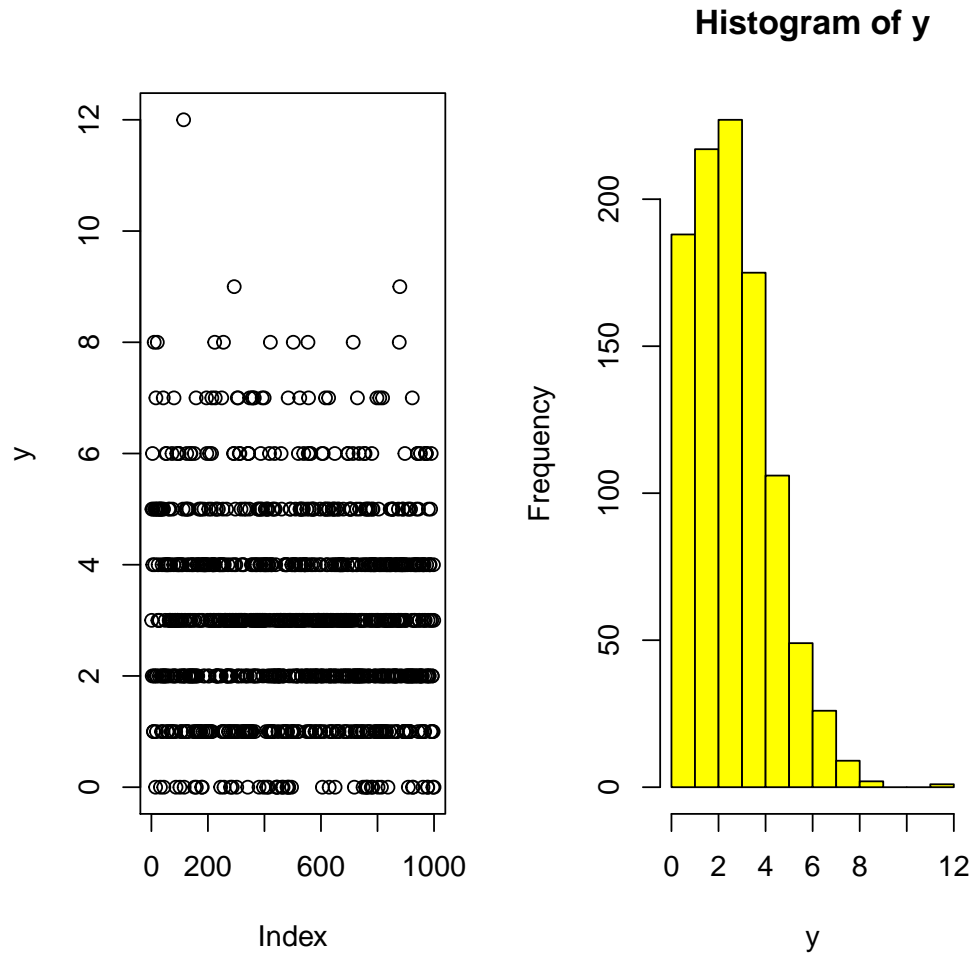
$$= \lambda$$

Similarily

$$E[Y^2] = \sum_{k=0}^{\infty} k^2 P(Y = k)$$

$$= \sum_{k=0}^{\infty} k^2\frac{\lambda^k}{k!}e^{-\lambda}$$

$$= \sum_{k=1}^{\infty} (k-1+1)\frac{\lambda^k}{(k-1)!}e^{-\lambda}$$

$$= \sum_{k=1}^{\infty} (k-1)\frac{\lambda^k}{(k-1)!}e^{-\lambda} + \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!}e^{-\lambda}$$

$$= \sum_{k=2}^{\infty} \frac{\lambda^2\lambda^{k-2}}{(k-2)!}e^{-\lambda} + \sum_{k=1}^{\infty} \frac{\lambda\lambda^{k-1}}{(k-1)!}e^{-\lambda}$$

