STK 210: Practical 6

1 Density Functions

In this practical session we are going to focus on two continuous density functions namely:

1. The uniform distribution, $X \sim UNIF(a,b)$
2. The exponential distribution, $X \sim EXP(\mu)$

It is important that you should revise the PDF, CDF, QUANTILE and RANDFUN functions discussed in Practical 5:

1.1 The Uniform Distribution, $X \sim UNIF(a,b)$

1.1.1 The Probability Density Function (pdf)

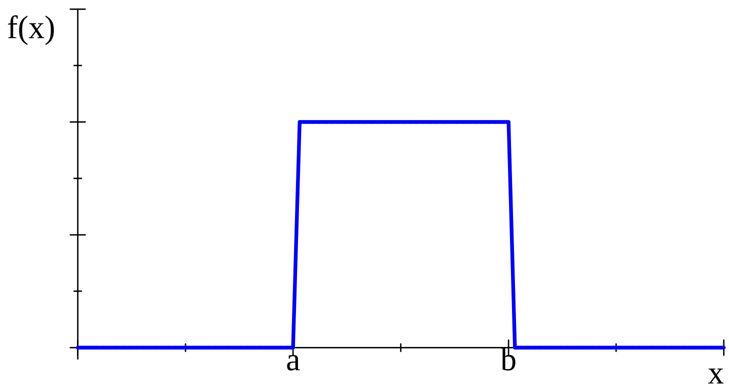
- The pdf:

$$f(x;a,b) = \begin{cases} \frac{1}{b-a} & \text{for } a \le x \le b \\ 0 & \text{elsewhere} \end{cases} \tag{1}$$

- Notation:

$$X \sim UNIF(a,b)$$

- Graph:



1.1.2 The Cumulative Distribution Function (CDF)

- For $x < a$:

$$F(x) = \int_{-\infty}^{x} 0 dt = 0$$

- For $a \le x \le b$:

$$\begin{aligned} F(x) &= \int_0^x \frac{1}{b-a} dt \\ &= \left. \frac{t}{b-a} \right|_\alpha^x = \frac{x-a}{b-a} \end{aligned}$$

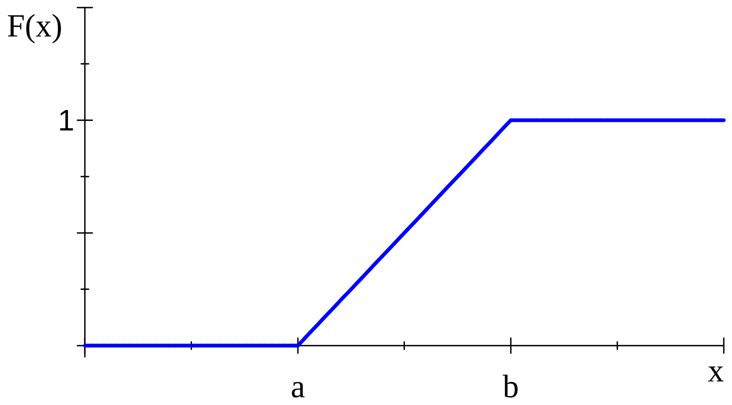
- For $x < a$:

$$F(x) = F(b) + F(x) = \frac{b-a}{b-\alpha} + \int_0^x 0dt = 1$$

- The CDF:

$$F(x) \begin{cases} 0 & \text{for } x < \alpha \\ \dfrac{x-a}{b-\alpha} & \text{for } a \leq x \leq b \\ 1 & \text{for } x > b \end{cases} \tag{2}$$

- Graph:



1.1.3 The Quantiles

- Let x_p denote the $p\,(100)^{th}$ quantile.
- The probability to the left of x_p is p therefore

$$\begin{aligned} p &= \int_{-\infty}^{x_p} \frac{1}{b-a} dt \\ &= \left. \frac{t}{b-a} \right|_a^{x_p} = \frac{x_p - a}{b-a}, \quad \text{for } a \leq x_p \leq b \end{aligned}$$

- Solve for x_p

$$\frac{x_p - a}{b-a} = p \implies : x_p - a = p\,(b-a)$$

$\therefore$

$$x_p = a + p\,(b-a) \quad \text{for } a \leq x_p \leq b \quad \text{Do you remember from STK110??} \tag{3}$$

1.1.4 The Moments

- Mean:

$$\mu = \frac{a+b}{2}$$

- Variance

$$\sigma^2 = \frac{(b-a)^2}{12}$$

Note: We are going to do moments in Chapter 4.

