CS 108 - Statistics, Sections B

Fall 2019, AUA

Homework No. 10

Due time/date: Section B: 10:32 AM, 02 December, 2019

Note: Please use **R** only in the case the statement of the problem contains (R) at the beginning. Otherwise, show your calculations on the paper. Supplementary Problems will not be graded, but you are very advised to solve them and to discuss later with TA or Instructor.

Problem 1: Asymptotic Confidence Intervals

a. Asymptotic CI for the Exponential Model

Here we want to construct an Asymptotic CI for λ in the $Exp(\lambda)$ Model. To that end, we first construct an Asyptotic CI for θ in the $\widetilde{Exp}(\theta)$ Model, from **Homework 09, Problem 2a**.

1. Find the MLE and Fisher Information for θ in the $\widetilde{Exp}(\theta)$ Model;

Note: You can just use the value of the Fisher Info for this model obtained in **Homework 07, Problem 1b**

2. Construct the $(1-\alpha)$-level Asymptotic CI for θ, using the MLE Asymptotic CI construction method (**Lecture 35**);
3. Using $\lambda = \dfrac{1}{\theta}$, construct a $(1-\alpha)$-level Asymptotic CI for λ

b. (R) Comparing CIs for the Exponential Model

Now, using **R**, let us compare CIs for $Exp(\lambda)$ Model Parameter λ, obtained using the Chebyshev Method (**Homework 09, Problem 2a**), and the Asymptotic Method, obtained in **Part a.** of this Problem. To that end, for each of the Methods, separately, do the following

- take a Sample Size $n = 100$;
- take a particular value of true λ, λ_{true} (the same for both cases);
- for i running from 1 to 120, do the following:
  - set the seed, depending on i, say, `set.seed(10*i)`
  - generate n samples from $Exp(\lambda_{true})$
  - construct the CI
  - if CI contains λ_{true}, plot a segment showing that CI over the point with x coordinate i in black; otherwise, plot it in red;
- plot a horizontal line passing through $y = \lambda_{true}$

Compare and explain the obtained graphs, do a conclusion.

Note: We are fixing the seed for these 2 methods, to have the same random sample both for Chebyshev CI and MLE CI.

(Supplementary) Think about plotting CIs obtained by these methods one over the other or side by side (for the ease of visual comparison).

Problem 2: Hypothesis Tesing

a. Choosing Null and Alternative Hypotheses

For the examples below, state the Null and Alternative Hypotheses. Describe a method (in general terms) to Test that Hypothesis.

1. We want to see if more than 50% of AUA students are visiting gyms regularly.
2. We want to see if the the Average Monthly Nominal Wage in Armenia is indeed $178430.0 AMD$, as stated at https://www.armstat.am/en/.
3. We want to see if an average person drinks more than 1.5l water daily.
4. We want to see if the percentage of man drivers in Armenia responsible for car accidents is higher than the percentage of woman drivers.
5. We want to see if the Right-Turn-on-Red (allowing drivers to turn right when the signal is Red) increases the number accidents[1].

b. Significance and Power

Assume $x_1, ..., x_{100}$ is an Observation from the model $\mathcal{N}(\mu, 4)$. We want to test the hypothesis

$$\mathcal{H}_0 : \ \mu = 0 \qquad vs \qquad \mathcal{H}_1 : \ \mu = 1.$$

We choose the Rejection Region

$$RR = \left\{ \overline{X} > 0.5 \right\}.$$

where

$$\overline{X} = \frac{X_1 + X_2 + ... + X_{100}}{100}.$$

The Sample Mean of our Observations x_k is $\overline{x} = 0.32$.

1. What is the Decision of the Test?
2. Calculate the Type I Error Probability;
3. Calculate the Type II Error Probability;
4. Calculate the power of the test.

c. Designing a Proportion Test

Based on the CI constructed for the $Bernoulli(p)$ Model by the Chebyshev Method, design a Test for Testing

$$\mathcal{H}_0 : \ p = p_0 \qquad vs \qquad \mathcal{H}_1 : \ p \neq p_0.$$

[1] See, for example, https://rosap.ntl.bts.gov/view/dot/1322/

d. Significance and Power, again

Assume we want to Test a Hypothesis about the Mean of the Normal Distribution, when the Variance is known. In particular, we want to design a test in the following way: we want to have both the Significance Level α (the Confidence Level $1-\alpha$), and we want to have a Power not less than $1-\beta$ (α, β are given). Our test is ($\mu_0 < \mu_1$ are given, fixed):

$$\mathcal{H}_0: \ \mu = \mu_0 \qquad vs \qquad \mathcal{H}_1: \ \mu = \mu_1.$$

If the number of the Sample is fixed, then we cannot control both Type I and II Errors Probabilities. On the other hand, we can choose the number of observations, n, to obtain the necessary bounds for error probabilities. So assume

$$X_1, ..., X_n \sim \mathcal{N}(\mu, \sigma^2)$$

is our Random Sample, and σ^2 is known. We will test the above hypothesis based on the Test Statistics

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}.$$

Our Rejection Region will be

$$RR = \{Z > c\}.$$

1. Choose the Critical Value c such that the Significance Level of our Test is α;
2. Calculate the Probability of the Type II Error under the obtained above Rejection Region;
3. Choose the number n so that the Power of the Test will be $\geq 1-\beta$;
4. (Suplementary) What will happen with n, if μ_1 will approach μ_0? Explain.