$$= \lambda^2\sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!}e^{-\lambda} + \lambda\sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}e^{-\lambda}$$
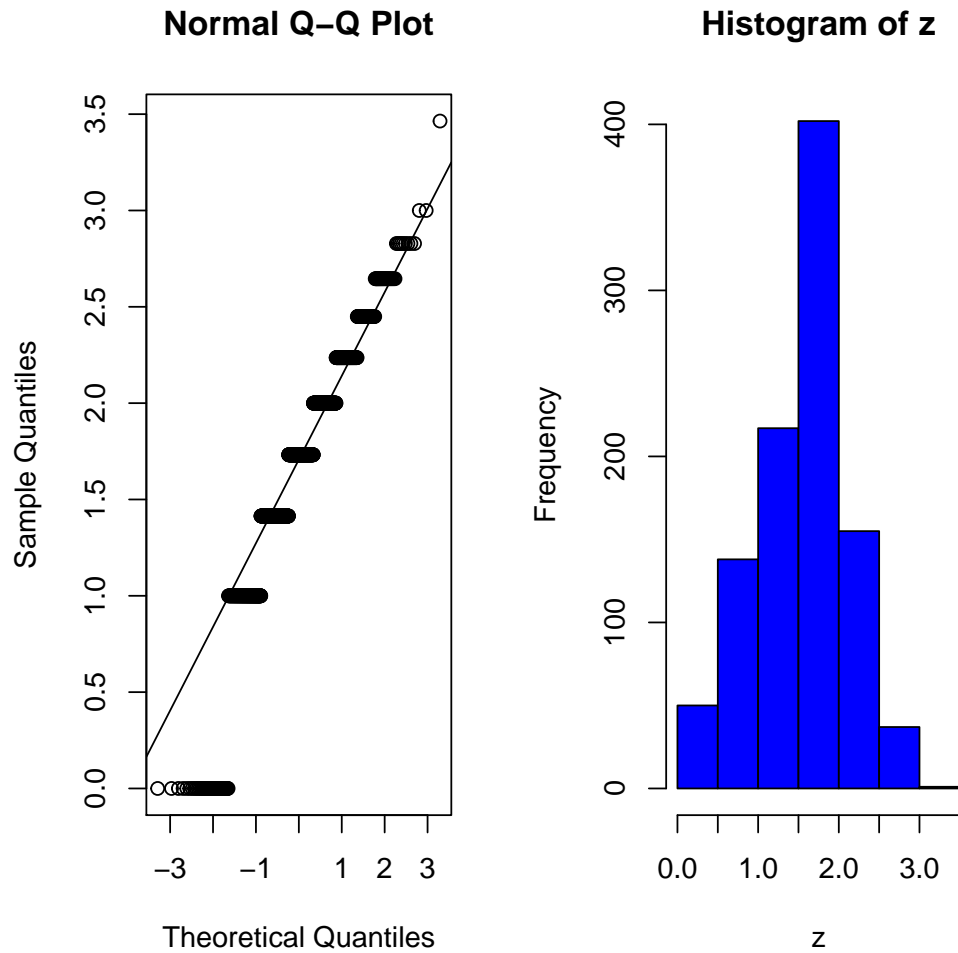
$$= \lambda^2 + \lambda$$

Thus

$$Var[Y] = E[Y^2] - (E[Y])^2$$
$$= \lambda^2 + \lambda - (\lambda)^2$$
$$= \lambda$$

(b) If $E[Y] = \lambda = 3, Var[Y] = \lambda = 1$ is not possible. Thus $\lambda = 3$ if $Var[Y]$ is also equal to 3.

(c) Here the variance can't be made equal to 1. Suppose $Var[Y] = 3$. Then we get $\lambda = 3$. Taking this parameter, the plot and histogram of the simulated random variables are:

Histogram of y

(d) The square root transformation gives

Normal Q–Q Plot

Sample Quantiles

Theoretical Quantiles

Histogram of z

Frequency

z

which seems very close to normal distribution.