1.2 The Exponential distribution, $X \sim EXP(\mu)$

1.2.1 The Probability Density Function (pdf)

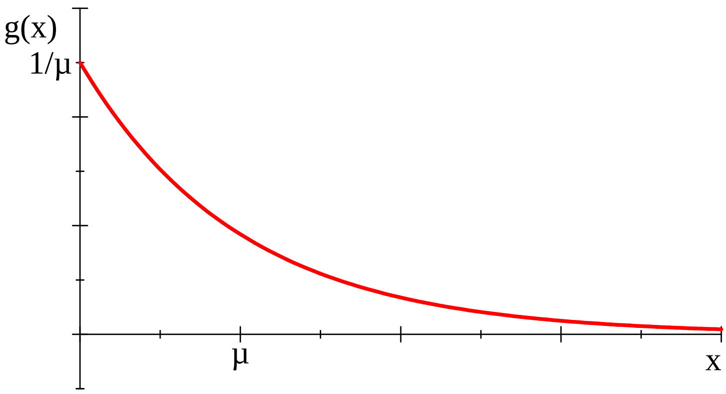
- The pdf:

$$g(x;\mu) = \begin{cases} \frac{1}{\mu} e^{-x/\mu} & \text{for } x > 0 \\ 0 & \text{elsewhere} \end{cases} \tag{4}$$

- Notation:

$$X \sim EXP(\mu)$$

- Graph:



1.2.2 The Cumulative Distribution Function (CDF)

- For $x < 0$:

$$G(x) = \int_{-\infty}^{x} 0 dt = 0$$

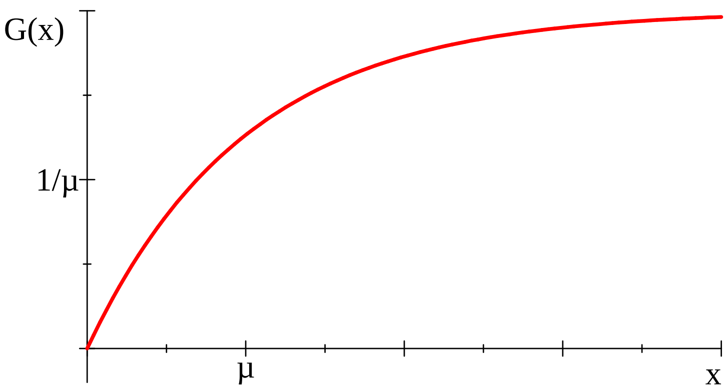
- For $x \geq 0$:

$$\begin{aligned} G(x) &= G(0) + \int_0^x \frac{1}{\mu} e^{-t/\mu} dt \\ &= -e^{-t/\mu}\Big|_0^x = -\left(e^{-x/\mu} - e^0\right) = 1 - e^{-x/\mu} \end{aligned}$$

- The CDF:

$$G(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 - e^{-x/\mu} & \text{for } x \geq 0 \end{cases} \tag{5}$$

- Graph:



1.2.3 The Quantiles

- Let x_p denote the $p\left(100\right)^{th}$ quantile.
- The probability to the left of x_p is p therefore

$$\begin{aligned} p &= \int_0^{x_p} \frac{1}{\mu} e^{-t/\theta} dt \\ &= -e^{-t/\mu}\Big|_0^{x_p} = 1 - e^{-x_p/\mu}, \quad \text{for } x_p \geq 0 \end{aligned}$$

- Solve for x_p

$$e^{-x_p/\mu} = 1 - p \implies : -x_p/\mu = \ln\left(1-p\right)$$

$\therefore$

$$x_p = -\mu \ln\left(1-p\right) = \mu \ln \frac{1}{1-p} \quad \text{for } x_p \geq 0 \tag{6}$$

1.2.4 The Moments

- Mean:

$$\mu$$

- Variance

$$\sigma^2 = \mu^2$$

Note: We are going to do moments in Chapter 4.

