Supplementary exercises 6.38 and 6.39 of IPS7e

Planning of a study on starting salaries for liberal arts major graduates. The parameter of interest is the population mean (μ), i.e. the mean salary in the population of recent graduates. The problem expects us to take the standard deviation of $s = 9000$ estimated in a pilot study as a known population value, i.e. $\sigma = 9000$. Although difficult to justify we will first work under this assumption, and then discuss dealing with unknown σ in the population in the note below.

Exercise 6.38

For known σ, we use the direct formula for sample size computed from a desired margin of error ($m = 400$), e.g. 8L-13:

$$n \geq (z^* \cdot \sigma/m)^2 = (1.96 \cdot 9000/400)^2 = 1944.8 \approx 1945.$$

We need at least 1945 subjects to achieve that. That seems of course quite infeasible, but it results from a pretty unrealistic expection for the margin of error ($m = 400$) when the population standard deviation is much higher ($\sigma = 9000$). Note that if we use the approximate value $z^* \approx 2$ instead of 1.96, as in the additional notes, the result becomes: $n \geq (2 \cdot 9000/400)^2 = 2025$.

For calculation in Minitab, we give the command (corresponding to the `Basic Stat-Power and Sample Size-Sample Size for Estimation` menu) and output:

```
MTB > SSCI;
SUBC>   NMean;
SUBC>     Sigma 9000 1;
SUBC>   Confidence 95.0;
SUBC>   IType 0;
SUBC>   MError 400.
```

Sample Size for Estimation

Method

Parameter	Mean
Distribution	Normal
Standard deviation	9000 (population value)
Confidence level	95%
Confidence interval	Two-sided

Results

Margin of Error	Sample Size
400	1945

Exercise 6.39

If we allow for a larger margin of error, the required sample size must drop down (because a larger sample size implies a smaller SE and margin of error). We redo the calculation with $m = 800$:

$$n \geq (z^* \cdot \sigma/m)^2 = (1.96 \cdot 9000/800)^2 = 486.02 \approx 487.$$

Because the sample size (n) enters into the SE as the square-root of n, increasing the margin of error by a factor of 2 (that is, doubling its original size) lowers the required sample size by a factor of $2^2 = 4$.

Added note (based on material in Chapter 7 and lecture 8)

The above formula requires σ to be known. In practice this is unrealistic if the value for σ is based on a pilot study only. Therefore we would instead want to estimate σ from the data when the actual analysis is carried out. Note that we still need a (guessed) value for σ in order to do the sample size calculation, but this value will not be assumed to be the known standard deviation when the data have been collected.

The only change in the formula is that z^* should be replaced by t^* from a t-distribution with $\text{df} = n-1$. With $n = 1945$ (for $m = 400$), $t^* = 1.96 = z^*$, so the calculation will be essentially the same (when z^* and t^* are entered into the formula with two decimals). Also for $\text{df} = 486$ (for $m = 800$), the difference from $t^* = 1.965$ to 1.960 from $\text{N}(0, 1)$ is so small that one would usually ignore it. Specifically if we redo the calculation above with the value 1.965 instead of 1.96,

$$n \geq (t^* \cdot \sigma/m)^2 = (1.965 \cdot 9000/800)^2 = 488.7 \approx 489.$$

We can use Minitab to calculate the relevant t^*-value (shown here for $m = 800$):

```
MTB > InvCDF .025;
SUBC>   T 486.
```

Inverse Cumulative Distribution Function

Student's t distribution with 486 DF

P(X ≤ x)	x
0.025	-1.96486

or to do the entire calculation, this time without ticking off the box for considering the standard deviation as known.

```
MTB > SSCI;
SUBC>   NMean;
SUBC>     Sigma 9000;
SUBC>   Confidence 95.0;
SUBC>   IType 0;
SUBC>   MError 800.
```

Sample Size for Estimation

Method

Parameter	Mean
Distribution	Normal
Standard deviation	9000 (estimate)
Confidence level	95%
Confidence interval	Two-sided

Results

Margin of Error	Sample Size
800	489