Extra exercise 8

Part (a)

In the Minitab code below we generate 3 columns of length 100, so that we can compare the findings from 3 replications of the procedure. Before doing that we set the seed (or base, in Minitab terminology). You need to use the same seed if you want to get exactly the same results. Next we display descriptive statistics and graphical summaries, as well as normal probability plots.

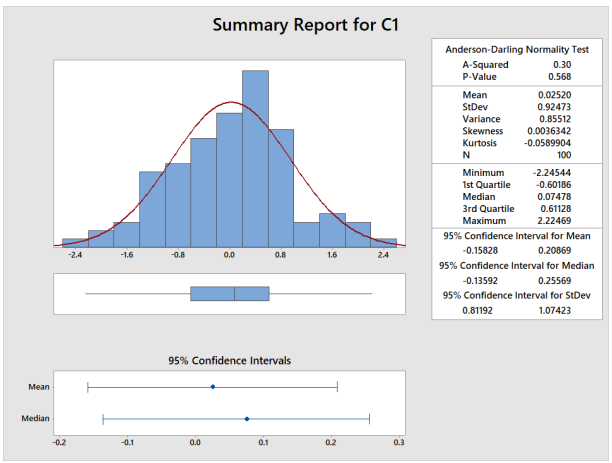
```
MTB > Base 140919.
MTB > Random 100 c1 c2 c3;
SUBC>   Normal 0.0 1.0.
MTB > Describe C1 C2 C3;
...
SUBC>   NMissing.
MTB > GSummary C1 C2 C3.
MTB > PPlot C1 C2 C3;
SUBC>   Normal;
SUBC>   Symbol;
SUBC>   FitD;
SUBC>   Grid 2;
SUBC>   Grid 1;
SUBC>   MGrid 1;
SUBC>   Panel.
```

