Statistical Inference

Lecturer: Abdol-Hossein Vahabie

Spring Semester 1401-1402

Writing Assignment III — Deadline 1402/02/18

1 True False

You're conducting a study on the effectiveness of two different treatments for a medical condition, and you want to compare their effects on patients' blood pressure. You randomly assign 50 patients to each treatment group, and measure their blood pressure before and after treatment. (5pts)

Which of the following statements is true or false, and why?

a) A bootstrap distribution can be created by resampling the patients within each treatment group.

b) If a two-sample t-test indicates that there is a significant difference in mean blood pressure between the two treatment groups, then we can conclude that one treatment is more effective than the other.

c) A larger sample size is always better when conducting a paired t-test.

d) In a paired analysis of blood pressure measurements, the null hypothesis would be that the mean difference between pre- and post-treatment blood pressure is zero.

e) A higher degree of variability within each treatment group would make it more difficult to detect a significant difference in mean blood pressure between the two groups.

f) If the mean difference in blood pressure between pre- and post-treatment measurements is greater in one treatment group than in the other, then we can conclude that that treatment is more effective.

g) In a paired analysis of blood pressure measurements, the differences between pre- and post-treatment blood pressure are used to calculate a confidence interval for the true mean difference in blood pressure.

h) To increase the power of a hypothesis test, we can decrease the significance level (alpha).

interval for the overall population proportion? Why or why not?

2 Hypothesis Test

A college claims that their electrical and computer graduates earn an average of \$98.6 K annually. A Ph.D student in statistics believes that this is an underestimate and plans to investigate the college's claim by collecting a sample of 52 randomly selected electrical and computer graduates of this college to test the hypothesis. (These data are provided in the table below.)

97.82138	98.67651	97.80384	97.92021	98.81304
97.96004	98.26793	98.97997	97.46191	99.37323
99.67803	98.70392	98.57253	97.94584	98.61963
97.94205	98.48125	97.81605	98.82261	98.75639
99.13282	99.63726	97.84347	98.94393	97.31063
97.97241	98.92979	97.78072	98.52047	97.67014
98.03187	98.25676	97.86675	96.95675	98.45049
98.60867	98.37789	97.95108	98.89890	99.01991
96.66481	99.05201	97.07024	98.04948	97.52770
97.85462	99.18324	98.64935	98.62164	96.86658
98.41857	98.71370			

The Income of samples is measured with the following summary of the data:

$$\mathrm{n} = 52, \bar{x} = 98.2932,\ \mathrm{s} = 0.6865$$

After checking the necessary conditions of the hypothesis test, answer the following questions. (10pts)

1. Write the null and alternative hypothesis.

2. Construct a 98% confidence interval for the mean income. And interpret this confidence interval. In this context, what does a 98% confidence level mean?

3. (R) Plot the histogram of the income and mark the mean income as a vertical line on top of the histogram plot. You must also mark the confidence intervals on the plot as two vertical lines.

4. Provide a two-side hypothesis test for a mean average income of \$98.6 K and use the information above to evaluate a test with significance level $\alpha = 0.02$.

3 ANOVA Test

Self-driving cars have gained popularity in recent years, but they still face challenges, one of which is finding the most efficient and safe path to reach their destination. To overcome this, reinforcement learning can be utilized to train self-driving cars to learn the best path.

Imagine that you are a data scientist working for a company that is developing self-driving cars. This company wants to create an intelligent and optimal path planner for their new product. To accomplish this goal, three teams simultaneously worked by utilizing reinforcement learning algorithms, Deep QNetwork (DQN), Q-learning, and Actor-Critic, to determine the optimal path to a specified destination. The company's end goal is to select the most effective algorithm for the job.

As part of your role, you are required to determine whether there is a significant difference in the travel times of self-driving cars trained using these algorithms. To achieve this, travel times (measured in minutes) were recorded for each algorithm, and the data is presented below for analysis.

