Chapter 4

Exponential population distribution

4.1 Definition

If the random variable X has an Exponential population distribution,

i.e. $X \sim Ex(\mu)$, then its pdf is given by

$$\boxed{f_x(x) = \frac{1}{\mu} e^{-\frac{x}{\mu}}} \qquad \text{for } x > 0$$

where $\mu > 0$ is the population mean of X.

Note sometimes the pdf is written using parameter $\lambda = \frac{1}{\mu}$, giving pdf, $f_x(x) = \lambda e^{-\lambda x}$.

An exponential distribution is often a good model for lifetimes of machines/light bulbs, and for waiting times or interarrival times to an event.

Examples of exponential distributions

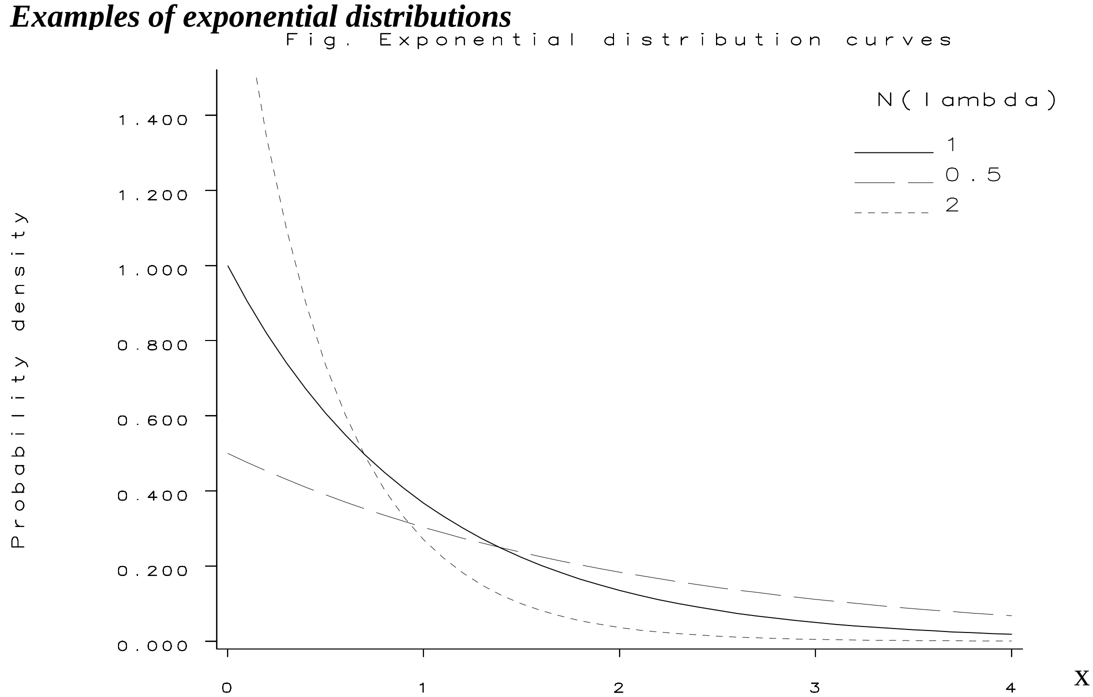


Figure: Showing different exponential distributions $f_x(x) = \lambda e^{-\lambda x}$

4.2 Population summary measures for an Exponential random variable

If $X \sim Ex(\mu)$ then the pdf of X is

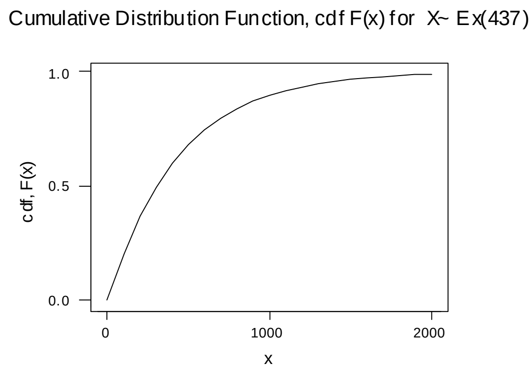
$$f_x(x) = \frac{1}{\mu} e^{-\frac{x}{\mu}} \qquad \text{for } x > 0 \text{ where } \mu > 0$$

