Report 2. SRS

Name of the project: Analysis of Average Winter Temperature in College Park, Maryland

The true parameter: the true average temperature (in °F) in College Park MD during the winter months, **μ**.

The population size N = 720

Calculate sample size *n* for 90% and 95% confidence levels and couple different *d*'s. Use true σ^2 for these calculations. Here is an idea on how to choose *d*. It is based on relative error. Take *r* = .05, *r* = .01 and *r* = .10. Since $r = \frac{\hat{\theta}-\theta}{\theta} = \left|\frac{d}{\theta}\right|$, **we get** $d = |r\theta|$

Population mean (μ) = 39.177 ˚F

Population variance (σ^2) = 91.7239

For r = 0.05, 0.01, 0.10:

$d_1 = |\, 0.01 \;*\; 39.177 \,|$ = **0.39177**

$d_2 = |\, 0.05 * 39.177 \,|$ = **1.95885**

$d_3 = |\, 0.10 \;*\; 39.177 \,|$ **= 3.9177**

For 90% confidence levels ($z_{\alpha/2} = z_{0.05} = 1.64$):

$$n_1 = \frac{1}{d^2/z^2\sigma^2 + 1/N} = \frac{1}{(0.39177^2/(1.64^2*91.7239)) + (1/720)} = 497.26 \sim \mathbf{498}$$

$$n_2 = \frac{1}{d^2/z^2\sigma^2 + 1/N} = \frac{1}{(1.95885^2/(1.64^2*91.7239)) + (1/720)} = 59.023 \sim \mathbf{60}$$

$$n_3 = \frac{1}{d^2/z^2\sigma^2 + 1/N} = \frac{1}{(3.9177^2/(1.64^2*91.7239)) + (1/720)} = 15.722 \sim \mathbf{16}$$

For 95% confidence levels ($z_{\alpha/2} = z_{0.025} = 1.96$):

$$n_4 = \frac{1}{d^2/z^2\sigma^2 + 1/N} = \frac{1}{(0.39177^2/(1.96^2*91.7239)) + (1/720)} = 548.105 \sim \mathbf{549}$$

$$n_5 = \frac{1}{d^2/z^2\sigma^2 + 1/N} = \frac{1}{(1.95885^2/(1.96^2*91.7239)) + (1/720)} = 81.444 \sim \mathbf{82}$$

$$n_6 = \frac{1}{d^2/z^2\sigma^2 + 1/N} = \frac{1}{(3.9177^2/(1.96^2*91.7239)) + (1/720)} = 22.248 \sim \mathbf{23}$$

Estimate your parameter of interest using SRS with *n*'s which you calculated above.

$\bar{y} = [1/n]*\sum_{i=1}^{n} y_i$

For 90% confidence level:

n_1 = **498**

$\bar{y}$ = [1/498]*[19498. 9] = 39.15442 ~ **39.2** °**F**

n_2 = **60**

$\bar{y}$ = [1/60]*[2342.4] = 39.0400 ~ **39.0** °**F**

n_3 = **16**

$\bar{y}$ = [1/16]*[609.1] = 38.06875 ~ **38.1** °**F**

For 95% confidence level:

n_4 = **549**

$\bar{y}$ = [1/549]*[21586.5] = 39.31967 ~ **39.3** °**F**

n_5 = **82**

$\bar{y}$ = [1/82]*[3204.4] = 39.07805 ~ **39.1** °**F**

n_6 = **23**

$\bar{y}$ = [1/23]*[913.7] = 39.72609~ **39.7** °**F**

Estimate variance of your estimator for these *n's*.

The estimated sample variance for SRS without replacement is:

$$\widehat{var}(\bar{y}) = (\frac{N-n}{N})\frac{s^2}{n} \text{ , } where\ s^2 = (\frac{1}{n-1})\sum_{i=1}^{n}(y_i - \bar{y})^2$$

<u>For 90% confidence level</u>:

n_1 = **498**

Sample variance (s^2) = 91.641

$$\widehat{var}(\bar{y}) = (\frac{720-498}{720})\frac{91.641}{498} = \mathbf{0.057\ °F}$$

n_2 = **60**

Sample variance (s^2) = 75.869

$$\widehat{var}(\bar{y}) = (\frac{720-60}{720})\frac{75.869}{60} = \mathbf{1.159\ °F}$$

n_3 = **16**

Sample variance (s^2) = 57.017

$$\widehat{var}(\bar{y}) = (\frac{720-16}{720})\frac{57.017}{16} = \mathbf{3.484\ °F}$$

<u>For 95% confidence level</u>:

n_4 = **549**

Sample variance (s^2) = 92.990

$$\widehat{var}(\bar{y}) = (\frac{720-549}{720})\frac{92.990}{549} = \mathbf{0.040\ °F}$$

n_5 = **82**

Sample variance (s^2) = 92.778

$$\widehat{var}(\bar{y}) = (\frac{720-82}{720})\frac{92.778}{82} = \mathbf{1.003\ °F}$$

n_6 = **23**

Sample variance (s^2) = 86.387

$\widehat{var}(\overline{y}) = (\frac{720-23}{720})\frac{86.387}{23}$ = **3.636 °F**

Calculate confidence intervals for these estimators.