Answer 4

(a) $Y \sim$ Laplace$(\mu, \sigma), y \in R$
Pdf of Y is
$f(y) = \frac{1}{2\sigma} e^{-\frac{|(y-\mu)|}{\sigma}}, y \in R$
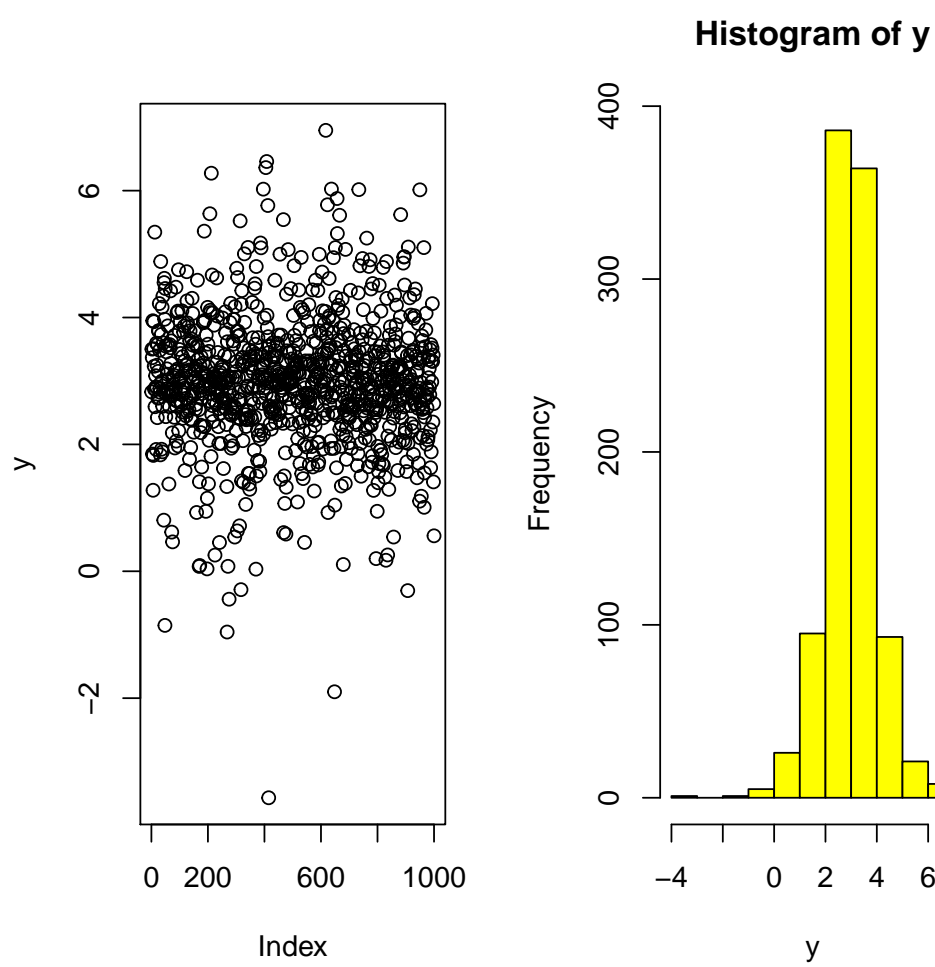Now the mean can be evaluated as

$$E[Y] = \int_{-\infty}^{+\infty} y f(y) \, dy$$

$$= \int_{-\infty}^{+\infty} y \frac{1}{2\sigma} e^{-\frac{|(y-\mu)|}{\sigma}} \, dy$$

$$= \frac{1}{2\sigma} \left[\int_{-\infty}^{\mu} y e^{\frac{y-\mu}{\sigma}} \, dy + \int_{\mu}^{+\infty} y e^{-\frac{y-\mu}{\sigma}} \, dy \right]$$

$$= \frac{1}{2\sigma} \left\{ \left[y\sigma e^{\frac{y-\mu}{\sigma}} \Big|_{-\infty}^{\mu} - \int_{-\infty}^{\mu} \sigma e^{\frac{y-\mu}{\sigma}} \, dy \right] + \left[y(-\sigma)e^{-\frac{y-\mu}{\sigma}} \Big|_{\mu}^{+\infty} - \int_{\mu}^{+\infty} (-\sigma)e^{-\frac{y-\mu}{\sigma}} \, dy \right] \right\}$$

