Started on	Monday, 22 July 2024, 12:19 PM
State	Finished
Completed on	Monday, 22 July 2024, 3:06 PM
Time taken	2 hours 46 mins
Marks	48/48
Grade	**10** out of 10 (**100**%)

Question 1

Correct

Mark 4 out of 4

A hypothesis test checks and compares the probability that some summary value of an observed sample was drawn from an assumed distribution of that summary value; in particular, this is compared against the probability of coming from a different distribution.
Let's repeat some basic terms needed to talk about such hypothesis tests. Match the correct terms.

- The random variable that maps a sample (drawn from an underlying distribution) to a single summary value of interest is called the [test statistic].
  Typical ones are the mean, the difference of means, or the count of certain occurrences.
- The distribution of the test statistic values over finite samples is called the [sampling distribution] of the test statistic.
- For the mean and difference of means, one can derive it from the distribution of the [population distribution] given that this is normally distributed and samples are drawn randomly.
- The hypothesis stating the assumed distribution is called the [null hypothesis].
- The hypothesis about the sample coming from a different distribution is called the [alternative hypothesis].
- Given the test statistic of a sample, the probability of obtaining that test statistic value is its [p-value].
- For interval scaled test statistics, the most extreme test statistic value(s) for which the null hypothesis is still accepted, are called [critical value].

Your answer is correct.

Correct

Marks for this submission: 4/4.

Question 2

Correct

Mark 4 out of 4

Let's repeat some basic terms needed to talk about hypothesis tests results. Match the correct terms.

- The probability of falsely accepting the null hypothesis is the type II error .
  It can only be calculated if a distribution for the alternative hypothesis is chosen.
- The probability α of falsely rejecting the null hypothesis is the type I error
- This is the probability that the distribution of the null hypothesis applies and a sample that leads to a reject is just by chance drawn from that distribution.
- A result is said to provide significant evidence against the null hypothesis, if the p-value of the result is <5%.
  In particular, the result allows to reject the null hypothesis for an α at least as low as 5%.
- A result is said to provide even highly significant evidence against the null hypothesis, if the p-value of the result is <1%.
- The probability of correctly rejecting the null hypothesis is the test power , and calculates as $1 - \beta$.

Your answer is correct.

Correct

Marks for this submission: 4/4.

Question 3

Correct

Mark 3 out of 3

Let's repeat some basic terms needed to talk about hypothesis test types. Match the correct terms.

- A test where deviations from the null hypothesis in both directions are allowed is called two-sided hypothesis test, e.g., if one wants to assess whether a modification has any effect.
  A test querying a deviation in a specific direction is a one-sided hypothesis test, e.g., if a positive effect is assessed.
- A test with a fixed reference distribution as null hypothesis is a one-sample hypothesis test.
  One where two samples are compared is a two-sample test.
- When the test statistic is assumed to follow a Student's t distribution, one can apply a t-test .
  When the test statistic is assumed to follow a normal distribution with known variance, one can apply a z-test .

Your answer is correct.

Correct

Marks for this submission: 3/3.

Question **4**

Correct

Mark 3 out of 3

Example: Assume we would like to know whether our modification to a ML method improves its accuracy (alternative hypothesis) or not (null hypothesis). We know the accuracy of the baseline is $80 \pm 3\%$ (which describes a distribution of accuracy over training splits)*. In a 4-fold cross validation we get an average of 83% accuracy, meaning:

- For our sample size of 4 accuracy values, the standard error of the mean of accuracies is $\frac{3}{\sqrt{4}} = 1.5$, and, thus,
- the z-score of our result wrt. the baseline distribution consequently is $\frac{83-80}{1.5} = 2$.
- Looking up the (right-sided) p-value for that z-score, we find that there only is a <2.5% probability that our modification brings no improvement compared to the baseline and results are obtained by chance (type I error).

We would like to communicate our results. However, to state that "We observed significant results." is meaningless to a reader without further context information.

Which of the following information must be provided in addition?

Some comments on the underlying assumptions:

** In this modelling case, our underlying population is the accuracies of training splits (and the population distribution is the distribution of accuracy values over training splits). This population is assumed to be normally distributed**. The test statistic that we consider, is the mean of accuracy values, which is a convenient choice: We know what the sampling distribution of the mean looks like, assuming accuracy values are normally distributed and the sample set of accuracy values is drawn randomly***.*

*** Note that it is a strong assumption that accuracy values are normally distributed over training splits! Even if (a) the underlying distribution of the population of the single test and training data points was normal (which usually is not the case in practice), and (b) the samples are drawn randomly; then, other than the mean, the accuracy (here looking at it as a kind of summary statistic) needs still not to be normally distributed.*

**** To be precise, we draw randomly from the training-test-splits; this only translates into a random sampling of accuracy values, if we assume that accuracy is normally distributed over training and test samples.*

Select one or more:

- [ ] a. the **type II error** probability (β) for this results
- [ ] b. the **type I error** probability (α) used to derive the significance claim
- [ ] c. the **p-value** of obtaining the observed results
- [x] d. the **test statistic**, here the mean of accuracy results
- [x] e. the **alternative hypothesis**, e.g., "has any effect" versus "has positive effect"
- [x] f. the **null hypothesis**, i.e., the assumed probability distribution of the test statistic over samples

Your answer is correct.

Correct

Marks for this submission: 3/3.

Question 5

Correct

Mark 3 out of 3

You conduct some independent tests of your blood pressure to check whether the average exceeds that of the average for people of your age.

Assume blood pressure is distributed normally over time. What does the statement "*The results show significant evidence that the null hypothesis can be rejected.*" tell you?