The following can be derived from the *population pdf* **:**

cdf	$F_x(x) = 1 - e^{-\frac{x}{\mu}}$
mean	$\mu_X = \mu$
standard deviation	$\sigma_X = \mu$
median, m	$m = 0.693\mu$
lower quartile, Q_1	$Q_1 = 0.288\mu$
upper quartile, Q_3	$Q_3 = 1.386\mu$
semi-interquartile range,	$\boldsymbol{SIR} = \frac{Q_3 - Q_1}{2} = 0.549\,\mu$
skewness, sk	$sk = \frac{\text{midquartile - median}}{SIR} = 0.262$

Derivation of the population cdf for $X \sim Ex(\mu)$

$F_x(x) = 1 - e^{-\frac{x}{\mu}}$ for $x > 0$



Derivation of the population median, *m* for $X \sim Ex(\mu)$

$$F_X(m) = 0.5 \Rightarrow 1 - e^{-\frac{m}{\mu}} = 0.5$$

$$\Rightarrow 1 - e^{-\frac{m}{\mu}} = 0.5$$

$$\frac{m}{\mu} = -\log_e(0.5) = 0.693$$

$$\Rightarrow \quad m = 0.693\mu$$

Similarly Q_1 and Q_3 can be derived.

4.3 Example

4.3.1 Data

The time period in days between major earthquakes (i.e. magnitude 7.5 or more on the Richter scale or casualties 1000 or more killed) was recorded from 16/12/1902 to 4/3/1977, a total of 63 earthquakes and therefore 62 time periods between earthquakes.

[From : Daly, Hand et al.(1995) p169 and Hand et al. 'Small data sets' p203, QUAKES.DAT]

840	157	145	44	33	121	150	280	434	736	584	887	263
1901	695	294	562	721	76	710	46	402	194	759	319	460
40	1336	335	1354	454	36	667	40	556	99	304	375	567
139	780	203	436	30	384	129	9	209	599	83	832	328
246	1617	638	937	735	38	365	92	82	220			

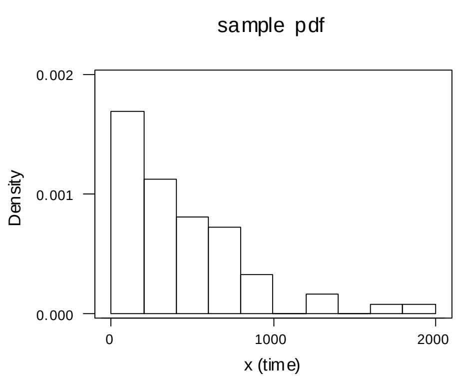
4.3.2 Compare the sample pdf with the population model pdf

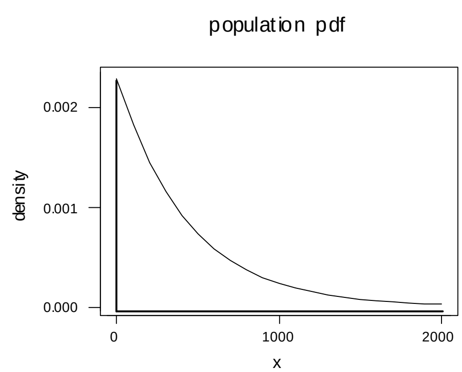
The sample pdf of the data is given on the left below. The sample mean of the data was 437.

The distribution appears Exponential so the data was modelled using the following distribution

$$X \sim Ex(437)$$

which is the population pdf given on the right below.



