Premedical course
Solution to MINITAB practical 5

Since this practical uses computer-generated 'random' numbers, actual answers may vary. The random number seed which was used for these solutions is shown for each question so that results may be replicated exactly. The irrelevant part of the results of 'Describe' and 'Tally' are not shown.

Practical 5a. Question 1.

```
MTB > Base 16807.
MTB > Random 100 c1;
SUBC>   Binomial 10 0.2.
MTB > Describe C1.
                N     MEAN   STDEV   SEMEAN
C1            100    2.050   1.298    0.130

MTB > let k1=sqrt(10*0.2*0.8)
MTB > print k1
K1       1.26491

MTB > Stem-and-Leaf C1;
SUBC>   Increment 1.
Stem-and-leaf of C1        N  = 100
Leaf Unit = 0.10

   14    0 00000000000000
   35    1 00000000000000000000
  (27)   2 000000000000000000000000000
   38    3 0000000000000000000000000
   14    4 000000000000
    2    5 00
```

The calculated mean, at 2.05, is close to the theoretical value of 2. The standard deviation of 1.30 is also close to the theoretical value of 1.26. The distribution is relatively symmetric but as it is limited by zero on the left it cannot be very close to the normal distribution. Indeed the mean minus two standard deviations would be a negative number and would therefore exclude no observations at all instead of the 2.5% which would be excluded by a normal distribution. In no instance out of the 100 simulated did as many as 6 people die; such an outcome is therefore not easy to accept as compatible with a death rate of 20% in the population.

It is possible to evalute the tail probability exactly rather than by simulating the process with random numbers; notice that we have to find the probability of 5 deaths or fewer and subtract that probability from unity.

```
MTB > CDF 5;
SUBC>   Binomial 10 .2.
       K  P( X LESS OR = K)
     5.00            0.9936
```

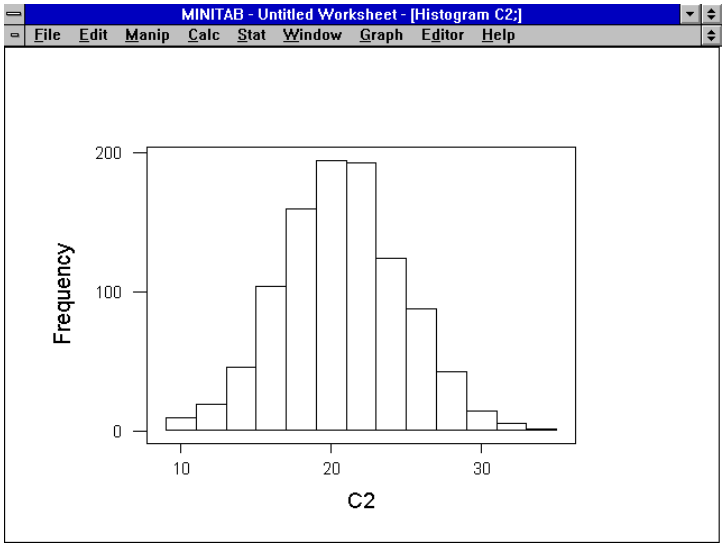
We obtain a P-value of 0.0064.

Practical 5a. Question 2

```
MTB > Base 16807
MTB > Random 1000 c2;
SUBC>   Binomial 100 0.2.
MTB > Describe C2.
                N     MEAN   STDEV   SEMEAN
C2           1000   20.213   4.008    0.127

MTB > let k2=k1*sqrt(10)
MTB > print k2
K2       4.00000

MTB > Histogram C2;
SUBC>   MidPoint;
SUBC>   Bar.
```




```
MTB > Tally C2;
SUBC>   Counts.
      C2  COUNT
      29      8
      30      6
      31      4
      32      1
      33      1
      N=   1000
```

This time the calculated mean, at 20.2, is close to the theoretical value of 20, and the standard deviation of 4.01 is very close to the theoretical value of 4. The distribution is very close to the Normal distribution. In only 12 instances out of the 1000 simulated did as many as 30 people die; such an outcome is therefore not easy to accept as compatible with a death rate of 20% in the population. We could say that the difference was significant ($P \approx 0.012$) in a one-sided test. The theoretical result using the method of the previous question gives $P = 0.0112$.

Alternatively we could calculate a z-statistic to give a probability of 0.0062 as follows:

```
MTB > CDF 30;
SUBC>   Normal 20 4.
  30.0000    0.9938
```

The Normal approximation therefore gives a probability which is rather too small.

Practical 5b. Question 1

```
MTB > Base 314159
MTB > Random 30 c1-c4;
SUBC> Chisquare 4.
MTB > Let c5=(c1+c2+c3+c4)/4
MTB > Stack (c1) (c2) (c3) (c4) (c6).
MTB > Describe c6 c5
               N     MEAN   STDEV   SEMEAN
C6           120    4.222   2.888    0.264
C5            30    4.222   1.421    0.259
```

The calculated mean does not change. The calculated standard deviation approximately halves because for samples of *n* it is reduced by a factor of the square root of *n*.

```
MTB > Stem-and-Leaf c6 c5;
SUBC>   Increment 1.
Stem-and-leaf of C6        N  = 120
Leaf Unit = 0.10

    4    0 3557
   24    1 00112233355667788999
   52    2 00011223333344445555566667 79
  (16)   3 0001112356777789
   52    4 000123355566789
   37    5 115666789
   28    6 12346789
   20    7 123489
   14    8 2347
   10    9 012568
    4   10
    4   11 02
    2   12
    2   13 3
    1   14
    1   15
    1   16 3

Stem-and-leaf of C5        N  = 30
Leaf Unit = 0.10

    1    1 9
    5    2 1157
   15    3 1223355578
   15    4 1357
   11    5 0024678
    4    6 335
```

The distribution of the original data is quite skew, but because of the central limit theorem the distribution of the means is more nearly normal.

Practical 5b. Question 2

The mean of the distribution may be deduced to lie between $4.222 \pm 1.96 \times 0.264$ (i.e. 3.70 and 4.73) with 95% confidence.

```
MTB > Base 314159
MTB > Random 1000 c7;
SUBC> Chisquare 4.
MTB > Describe c7

               N     MEAN   STDEV   SEMEAN
C7          1000   4.0874   2.8443   0.0899
```

With the larger sample we may be more precise about the mean of the distribution. It may now be deduced to lie between $4.087 \pm 1.96 \times 0.090$ (i.e. 3.91 and 4.26) with 95% confidence. The width of the confidence interval has shrunk by a factor of 0.264/0.090, which is 2.93. This is what we would expect: the square root of the ratio of sample sizes is 2.89.