Solutions to Assignment #3

1.

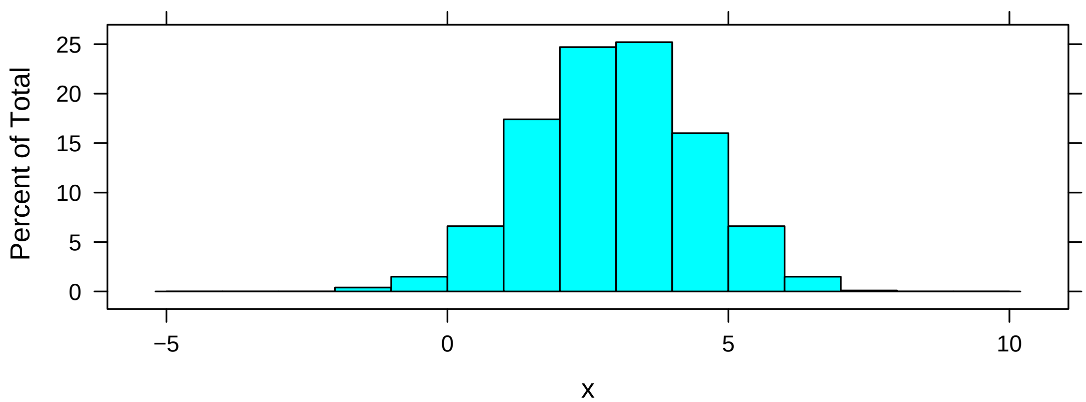
Solution:

Normal sample from N$(3, 2)$.

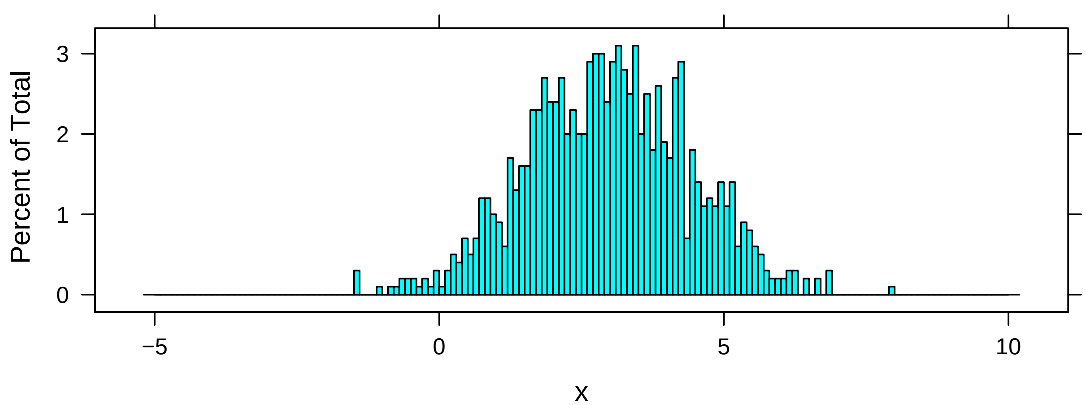
```
> x = rnorm(1000, 3, sqrt(2))
```

Two histograms with different width intervals.

```
> library(lattice)
> print(histogram(~x, breaks = seq(-5, 10, 1)))
```

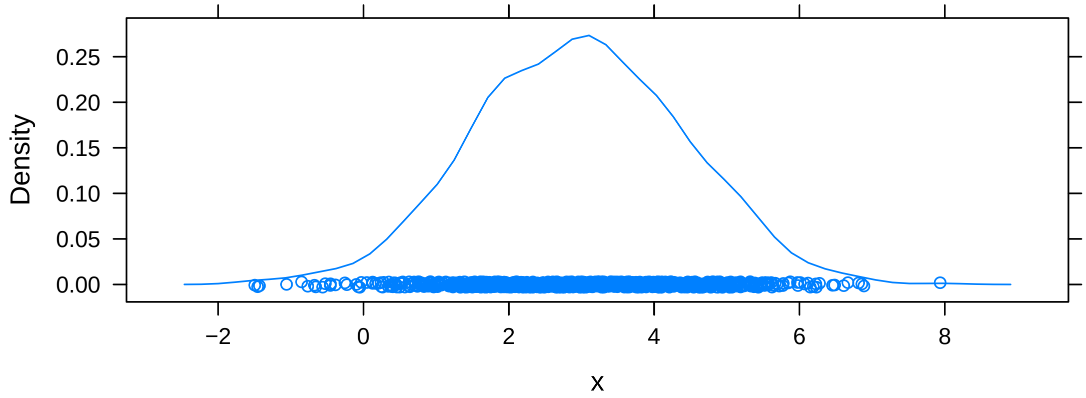



```
> print(histogram(~x, breaks = seq(-5, 10, 0.1)))
```