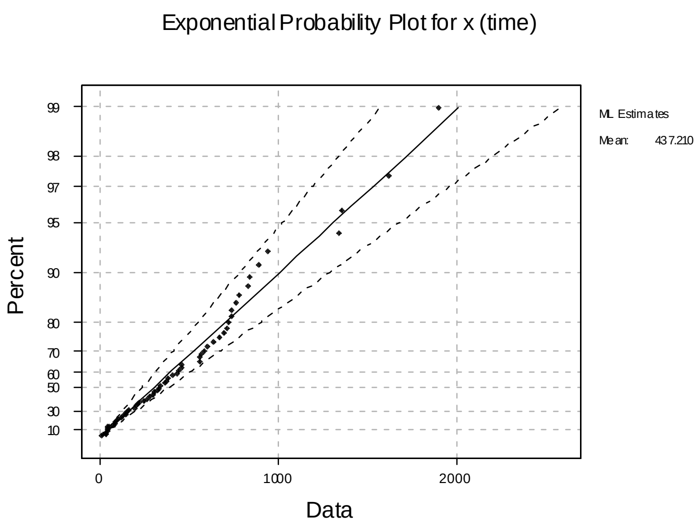


4.3.3 Compare the summary measures for the model with those of the sample.

Summary measure	Sample	Model
Mean	437	437
Standard deviation	400	437
Median	332	303

Lower Quartile	127	126
Upper Quartile	674	606
SIR	274	240
Skewness	0.13	0.26

4.3.4 Compare the sample with an Exponential distribution using a Probability Plot



If the Exponential model is OK, then the points plotted should roughly follow the middle line, lying within the two dashed bands. Here Exponential model appears OK, but not quite perfect.

4.3.5 Conclusion:

The Exponential model for the data appears adequate from the following:

i) the sample and model pdfs
ii) the sample and model summary measures
iii) the Exponential probability plot

4.3.6 Applying the model

The model could now be used to obtain estimated answers to questions relating to random variable X, the time in days between major earthquakes, assuming that

$$X \sim Ex(437)$$

Q1 Estimate the probability that the time between major earthquakes is less than 1 year.

$$p(X \le 365) = F_X(365) = 1 - e^{-\frac{365}{\mu}} = 1 - e^{-\frac{730}{437}} = 0.566 \qquad \text{i.e. 56.6\% chance}$$

Q2 Estimate the probability that the time between major earthquakes is more than 2 years.

$$p(X \le 730) = F_x(730) = 1 - e^{-\frac{730}{\mu}} = 1 - e^{-\frac{730}{437}} = 0.812$$

$$p(X > 730) = 1 - 0.812 = 0.188 \qquad \text{i.e. 18.8\% chance}$$

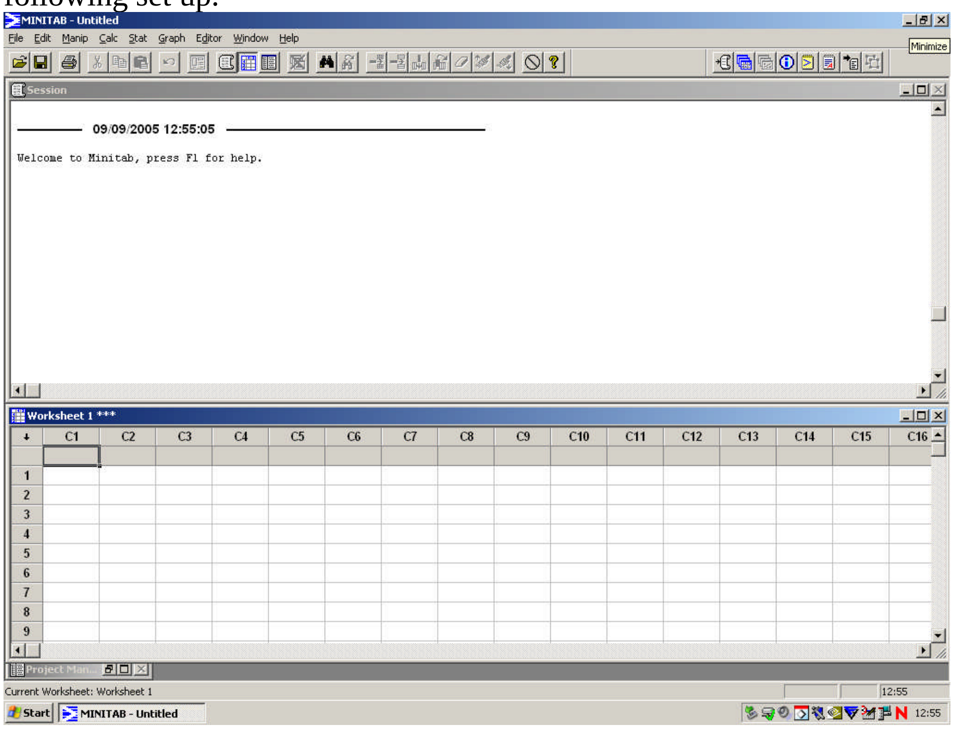
Q3 Estimate the probability the time between major earthquakes is between 1 and 2 years.