Q-learning	1000	1100	700	800	500	700
Actor-Critic	890	650	1100	900	400	350
Deep Q-Network (DQN)	1200	1000	980	900	750	800

Write the Null and Alternative hypotheses for testing for a difference between the average travel time of three different Algorithms and do calculation and conduct analysis using one-way ANOVA. (It's important to complete the ANOVA table.) (15pts)

4 Hypothesis Test

1. A coffee shop owner wants to know if there is a relationship between the time of day and the type of drink ordered by customers. The following table shows the observed frequencies for the different combinations of time of day (morning, afternoon, evening) and drink type (coffee, tea, smoothie):

	Coffee	tea	smoothie	total
morning	75	30	15	120
afternoon	60	45	15	120
evening	30	15	45	90
total	165	90	75	330

   Test the hypothesis of whether there is a relationship between time of day and drink type, using a significance level of 5. (10pts)

2. A nutritionist wants to study the calorie intake of people who regularly eat at fast food restaurants. She randomly sampled 50 individuals who eat at fast food restaurants at least twice a week and recorded their daily calorie intake. The table below shows the summary statistics of the sample:

n	$\bar{x}$	s	min	max
50	2100	500	1200	3400

   The nutritionist believes that the average daily calorie intake of people who regularly eat at fast food restaurants is 2500 calories per day. Is there sufficient evidence to reject this claim at the 5% level of significance? (15pts)

   a) Write the null and alternative hypotheses in symbols and words.

   b) Check the necessary conditions and calculate the test statistic, t, and the associated degrees of freedom.

   c) Find and interpret the p-value in this context.

   d) What is the conclusion of the hypothesis test?

   e) If you were to construct a 95% confidence interval that corresponded to this hypothesis test, would you expect 2500 calories to be in the interval?

5 R Programming

A financial institution has produced two investment products with two different strategies. These products are recommended for investors looking for rapid gains in volatile markets. For a comparative study, three financial advisors with the following decision-making strategies have recommended these two products to a number of clients (each advisor to 100 clients) to compare their effectiveness:

- Advisor A: win-stay lose-shift approach
- Advisor B: Fully Random approach
- Advisor C: First, recommends 10 times each product. Then, he recommends the product with the maximum effect 7 times, followed by randomly selecting 3 recommendations. This process is repeated by again recommending the product with the maximum effect 7 times, followed by another 3 random selections. (This cycle should continue until the last trial is completed.)

To check the effectiveness of the investment product, according to industry standards, the investment returns of the clients will be analyzed after one year. We consider the investment returns (in percentage) up to a desired level as a reward, otherwise, a penalty is applied. The investment product with more rewards is considered better and more effective.

To receive this Reward, you must use the following function in your code:

Listing 1: The Reward function R code that you should use for this problem.

```
Get_Reward <- function(Product, Student_ID) {
 compute_reward <- function(alpha, beta, scale, offset) {
  stochastic_num <- rbeta(1, alpha, beta)
  return(stochastic_num * scale + offset)
}
suffix <- (as.numeric(substring(as.character(Student_ID), 7)) %% 3) + 1
alpha1 <- suffix
beta1 <- 5
alpha2 <- 2
beta2 <- alpha1

if (Product == 1) {
   return(compute_reward(alpha1, beta1, -14, 7))
} else if (Product == 2) {
   return(compute_reward(alpha2, beta2, 14, -7))
} else {
   print("Product number must be 1 or 2")
}
}
```

Win-Stay-Lose-Shift (WSLS) is a simple strategy often used in reinforcement learning. It is a type of ”decision rule” that an agent can use to decide what action to take in response to a particular state of the environment. In this strategy, we consider a reference for reward. If the reward is greater than this reference, it is considered a win and otherwise a loss. (in this question reference = 0)

We use the following rule to make a decision:

$$\text{If win: } \begin{cases} p\left(a_{t+1} == a_t \mid \text{ win }\right) = P_{SW} = P(\text{ stay — win }) \\ p\left(a_{t+1} \neq a_t \mid \text{ win }\right) = 1 - P_{SW} = P(\text{ shift — win }) \end{cases}$$
$$\text{If lose: } \begin{cases} p\left(a_{t+1} == a_t \mid \text{ lose }\right) = 1 - P_{SL} = P(\text{ stay — lose }) \\ p\left(a_{t+1} \neq a_t \mid \text{ lose }\right) = P_{SL} = P(\text{ shift — lose }) \end{cases}$$

In this exercise, consider the probabilities $P_{SW} = 0,8$ and $P_{SL} = 0,7$. To clarify, when we take action a_t and it results in a positive reward, there's an 80% chance we will choose the same action again in the next step, and a 20% chance we will choose a different action. On the other hand, if the previous action resulted in a negative reward, there's a 70% chance we will shift to a different action, while there's a 30% chance we will stick with the same action.

