CS306: Data Analysis and Visualization
# Lab 6: Report

Rajdeep Pinge 201401103
Aditya Joglekar 201401086

Objective:

To study how t-test can be applied to compare data, and make informed decisions about the given problem.

Problem Description:

Visual attention mechanisms guide where a typical human being will focus his/her attention to when looking at a real world scene (or an image). In digital video processing, this is represented by a saliency map in which a higher value indicates that the corresponding region in the image attracts more attention (likewise, a smaller val99ue in saliency map implies that the corresponding area attracts lower eye attention). An example is shown below. The 'input image' represents a real-world scene while the 'Saliency map' shows the corresponding regions of attention. Notice how human eyes tend to focus only on certain objects (eg. people, boat etc.) and ignore others (eg. sea, skyline).
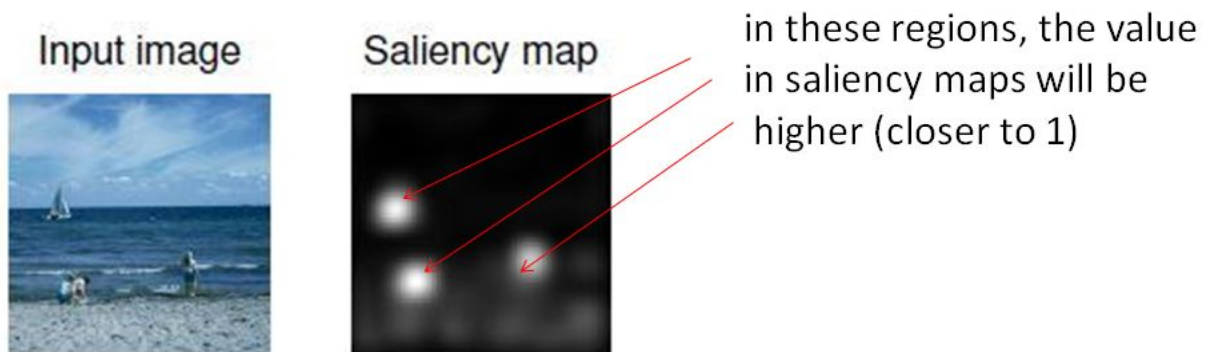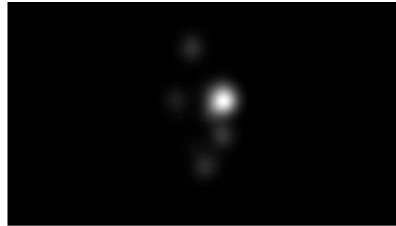

In this experiment, your goal is to analyze if processing of video frames affects where humans will look at.
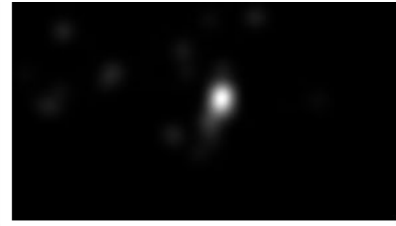
Q1. Tunnel

Original frame (Tunnel)   saliency_map_Tunnel_processing1   saliency_map_Tunnel_processing2

The data has been given for 2 saliency maps. Original frame corresponds to the real scene. The first saliency map is obtained when the original frame processed by algorithm 1 is viewed by observers. Similarly, the second saliency map is obtained when the original frame processed by algorithm 2 is viewed by the same set of observers.

Your goal is to find if the saliency map due to algorithm 1 is statistically different from that due to algorithm 2. Load data_lab6.mat. You will see the data 'saliency_map_Tunnel_processing1' and 'saliency_map_Tunnel_processing2'.

1. Which t-test, paired or unpaired is suitable, in this case? Give reasons for your answer.

In the given experiment, we use two algorithms on the same underlying image frame. Here, we treat the saliency maps of the two algorithms as the samples taken from the given image.

The two samples are thus not independent. Hence, we will choose the paired t-test.

Note - Here we only consider the mean intensity of the two pictures not taking into account the spatial distribution. We are only looking at the mean intensity in the saliency maps. A greater intensity (in absolute terms) means that the map contains more white spots i.e. more points of focus and hence intensity pixels of higher value. Which means on an average the algorithm leads to humans paying more attention for that sample.

Note- Before, performing the t-test, we also calculate the Effect size as a sanity check.