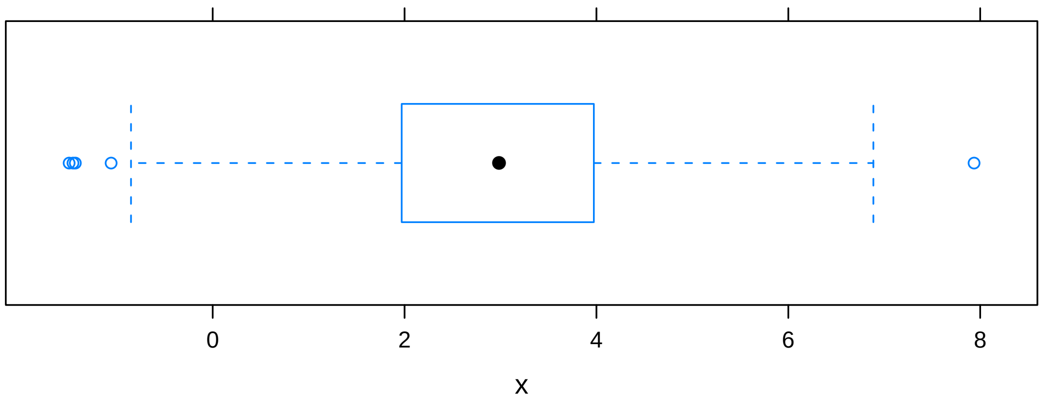
A density plot for comparison.

```
> print(densityplot(~x))
```



A box-and-whisker plot.

```
> print(bwplot(~x))
```



Some numerical summaries.

```
> fivenum(x)

[1] -1.496412  1.969921  2.984676  3.973052  7.936133

> mean(x)

[1] 2.970241

> sd(x)

[1] 1.428932

> var(x)

[1] 2.041847
```

```
> median(x)

[1] 2.984676
```

2. Do Exercise 5.5.5 using R.

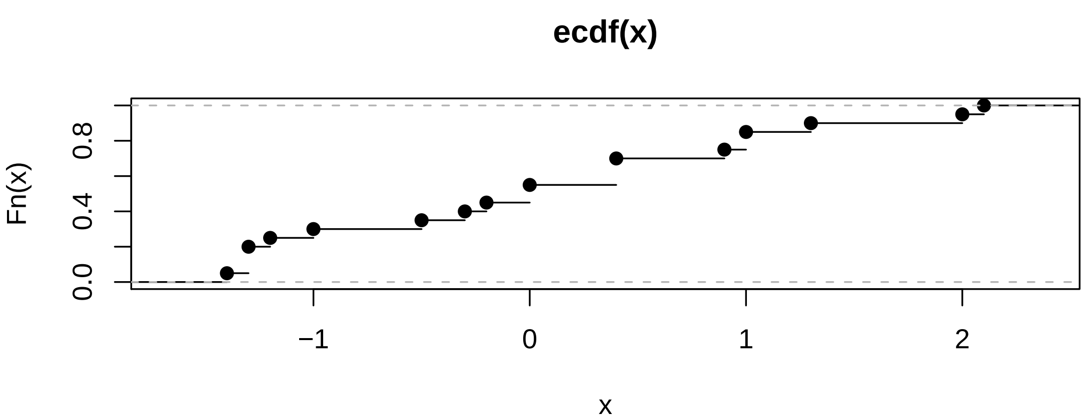
Solution: The empirical function is the proportion of data less than or equal to x.

```
> x = c(1, -1.2, 0.4, 1.3, -0.3, -1.4, 0.4, -0.5, -0.2, -1.3, 0,
+     -1, -1.3, 2, 1, 0.9, 0.4, 2.1, 0, -1.3)
> plot(ecdf(x))
> median(x)

[1] 0

> q0.25 = quantile(x, 0.25)
> q0.75 = quantile(x, 0.75)
> iqr = q0.75 - q0.25
> F1 = sum(x <= 1)/20
> print(c(q0.25, q0.75, iqr, F1))

   25%    75%    75%
-1.050  0.925  1.975  0.850
```