For 90% confidence level:

$$CI_{90} = \overline{y} \pm t_{0.05,\,n-1}\sqrt{\widehat{var}(\overline{y})}$$

n_1 = 498

$CI_{90} = 39.2 \pm 1.65\sqrt{0.057}$ = (**38.81, 39.59**)

n_2 = 60

$CI_{90} = 39.0 \pm 1.67\sqrt{1.159}$ = (**37.20 , 40.80**)

n_3 = 16

$CI_{90} = 38.1 \pm 1.75\sqrt{3.484}$ = (**34.83, 41.37**)

For 95% confidence level:

$$CI_{95} = \overline{y} \pm t_{0.025,\,n-1}\sqrt{\widehat{var}(\overline{y})}$$

n_4 = 549

$CI_{95} = 39.3 \pm 1.96\sqrt{0.040}$ = (**38.91, 39.69**)

n_5 = 82

$CI_{95} = 39.1 \pm 1.99\sqrt{1.003}$ = (**37.11, 41.09**)

n_6 = 23

$CI_{95} = 39.7 \pm 2.07\sqrt{3.636}$ = (**35.75, 43.65**)

Choose the optimal sample size *n* among the ones calculated above. The best sample size should be between 10% – 20%. Definitely it should be a 'large sample' size *n* > 40. If you have several such *n*, choose the one which produces the smaller CI or has a smaller α level.

Among the calculated sample sizes above, the optimal sample size is **n = 82** with 95% confidence interval **(37.11, 41.09).**

Since N = 720, our sample size should be larger than 72 (10% = 720*0.10 = 72) and smaller than n = 144 (20% = 720*0.20 = 144). The only sample size that falls between this range is n = 82.

Does your choice of best estimator guarantee the nominal confidence level? To answer this question, take 100 samples of size *n* where *n* has been selected above. For each sample, compute the difference between the parameter and its estimator. Compare these differences with *d*. How many samples have the difference less than *d*? Does it agree with the nominal confidence level? Justify your answer.

For sample size n = 82, d **=** 1.95885.

There were 5 samples with a difference between μ and $\bar{y}$ that was more than d. The other 95 samples had a difference less than d. Since we used the 95% confidence level to choose n = 82, we would expect 95% of the samples to have a d that is less than or equal to 1.95885. As shown by the 100 samples we took, our estimator using n = 82 guarantees this nominal confidence level.

Output for the 100 samples:

```
  [1] 0.348951220 1.470560976 1.485536585 0.530658537 1.469341463 2.159585366 0.714804878 1.446170732 0.208707317 0.523000000 2.031878049 0.559585366
 [13] 0.114463415 0.558707317 0.386756098 1.153829268 0.663585366 0.855048780 0.333975610 1.051048780 0.623341463 0.182756098 0.460804878 0.193731707
 [25] 1.092512195 0.551048780 0.031878049 1.369682927 0.481536585 0.527878049 0.473341463 0.547731707 0.959926829 1.091634146 1.801048780 0.635195122
 [37] 1.549829268 1.254707317 0.505926829 1.809926829 0.179097561 0.308707317 0.134317073 0.559926829 0.721780488 0.032756098 0.164463415 0.176658537
 [49] 1.409926829 0.487634146 1.404707317 1.618463415 1.593731707 0.674560976 0.506268293 1.709585366 0.318463415 0.179439024 0.174560976 0.752268293
 [61] 0.908365854 0.640414634 1.329097561 0.681878049 1.063585366 0.578219512 0.256268293 3.172121951 1.527878049 0.520902439 0.091292683 0.780658537
 [73] 1.427878049 0.992512195 2.141634146 2.241634146 0.208365854 0.842512195 0.180317073 1.058707317 1.099829268 0.796170732 1.586756098 1.427000000
 [85] 0.091634146 0.107146341 0.866902439 0.086756098 1.355926829 1.187975610 0.637634146 0.285195122 0.380317073 0.612024390 1.095292683 0.809926829
 [97] 0.944951220 0.007146341 0.179439024 0.374219512
[1] 5
```