2. Based on your answer, perform the t-test and analyze the differences between the two saliency maps.

Tunnel images-

Hypothesis

Null hypothesis H_0: μ_1 - μ_2 = 0
Alternate hypothesis H_a: μ_1 - μ_2 ≠ 0

Where μ_1: mean of the saliency map generated by algorithm 1
μ_2: mean of the saliency map generated by algorithm 2

Paired t-test results

Mean of difference = 0.0020 (low)

Standard Deviation of difference = 0.0220

Effect size = 0.0919 (low) < 0.2

t-score = 132.3175

P-value = 1.0000

Degree of Freedom = n - 1 = 2073599

Unpaired t-test results

algorithm 1:	Mean = 0.0111	Standard Deviation(s1) = 0.0619
algorithm 2:	Mean = 0.0131	Standard Deviation(s2) = 0.0627

We assume that sampling distribution of the mean is normal due to Central Limit Theorem. Hence 1st assumption of t test is valid.

Here the standard deviation and consequently, the variance is of equal order. This satisfies the 2nd assumption for t test

Degree of Freedom = n_1 + n_2 - 2 = 4147198

We assume that independence of degree of freedom, the 3rd assumption for t test is satisfied.

Difference of mean = 0.0020   s = min(s1, s2) = 0.0619

Effect size = (Difference of mean) / s = 0.0326 (low) < 0.2

t-score = 33.0254

p-value = 1.0000

Observations:

1. The effect size which has been calculated prior to performing statistical tests is very low, indicating that there is not much difference between the samples which are generated by two different algorithms.
2. Notice that we have got a very high value of t-score.
3. The p-value obtained as a result is 1. This means that the t-score is located almost at the extreme right of the t- distribution.
4. Hence, for any alpha (α) value, our null hypothesis will be rejected. This is because of the very large t-score and p-values obtained.
5. Both versions of t-test are overwhelmingly predicting a 'statistically-significant' difference between the two algorithms.
6. But, the actual difference in mean intensity is very less.
7. This actually exposes a big problem with the statistical tests in that they are very much influenced by the sample size. For larger sample size, it is much more likely that the null hypothesis will be rejected.
8. Here, the sample size is huge as we have included every pixel of the images. Thus, the t-test value is large which leads to a very large p-value (1.000). This is an example of how tests can predict statistically significance when it is not present contextually (physically). We realized that this is similar to the example quoted in a previous lecture where aspirin was found to be "statistically significant" when in reality that was not the case.

3.Does your answer in previous part agree with visual inspection of the saliency maps?

No. it doesn't match with the visual inspection.

Note that we have ignored the spatial distribution of the intensities in calculation. But upon visual inspection of the two images, the average intensities visibly turn out to be more or less the same.

Notice that the two pictures have slightly different intensity distribution spatially, but the mean obtained experimentally is almost the same. In that sense, the answer in part 2 which rejects null hypothesis will contradict what we expect from visual inspection and calculation of Effect Size i.e. there is not much significance in the two algorithms.

4. Notwithstanding your answer in part 1, suppose you are asked to perform an unpaired t-test to compare the two given saliency maps. What possible information can this provide?

If we use equal sample sizes in the unpaired t-test formula, it reduces to a form similar to the Effect Size.

We obtain the following expression,

(Mean_x1 - mean_x2)sqrt(n)/sqrt(s1^2 + s2^2). Note the similarity to Effect Size formula.

Thus the information obtained will be similar to that of the Effect Size.

In paired t-test, we take the mean of the difference between the intensity data points associated with the same pixel. But, in the unpaired t-test, we obtain the following expression,

(Mean_x1 - mean_x2)sqrt(n)/sqrt(s1^2 + s2^2). Note the similarity to Effect Size formula.
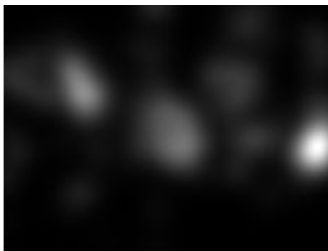
Thus, if the standard deviations of the two datasets is less, then we can say that the two means are more representative of the corresponding distribution. In that case, get an accurate estimate of how different the two algorithms are.

Q2.

Repeat all 4 parts for the following data.

