Supplementary exercise 4.10 of IPS7e
(continuation of Supplementary exercise 4.9)

Simulation of draws of Internet users' age group; specifically whether a randomly selected Internet user is of age 18-29 years. The (true) probability of this event is assumed to be 0.3.

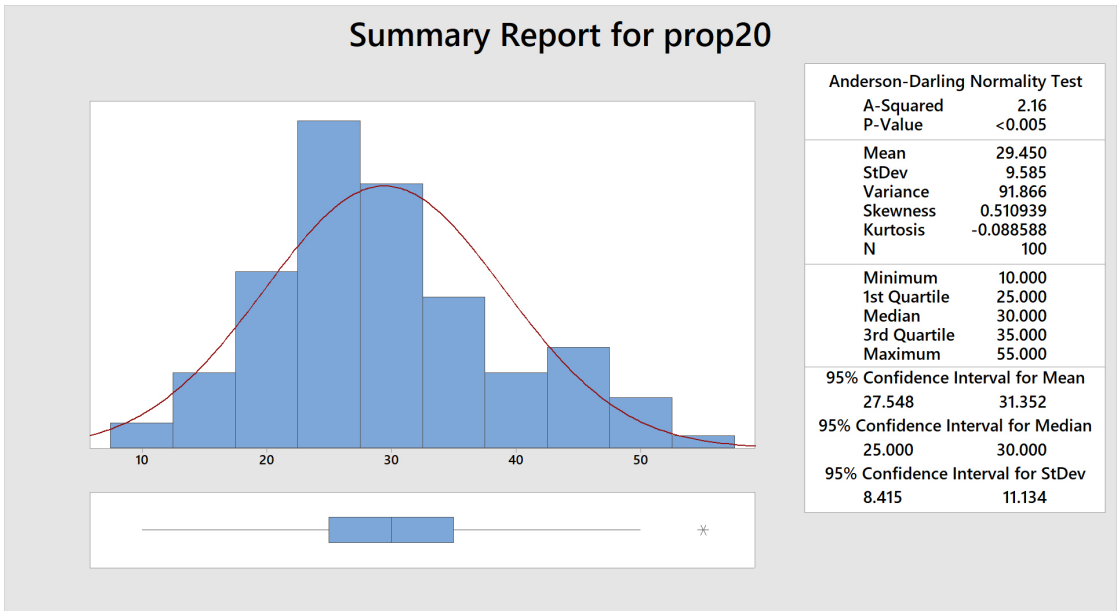
The following commands in Minitab will simulate 100 draws of 20 Internet users and convert to percentages, and then do the same for draws of samples of size 320. Then, we describe the distributions by simple numerical and graphical summaries; the number of bins in the histograms have been adjusted to 10.

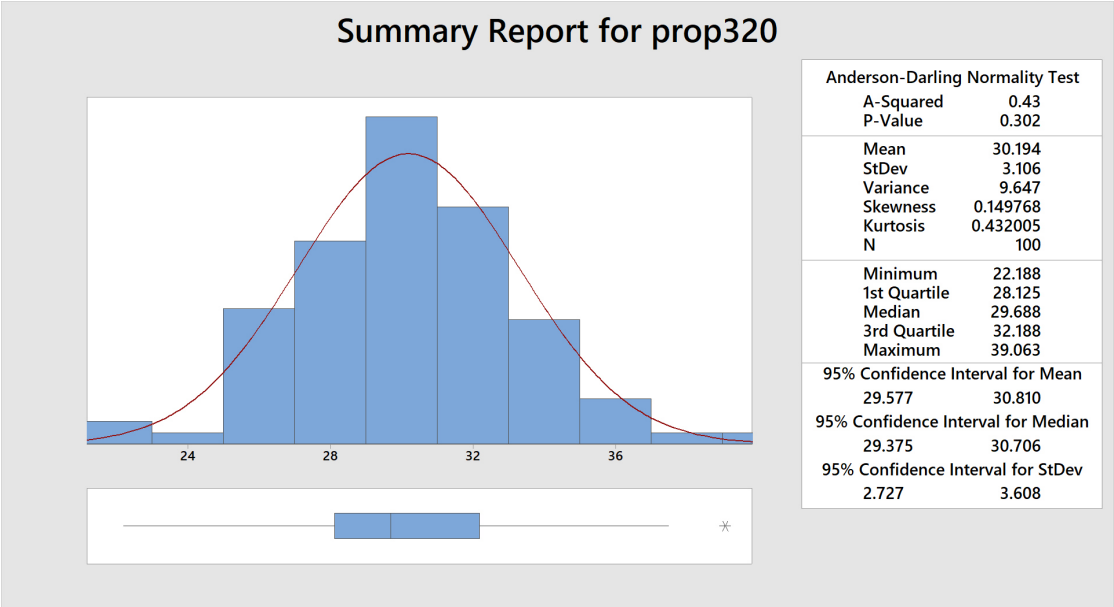
```
MTB > Base 160918.
MTB > Random 100 c1;
SUBC>   Binomial 20 .3.
MTB > Name C2 'prop20'
MTB > Let 'prop20' = C1/20*100
MTB > Random 100 c3;
SUBC>   Binomial 320 .3.
MTB > Name C4 'prop320'
MTB > Let 'prop320' = C3/320*100
MTB > Describe 'prop20' 'prop320'.
MTB > GSummary 'prop20' 'prop320'.
```

Descriptive Statistics: prop20, prop320

Statistics

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
prop20	100	0	29.450	0.958	9.585	10.000	25.000	30.000	35.000	55.000
prop320	100	0	30.194	0.311	3.106	22.188	28.125	29.688	32.188	39.063





Comments:
Both distributions look reasonably symmetric and bell-shaped. They are both centered approximately around 30% (or 0.3) with their means being close to 30, but the spread of the distribution for sample size 320 is much smaller. Theoretically (we will see these formulae in Session 5), the standard deviation should be $\sqrt{320/20} = 4$ times larger for sample size 20 than for sample size 320, and althogh the actual ratio is somewhat lower ($9.585/3.106 = 3.1$) this can be attributed to random variation (e.g., try with another seed).