Interval Estimation of the Population Mean *R Programming*	14 of June 2013 Al-Ahmadgaid B. Asaad *alstated@gmail.com*

Interval estimation of the population mean can be computed from functions of the following R packages:

- stats - contains the `t.test`;

- TeachingDemos - contains the `z.test`; and,

- BSDA - contains the `zsum.test` and `tsum.test`.

The `t.test` of the stats package is a student's t test, and is use when raw dataset is given. The same case for `z.test`, but this function is specifically for z-test of known population standard deviation. When dataset is not given and only the summary statistics (mean, and standard deviation) are presented, then the appropriate functions are `zsum.test` or `tsum.test`. Note that, `t.test` and `tsum.test` are functions of the same statistical test, and that of `z.test` and `zsum.test`. Consider the example below,

Example 1. The 2012-2013 SASE scores of the 33 random students from College of Science and Mathematics (CSM) of MSU-IIT were recorded: 84, 93, 101, 86, 82, 86, 88, 94, 89, 94, 93, 83, 95, 86, 94, 87, 91, 96, 89, 79, 99, 98, 81, 80, 88, 100, 90, 100, 81, 98, 87, 95, and 94. The population of these scores are believe to be normally distributed with 6.8 standard deviation. Determine and interpret the 95% and 99% confidence interval of the population mean.

From the data, we obtain the following information: (i) the sample size is more than 30, and (ii) the population standard deviation is known. Therefore, the appropriate test is z-test. And the function to use is `z.test`, that is

```
> scores <- c(84, 93, 101, 86, 82, 86, 88, 94, 89, 94, 93,
+             83, 95, 86, 94, 87, 91, 96, 89, 79, 99, 98,
+             81, 80, 88, 100, 90, 100, 81, 98, 87, 95, 94)
> #For 95% Confidence Interval
> library(BSDA)
> z.test(scores, sigma.x = 6.8)


        One-sample z-Test


data: scores
z = 76.3126, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
```

```
 88.01327 92.65340
sample estimates:
mean of x
 90.33333


> #For 99% Confidence Interval
> z.test(scores, sigma.x = 6.8, conf.level = 0.99)


        One-sample z-Test


data: scores
z = 76.3126, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
99 percent confidence interval:
 87.28425 93.38241
sample estimates:
mean of x
 90.33333
```

Interpretation: The true mean of all SASE scores in the school year 2013-2014 from CSM is likely between 88.01327 and 92.65340 (95% CI). And the true mean of all SASE scores for the said college and school year is likely between 87.28425 and 93.38241 (99% CI).

Aside from the confidence interval, the function returns also the computed z-statistics with p-value, and as well as the point estimate of the mean. To get rid of this, one can add a suffix `$conf.int` to the function to extract the confidence interval only.

```
> z.test(sam.scores, sigma.x = 6.8, conf.level = 0.95)$conf.int
[1] 88.01916 92.65929
attr(,"conf.level")
[1] 0.95
>
> z.test(sam.scores, sigma.x = 6.8, conf.level = 0.99)$conf.int
[1] 87.29014 93.38831
attr(,"conf.level")
[1] 0.99
```

Example 2. The following data (341, 345, 338, 339, 340, 343, 341, 343, 341, 328, 343, 347, 337, 348, and 339) are random samples from normally distributed population. Compute and interpret the 90% confidence interval.

The appropriate test for this is t-test since the sample size is small, $n < 30$, and the population variance is unknown. And thus,

```
> data <- c(41, 345, 338, 339, 340, 343, 341, 343, 341, 328,
+           343, 347, 337, 348, 339)
> t.test(data, conf.level = 0.9)$conf.int
[1] 285.5911 356.1423
```

```
attr(,"conf.level")
[1] 0.9
```

Interpretation: The true mean of the population of the given data above is likely between 285.5911 and 356.1423 (90% CI).

Often in textbooks, however, we are presented with summary statistics of the data like the next example below from *Simplified Biostatistics* by Abubakar S. Asaad.

Example 3. The biostatistician took a random sample of 49 patients from a list of all patients ever admitted to the hospital within a three-month period and the number of drugs prescribed per admission was determined for each. The average drug per case was found to be 7.5 with standard deviation of 2.5. Calculate and interpret the 95% confidence interval for true mean of all the patients ever admitted to the hospital.

In this example, no dataset is given, but we have the computed mean = 7.5 of this dataset, standard deviation = 2.5, and sample size = 49. Thus, to compute for the interval estimate of the population mean in R, we use the `zsum.test`

```
> zsum.test(mean.x = 7.5, sigma.x = 2.5, n.x = 49)$conf.int
[1] 6.800013 8.199987
attr(,"conf.level")
[1] 0.95
```

Interpretation: The true mean of all the patients ever admitted to the hospital is likely between 6.800013 and 8.199987 (95% CI).

The `tsum.test` function is used in situation like in Example 3, but this time the population variance should be unknown and the sample size should be less than 30.

References

Asaad, Abubakar S. (2011). *Simplified Biostatistics*. Manila: Rex Book Store, Inc..

Labels

Descriptive Statistics, R, Tutorial