Original image(dani_belgium)   Saliency_Map_dani_belgium_processing1   Saliency_Map_dani_belgium_processing2

In this case, we have a new image which was again processed by algorithm 1 and 2, and saliency maps were obtained. Load data_lab6.mat. You will see the data 'Saliency_Map_dani_belgium_processing1' and 'Saliency_Map_dani_belgium_processing2'.

1. Which t-test, paired or unpaired is suitable, in this case? Give reasons for your answer.

In the given experiment, we use two algorithms on the same underlying image frame. Here, we treat the saliency maps of the two algorithms as the samples taken from the given image.

The two samples are thus not independent. Hence, we will choose the paired t-test.

Note - Here we only consider the mean intensity of the two pictures not taking into account the spatial distribution. We are only looking at the mean intensity in the saliency maps. A greater intensity (in absolute terms) means that the map contains more white spots i.e. more points of focus and hence intensity pixels of higher value. Which means on an average the algorithm leads to humans paying more attention for that sample.

Note- Before, performing the t-test, we also calculate the Effect size as a sanity check.

2. Based on your answer, perform the t-test and analyze the differences between the two saliency maps.

Belgium images-

Hypothesis

Null hypothesis H_0: $\mu_1 - \mu_2 = 0$

Alternate hypothesis H_a: $\mu_1 - \mu_2 \neq 0$

Where μ_1: mean of the saliency map generated by algorithm 1

μ_2: mean of the saliency map generated by algorithm 2

Paired t-test results

Mean of difference = 0.0029

Standard Deviation of difference = 0.1351

Effect size = 0.0216 (low) < 0.2

t-score = 19.2325

P-value = 1.0000

Degree of Freedom = 788224

Unpaired t-test results

| algorithm 1: | Mean = 0.0914 | Standard Deviation(s1) = 0.1251 |
| algorithm 2: | Mean = 0.0943 | Standard Deviation(s2) = 0.1490 |

We assume that sampling distribution of the mean is normal due to Central Limit Theorem. Hence 1st assumption of t test is valid.

Here the standard deviation and consequently, the variance is of equal order. This satisfies the 2nd assumption for t test

Degree of Freedom = 1576448

Difference of mean = 0.0029   s = min(s1, s2) = 0.1251

Effect size = (Difference of mean) / s = 0.0233 (low) < 0.2

t-score = 13.3548

p-value = 1.0000

Observations:

1. The effect size which has been calculated prior to performing statistical tests is very low, indicating that there is not much difference between the samples which are generated by two different algorithms.
2. Notice that we have got a very high value of t-score.
3. The p-value obtained as a result is 1. This means that the t-score is located almost at the extreme right of the t- distribution.
4. Hence, for any alpha (α) value, our null hypothesis will be rejected. This is because of the very large t-score and p-values obtained.
5. Both versions of t-test are overwhelmingly predicting a 'statistically-significant' difference between the two algorithms.
6. But, the actual difference in mean intensity is very less.
7. This actually exposes a big problem with the statistical tests in that they are very much influenced by the sample size. For larger sample size, it is much more likely that the null hypothesis will be rejected.
8. Here, the sample size is huge as we have included every pixel of the images. Thus, the t-test value is large which leads to a very large p-value (1.000). This is an example of how tests can predict statistically significance when it is not present contextually (physically). We realized that this is similar to the example quoted in a previous lecture where aspirin was found to be "statistically significant" when in reality that was not the case.

3. Does your answer in previous part agree with visual inspection of the saliency maps?

T- test predicts there is significant difference between the saliency maps given by two algorithms.

This means that the t-test makes the prediction that the means of the intensities due to the two algorithms is different. It is important to bear in mind that it is only making a claim about the mean intensities. We have already acknowledged the fact that the above test simply ignores the spatial information in the image.

Unfortunately, when humans interpret the image, they interpret it "spatially". But here we need to check if the visual perception agrees with the claim made by the t-test which actually is not about the spatial distribution.

Spatially as visible to human eyes, the two intensity plots are very different. So we might carelessly say that it agrees with the t-test hypothesis. However, we must turn a blind eye to the spatial distribution as the t-test does not make any claim regarding that.

However, even visually we feel that there is more of "white" in the image of algorithm 2 than the image of algorithm 1. Thus we feel as if there might be a difference between the two algorithms

Thus, the answer in Q2 agrees with the visual intuition in case of data related to the image in this section.