2 Exercise: Probability Densities

1. Let

$$X = \text{flight time of airplane from Johannesburg to Cape Town}$$

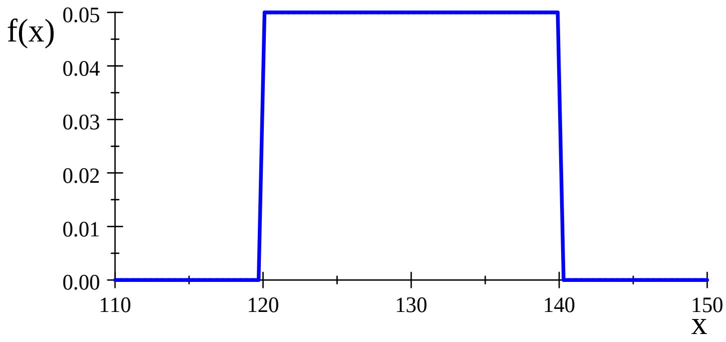
   - $X \sim UNIF\left(\alpha, \beta\right)$ is a continuous random variable.
   - **Assume**:
     - The flight time is within the interval 120 to 140 minutes.
     - The probability of a flight time within any one-minute interval is the same as the probability of a flight time within any other one-minute interval.
   - The **sample space** is:

$$S = \{x | 120 \leq x \leq 140\}$$

   - The probability density function (pdf) is :

$$f\left(x; 120, 140\right) = \begin{cases} \dfrac{1}{140-120} = \dfrac{1}{20} = 0.05 & \text{for } 120 \leq x \leq 140 \\ 0 & \text{elsewhere} \end{cases} \tag{7}$$

- Graph:



(a) Use the CDF function in SAS to calculate the following probabilities:

Syntax: `CDF('UNIFORM',x,`α`,`β`)`

i. Calculate the probability that the flight time will be at most 130 minutes i.e. calculate $P(X \leq 130)$. 0.5

ii. Calculate the probability that the flight time will be between 130 and 135 minutes i.e. calculate $P(130 \leq X \leq 135)$. 0.25

iii. Calculate the probability that the flight time will be more than 125 minutes i.e. calculate $P(X > 125)$. 0.75

(b) Use the QUANTILE function in SAS to calculate the following:

Syntax: `QUANTILE('UNIFORM',p,`α`,`β`)`

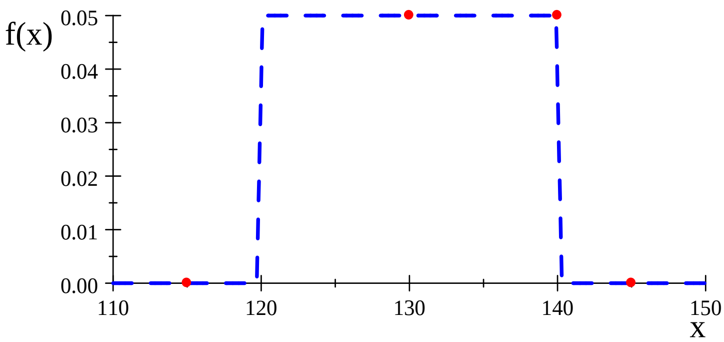
i. Q_1 125

ii. me 130

iii. Q_3 135

iv. IQR 10

(c) Use the pdf function to calculate the value of the density function at the following values of x.



(i)	$x = 115$	0
(ii)	$x = 130$	0.05
(iii)	$x = 140$	0.05
(iv)	$x = 145$	0

Syntax: `pdf('UNIFORM',x,`α`,`β`)`

(d) Calculate the value of the following:

i. The mean of X, i.e. $\mu = \dfrac{\alpha + \beta}{2}$. 130

ii. The variance of X, i.e. $\sigma^2 = \dfrac{(\beta - \alpha)^2}{12}$. 33.333

iii. The standard deviation of X, i.e. $\sigma = \sqrt{\text{Var}(X)}$. 5.7735

iv. The coefficient of variation, i.e. $CV = \dfrac{\sigma}{\mu} \times 100$. 4.4411

(e) Instead of using the PDF, CDF and the QUANTILE functions we are going to calculate the following the from basic mathematical principals.

i. $P(X \leq 130)$ use equation (2) See Q1(a)i

ii. $P(130 \leq X \leq 135)$ use equation (2) See Q1(a)ii

iii. Q_1 use equation (3) See Q2(b)i

iv. Q_3 use equation (3) See Q2(b)iii

v. $P(X = 130)$ use equation (1) See Q2(c)ii

(f) Use the RANDSEED call with a seed of $\boxed{789}$ to generate a sample of size $n = 10$ from $X \sim UNIF(120, 140)$. Store the generated values in the vector $\mathbf{q}$ and print.