$$= \frac{1}{2} \left\{ \left[\mu - \sigma e^{\frac{y-\mu}{\sigma}} \Big|_{-\infty}^{\mu} \right] + \left[\mu + (-\sigma)e^{-\frac{y-\mu}{\sigma}} \Big|_{\mu}^{+\infty} \right] \right\}$$

$$= \frac{1}{2} [(\mu - \sigma) + (\mu + \sigma)]$$

$$= \mu$$

Similarily $E[X^2]$ can be evaluated and it comes out to be $\mu^2 + 2\sigma^2$. Thus

$$Var[Y] = \mu^2 + 2\sigma^2 - \mu^2$$
$$= 2\sigma^2$$

(b) The set of parameters (μ, σ) that make $E[Y] = 3$ and $Var[Y] = 1$ are $(3, \frac{1}{\sqrt{2}})$.
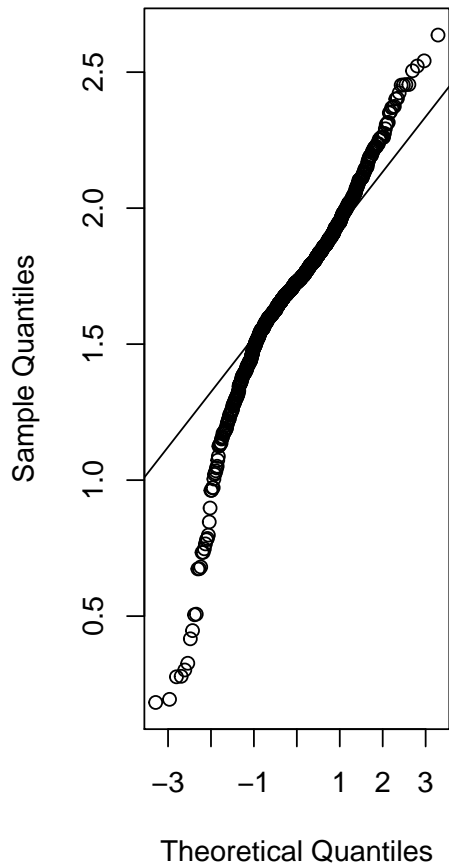
(c) If $E[Y] = \mu = 3, Var[Y] = 2\sigma^2 = 1$, the plots of the simulated random variables and the histogram are given as:
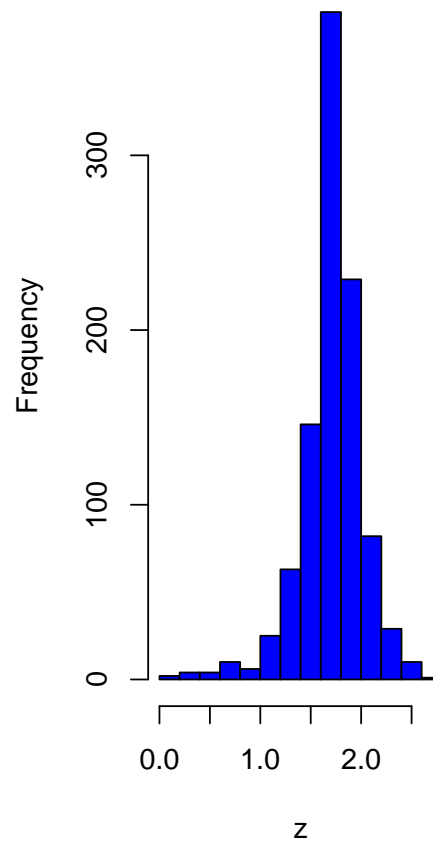
13