4. Notwithstanding your answer in part 1, suppose you are asked to perform an unpaired t-test to compare the two given saliency maps. What possible information can this provide?

If we use equal sample sizes in the unpaired t-test formula, it reduces to a form similar to the Effect Size.

We obtain the following expression,

$(Mean\_x1 - mean\_x2)sqrt(n)/sqrt(s1\char`^2 + s2\char`^2)$. Note the similarity to Effect Size formula.

Thus the information obtained will be similar to that of the Effect Size.

In paired t-test, we take the mean of the difference between the intensity data points associated with the same pixel. But, in the unpaired t-test, we obtain the following expression,

$(Mean\_x1 - mean\_x2)sqrt(n)/sqrt(s1\char`^2 + s2\char`^2)$. Note the similarity to Effect Size formula.

Thus, if the standard deviations of the two datasets is less, then we can say that the two means are more representative of the corresponding distribution. In that case, get an accurate estimate of how different the two algorithms are.

Conclusion-

1. We felt that the t-test was inadequate to perform the analysis required in the question.
2. This is because the t-test only considers the mean statistic of the given data. It can only make claims regarding the similarity of the mean. It does not provide us any information about the distribution of the actual data.
3. However, we are dealing with image data where the spatial relation between pixels is very important. When we reduce it to a mean statistic for the above test we are losing out on a lot of information.
4. For instance, visually we can immediately say that image 1, the algorithms behave similar while they are different with respect to the second image.
5. Also, the data has a lot of samples. As a result, the t-test gives extreme results.

Alternative approaches

1. Row-wise t-test:

   Method - In this method, we have applied t test for each row and calculated the t-value for each row.

   1.1: Taking average of t-scores for all the rows

   In this case we use the t-scores directly.

   Tunnel:

   Degree of freedom = 1919

   Paired t-test:

   t-score = 4.6213

   p-value = 1.0000

   In this case, there is no significant difference with the answer obtained from the normal method. Hence no new information is got from this test.

   Unpaired t-test:

   t-score = 3.2203

p-value = 0.9993

Here we get a different answer. Now, if we take significance level $\alpha = 0.0005$, then we can say that the samples follow the null hypothesis. i.e the samples are similar. This is the result that we get from the visual inspection as well. Although, this result may not be entirely perfect because we are not considering the spatial information of the image.

Belgium:
Degree of freedom = 1024
Paired t-test:
      t-score = 0.9738
      p-value = 0.8348

Unpaired t-test:
      t-score = 1.3326
      p-value = 0.9085

If we take $\alpha = 0.05$, then both the above tests show that the samples follow the null hypothesis. This is contradictory to visual analysis. This may be because we don't consider the spatial information in this test. It may also be because using t values as it is, the positive and negative t scores cancel each other which may lead to wrong answer.

1.2: Taking average of absolute values of t-scores for all the rows
     In this case we use take the absolute value of t-score. This is similar to taking single tailed t distribution.

Tunnel:
Paired t-test:
      t-score = 11.3542
      p-value = 1.0000

Unpaired t-test:
      t-score = 8.3945
      p-value = 1.0000

These answers match with the normal test and hence there is no significant use of this approach.

Belgium:
Paired t-test:

    t-score = 10.6514
    p-value = 1.0000

Unpaired t-test:

    t-score = 9.3983
    p-value = 1.0000

These answers match with the normal test and hence there is no significant use of this approach.

2. Taking random samples from the intensity maps to apply t-test

We saw that applying the t-test to the whole dataset is a bad idea. It gives us extreme t-score values.

Instead, we take random samples from the datasets. This brings in the possibility of the t-scores depending on the current sample. Thus, we run the test for a number of iterations to get a mean t- score value.

Tunnel dataset

For sample size= 100, iterations=1000, alpha= 0.05

We get

T-score = 0.7281

P-value = 0.23

Put full code of random sample
Result: The result is *not* significant at p < .05.

Also, the variance was not too much. Further, running the code a number of times gives the same mean t-score. This seems to indicate that the random sampling method is effective in capturing the typical t-value of a dataset.

The same thing repeated for a larger sample size: 1000000

Mean t-score : 7.2306

The P-Value is < .00001.

The result is significant at p < .05.

