Proceedings of the International Workshop on the Scientific approach to the Acheiropoietos Images, ENEA Frascati, Italy, 4-6 May 2010

A robust statistical analysis of the 1988 Turin Shroud radiocarbon dating results

G. Fanti[1], F. Crosilla[2], M. Riani[3], A.C. Atkinson[4]

[1]Department of Mechanical Engineering University of Padua, Italy, giulio.fanti@unipd.it.
[2]Department of Geo-Resources and Territory, University of Udine, Italy fabio.crosilla@uniud.it.
[3]Department of Economics, University of Parma, Italy, mriani@unipr.it
[4]Department of Statistics, London School of Economics, London WC2A 2AE, UK, a.c.atkinson@lse.ac.uk.

Abstract
Using the 12 published results from the 1988 radiocarbon dating of the TS (Turin Shroud), a robust statistical analysis has been performed in order to test the conclusion by Damon et al. (*1998*) *that the TS is mediaeval. The 12 datings, furnished by the three laboratories, show a lack of homogeneity. We used the partial information about the location of the single measurements to check whether they contain a systematic spatial effect. This paper summarizes the results obtained by Riani et al. (2010), showing that robust methods of statistical analysis can throw new light on the dating of the TS.*

Keyword: ANOVA, Forward Search, Robust methods, t-statistics, Turin Shroud.

1. INTRODUCTION

The results of the 1988 radiocarbon dating [1] of the TS were published as providing conclusive evidence that the linen fabric dates from between 1262 and 1384 AD, with a confidence level of 95%.

However, after publication of the result, many speculated that the sample had been contaminated due to the fire of 1532 which seriously damaged the TS, or to the sweat of hands impregnating the linen during exhibitions, others that the date was not correct due to the presence of medieval mending and so on. We give references to some of these concerns in Section 7.

The purpose of this paper is to summarize the results obtained in Ref. 2 which show how robust methods of statistical analysis, in particular the combination of regression analysis and the forward search [3] combined with computer power and a liberal use of graphics, can help to shed new light on results that are a source of scientific controversy. Throughout we analyse only numbers from the data given in Ref. 1.

2. DESCRIPTION OF THE DATA

The samples for radio carbon dating were taken from a strip of material cut from one corner of the TS. The strip was divided into five parts; the three parts on the right of Figure 1 were sent to laboratories in Arizona, Oxford and Zurich. Arizona also received the fourth, smaller, part on the left. A larger part on the left of Figure 1 was taken by the Arcidiocesi of Turin as a "Riserva".

Figure 2 indicates the cutting of the strip in question.

These samples were divided into a total of 12 sub-samples for which datings were made. The resulting dates ranged from 591 BP for a reading from Arizona, to 795 BP from Oxford.

3. HETEROGENEITY ANALYSIS

Damon et al. [1] noticed that the data show some heterogeneity, which they assessed using a chi-squared test. In this section we instead use the analysis of variance to test whether these 12 observations can be considered as homogeneous, i.e. as 12 repeated measurements coming from a single unknown quantity.

More formally, a general model for observation j at site i is

$$y_{ij} = \mu_i + \sigma v_{ij} \varepsilon_{ij} \quad (i = 1, 2, 3; j = 1, ..., n_i), \quad (1)$$

where the errors ε_{ij} have a standard normal distribution.

Our central concern is the structure of the μ_i; at this point whether they are all equal. However, before proceeding to the test this hypothesis we need to establish the error structure. Riani et al. [2] suggest the three following possibilities

