Robustness of Type I Error Rate of Hotelling's T^2 and Alternative Tests to Outliers

Serge G. Demers[1,2] , Richard Wolfe[1] & Bruno D. Zumbo[3]
OISE/University of Toronto[1], Laurentian University[2], The University of British Columbia[3]
for contact information : sdemers@laurentian.ca or bruno.zumbo@ubc.ca

RESEARCH QUESTION

- To investigate the robustness of the Hotelling's T^2 test, a robust version of the test, and an outlier detection and removal strategy in the presence of various types and quantities of outliers.

HOTELLING'S T^2 TEST

- Hotelling's T^2 test is the multivariate equivalent of the t^2 test.
- It is an inferential test used to compare a multivariate data set to either another data set or to a fixed value.
- The calculation of Hotelling's T^2 test uses the sample mean and covariance structure.
- The general problem is that both the sample mean and the standard deviation are inflated by the outliers.
- This inflation of the values masks the presence of the outliers, rendering their identification impossible without the use of robust statistics.
- As such, Hotelling's T^2 test is susceptible to biased results when outliers are present in the data set.

WILLEMS et al.'s ROBUST TEST (2002)

- Willems et al. (2002) contend that the traditional Hotelling's T^2 test suffers badly from the effect of outliers.
- They developed a robust version of the test using a highly robust reweighted MCD estimator for both the mean and the covariance matrix.
- This causes the robust test to have a different distribution from that of the classical test.
- They developed an approximate distribution for this robust statistic using a Monte Carlo study.
- Willems et al.'s test makes use of all the available data, giving less importance to the more outlying points.
- Their Monte Carlo study showed that, for a one-sample test, the loss of power for the robust test was acceptable, even for small sample sizes such as 10 or 20 observations.
- Their study also showed that the hypotheses based on their robust T^2 test should yield good Type I error probability.

KOSINSKI'S OUTLIER DETECTION AND REMOVAL METHOD

- His approach to outlier identification is to partition the data into 'good' and 'bad' groups.
- As a starting point to his algorithm, Kosinski uses elemental partitions of a *p*-variate data set.
- Each elemental partition contains only p+1 observations at first. The number of such elemental partitions is a function of the number of observations in the data set, the number of variates, as well as the desired alpha-level.
- Each of the elemental partitions then has a forward search conducted on it in order to continuously add observations, thereby increasing the size of the 'good' partition.
- The forward search is conducted based on an ordering of the Mahalanobis distances of each observation from the center of the partition under study.

SIMULATION STUDY

METHODOLOGY

- Simulated data are based on the Ontario SIMS (1980-82).
- Outliers are generated using a mixed normal contamination model with a variance inflation approach.
- Outliers are assigned in one of three ways : 1) an entire row of data, (2) a single cell, (3) each cell for an individual has a random chance of containing an outlier.
- 1000 replications for each combination of conditions
- The investigation is focused on the effect of small sample sizes and various outlier patterns on the three methods.
- Robustness is based on Serlin's (2000) criteria.

Variables in this simulation study:

- *5 sample sizes* : 10, 20, 30, 50 and 100.
- *4 contamination rates* : 0%, 5%, 10% and 15%.
- 5 *levels of number of variates* : from 2 to 6 variates in the data set.
- *3 alpha levels* : .1, .05 and .01.
- *3 outlier scopes* : single cell as an outlier, entire row as an outlier and outliers randomly assigned within an individual.
- *3 statistical tests* : Hotelling's T^2 test, Willem's et al.'s robust T^2 test, and Kosinski's outlier detection and remo-val method followed by a traditional Hotelling's T^2 test.

RESULTS

Table 1
Number of Non-Robust Results by Contamination Rate and Outlier Scope

Outlier Scope	Contamination Rate	Traditional T^2 Test	Robust T^2 Test	Outlier Removal Method
No outliers	0%	2	10	27
Individual as Outlier	5%	0	7	2
	10%	0	5	0
	15%	0	4	0
Single Variate as Outlier	5%	5	9	15
	10%	3	10	8
	15%	0	8	3
Outliers Randomly Assigned	5%	3	12	10
	10%	0	6	4
	15%	0	4	0

SUMMARY

- The traditional T^2 test consistently outperformed the other two under all of the examined conditions.
- The situation where a single variate contains an outlier was the most problematic for all three examined methods.
- Randomly assigned outliers also produced generally less robust results than the traditional view of outliers where an entire row is contaminated.
- The outlier removal method performed most poorly for the situation with no contamination.

DISCUSSION

- The robust T^2 test produced invalid results in situations where the number of variates was significant with respect to the sample size, although the program did not indicate a problem.
- Given my findings, what method should researchers use?
  - The traditional T^2 test is the best candidate.
  - The robust T^2 test performs well once the sample size gets large (between 50 and 100 for most conditions).
  - The outlier detection and removal method is extremely computer-intensive, and does not yield robust results; as such, this method should be avoided.