For a larger sample size, the t-score is large and hence the test results are extreme. But again, repeating the experiment for a large number of times gives the same t-value more or less.

Thus, it seems that the random sampling method does give us a average idea about the performance of the sampling on the dataset.
Also, the t-test is highly dependent on what sampling sizes we choose. This is a huge drawback of the test.

This solves the problem of large sample sizes and also considers the possibility that the sample might be obtained by chance by running the test for a large number of iterations.

3. Block wise t-test

Here , both the saliency image matrices have been split into 10x10 pixel blocks. The t-test is performed on each such block and then finally the mean of all t-scores is taken to find p value.

Here the t-score for the Tunnel image comes out to be 1.6042e+14. This is because at certain places, the effect size blows up due to very small value of standard deviation. This is impractical and hence this approach should stop here.

Codes:

Normal approach

Tunnel:

```
clear;
close all;

load('data_lab6');

% converting matrix to array
tunnel1 = saliency_map_Tunnel_processing1(:);
tunnel2 = saliency_map_Tunnel_processing2(:);

%%%%%%%%%%%%%%%%%%%%%%%% paired t test %%%%%%%%%%%%%%%%%%%%%%%%%%%%

% take difference of sample values
diff_tunnel = tunnel2 - tunnel1;

% find mean of difference of sample values
mean_tunnel = mean(diff_tunnel);

% find standard deviation of difference of sample values
sd_tunnel = std(diff_tunnel);

% calculate degree of freedom for the paired test sample which is (n-1)
deg_of_freedom = length(diff_tunnel) - 1;

% calculate effect size
effect_size = mean_tunnel / sd_tunnel;

% find t value for paired t test
paired_t = effect_size * sqrt(deg_of_freedom + 1);


%%%%%%%%%%%%%%%%%%%% unpaired t test %%%%%%%%%%%%%%%%%%%%%%%%

tunnel1_mean = mean(tunnel1);
tunnel1_sd = std(tunnel1);
n1 = length(tunnel1);
```

```
tunnel2_mean = mean(tunnel2);
tunnel2_sd = std(tunnel2);
n2 = length(tunnel2);


s = min(tunnel1_sd, tunnel2_sd);

effect_size_unpaired = (tunnel2_mean - tunnel1_mean) / s

[unpaired_t, df_unpaired] = calculate_t(tunnel1_mean, tunnel2_mean, tunnel1_sd,
tunnel2_sd, n1, n2)
```

Belgium:

```
clear;
close all;

load('data_lab6');

% converting matrix to array
belgium1 = Saliency_Map_dani_belgium_processing1(:);
belgium2 = Saliency_Map_dani_belgium_processing2(:);

%%%%%%%%%%%%%%%%%%%%%% paired t test %%%%%%%%%%%%%%%%%%%%%%%%%%%%

% take difference of sample values
diff_belgium = belgium2 - belgium1;

% find mean of difference of sample values
mean_belgium = mean(diff_belgium);

% find standard deviation of difference of sample values
sd_belgium = std(diff_belgium);

% calculate degree of freedom for the paired test sample which is (n-1)
deg_of_freedom = length(diff_belgium) - 1;

% calculate effect size
effect_size = mean_belgium / sd_belgium

% find t value for paired t test
paired_t = effect_size * sqrt(deg_of_freedom + 1)


%%%%%%%%%%%%%%%%%%%%% unpaired t test %%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```matlab
% sample 1
belgium1_mean = mean(belgium1);
belgium1_sd = std(belgium1);
n1 = length(belgium1);

% sample 2
belgium2_mean = mean(belgium2);
belgium2_sd = std(belgium2);
n2 = length(belgium2);

% to find effect size
s = min(belgium1_sd, belgium2_sd);

effect_size_unpaired = (belgium2_mean - belgium1_mean) / s

% calculate unpaired t value
[unpaired_t, df_unpaired] = calculate_t(belgium1_mean, belgium2_mean,
belgium1_sd, belgium2_sd, n1, n2)
```