1. **Unweighted Analysis**. Standard analysis of variance: all $v_{ij} = 1$

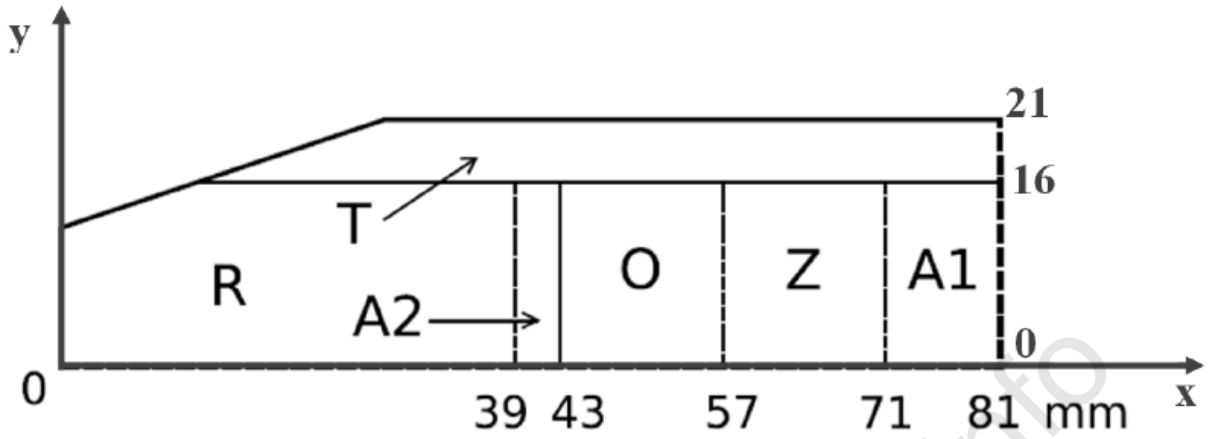

Figure 1. Diagram showing the piece removed from the TS and how it was partitioned. T: trimmed strip. R: retained part called "Riserva". O, Z, A1, A2: subsamples given to Oxford, Zurich, and Arizona (two parts) respectively.

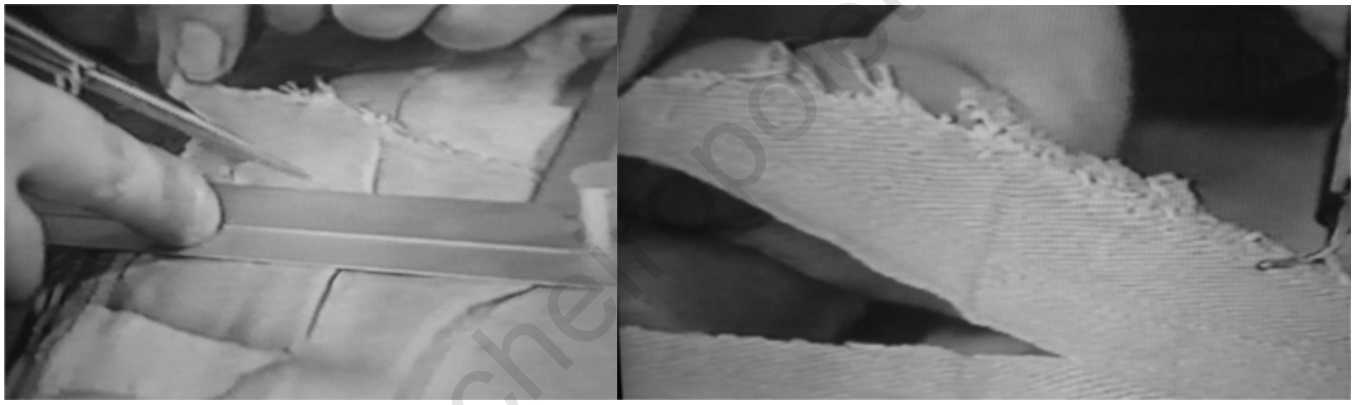

Figure 2. Cutting of the linen strip from the TS for the 1988 radiocarbon dating. (G. Riggi di Numana, Fototeca 3M).

2. **Original weights**. We weight all observations by $1/v_{ij}$, where the v_{ij} are the standard errors published by Damon et al. [1], that is, we perform an analysis of variance using responses:

$$z_{ij} = y_{ij}/v_{ij} \,. \qquad (2)$$

3. **Modified weights for Arizona**. This last formulation takes into account the fact that according to Damon et al. the standard errors for Arizona, unlike the two other laboratories, include only two of the three sources of error.