(g) Use the RANDSEED call with a seed of $\boxed{987}$ to generate a sample of size $n = 1000$ from $X \sim UNIF(120, 140)$. Store the generated values in the vector $\mathbf{r}$.

i. Print the observations number 101 to 110 of the generated sample.

ii. Create a SAS dataset **Airplane** from the vector $\mathbf{r}$ with the 1000 values generated from $X \sim UNIF(120, 140)$. Use the variable name **time**.

iii. Use **PROC UNIVARIATE** to calculate the following from the empirical distribution of X.

Note: Check all the values below in the Output.

A. $\bar{x} = 129.9488$, $s^2 = 33.5428$, $s = 5.7916$ and $CV = \dfrac{s}{\bar{x}} \times 100 = 4.4568$

Note: Compare answers with Question 1(d).

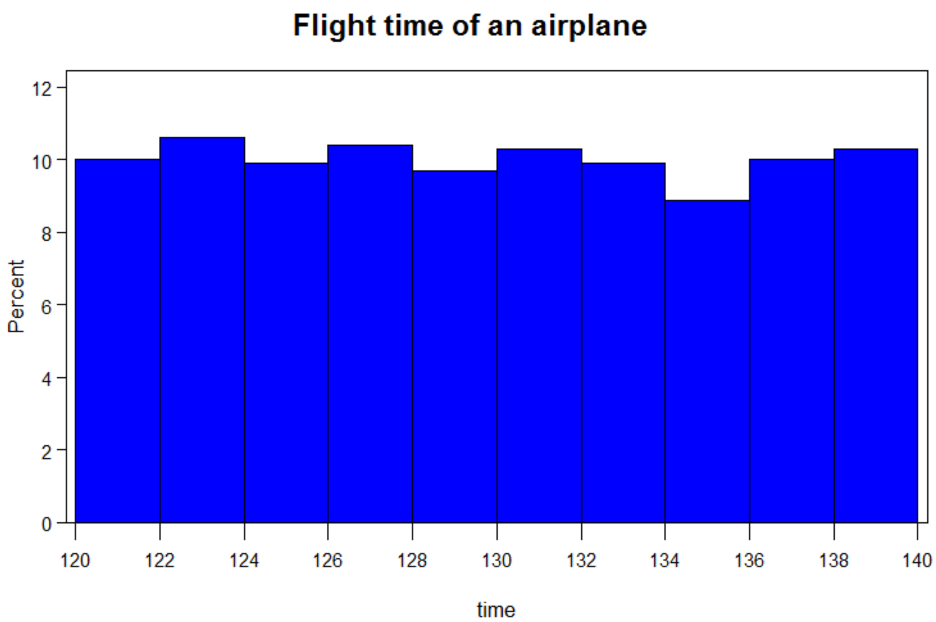
B. $Q_1 = 124.845$, $me = 129.929$, $Q_3 = 135.187$ and $\text{IQR} = Q_3 - Q_1 = 10.3419$

Note: Compare answer with Question 1(b).

C. Draw a graph by making use of the HISTOGRAM statement in PROC UNIVARIATE

```
histogram / endpoints=120 to 140 by 2 cfill=blue;
```

Graph:



(h) Determine the following probabilities from the **empirical distribution**.