Function used to calculate unpaired t test t score

```matlab
% function to calculate degree of freedom and t value for given set of data

function [t, df] = calculate_t(x1, x2, s1, s2, n1, n2)

% parameters
    % x1 : mean of first distribution
    % x2 : mean of second distribution
    % s1 : standard deviation of first distribution
    % s2 : standard deviation of second distribution
    % n1 : number of sample points of first distribution
    % n2 : number of sample points of second distribution

% return values
    % df : degrees of freedom
    % t : t value of the data


    % calculating degrees of freedom
    df = n1 + n2 - 2;

    % calculating t value
    t = (x1 - x2) ./ ( sqrt( ((s1.*s1.*(n1-1) + s2.*s2.*(n2-1))./df) .* (1/n1
+ 1/n2) ) );
end
```

Row-wise t-test:

```matlab
clear;
close all;

load('data_lab6');

% converting matrix to array
tunnel1 = saliency_map_Tunnel_processing1';
tunnel2 = saliency_map_Tunnel_processing2';

%%%%%%%%%%%%%%%%%%%%%% paired t test %%%%%%%%%%%%%%%%%%%%%%%%%%%

% take difference of sample values
diff_tunnel = tunnel2 - tunnel1;

% find mean of difference of sample values
mean_tunnel = mean(diff_tunnel);

% find standard deviation of difference of sample values
sd_tunnel = std(diff_tunnel);

% calculate degree of freedom for the paired test sample which is (n-1)
deg_of_freedom = length(diff_tunnel) - 1;

% calculate effect size
effect_size = mean_tunnel ./ sd_tunnel;

% find t value for paired t test
paired_t = effect_size .* sqrt(deg_of_freedom + 1);


t_pair = mean(paired_t(:), 'omitnan')

%%%%%%%%%%%%%%%%%%%% unpaired t test %%%%%%%%%%%%%%%%%%%%%%

tunnel1_mean = mean(tunnel1);
tunnel1_sd = std(tunnel1);
n1 = length(tunnel1);

tunnel2_mean = mean(tunnel2);
```

```
tunnel2_sd = std(tunnel2);
n2 = length(tunnel2);


s = min(tunnel1_sd, tunnel2_sd);

effect_size_unpaired = (tunnel2_mean - tunnel1_mean) ./ s;

[unpaired_t, df_unpaired] = calculate_t(tunnel1_mean, tunnel2_mean, tunnel1_sd,
tunnel2_sd, n1, n2);

t_unpair = mean(unpaired_t, 'omitnan')
```

Random sample t-test

```
clear
close all;
load('data_lab6.mat')

tunnel1 = saliency_map_Tunnel_processing1(:); % unroll
tunnel2= saliency_map_Tunnel_processing2(:);

ite=1000;

t_arr= zeros(ite,1);

sample_size=1000;

for i=1:ite


% take random samples from the two intensity datasets
diff= randsample(tunnel2,sample_size) - randsample(tunnel1,sample_size);
d= mean(diff);
s= std(diff);
sample_size= length(diff);


t_arr(i)= d*sqrt(sample_size)/s; % t- score for the given samples
```

```
end

mean(t_arr) % mean value of t-score
```

Block wise t-test

```matlab
clear;
close all;

load('data_lab6');

s1 = saliency_map_Tunnel_processing1;
s2 = saliency_map_Tunnel_processing2;

% converting matrix to array
tunnel1 = s1;
tunnel2 = s2;

%%%%%%%%%%%%%%%%%%%%%%%% paired t test %%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% take difference of sample values
diff_tunnel = tunnel2 - tunnel1;

chunk = 10;
[rows, cols] = size(diff_tunnel);

xiter = rows/chunk;
yiter = cols/chunk;

paired_t_arr = zeros(xiter,yiter);

for j = 1 : yiter
    for i = 1 : xiter
    sample_diff = diff_tunnel(10*(i-1)+1:10*(i-1)+10 ,
10*(j-1)+1:10*(j-1)+10);

    % find mean of difference of sample values
    mean_tunnel = mean(sample_diff(:));

    % find standard deviation of difference of sample values
    sd_tunnel = std(sample_diff(:));

    % calculate degree of freedom for the paired test sample which is (n-1)
    deg_of_freedom = chunk * chunk - 1;
```

```matlab
        % calculate effect size
        effect_size = mean_tunnel / sd_tunnel;

        % find t value for paired t test
        if ~isnan(effect_size)
            paired_t_arr(i, j) = effect_size * sqrt(deg_of_freedom + 1);
        end

        sum(paired_t_arr(:))
        end
end


mean(paired_t_arr(:))
```