MTH 656 Statistics Method, 2012 Fall

Shuowen Wei’s Project

Introduction

In this project, I will do the simulations of confidence intervals and tests of hypotheses in small samples from normal populations. So, first of all I will generate many (say, 1000) small samples using the “random number generation” of excel, then compute their sample means, sample standard variances. Then based the theorem that the small samples’ means are subject to t-distribution, I will find the confidence interval for each sample and do the hypotheses test to see if the particular sample’s mean fall into this interval or not, count the frequency of successes. Then I will draw a histogram for all of these samples and to see their curve is similar to t-distribution or not. In the end, I will conduct the confidence interval of the each sample’s variance, based on the theorem that their variance is subject to chi-square distribution, then I will do the hypotheses tests for each sample’s variance to see whether it falls into its confidence interval.

In “random number generation” of “data analysis” in excel, I set the number of variables to n=15 and the number of random rows to 1000. The population mean I set is 10 and the population standard deviation is 4.

The table below is just a small section of the output in the simulation:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| individual value results | | |  |  |  |  |  |  |
| 1 | 2 | 3 |  | 14 | 15 | sampl avg | sampl sd | est sd avg |
| 10.97672 | 10.0892 | 13.74131 | … … | 12.2699 | 10.55702 | 9.155143 | 4.781918 | 1.234686 |
| 9.01288 | 8.805793 | 9.923039 | … … | 6.518577 | 11.07348 | 10.44721 | 3.32134 | 0.857566 |
| 6.764959 | 8.713965 | 15.80556 | … … | 5.834296 | 15.02673 | 9.486015 | 3.851498 | 0.994452 |
| 13.33363 | 15.67614 | 13.15338 | … … | 9.98669 | 13.48813 | 9.955963 | 3.680688 | 0.95035 |
| 3.695437 | 19.55486 | 20.80713 | … … | 7.640743 | 9.70184 | 10.45056 | 5.447499 | 1.406538 |
| 11.04495 | 11.84845 | 12.31353 | … … | 5.454564 | 11.09475 | 10.77482 | 2.649223 | 0.684027 |
| 15.06948 | 14.27841 | 16.17643 | … … | 9.214856 | 7.378482 | 11.45874 | 4.539058 | 1.17198 |
| 0.423821 | 15.38492 | 9.612774 | … … | 15.90926 | 7.690447 | 9.160104 | 4.523814 | 1.168044 |
| … | … | … | … | … | … | … | … | … |

Then I computed the sample mean (“sampl avg” in the table), sample standard deviation (“sampl avg” in the table) and the estimate sample standard deviation mean (“est sd avg” in the table). And there are 1000 rows in total.

The first part is the confidence intervals for different alpha levels. I choose two alpha levels, 95% and 80%. After looking for the corresponding multipliers of the two different alpha levels, we have the two different confidence intervals as below:

:

where n=15, degree of freedom is and

:

where n=15, degree of freedom is and .

Hence, a small section of the simulation and test results are as follows:

(population mean: 10)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Mean 95% | | |  | Mean 80% | | |
| lower95 | upper95 | mhu in int95 |  | lower80 | upper80 | mhu in int10 |
| 6.506742 | 11.80354 | 1 |  | 7.49449 | 10.8158 | 1 |
| 8.607733 | 12.28669 | 1 |  | 9.293786 | 11.60064 | 1 |
| 7.352914 | 11.61912 | 1 |  | 8.148476 | 10.82355 | 1 |
| 7.917464 | 11.99446 | 1 |  | 8.677743 | 11.23418 | 1 |
| 7.433537 | 13.46759 | 1 |  | 8.558767 | 12.34236 | 1 |
| 9.307582 | 12.24206 | 1 |  | 9.854803 | 11.69483 | 1 |
| 8.944846 | 13.97264 | 1 |  | 9.88243 | 13.03506 | 1 |
| 6.65465 | 11.66556 | 1 |  | 7.589085 | 10.73112 | 1 |
| 8.480775 | 13.11284 | 1 |  | 9.344563 | 12.24905 | 1 |
| 8.44148 | 13.05381 | 1 |  | 9.301588 | 12.1937 | 1 |
| 7.108381 | 12.1186 | 1 |  | 8.042687 | 11.18429 | 1 |
| 8.949097 | 12.47557 | 1 |  | 9.606715 | 11.81795 | 1 |
| … | … | … |  | … | … | … |

And the simulation results are as follows:

|  |  |  |  |
| --- | --- | --- | --- |
| fraction95 in c.i. |  |  | fraction95 in c.i. |
| 0.935 |  |  | 0.787 |

The first result shows that there are 93.5% of the samples’ means fall in the confidence interval, which is very close to the expected value 95%.