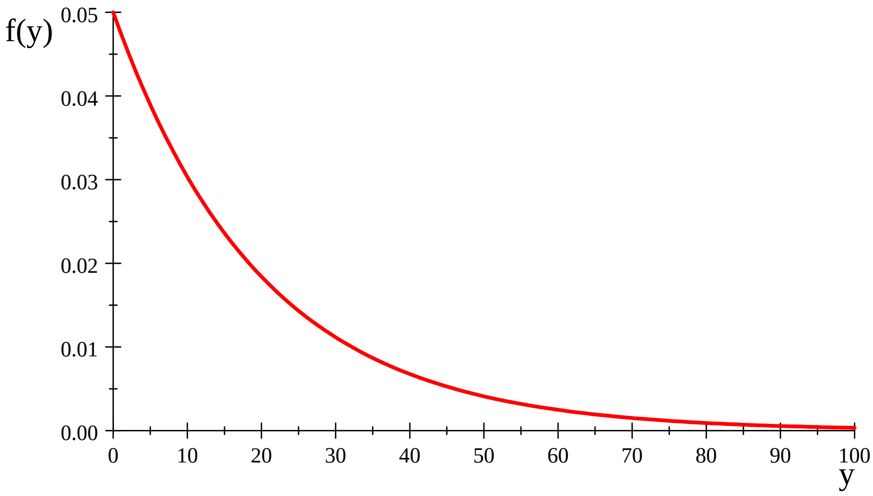
i. $P(X \leq 130)$ 0.506

ii. $P(130 \leq X \leq 135)$ 0.238

iii. $P(X > 125)$ 0.739

2. The lifetime, Y, of an alkaline battery has an exponential distribution with a mean lifetime of $\theta = 20$ hours. The probability density function is as follows

$$f(y) = \begin{cases} \frac{1}{20}e^{-y/20} & y > 0 \\ 0 & \text{elsewhere} \end{cases}$$
$$Y \sim EXP(\theta)$$



(a) Use the CDF function in SAS to calculate the following probabilities:

Syntax: $\boxed{\texttt{CDF('EXPONENTIAL',x,}\theta\texttt{)}}$

i. $P(Y \leq 40)$ 0.8646647

ii. $P(10 \leq Y < 30)$ 0.3834005

iii. $P(Y > 25)$ 0.2865048

iv. $P(|Y - 20| \leq 5)$ 0.1858618

v. $P(Y > 35|Y > 10)$ 0.2865048

Note: Compare answer of 2(a)v with 2(a)iii. What do we call this property?

(b) Use the QUANTILE function in SAS to calculate the following:

Syntax: $\boxed{\texttt{QUANTILE('EXPONENTIAL',p,}\theta\texttt{)}}$

i. P_{10} 2.1072103

ii. Q_1 5.7536414

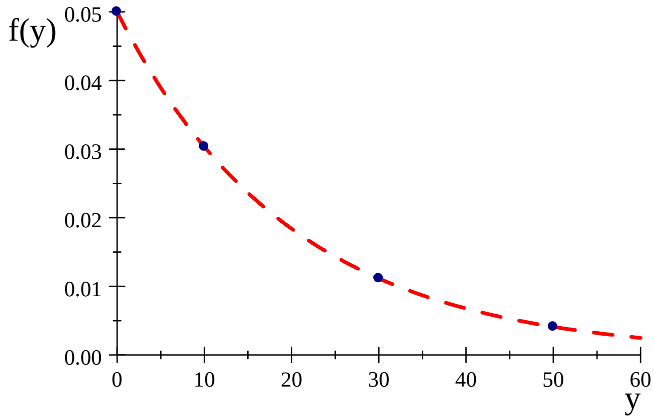
iii. me 13.862944

iv. Q_3 27.725887

v. P_{90} 46.051702

vi. IQR 21.972246

(c) Use the pdf function to calculate the value of the density function at the following values of y.



(i)	$y = 0$	0.05
(ii)	$y = 10$	0.0303265
(iii)	$y = 30$	0.0111565
(iv)	$y = 50$	0.004104

Syntax: $\boxed{\texttt{pdf('EXPONENTIAL',x,}\theta\texttt{)}}$

(d) Calculate the value of the following:

i. The mean of Y, i.e. $\mu = E(Y)$. 20

ii. The variance of Y, i.e. $\sigma^2 = \text{Var}(Y)$. 400

iii. The standard deviation of Y, i.e. $\sigma = \sqrt{\text{Var}(Y)}$. 20

iv. The coefficient of variation, i.e. $CV = \frac{\sigma}{\mu} \times 100$. 100

(e) Instead of using the PDF, CDF and the QUANTILE functions we are going to calculate the following the from basic mathematical principals.

i. $P(Y \leq 40)$ use equation (5) See Q2(a)i

ii. $P(10 \leq Y < 30)$ use equation (5) See Q2(a)ii

iii. $P(Y > 25)$ use equation (5) See Q2(a)iii

iv. me use equation (6) See Q2(b)iii

v. Q_3 use equation (6) See Q2(b)iv

vi. $P(Y = 10)$ use equation (4) See Q2(c)ii

(f) Use the RANDSEED call with a seed of $\boxed{123}$ to generate a sample of size $n = 10$ from $Y \sim EXP(20)$. Store the generated values in the vector $\mathbf{v}$ and print.