$p(365 < X < 730) = p(X \le 730) - p(X \le 365) = F_X(730) - F_X(365) =$

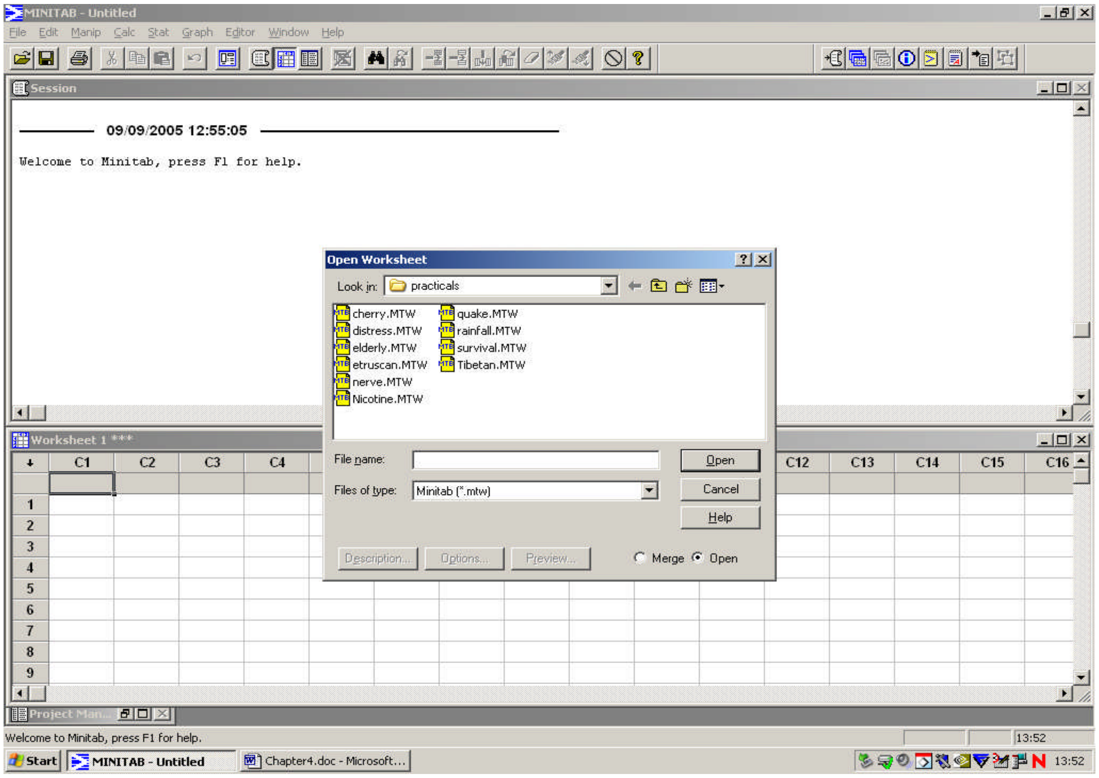
0.812-0.566=0.246

4.4 MINITAB commands for fitting population distributions

MINITAB is a command and menu driven package. Bu clicking the MINITAB icon you enter the following set up.



Commands are entered in the top half of the screen while the bottom half is for data. The menu options are obtained by clicking the relevant menus on the top of the MINITAB. For example by asking to import a MINITAB data file .MTW the follow screen will appear.