1. In this case, the environment gives random rewards. What do you think is the reason? (2pts)

2. Implement this problem in the R environment with the help of the given functions. Run each method once for a hundred customers (trial 100). Visualize the results using a reward per trial chart and perform a thorough analysis of the outcomes. (6pts)

3. To obtain accurate results, it's important to repeat the previous section multiple times and record the outcomes. Start by providing a rationale for why this is necessary. Then, Repeat this once 5 times and another time 20 times. and plot the result as a reward per trial. plots should also include confidence intervals with a significance level of $\alpha = 0.05$. Finally, Compare the three strategies. (7pts)

4. Create a boxplot that illustrates the reward distribution for the three advisors in the experiment's 10 replications. Focus solely on the last client's data (100th trial) when generating this plot. (3pts)

5. Some experts have claimed that the first product is more effective than the second product. Check the correctness of this claim using the hypothesis test. (7pts)

6 R Programming

You should develop a deep learning model to classify images of handwritten digits (0-9) from the MNIST dataset. After training your model on a large dataset, you want to evaluate its performance on a new dataset of 5000 images. You randomly sample 500 images from the new dataset and use your model to classify them as either a digit from 0 to 9. You record the predicted class and the actual class for each image in an R dataframe.

Now, you want to assess whether your model's performance on this new dataset is significantly different from what would be expected by chance. To do this, perform a goodness-of-fit test in R, using the chi-squared test statistic and a significance level of 0.05. (20pts)

Listing 2: here is the code you can use for Deep learning part

```
#Load the MNIST dataset
mnist <- dataset_mnist()
x_test <- mnist$test$x
y_test <- mnist$test$y

#Load your trained deep learning model
model <- load_model_hdf5("path/to/model.h5")

#Use the model to make predictions on a new dataset of 5000 images
new_data <- x_test[1:5000, , , ] # extract 5000 images from the test set
predictions <- predict(model, new_data) # make predictions on the new data

#Extract the predicted and actual classes for the first 500 images
predicted <- apply(predictions[1:500, , ], 1, which.max) - 1 # subtract 1 to convert to
    digits 0-9

actual <- y_test[1:500] # extract actual labels from the test set

#Store the predicted and actual classes in a dataframe
df <- data.frame(predicted = predicted, actual = actual)
```

Required Document

Please upload a file in ZIP format (not RAR) to the elearn platform.

General Rules

- Please upload a file in ZIP format (~~not RAR~~) to the elearn platform(`https://elearn5.ut.ac.ir/course/view.php?id=14838`).
- You are allowed a total grace period of 3 days to submit late assignments for all of your exercises.
- It is prohibited to use ~~handwritten material~~ and only material produced through typing in the HW template is permissible.
- Utilizing a LaTeXto compose the report will grant an additional 5 points.
- If you submit your exercise in LaTeXformat, please include your complete LaTeXproject as a Zip file along with the pdf and r codes.
- You must make a decision to select and address either one of the two R inquiries (answering both questions will not earn any points).

Deadline

Wednesday 23:59. 1402/02/18.

Contact Information

Please direct your questions regarding Homework 3 only to the teaching assistants, Sarmad Zandi and Melika Sadeghi, through the course mail (statistical.inference.ut@gmail.com). Use ”HW3” as the subject line.

Good Luck

Attachment

In this assignment, you will have the opportunity to explore important concepts in statistics, including the Central Limit Theorem (CLT) , Confidence intervals (CI), Hypothesis testing, and perform statistical tests for numerical, categorical, and image data.

The central limit theorem is a fundamental idea in statistics that tells us that as the sample size increases, the distribution of sample means will approach a normal distribution, regardless of the shape of the population distribution. This allows us to make inferences about a population from a sample, which is crucial for statistical analysis.

Confidence intervals are a range of values that we can be reasonably certain includes the true population parameter. They are calculated based on the sample data and the level of confidence chosen by the analyst. Confidence intervals are used to quantify the uncertainty around a point estimate of a population parameter.

Hypothesis testing is the process of making a decision about a population parameter based on a sample. It involves setting up a null hypothesis (a statement of no effect or no difference) and an alternative hypothesis (a statement of an effect or difference) and then collecting data to see which hypothesis is supported by the evidence.

Throughout this assignment, you will also have the opportunity to perform statistical tests for numerical, categorical, and image data. These tests will allow you to make inferences about population parameters based on sample data and will help you develop your skills in statistical analysis. by the end of this assignment, you will have a deeper understanding of these important concepts and their practical applications in statistics. (For more information, I suggest you read the attached links or other resources first.)