Select one or more:

- [ ] a. The type II error β (false rejection of H_1) is <1%.
- [x] b. The type I error α (false rejection of H_0) is <5%.
- [ ] c. You have an unusually high average blood pressure.
- [ ] d. You have an unusual average blood pressure.
- [x] e. There is only a <5% probability that you have no unusually high blood pressure and just by chance drew an unusual sample.
- [ ] f. The type II error β (false rejection of H_1) is <5%.
- [ ] g. There is only a <5% probability that you have no unusual blood pressure and just by chance drew an unusual sample.
- [x] h. The probability of falsely rejecting the null hypothesis is very low.
- [ ] i. The type I error α (false rejection of H_0) is <1%.
- [ ] j. The probability of correctly accepting the alternative hypothesis (power of the test) is very high.

Your answer is correct.

Correct

Marks for this submission: 3/3.

Question 6

Correct

Mark 4 out of 4

A proper formulation of the previous test result would be the following. Which information gives insights into which aspect of the hypothesis test?

"The observed results provide

- ***significant** evidence (=>* *low p-value, type I error <5%* *)*
- *that the **average accuracy** (=>* *test statistic* *)*
- *is not **the same as that of the baseline, which is assumed to be distributed** with standard deviation 3 around a mean of 80% (in short:* 80 ± 3*) (=>* *null hypothesis, one-sample test* *),*
- *but that our modification has **a positive effect** (=>* *alternative hypothesis, right-sided test* *)."*

p-value of less than 2.5%

Your answer is correct.

Correct

Marks for this submission: 4/4.

Question 7
Correct
Mark 3 out of 3

Before applying a specific hypothesis test technique, it is quite important to check and know that all preliminaries for this technique are fulfilled. In the lecture, we had a glimpse at two types of tests with differently strict but generally very restrictive preliminaries.
What preliminaries do they have in common for being applicable to testing a hypothesis?

Select one or more:

- [ ] a. If doing a two-sample test: The sample sizes are equal.
- [x] b. The data must have interval scale, i.e., be comparable and allow for measuring a relative distance between them.
- [x] c. The test statistic is normally distributed if all nuisance parameters like variance are known. This follows automatically for the mean as test statistic, if random samples are drawn from a normally distributed population. But else not!
- [x] d. The samples are drawn randomly. Otherwise the assumption that the sampling distribution is normal is violated.
- [x] e. If testing the mean: The underlying population of the samples is normally distributed.
- [ ] f. The data must have relative scale, i.e., be comparable real values with absolute zero.

Your answer is correct.

Correct
Marks for this submission: 3/3.

Question 8
Correct
Mark 3 out of 3

What are the differences of z- and t-test for practical application?

z- vs. t-test

	z-test	t-test
Applicable if test statistic variance is...	known	unknown
Reads p-values from ...	normal distribution	Student's t-distribution
Recommended for sample sizes of ...		<=50

Correct
Marks for this submission: 3/3.

Question 9

Correct

Mark 6 out of 6

Observations: A coin shows an unequal number of tails and heads when thrown k times.

The **null hypothesis** is that the coin is [fair] (count of tails is normally distributed around half the number of throws), the **alternative hypothesis** is that the coin is [unfair] (count of heads much higher or lower than half the number of throws).

This is a [two-sided] [one-sample] hypothesis test of the test statistic [count] with sample size [1].

Correct

Marks for this submission: 6/6.

Question 10

Correct

Mark 1 out of 1

Again the coin example, but with as alternative hypothesis the claim that the coin shows more often heads.

This is a [one-sided (right-sided)] hypothesis test.

Correct

Marks for this submission: 1/1.

Question 11

Correct

Mark 4 out of 4

Null hypothesis: A manufacturer claims that their lamps last for 10k hours usage on average (normally distributed).

We don't believe them and formulate the alternative hypothesis that lamps last less than 10k hours on average, and test it on 30 lamps.

This is a [one-sided (left-sided)] [one-sample] hypothesis test of the test statistic [mean] with sample size [30].

Correct

Marks for this submission: 4/4.

Question 12

Correct

Mark 6 out of 6

Observations: Average weight of two groups of penguins from two different penguin colonies, with group sizes n_1 and n_2 respectively.

The null hypothesis is that the average weight of a penguins in the colonies does [not differ] (= sampling distribution of the difference of means has mean 0).

The alternative hypothesis is that the weight does [differ].

This is a [two-sided] [two-sample] hypothesis test of the test statistic [difference of means] with sample sizes [n1,n2].

We assume that weight of penguins is normally distributed, and sampling of the groups is done randomly. Thus, if the weight variance of each colony is known, the sampling distribution of the difference of means follows a [normal distribution] and a [z-test] can be applied.

If the variances are estimated from the samples, the difference of their means follows a [Student's t distribution] and a [t-test] can be applied.

Correct

Marks for this submission: 6/6.

Question 13

Correct

Mark 4 out of 4

Observation: When applying a modification to a baseline ML method, the average accuracy in a 5-fold cross-validation is better than that in a 5-fold cross-validation of the baseline.

Based on this, we claim that our method is better than the baseline (alternative hypothesis).

This is a [one-sided (right-sided)] [two-sample] hypothesis test of the test statistic [difference of means] with sample sizes 5 and 5.

Assuming that the accuracy is normally distributed over train-test-splits, we can apply a [t-test].

Correct

Marks for this submission: 4/4.

◀ 02. Quiz - Stochastic and Statistics

Jump to...

04. Quiz - Bayesian networks ▶