3. For a statistical model $\{f_\theta(s) : \theta \in \Omega\}$ with sufficient statistic $T(s)$, prove that for any two parameter values $\theta_1, \theta_2 \in \Omega$ and for any two data samples s_1 and s_2 such that $T(s_1) = T(s_2)$ that

$$\frac{L(\theta_1 \mid s_1)}{L(\theta_2 \mid s_1)} = \frac{L(\theta_1 \mid s_2)}{L(\theta_2 \mid s_2)} .$$

Solution: Let $T(s)$ be a sufficient statistic and let s_1 and s_2 be two samples such that $T(s_1) = T(s_2)$. Then, by the definition of sufficiency,

$$\frac{L(\theta \mid s_1)}{L(\theta \mid s_2)} = c(s_1, s_2)$$

for all θ. In particular, this ratio is equal if we plug in θ_1 or θ_2. Hence,

$$\frac{L(\theta_1 \,|\, s_1)}{L(\theta_1 \,|\, s_2)} = \frac{L(\theta_2 \,|\, s_1)}{L(\theta_2 \,|\, s_2)}$$

and if we rearrange factors,

$$\frac{L(\theta_1 \,|\, s_1)}{L(\theta_2 \,|\, s_1)} = \frac{L(\theta_1 \,|\, s_2)}{L(\theta_2 \,|\, s_2)}$$

which completes the proof.

4. Let $X_1, \ldots, X_n$ be an i.i.d. sample of size n from a Uniform$(-\theta, \theta)$ distribution.

   (a) Find a minimal sufficient statistic for θ.

   Solution: The likelihood function is

$$\begin{aligned} L(\theta \,|\, s) &= \prod_{i=1}^{n} \left(\frac{1}{2\theta}\right) 1_{\{-\theta \le x_i \le \theta\}} \\ &= \prod_{i=1}^{n} \left(\frac{1}{2\theta}\right) 1_{\{|x_i| \le \theta\}} \\ &= (2\theta)^{-n} 1_{\{\max_i |x_i| \le \theta\}} \end{aligned}$$

   By the factorization theorem, $\max_i |x_i|$ is a sufficient statistic. Furthermore, for samples $s_1 = (x_1, \ldots, x_n)$ and $s_2 = (y_1, \ldots, y_n)$, if the likelihood ratio

$$\frac{L(\theta \,|\, s_1)}{L(\theta \,|\, s_2)} = \frac{(2\theta)^{-n} 1_{\{\max_i |x_i| \le \theta\}}}{(2\theta)^{-n} 1_{\{\max_i |y_i| \le \theta\}}}$$

   does not depend on θ, it must be the case that the numerator and dominator are non-zero for the same range, so that $\max_i |x_i| = \max_i |y_i|$ which implies that $\max_i |x_i|$ is a minimal sufficient statistic as well.

   (b) Find a maximum likelihood estimate of θ.

   Solution: Note that $L(\theta \,|\, s) = (2\theta)^{-n} 1_{\{\max_i |x_i| \le \theta\}}$ is zero for $\theta < \max_i |x_i|$, is positive for $\theta = \max_i |x_i|$, and is a decreasing function in the interval $(\max_i |x_i|, \infty)$. This latter statement can be shown explicitly by either taking the derivative

$$\frac{\mathrm{d}L(\theta \,|\, s)}{\mathrm{d}\theta} = -2n(2\theta)^{-n-1} < 0 \quad \text{for } \theta \ge \max_i x_i$$

   or by showing for any numbers θ_1, θ_2 where $\max_i |x_i| \le \theta_1 < \theta_2$ that $L(\theta_1 \,|\, s)/L(\theta_2 \,|\, s) = (\theta_2/\theta_1)^n > 1$ for $n \ge 1$.

5. Let α_0 be a known fixed constant and let $X_1, \ldots, X_n$ be an i.i.d. sample from a Gamma distribution with density for a single observation

$$f_\theta(x) = \frac{\theta^{\alpha_0}}{\Gamma(\alpha_0)} x^{\alpha_0 - 1} \mathrm{e}^{-\theta x}, \quad \text{for } x > 0,\ \theta > 0.$$

(a) Find a minimal sufficient statistic for θ.

