Report 2. SRS.

Group #1

Predicting Mean Maximum Rainfall in Mount Ginini

Group member name	Contribution
Richard Liang	16.7%
Rhea Rakheja	16.7%
Olivia Majedi	16.7%
Priyanka Kishore	16.7%
Michael Strobel	16.7%
Jiadong Li	16.7%

- The true parameter (the mean max temperature):
  μ = 19.1129 degrees celsius

  The population size N = 745

- 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 $r = |\frac{\hat{\theta} - \theta}{\theta}| = |\frac{d}{\theta}|$, we get $d = |r\theta|$.

r value	Approximate d value
.05	0.9556
.01	0.1911
.10	1.9112

Confidence Interval	Approximate d Value	Rounded Sample Size, n
90%	0.9556	60
	0.1911	511
	1.911	16
95%	0.9556	82
	0.1911	564
	1.911	23

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

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

Confidence Interval	Rounded Sample Size, n	Sample Mean for Daily High Temperatures	Variance for Sample Mean
90%	60	19.1083	17.4108
	511	19.1212	22.9045
	16	20.2562	9.6799
95%	82	19.9549	28.5167
	564	19.2234	21.6231
	23	17.3522	18.3060

• Calculate confidence intervals for these estimators.

Confidence Interval	Rounded Sample Size, n	Sample Mean for Daily High Temperatures	Confidence Interval for Sample Mean
90%	60	19.1083	[18.2587, 19.9580]
	511	19.1212	[18.9254, 19.3170]
	16	20.2562	[18.9907, 21.5218]
95%	82	19.9549	[18.8645, 21.0452]
	564	19.2234	[19.0336, 19.4133]
	23	17.3522	[15.6308, 19.0735]

- 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.

  We chose the sample size n=82 for the 95% confidence interval. This falls nicely in between the 10% and 20% guidelines and provides a smaller CI than 60. And this uses the 95% CI, which gives a smaller alpha.

  A sample size greater than 500 is unrealistic in the real world and samples far too large of a proportion of our population. The sample sizes 16 and 23 are too low to give an accurate estimate.

- 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.

  Out of the 100 samples taken, 95 of the samples had a difference between the parameter and its estimator less than our d value of 0.9556. This does agree with the nominal confidence level because we had 95 samples with a d less than our d-value of .95556 for a 95% confidence interval. The proportion of samples with d less than the d-value matches the expected proportion.

The code used for this assignment is included in the following pages.