The following notation will be used:

> indicates click a menu option

| indicates an entry on a form [with prompt in lower case and **ENTRY** to be typed in **'UPPER CASE BOLD'** inside '']

4.4.1 To obtain the results in the Earthquake Exponential example above the following MINITAB commands were used:

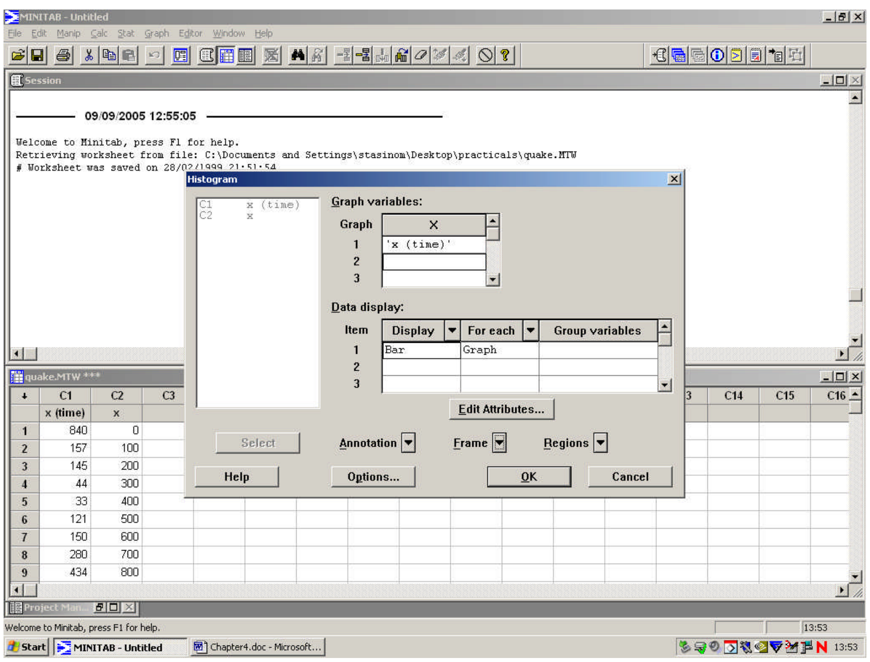
In C1 type the raw data values 840, 157, 145 ….., 220
In C2 type new values 0, 100, 200, 300, ……2000

x (time)	x	density, f(x)	cdf, F(x)
840	0	0.0022883	0.000000
157	100	0.0018203	0.204539
145	200	0.0014480	0.367241
44	300	0.0011518	0.496665
33	400	0.0009162	0.599616
121	500	0.0007288	0.681510
150	600	0.0005797	0.746654
280	700	0.0004612	0.798473
434	800	0.0003668	0.839693

736	900	0.0002918	0.872482
584	1000	0.0002321	0.898564
887	1100	0.0001846	0.919312
263	1200	0.0001469	0.935816
1901	1300	0.0001168	0.948944
695	1400	0.0000929	0.959387
294	1500	0.0000739	0.967694
562	1600	0.0000588	0.974302
721	1700	0.0000468	0.979558
76	1800	0.0000372	0.983739
710	1900	0.0000296	0.987065
46	2000	0.0000235	0.989711
402			
194			
.			
.			
.			
82			
220			

4.4.2 To produce the sample pdf plot from C1 with annotation and frame:

> Graph > Histogram > Simple



| Graph Variables **’C1’**

\> Scale > Y-Scale Type | ⦿ Density | OK
\> Labels > Titles/Footnotes | Title **’SAMPLE PDF’** > OK
\> Data View > Data Display ✓ Bars > OK
\> OK

To frame the graph
Edit the y-scale of the histogram by double clicking on it.
\> Scale | ⦿ Number of ticks **’3’ |** Minimum **‘0’** | Maximum **‘0.002’**
\> OK
Edit the x-scale by double-clicking on it
\> Scale | ⦿ Position of ticks **‘0 1000 2000’**
\> Binning | ⦿ Cutpoint
\> OK

4.4.3 To produce the Exponential probability plot:

\> Graph > Probability Plot > Single
| ‘C1’ |
\> Distribution > Exponential > OK
\> OK

4.4.4 To calculate the values of the Exponential pdf *f*(*x*) in C3 corresponding to values of x in C2

\> Calc > Probability Distributions > Exponential
| ⦿ Probability Density | Scale **‘437’** | Input Column **‘C2’** | Optional Storage **‘C3’**
\> OK
Note to produce the cdf F(x) use | ⦿ Cumulative Probability |