Reference 2 shows that irrespective of the kind of ANOVA which is used, while the test for homogeneity of the variances among the 3 laboratories never turns out to be significant (the minimum p-value is greater than 0.3), the test for homogeneity of means is always significant at the 5% level.

Christen [4] used these data as an example of Bayesian outlier detection with a mean shift outlier model (Abraham and Box [5]) in which the null model was that the data were a homogeneous sample from a single normal population. He found that the two extreme observations, 591 and 795 were indicated as outlying. When these two observations were removed, the data appeared homogeneous, with a posterior distribution of age that agreed with the conclusion of Damon et al. [1].

4. SPATIAL HETEROGENEITY

We have appreciable, but only partial, knowledge of the spatial layout of the samples from Damon et al. [1]. Three pieces were dated by Oxford, four by Arizona and five by Zurich. However it is not known how the samples in Figure 1 were divided within the laboratories, nor is it known whether the four readings from Arizona came only from A1 or from A1 and A2.

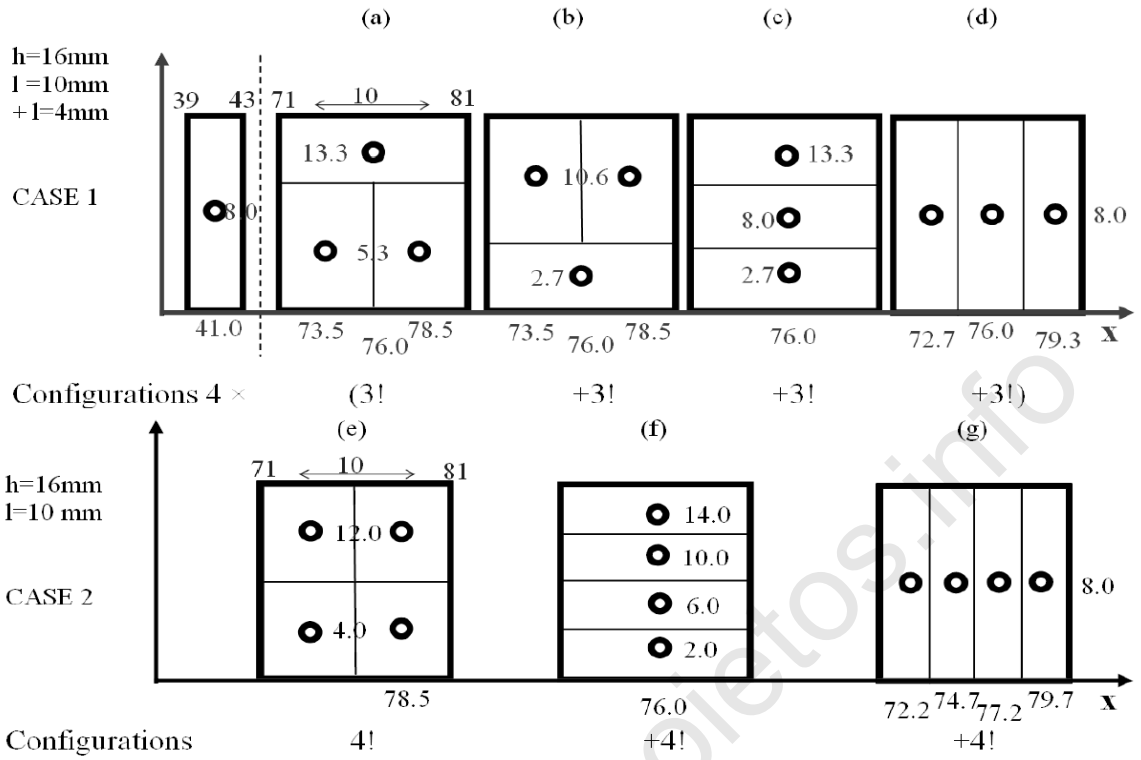

Figure 3. Arrangements investigated for the Arizona sample. The image on top assumes that Arizona dated both pieces (A1 and A2). The image at the bottom assumes that Arizona only dated piece A1. Total number of cases considered is 168 = 96+72.