Histogram of y

14

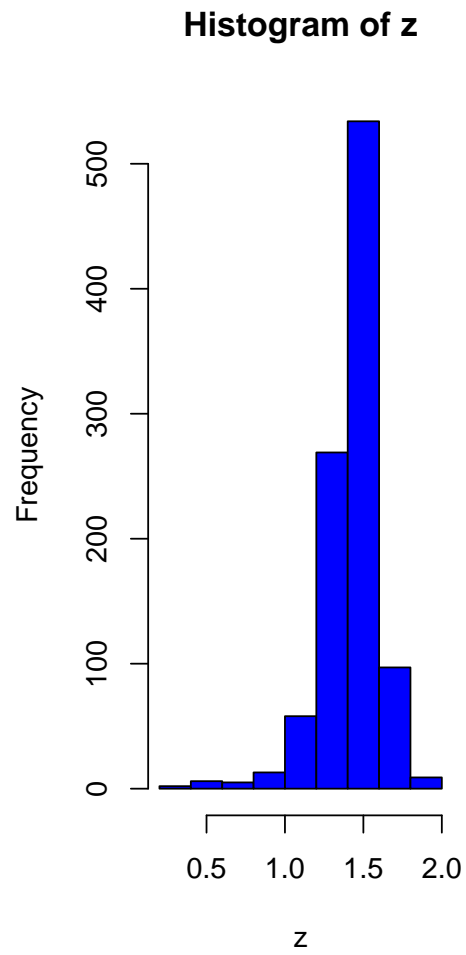(d) The square root power transformation gives the following plots:
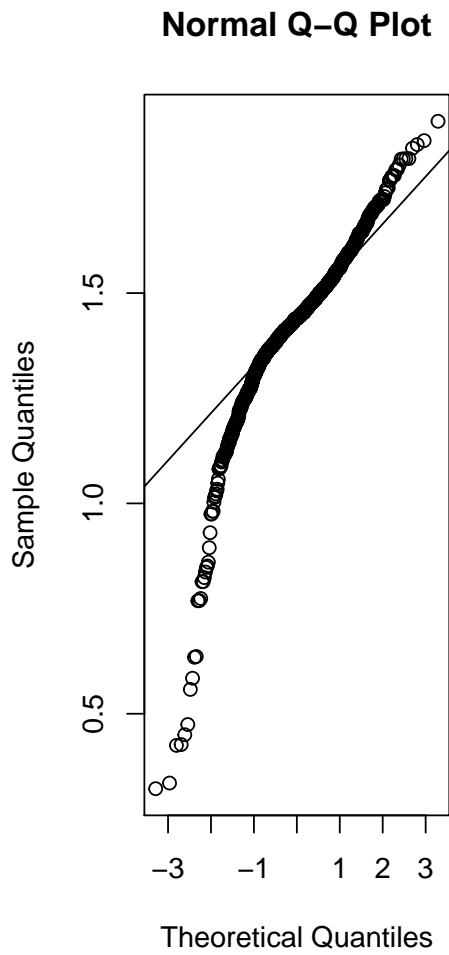
Which Do You Think Is Best?