Problem 3: Testing in Practice

a. Testing for the Normal Mean

On a new highway, the speed limit is set to 70 km/h. The police department is claiming that the average speed is higher than 70 km/h on that highway, and wants to put an aragachaph on that highway. In support of their claim, the department is providing a data from randomly chosen 10 cars: the recorded speeds are (in km/h):

66, 79, 80, 74, 81, 79, 65, 78, 77, 69.

Is this data providing a sufficient evidence against the Hypothesis that the average speed is 70 km/h? Use 1% significance level. You can assume that the car speeds are normally distributed.

1. Do the Testing by hand, i.e., by calculating Test Statistics value and comparing with Critical Values;
2. (**R**) Do the Testing using CIs;
3. (**R**) Do the Testing using p-Values

b. Testing for the Proportion

A company claims that 83% of customers buying their product are satisfied with that product. Assume we have a sample of 74 persons, chosen randomly, who bought their product, and only 52 of them are satisfied. Perform a Test of Significance Level 0.05 to see if the data is supporting the company's claim.

c. Testing for the equality of Proportions

The following example is from Wikipedia page for A/B Testing, https://en.wikipedia.org/wiki/A%2FB_testing.

> A company with a customer database of 2,000 people decides to create an email campaign with a discount code in order to generate sales through its website. It creates two versions of the email with different call to action (the part of the copy which encourages customers to do something — in the case of a sales campaign, make a purchase) and identifying promotional code.
>
> - To 1,000 people it sends the email with the call to action stating, "Offer ends this Saturday! Use code A1",
> - and to another 1,000 people it sends the email with the call to action stating, "Offer ends soon! Use code B1".
>
> All other elements of the emails' copy and layout are identical. The company then monitors which campaign has the higher success rate by analyzing the use of the promotional codes. The email using the code A1 has a 5% response rate (50 of the 1,000 people emailed used the code to buy a product), and the email using the code B1 has a 3% response rate (30 of the recipients used the code to buy a product). The company therefore determines that in this instance, the first Call To Action is more effective and will use it in future sales. A more nuanced approach would involve applying statistical testing to determine if the differences in response rates between A1 and B1 were statistically significant (that is, highly likely that the differences are real, repeatable, and not due to random chance).

Now, do that nuanced approach:

- perform a significance test at 5%, by showing your calculations;
- (**R**) Use **R**-s `prop.test` to perform the Test. Give your decision based on the output.

d. Testing for the equality of Means

Assume we want to assess if a full-time job is affecting negatively on the average grade for Statistics. Say, our survey shows the following results:

	No. of Students	Average grade	Grades sd
Students without a f/t job	34	75.2	17.8
Students with a f/t job	25	77.1	16.9

- Test, at the 5% level, our Hypothesis, by showing your calculations;
- (**R**) Use **R**-s `t.test` to perform the Test. Give your decision based on the output.

Note: You can assume the grades are Normally Distributed. Also, you can consider two cases: Variances are equal or not.

Problem 4: Philosophical

a. What is wrong?

Assume I have a Data from a Normal Distribution, stored in the `obs` variable:

```
print(obs)
```

```
##  [1]  2.9925  5.7408 -3.6110  7.4327  2.7305  4.1735  6.8206  1.3712  5.5738
## [10]  4.6083  4.1610  5.2390  1.3901  9.3370  6.0749 -3.2101 -0.4175  0.6467
## [19] -0.4276  3.6057
```

with a Sample Mean

```
cat("Mean of a Sample is equal to: ", mean(obs))
```

```
## Mean of a Sample is equal to:  3.211605
```

Because I know that my sample comes from a Normal, I am making a Hypothesis:

$$H_0: \ \mu = 3.211605 \qquad vs \qquad H_1: \ \mu \neq 3.211605.$$

I want to test it at 1% Significance Level. I am using the t-test, since I do not know the value of σ:

```
a <- 0.01
mu0 <- 3.211605
crit.value <- qt(1-a/2, df = 20-1)
cat("The Critical Value is: ", crit.value, "\n")
```

```
## The Critical Value is:  2.860935
```

```
t = (mean(obs)-mu0)/(sd(obs)/sqrt(20))
cat("The value of T-Statistics is: ", t, "\n")
```

```
## The value of T-Statistics is:  0
```

```
abs(t)>crit.value
```

```
## [1] FALSE
```

So I am Not Rejecting $\mathcal{H}_0$.

Or, which is the same, I am just running **R**-s `t.test`:

```
t.test(obs, mu = 3.211605, conf.level = 0.99)
```

```
##
##  One Sample t-test
##
## data:  obs
## t = 0, df = 19, p-value = 1
## alternative hypothesis: true mean is not equal to 3.211605
```

```
## 99 percent confidence interval:
##  1.01337 5.40984
## sample estimates:
## mean of x
##  3.211605
```

So, by the result of the Test, $p > 0.01$, or μ_0 is in the CI, so I am not Rejecting $\mathcal{H}_0$.

What am I doing wrong?