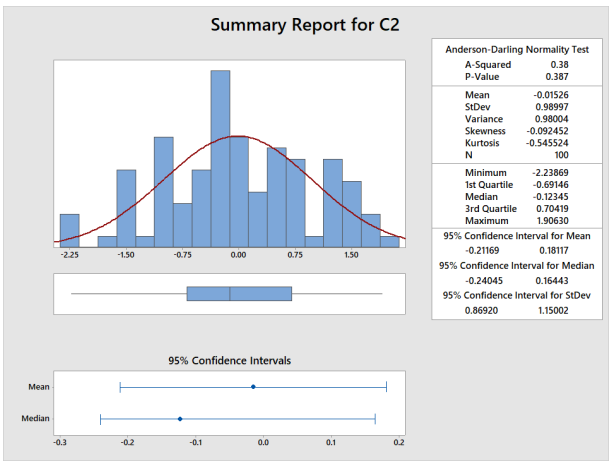
Descriptive Statistics: C1, C2, C3

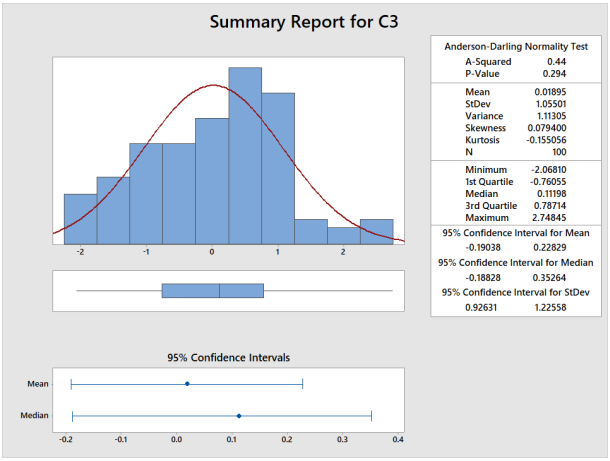
Statistics

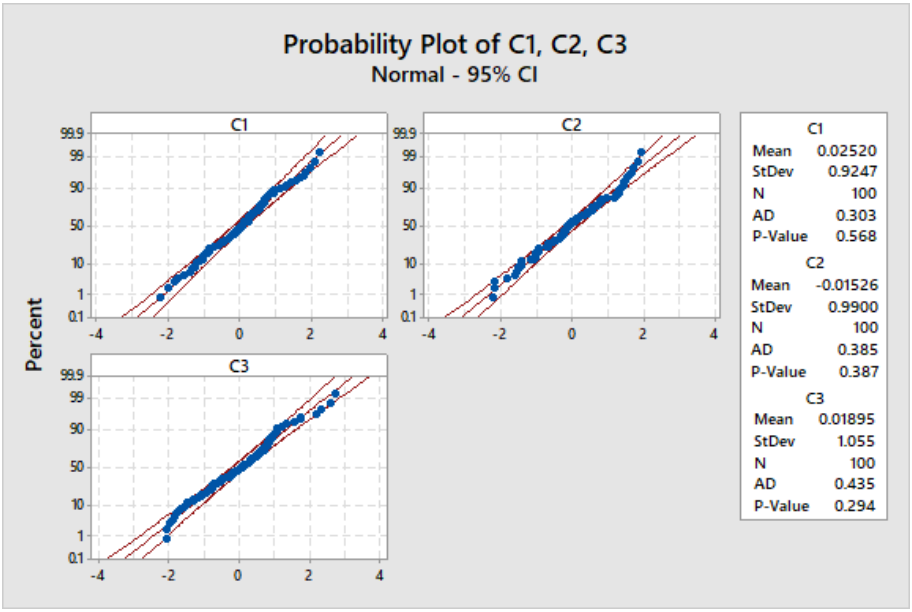
Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
C1	100	0	0.0252	0.0925	0.9247	-2.2454	-0.6019	0.0748	0.6113	2.2247
C2	100	0	-0.0153	0.0990	0.9900	-2.2387	-0.6915	-0.1234	0.7042	1.9063
C3	100	0	0.019	0.106	1.055	-2.068	-0.761	0.112	0.787	2.748

Variable	Skewness	Kurtosis
C1	0.00	-0.06
C2	-0.09	-0.55
C3	0.08	-0.16









Comments:
All 3 columns generated seem to conform well with the standard normal distribution; note that the skewness and kurtosis are both close to zero. The histograms are not perfect but deviations from the normal curve are only minor. All normal probability plots are reasonably straight lines, with only a few points off. The normality tests are all far from significant. Try instead with the seed (base) set at 140920 to see some less nice results. Both sets of results are evidently valid simulations from a standard normal distribution.

The reason that the distributions look so good is that with 100 points the approximation of the data to the normal distribution is quite good. Try repeating with $n = 30$ observations...

Part (b)
We use the same approach in Minitab as described above, with simulated data from a uniform distribution (0,1).

```
MTB > Base 140919.
MTB > Random 100 c4 c5 c6;
SUBC>   Uniform 0.0 1.0.
```

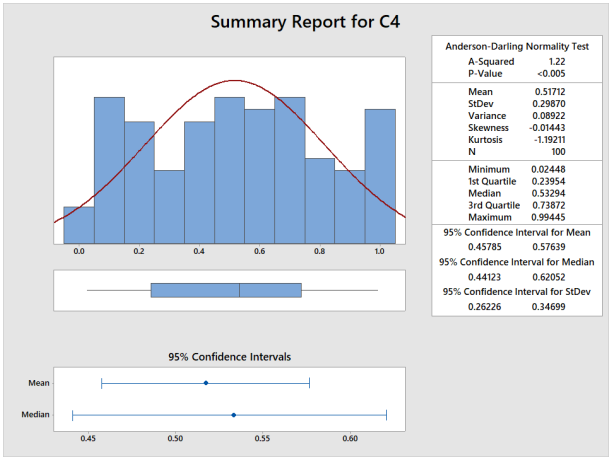
```
MTB > Describe C1 C2 C3;
...
MTB > GSummary C4 C5 C6.
MTB > PPlot C4 C5 C6;
SUBC>   Normal;
...
```