Normal Q–Q Plot

Histogram of z

15
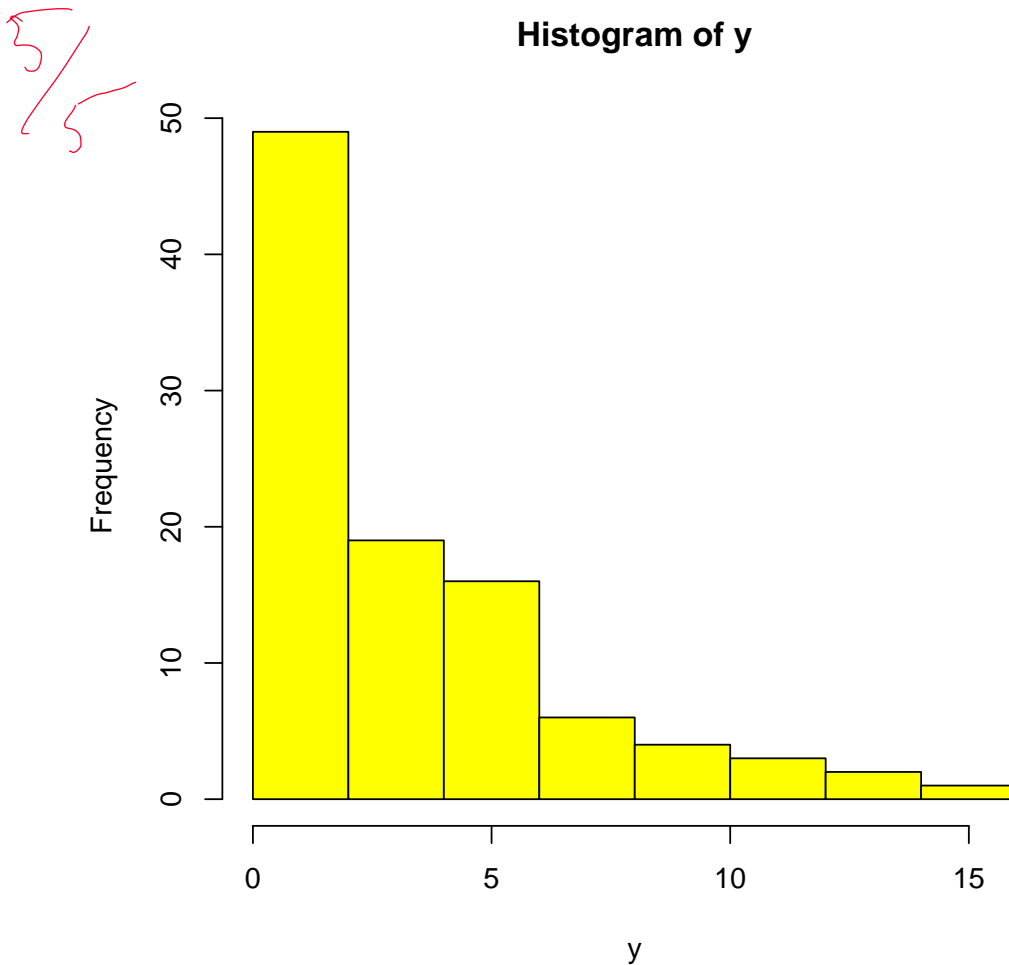
The cube root power transformation gives the following plots:

Normal Q–Q Plot

Sample Quantiles

Theoretical Quantiles

Histogram of z

Frequency

z

Answer 5

(a) Plotting the histogram for the interarrival data:

Histogram of y

From the histogram, it is clear that the support is R^+. The shape of the ditribution indicates that it is exponential. Now the exponential distribution has one parameter μ with the pdf as:

$f(x) = \frac{1}{\mu}e^{-\frac{x}{\mu}}$

The theoretical mean and variance are given as:

$E[X] = \mu$ and $Var[X] = \mu^2$ The sample mean and variance are

```
[1] 3.285651
```

```
[1] 11.05406
```

respectively.

Comparing the sample mean and the theoretical mean, we get the parameter value $\mu =$

3.285651

(b) The data clearly shows that all the sample values are greater than zero. They lie in the following range:

```
[1] 0.0405379
```

```
[1] 15.35806
```