The second example shows there 78.7% of the those samples’ means fall in the confidence interval, which is also very close to expected value 80%.

And histogram table and chart are as follows:

|  |  |
| --- | --- |
| *Bin* | *Frequency* |
| -28.5 | 0 |
| -26.5 | 1 |
| -24.5 | 0 |
| -22.5 | 0 |
| -20.5 | 0 |
| -18.5 | 0 |
| -16.5 | 0 |
| -14.5 | 0 |
| -12.5 | 0 |
| -10.5 | 0 |
| -8.5 | 0 |
| -6.5 | 0 |
| -4.5 | 4 |
| -2.5 | 19 |
| -0.5 | 59 |
| 1.5 | 195 |
| 3.5 | 494 |
| 5.5 | 1183 |
| 7.5 | 2027 |
| 9.5 | 2756 |
| 11.5 | 2995 |
| 13.5 | 2448 |
| 15.5 | 1582 |
| 17.5 | 811 |
| 19.5 | 285 |
| 21.5 | 110 |
| 23.5 | 25 |
| 25.5 | 3 |
| 27.5 | 2 |
| 29.5 | 0 |
| 31.5 | 1 |
| More | 0 |

The second part of this project is to conduct the confidence interval and tests of hypotheses for the population variance.

With using the same outcomes, I still choose the two different alpha levels to be 95% and 80%, after looking for the corresponding multipliers of the two different alpha levels, we have the two different confidence intervals for the population variance as below:

:

where n=15, degree of freedom is , and

:

where n=15, degree of freedom is , and .

Hence, a small section of the simulation and test results are as follows:

(population variance: 16)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Vairance 95% | | |  | Vairance 80% | | |
| lower95 | upper95 | mhu in int95 |  | lower80 | upper80 | mhu in int80 |
| 15.158597 | 56.87517 | 1 |  | 15.19803 | 41.09804 | 1 |
| 7.3127591 | 27.437527 | 1 |  | 7.331784 | 19.82638 | 1 |
| 9.8336335 | 36.895867 | 1 |  | 9.859216 | 26.66098 | 1 |
| 8.980753 | 33.695853 | 1 |  | 9.004117 | 24.34865 | 1 |
| 19.672021 | 73.809571 | 0 |  | 19.7232 | 53.33485 | 0 |
| 4.6525585 | 17.456435 | 1 |  | 4.664662 | 12.61403 | 0 |
| 13.657967 | 51.244796 | 1 |  | 13.6935 | 37.02953 | 1 |
| 13.566385 | 50.90118 | 1 |  | 13.60168 | 36.78123 | 1 |
| 11.592588 | 43.495479 | 1 |  | 11.62275 | 31.42986 | 1 |
| 11.494018 | 43.125643 | 1 |  | 11.52392 | 31.16262 | 1 |
| 13.562645 | 50.887147 | 1 |  | 13.59793 | 36.77109 | 1 |
| 6.71913 | 25.210227 | 1 |  | 6.73661 | 18.21693 | 1 |
| 10.087007 | 37.846528 | 1 |  | 10.11325 | 27.34793 | 1 |
| … | … | … |  | … | … | … |

And the simulation results are as follows:

|  |  |  |  |
| --- | --- | --- | --- |
| fraction95 in c.i. |  |  | fraction80 in c.i. |
| 0.871 |  |  | 0.801 |

The first result shows that there are 87.1% of the samples’ variances fall in the confidence interval, which is close to the expected value 95%.

The second example shows there 80.1% of the those samples’ variances fall in the confidence interval, which is very close to expected value 80%.

Conclusion, bases on the simulations’ results above, we verify that the theorems in the book are right. The expected many samples’ means and variances are subject t-distribution and chi-square distribution when those samples’ sizes are small and chosen from a normal distribution population.