4.4.5 To produce the Exponential pdf *f*(*x*) plot with annotation and frame:

\> Graph > Histogram > With Fit
| Graph Variables **’C1’**
\> Scale > Y-Scale Type | ⦿ Density > OK
\> Labels > Titles/Footnotes | Title **’POPULATION PDF’** > OK
\> Data View > Data Display ☐ Bars > Distribution > Exponential > OK
\> OK
or
\> Graph > Scatterplot > With Connect Line

| Y ‘C3’ | X ‘C2’ | Data Display ☐ Symbols ✓ Connect Line | OK
> OK

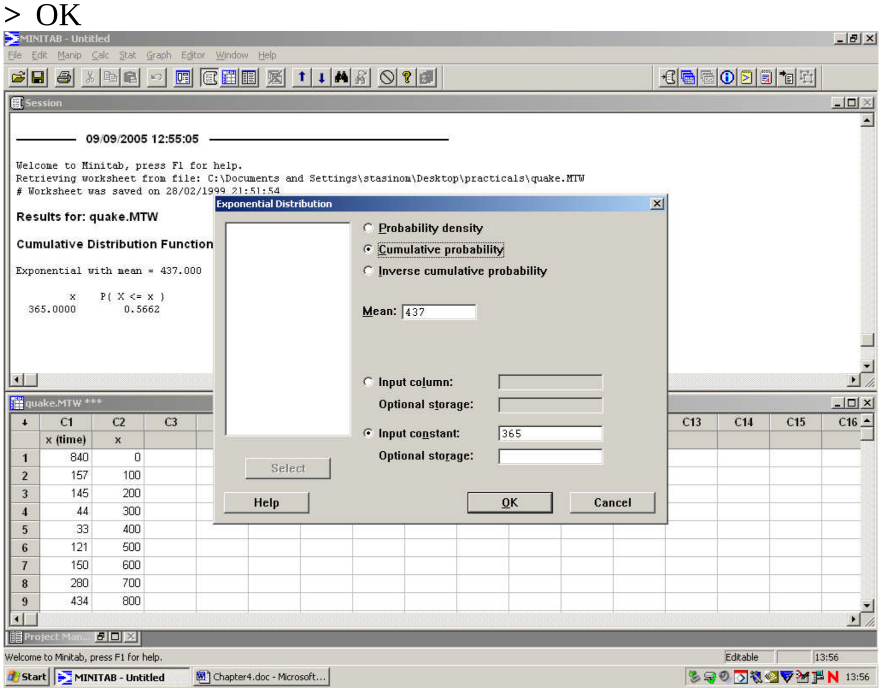
4.4.6 To produce the Exponential cdf $F(x)$ plot with annotation and frame:

> Graph > Empirical cdf > Single
| Graph Variables **’C1’**
> Distribution > Exponential > Data Display | ⦿ Distribution fit only >OK
> Scale > Y-Scale Type | ⦿ Probability > OK
> Labels > Titles/Footnotes | Title **’POPULATION CDF’** > OK
> OK

4.5 MINITAB commands for calculating pdf, cdf and inverse cdf values

4.5.1 Calculating cdf $F_X(x) = (X \le x) = p$ using MINITAB

Q1 To show that $p(X \le 365) = F_X(365) = 0.566$ when $X \sim Ex(437)$
> Calc > Probability Distributions > Exponential
| ⦿ Cumulative Probability | Scale **‘437’** | Input constant **‘365’**
> OK



Cumulative Distribution Function

```
Exponential with mean = 437.000

          x      P( X <= x)
   365.0000        0.5662
```

4.5.2 Calculating inverse cdf $F_X^{-1}(p)$ using MINITAB

Q2 To show that the median value of X is m = 302.9 when $X \sim Ex(437)$

$F_X(m) = p(X \le m) = 0.5$ so that $m = F_X^{-1}(0.5)$

\> Calc > Probability Distributions > Exponential

| ⦿ Inverse Cumulative Probability | Scale **'437'** | Input constant **'0.5'**

\> OK

Inverse Cumulative Distribution Function

```
Exponential with mean = 437.000

 P( X <= x)          x
    0.5000      302.9053
```