i.e. the sample data has the support as $[0.0405379, 15.35806]$. This is a subset of the theoretical support $R^+ \cup \{0\}$.
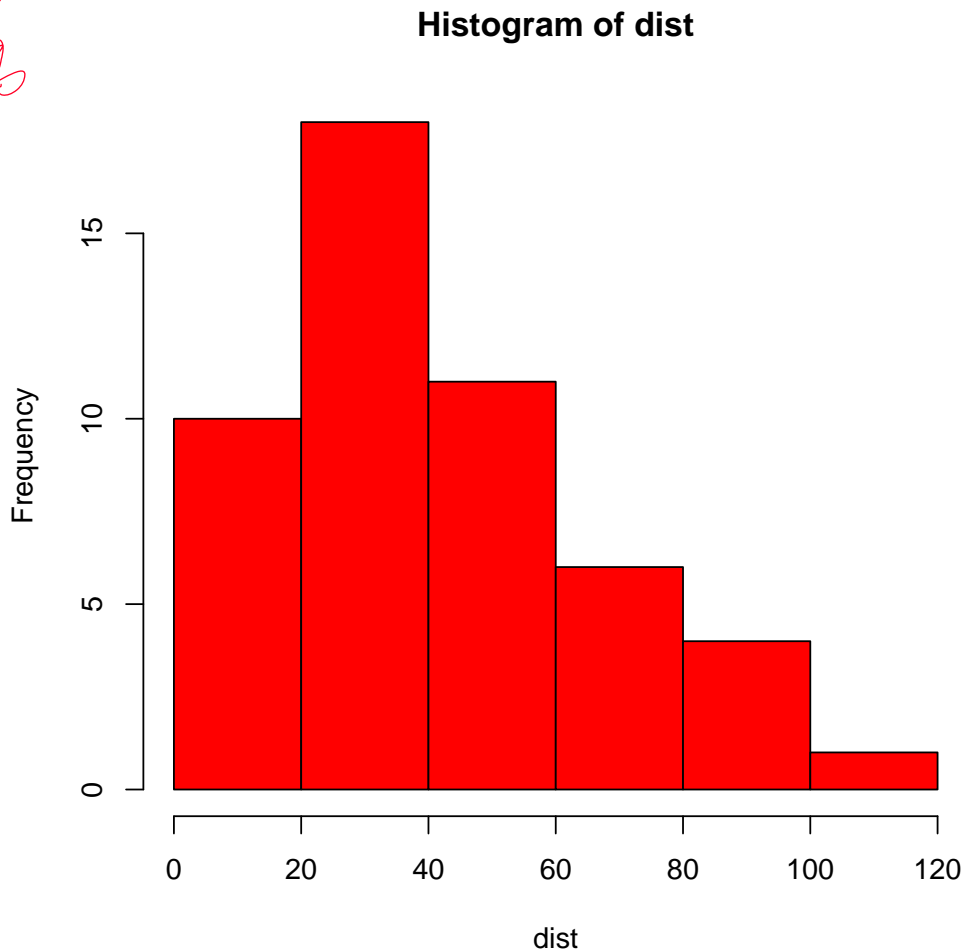
Further, the theoretical mean μ is taken to be equal to the sample mean under the proposed model. The theoretical variance μ^2 is again taken equal to the sample variance under the proposed model.

(c) Now $P(X > x) = e^{-\frac{x}{\mu}}$. Thus $P(X > 15)$ comes out to be

```
[1] 0.0104067
```

Answer 6

4/5 (handwritten)

(a) Plotting the histogram for the cars data:

Histogram of dist

From the histogram, it is clear that the support is R^+. The shape of the distribution indicates that gamma distribution would be appropriate.. Now the gamma distribution has two parameters α and β with the pdf as:

$f(x) = \frac{\beta^\alpha}{\Gamma\alpha} e^{-\beta x} x^{(\alpha-1)}, x > 0$

But Data Are Discrete! (handwritten)

The theoretical mean and variance are given as:

$E[X] = \frac{\alpha}{\beta}$ and $Var[X] = \frac{\alpha}{\beta^2}$ The sample mean and variance are

```
[1] 42.98
```

```
[1] 664.0608
```

respectively.

Comparing the sample mean and variance with the theoretical mean and variance, we get the parameter values as

```
[1] 2.781794
```

```
[1] 0.06472299
```

(b) The data clearly shows that all the sample values are greater than zero. They lie in the following range:

```
[1] 2
```

```
[1] 120
```

i.e. the sample data has the support as [2,120]. This is a subset of the theoretical support R^+.

Further, the theoretical mean is taken to be equal to the sample mean under the proposed model. The theoretical variance is again taken equal to the sample variance under the proposed model.

(c) $P(X > 80)$ comes out to be

```
[1] 0.08919011
```

The observed frequency is

```
[1] 0.1
```

Answer 7

(a)
```
       [,1]
[1,]    5
[2,]    2
[3,]    0
[4,]   16
```

(b)
```
       [,1]
[1,]    6
```

(c)
```
       [,1] [,2] [,3]
[1,]    4    2   -2
[2,]    2    1   -1
[3,]   -2   -1    1
```

(d)
```
     [,1] [,2] [,3]
[1,]   92   -9   20
[2,]   -9   31    8
[3,]   20    8   18
```

(e) `[1] 141`

(f)
```
              [,1]        [,2]        [,3]
[1,]   0.01720655  0.01121560 -0.02410310
[2,]   0.01121560  0.04374782 -0.03190526
[3,]  -0.02410310 -0.03190526  0.09651689
```