Assign2

Frank Guo

2024-03-26

Answer Sheet for Stats 707 Assignment 2

Question 1: [30 Marks].

It is well known that many people try to improve their chances of winning a lottery by using lucky numbers or buying their tickets from certain lucky stores. To estimate what proportion of population would buy from a lucky store, a question was asked in a class survey. Participants were asked to respond Yes/No to the following question:
"If you learned that a certain store has sold the winning lotto ticket a few times in the past year, would you be tempted to buy a ticket from this store?"
In the survey, 12 out of 43 respondents said 'Yes'. These responses are stored in an Excel file luckylotto.CSV on Canvas.

Task 1:

a. Based on this data, what will be your estimate for the proportion of people in the whole population that would consider buying their tickets this way? [5 Marks].
b. Based on the key concepts of Statistical Inference that you have learned so far, Explain why this estimate can be considered to be a good estimate? [5 Marks].

```
## [1] 0.2790698
```

Answer for Task 1:

a:

Use mean of response_code as an estimator for the proportion of people in the whole population. the mean of response_code equels the proportion of sample. Hence, use ***Sample Mean*** as the estimator of ***population mean***.

b:.
Use the mean of response_cod as an estimator of proportion of whole population becarse of:
the sample is random and indenpent. the sample size is 45, larnger than 30. according Law of Large Numbers, with larger sample sizes, the sample mean becomes increasingly closer to the true population mean. the sample mean is more likely to provide an accurate estimate of the population mean when the sample size is large.

Task 2:

a. Find the confidence intervals for your estimate at the 95% and the 99% levels and interpret each. Specify what the 'confidence' actually means and what it doesn't. [6 Marks].
b. Explain why the 99% confidence interval is wider than the 95% one? [4 Marks].

```
t1 <- t.test(sample,conf.level = 0.95)
print('confidence intervals for 95%:')
```

```
## [1] "confidence intervals for 95%:"
```

```
t1$conf.int
```

```
## [1] 0.1393953 0.4187443
## attr(,"conf.level")
## [1] 0.95
```

```
t2 <- t.test(sample, conf.level = 0.99)
print('confidence interval for 99%:')
```

```
## [1] "confidence interval for 99%:"
```

```
t2$conf.int
```

```
## [1] 0.09233251 0.46580702
## attr(,"conf.level")
## [1] 0.99
```

Ansewer for Task2:

a..
The confidence interval at 95% confident level is : 0.1393953 0.4187443.
the interval has 95% probability covers the true mean (in ths case also is the true proportion) of population. We have the confidence that there 95% probability that the true mean will be in this interval.
The confidence interval at 99% confident level is : 0.09233251 0.46580702.
That means that the interval has 99% probability covers the tree mean(in ths case also is the true proportion) of population.We have the confidence that there 95% probability that the true mean will be in this interval.
The "confidence" here means the level of certainty that the true parameter(the mean or proportion of population) will be covered in this interval.It is not the probability of the true parameter occurring within the interval, but the coverage probability across repeated sampling.

b..
Asking for bigger probability to cover the true population,the range of values that are likely to include the true population mean should be expended. That is why the interval of 99% confidence level is wider than 95% confidence level.

Task 3:

It is of interest to see if the proportion of people in the population who would consider buying their ticket from a lucky store is 0.5, or if it is less than that. Use a t-test to answer this question.
a. Clearly define the null and alternative hypotheses that are most appropriate for this question.[4 Marks].
b. Perform the t-test and paste your output in your report.[2 Marks]. c. Interpret the findings of the test. [4 Marks].

```
tr <- t.test(sample,alternative='less',mu=0.5,conf.level = 0.95)
print(tr)
```

```
##
##  One Sample t-test
##
## data:  sample
## t = -3.1921, df = 42, p-value = 0.001337
## alternative hypothesis: true mean is less than 0.5
## 95 percent confidence interval:
##       -Inf 0.3954802
## sample estimates:
## mean of x
```

```
## 0.2790698
```

Answer for Task3:

a..

Null hypotheses H0: the proportion(mean) = 0.5 Alternative hypotheses HA: the proportion < 0.5

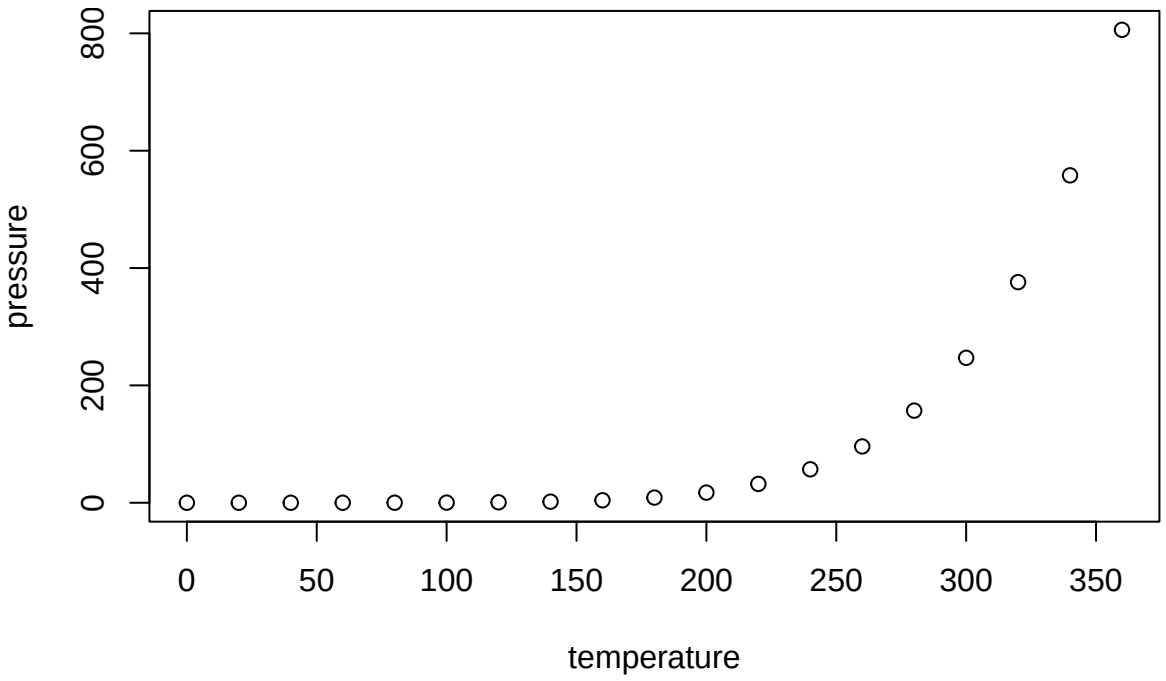
b..

The P value is 0.001337461 at 95% confidence level.
Reject the Null jhypotheses H0.
One Sample t-test.

data: sample
t = -3.1921, df = 42, p-value = 0.001337.
alternative hypothesis: true mean is less than 0.5.
95 percent confidence interval:
-Inf 0.3954802.
sample estimates:
mean of x
0.2790698

c.. Evidence in favour of H0. We reject H0 when the p-value is very small, i.e. there is very little evidence in the data to support H0. ## Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.