4.5.3 Calculating pdf $f_X(x)$ using MINITAB

Q3 To show that the pdf $f_X(365) = 0.001$ when $X \sim Ex(437)$

\> Calc > Probability Distributions > Exponential

| ⦿ Probability Density | Scale **'437'** | Input constant **'365'**

\> OK

Probability Density Function

```
Exponential with mean = 437.000

         x      P( X = x)
  365.0000        0.0010
```

PRACTICAL 4 Population Distributions

Q1 The times X between major earthquakes are stored in column C1 of QUAKE.MTW
Column C2 stores values 0,100,200,…., 2000.
[From: Hand et al. 'Small data sets', p203, data set 255]

a) Open MINITAB and input the saved worksheet QUAKE.MTW by
\> File > Open Worksheet
and finding the file K:\SCTMS\SOM\MA2010\PRACTICALS\QUAKE.MTW

b) Obtain **descriptive statistics** for the variable X (times between major earthquakes).

c) Follow the MINITAB commands given in the handout to obtain

i) a **histogram** of the variable X.

ii) an Exponential **probability plot** to check whether X can be modelled using an Exponential distribution.
Does an Exponential distribution provide a suitable model for the distribution of X?

iii) **calculate** the values of the **pdf of an Exponential** distribution with the population mean equal to the data sample mean (i.e. 437),
hence **plot** the **pdf** of the fitted Exponential, ***Ex*(437)** distribution.

iv) **calculate** the values of the **cdf of an Exponential** distribution with the population mean equal to the data sample mean (i.e.437),
hence **plot** the **cdf** of an Exponential, ***Ex*(437)** distribution

d) Use the MINITAB commands for calculating pdf, cdf and inverse cdf values
[i.e. **>** **Calc** **>** **Probability Distributions** **>** **Exponential** etc]
to calculate the following for the model $X \sim Ex(437)$:

i) $p(X \leq 365) = F_X(365)$

ii) $p(X \leq 730)$

iii) m the median of X

iv) Q_1 the lower quartile of X

v) Q_3 the upper quartile of X

Q2 The time intervals between successive pulses along a nerve fibre were measured in seconds. There were 800 pulses leading to 799 time intervals, stored in column C1 of worksheet NERVE.MTW. [From: Hand et al. ‘Small data sets’, p124, data set 160].

a) Open MINITAB and input the saved worksheet NERVE.MTW by
> File > Open Worksheet and finding the file NERVE.MTW from K:\etc

b) Obtain descriptive statistics for the variable X (time intervals between pulses]

c) Use MINITAB commands to obtain

i) a simple **histogram** of the variable X.

ii) Exponential **probability plot** to check whether *X* can be modelled using an Exponential distribution. Does an Exponential distribution provide a suitable model for the distribution of *X*?

ii) input **new values for *X*** (equally spaced along the sample range of *X*) and **calculate** the corresponding values of the **pdf** of an Exponential distribution (with the population mean equal to the data sample mean) and
hence **plot** the fitted Exponential distribution **pdf**.

Q3 The survival times *X* of patients suffering from a chronic disease measured in days were recorded for 43 patients, and stored in column C1 of worksheet SURVIVAL.MTW. [From: Hand et al. 'Small data sets', p140, data set 179].

a) Open saved worksheet SURVIVAL.MTW and **repeat b) and c) i) and ii) from Q2**
Does the Exponential distribution provide a suitable model for the distribution of *X*?

b) Use the **alternative probability plots** in MINITAB to see if you can find a suitable alternative model for the distribution of *X*, i.e. by > Graph > Probability Plot etc.

Q4 The maximum daily rainfall *X* was recorded for each of 47 successive years in Turramurra, Sydney, Australia, and stored in column C1 of worksheet RAINFALL.MTW. [From: Hand et al. 'Small data sets', p123, data set 157].

a) Open saved worksheet RAINFALL.MTW and **repeat b) and c) i) and ii) from Q2**
Does the Exponential distribution provide a suitable model for the distribution of *X*?

b) Use the **alternative probability plots** in MINITAB to see if you can find a suitable alternative model for the distribution of *X*, i.e. by > Graph > Probability Plot etc.