Solution: The likelihood is

$$\begin{aligned} L(\theta \mid s) &= \frac{\theta^{n\alpha_0}}{(\Gamma(\alpha_0))^n} \left(\prod_{i=1}^n x_i^{\alpha_0 - 1} \right) \mathrm{e}^{-\theta \sum_{i=1}^n x_i} \\ &= \left(\frac{\theta^{n\alpha_0}}{(\Gamma(\alpha_0))^n} \mathrm{e}^{-\theta \sum_{i=1}^n x_i} \right) \left(\prod_{i=1}^n x_i^{\alpha_0 - 1} \right) \end{aligned}$$

and by the factorization theorem, $T(s) = \sum_{i=1}^n x_i$ is sufficient. (Note that if α_0 were unknown, then the sum would not be sufficient.)
If $s_1 = (x_1, \ldots, x_n)$ and $s_2 = (y_1, \ldots, y_n)$ are two samples and if the ratio

$$\frac{L(\theta \mid s_1)}{L(\theta \mid s_2)} = \left(\mathrm{e}^{-\theta(\sum_{i=1}^n x_i - \sum_{i=1}^n y_i)} \right) \left(\prod_{i=1}^n (x_i / y_i)^{\alpha_0 - 1} \right)$$

does not depend on θ, then it must be the case that $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i$. Thus, $T(s) = \sum_{i=1}^n x_i$ is a minimal sufficient statistic.

(b) Find a maximum likelihood estimate of θ.

Solution:

$$\begin{aligned} L(\theta \mid s) &= \frac{\theta^{n\alpha_0}}{(\Gamma(\alpha_0))^n} \left(\prod_{i=1}^n x_i^{\alpha_0 - 1} \right) \mathrm{e}^{-\theta \sum_{i=1}^n x_i} \\ \ell(\theta \mid s) &= n\alpha_0 \log(\theta) - n \log(\Gamma(\alpha_0)) + (\alpha_0 - 1) \sum_{i=1}^n \log(x_i) - \theta \sum_{i=1}^n x_i \\ \frac{\partial \ell(\theta \mid s)}{\partial \theta} &= \frac{n\alpha_0}{\theta} - \sum_{i=1}^n x_i = 0 \\ \hat{\theta} &= \frac{\alpha_0}{\bar{x}} \end{aligned}$$

6. Let $X_1, \ldots, X_n$ be an i.i.d. sample from a Beta$(\theta, 1)$ distribution.

(a) Find a minimal sufficient statistic for θ.

Solution: The likelihood for the sample is

$$L(\theta \mid s) = \theta^n \left(\prod_{i=1}^n x_i \right)^{\theta - 1}$$

so $\prod_{i=1}^n x_i$ is a sufficient statistic by the factorization theorem. Note that $\prod_{i=1}^n x_i = \exp(\sum_{i=1}^n \log(x_i))$ which implies that $\sum_{i=1}^n \log(x_i)$ is also a sufficient statistic.
To show this is a minimal sufficient statistic, let $s_1 = (x_1, \ldots, x_n)$ and $s_2 = (y_1, \ldots, y_n)$ be two samples. If the likelihood ratio

$$\frac{L(\theta \mid s_1)}{L(\theta \mid s_2)} = \left(\prod_{i=1}^n (x_i / y_i) \right)^{\theta - 1}$$

does not depend on θ, it must be that

$$\prod_{i=1}^{n} x_i = \prod_{i=1}^{n} y_i$$

which proves that $T(s) = \prod_{i=1}^{n} x_i$ is a minimal sufficient statistic.

(b) Find a maximum likelihood estimate of θ.

Solution:

$$\begin{aligned} L(\theta \mid s) &= \theta^n \left(\prod_{i=1}^{n} x_i \right)^{\theta - 1} \\ \ell(\theta \mid s) &= n \log(\theta) + (\theta - 1) \sum_{i=1}^{n} \log(x_i) \\ \frac{\partial \ell(\theta \mid s)}{\partial \theta} &= \frac{n}{\theta} + \sum_{i=1}^{n} \log(x_i) \\ \hat{\theta} &= -\frac{n}{\sum_{i=1}^{n} \log(x_i)} \end{aligned}$$