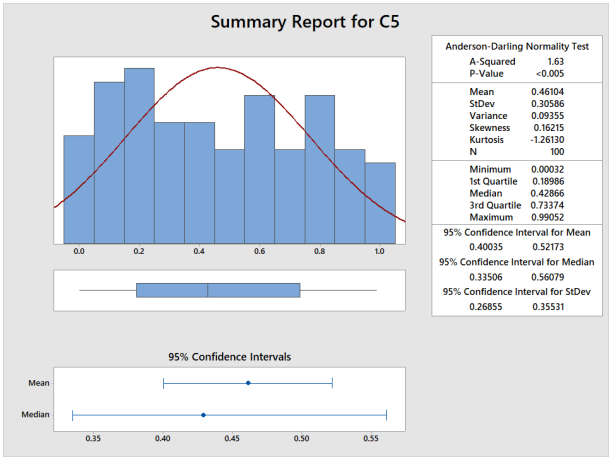
Descriptive Statistics: C4, C5, C6

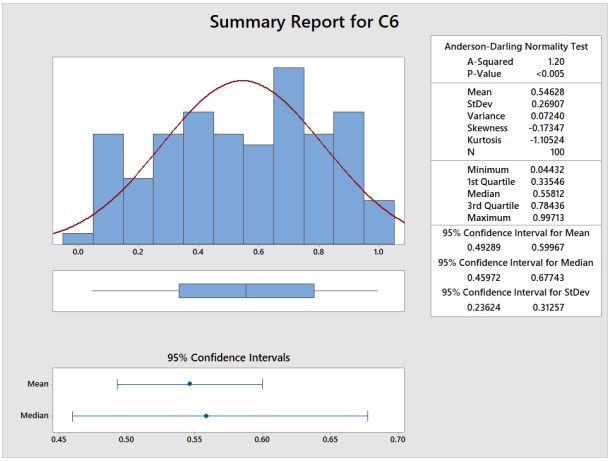
Statistics

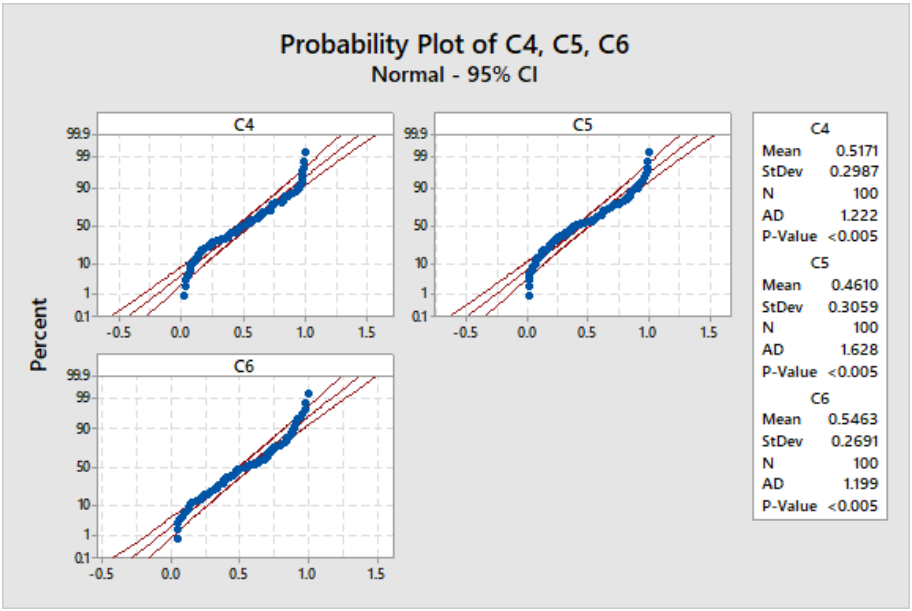
Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
C4	100	0	0.5171	0.0299	0.2987	0.0245	0.2395	0.5329	0.7387	0.9945
C5	100	0	0.4610	0.0306	0.3059	0.0003	0.1899	0.4287	0.7337	0.9905
C6	100	0	0.5463	0.0269	0.2691	0.0443	0.3355	0.5581	0.7844	0.9971

Variable	Skewness	Kurtosis
C4	-0.01	-1.19
C5	0.16	-1.26
C6	-0.17	-1.11









Comments:
All 3 columns generated are fitted poorly by the normal distribution.

The histograms lack tails (going down like in a normal distribution), and the kurtosis is negative and quite strong for all distributions. The normal plots all have the same characteristic shape with curves that bend downwards to the left and upwards to the right - these patterns reflect the too short left and right tails of the distribution. Accordingly, it should not surprise us that the A-D normality tests are all clearly significant, thus offering statistical evidence against a normal distribution.