# Paired Student t-test

Mikhail M.Meskhi

March 21, 2018

**Abstract**

The following is a experimental setup of Welch's Student Paired t-test for unequal variance samples. $H_0$ = same means, $H_a \neq$ same means. Below are the tables with corresponding $p$-values and results.

## Experimental Setup

**Mean Data Accuracy and Variance**

| Domain Adaptation Algorithm | Datasets | |
|---|---|---|
| | Supernova Ia | Mars Landforms |
| Source Model | 69.13 (0.00) | 74.36 (9.40) |
| Subspace Alignment | 62.56 (7.98) | 85.16 (2.65) |
| JDOT SVM | 77.57 (0.13) | 85.2 (0.59) |
| JDOT NN | 69.05 (0.08) | 80.96 (0.06) |
| DANN | 80.4 (0.3) | 88.61 (0.22) |
| TJM | 65.56 (0.01) | 82.28 (0.03) |
| JDA | 70.64 (0.03) | 80.40 (0.03) |
| ARTL | 66.21 (0.01) | 88.12 (0.02) |
| GFK | 63.98 (0.02) | 83.56 (0.04) |
| Bayesian DA | **86.17** (**0.35**) | **90.81** (**1.49**) |

The above table will be used to conduct a student paired t-test with the following pair form: **{Bayesian Domain Adaptation, ClsX}**.

### Student paired t-test:

The following t-test is used when two samples are independent from each other. Using the following formula we calculate the $t$-value and then find the corresponding $p$-value to determine statistical significance for rejecting the null hypothesis.

$$t' = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{S_1^2}{N_1} + \frac{S_2^2}{N_2}}}$$

To calculate degrees of freedom $df$ we use $\Delta$:

$$\Delta = \frac{(\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2})}{\frac{s_1^4}{N_1^2 v_1} + \frac{s_2^4}{N_2^2 v_2}}$$

Testing whether or not we get a significant $p$-value to reject the null-hypothesis for [BDA, GFK] on the Supernova dataset is as follows:

### Formulation:

1. $N = 10$
2. $\bar{x} = 86.17$
3. $\bar{y} = 63.98$
4. $S_1^2 = 0.1225$
5. $S_2^2 = 0.0004$
6. $\Delta \approx 9$
7. $t' \approx 200$

### Interpretation:

With $t' \approx 200$ and $\Delta \approx 9$ we find our $p$-value to be $< 2.2e - 16$. Thus, concluding that we reject the null-hypothesis due to extreme statistical significance against it.

To avoid redundancy, the tables below have the $p$-values for all the following paired tests. And in order to avoid inflated Type I error, I apply the Bonferroni Adjustment to fix the FWER (family wise error rate). The tables below display the adjusted $p$-values and the their significance on $H_0$.

### Conclusion

It is evident from the tables above that the pairwise t-test has concluded extreme statistical significance between our proposed Bayesian Domain Adaptation classifier and other state of the art classifiers on these two datasets.

| Classifier | $t$-value | *d.f.* | Adjusted $p$-value | Result |
|---|---|---|---|---|
| BDA vs. GFK | 200 | 9 | $9.047785e-18$ | $H_0$ Rejected |
| BDA vs. ARTL | 180 | 9 | $2.325296e-17$ | $H_0$ Rejected |
| BDA vs. JDA | 140 | 9 | $2.321808e-16$ | $H_0$ Rejected |
| BDA vs. TJM | 186 | 9 | $1.938281e-17$ | $H_0$ Rejected |
| BDA vs. DANN | 40 | 18 | $7.282950e-11$ | $H_0$ Rejected |
| BDA vs. JDOT NN | 151 | 10 | $9.608068e-17$ | $H_0$ Rejected |
| BDA vs. JDOT SVM | 73 | 11 | $3.742138e-14$ | $H_0$ Rejected |
| BDA vs. SA | 9 | 9 | $7.142476e-06$ | $H_0$ Rejected |

Table 1: Statistical Significance on Supernova

| Classifier | $t$-value | *d.f.* | Adjusted $p$-value | Result |
|---|---|---|---|---|
| BDA vs. GFK | 15 | 9 | $8.754559e-08$ | $H_0$ Rejected |
| BDA vs. ARTL | 6 | 9 | $3.002099e-04$ | $H_0$ Rejected |
| BDA vs. JDA | 22 | 9 | $3.905075e-09$ | $H_0$ Rejected |
| BDA vs. TJM | 18 | 9 | $1.943669e-08$ | $H_0$ Rejected |
| BDA vs. DANN | 5 | 9 | $6.835074e-04$ | $H_0$ Rejected |
| BDA vs. JDOT NN | 21 | 9 | $5.783850e-09$ | $H_0$ Rejected |
| BDA vs. JDOT SVM | 11 | 12 | $2.251948e-06$ | $H_0$ Rejected |
| BDA vs. SA | 17 | 14 | $1.756968e-04$ | $H_0$ Rejected |

Table 2: Statistical Significance on Mars