(g) Use the RANDSEED call with a seed of $\boxed{123}$ to generate a sample of size $n = 1000$ from $Y \sim EXP(20)$. Store the generated values in the vector $\mathbf{w}$.

i. Print the observations number 251 to 260 of the generated sample.

ii. Create a SAS dataset **Battery** from the vector $\mathbf{w}$ with the 1000 values generated from $Y \sim EXP(20)$. Use the variable name **hours**.

iii. Use **PROC UNIVARIATE** to calculate the following from the empirical distribution of Y.

Note: Check all the values below in the Output.

A. $\bar{y} = 19.3043$, $s^2 = 379.2230$, $s = 19.4736$ and $CV = \frac{s}{\bar{y}} \times 100 = 100.8773$

Note: Compare answers with Question 2(d).

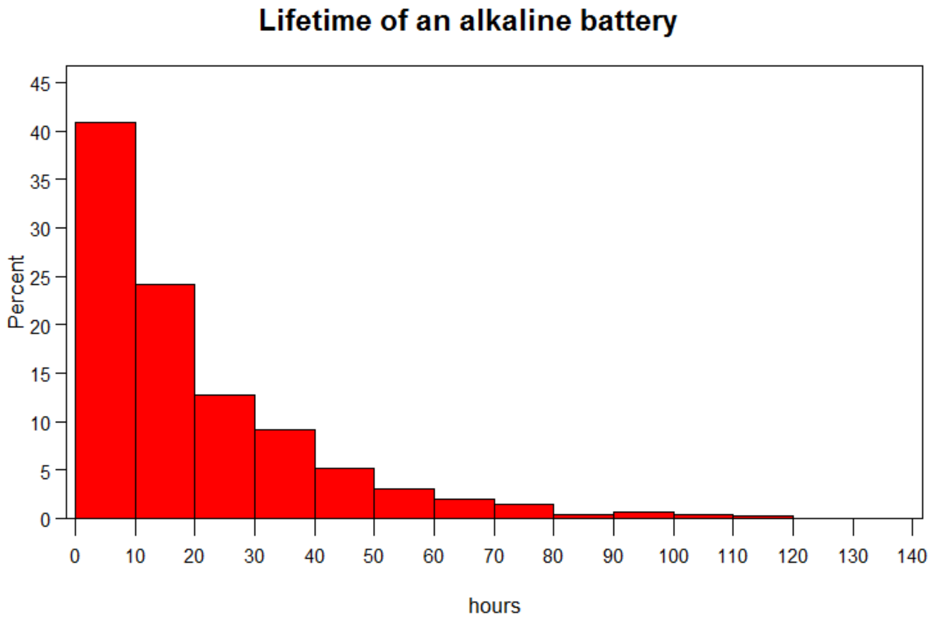
B. $P_{10} = 1.90539$, $Q_1 = 5.3396$, $me = 12.6561$, $Q_3 = 26.3619$, $P_{90} = 44.9846$ and $\text{IQR} = Q_3 - Q_1 = 21.0222$

Note: Compare answer with Question 2(b).

C. Draw a graph by making use of the HISTOGRAM statement in PROC UNIVARIATE

```
histogram / endpoints=0 to 140 by 10 cfill=red;
```

Graph:



(h) Calculate the following probabilities from the **empirical distribution** of Y.

i. $P(Y \leq 40)$ 0.869 (See Q2(a)i)

ii. $P(10 \leq Y < 30)$ 0.369 (See Q2(a)ii)

iii. $P(Y > 25)$ 0.280 (See Q2(a)iii)

iv. $P(|Y - 20| \leq 5)$ 0.154 (See Q2(a)iv)

v. $P(Y > 35 | Y > 10)$ 0.291 (See Q2(a)v)

- Since this is a conditional probability i.e. given the lifetime is greater than 10 hours, we need to work with a **subset** of the dataset.
- To accomplish this, we can use a new PROC FREQ or PROC MEANS with a WHERE statement i.e.

```
where hours>10;
```

- This will ensure that only lifetimes greater than 10 will be selected.