On the assumption that the four readings from Arizona all came from A1, Walsh [6] showed evidence for the regression of age on the known centre points of the pieces of fabric. Ballabio [7], as well reviewing earlier work, introduced a second spatial variable into the analysis, the values of both variables depending on how the division into subsamples was assumed to have been made. He was defeated by the number of possibilities.

The possible configurations for the subsamples from Arizona are shown in Figure 3. If we also consider all possible plausible ways in which cuts could have been made by the laboratories of Oxford and Zurich, we end up with 96 and 23 configurations. In summary there are 387,072 possible cases to analyse.

5. MULTIPLE REGRESSION

To try to detect any trend in the age of the material we fit a linear regression model in x_1 (longitudinal) and x_2 (transverse) distances. The analysis is not standard. Riani et al. [2] permute the values of x_1 and x_2 and perform all 387,072 analyses.

The question is how to interpret this quantity of numbers. Without any trend in the longitudinal and transverse directions we expect to obtain a distribution of *t*-statistics for the regression coefficients which is centred around zero and we approximately expect to obtain half of the 387072 configurations with a positive value of the *t*-stat and the other half with negative values. The top panel of Figure 4 (taken from Ref. 2) shows the distribution of the *t*-statistic for x_2. This has a *t* like shape centred around 0.5. The bottom panel of Figure 4, the *t*-statistic for x_1, is however quite different, showing two peaks. The larger peak is centred around −2.9 whereas the thinner peak is centred around −1. It is also interesting to notice that for each of the 387,072 configurations we obtain a negative value of the *t*-statistic for the longitudinal coordinate.

As we have shown that x_2 is not significant (even if it is surprisingly not centered around 0), we continue our analysis with a focus on x_1. In particular, we want to discover what feature of the data leads to the bimodal distribution in Figure 4. If we consider the longitudinal projections of the 387,072 configurations we obtain 42,081 possibilities.

Summarising the results in Ref 2 which performs a detailed analysis of all these longitudinal configurations, it comes out that inference about the slope of the relationship depends critically on whether configuration A2 (see Figure 1) was analysed. More precisely, the only configurations which give rise to non-significant values of the t-statistic are those associated with:

1) configuration A2 (that are based on the assumption that Arizona dated both A1 and A2), see Figure 1.
2) the response at the longitudinal coordinate $x_1 = 41$ is y=591 or y=690.

We now analyse the data structure, taking typical members inside the configurations 41-591 and of 41-690 and look at some simple diagnostic plots.

To determine whether the proposed data configuration 41-591 is plausible we look at residuals from the fitted regression model. In order to overcome the potential problem of masking (when one outlier can cause another to be hidden) we use a forward search [3] in which subsets of m carefully chosen observations are used to fit the regression model and see what happens as m increases from 2 to 12. Figure 6 shows a forward plot of the residuals of all observations, scaled by the estimate of sigma at the end of the search, that is when all 12 observations are used in fitting. The plot shows the pattern typical of a single outlier, here 41-591 which is distant from all the other observations until $m = n$, when it affects the fitted model.

The conclusion from this analysis is that whether one of the lower y values, 591 or 606, or one of the higher y values, 690 or 701, from Arizona is assigned to $x_1 = 41$, an outlier is generated, indicating an implausible data set. The comparable plots when it is assumed that Arizona only analysed A1 are quite different in structure. There is a stable scatter of residuals in the left-hand panel as the forward search progresses, with no especially remote observation. We conclude, that there is statistical evidence that Arizona only analysed A1 and that there is a significant trend in the longitudinal coordinates.

6. CONCLUSIONS

The Shroud data relative to the 1988 radiocarbon dating show surprising heterogeneity. This leads us to conclude that the twelve measurements of the age of the TS cannot be considered as repeated measurements of a single unknown quantity.

The presence of a linear trend explains the difference in means that was found using the ANOVA test.

The evidence of the heterogeneity together with the evidence of a strong linear trend lead us to conclude that the statement of Damon et al.: "*The results provide conclusive evidence that the linen of the Shroud of Turin is mediaeval*" [1] needs to be reconsidered in the light of the evidence produced by our use of robust statistical techniques.

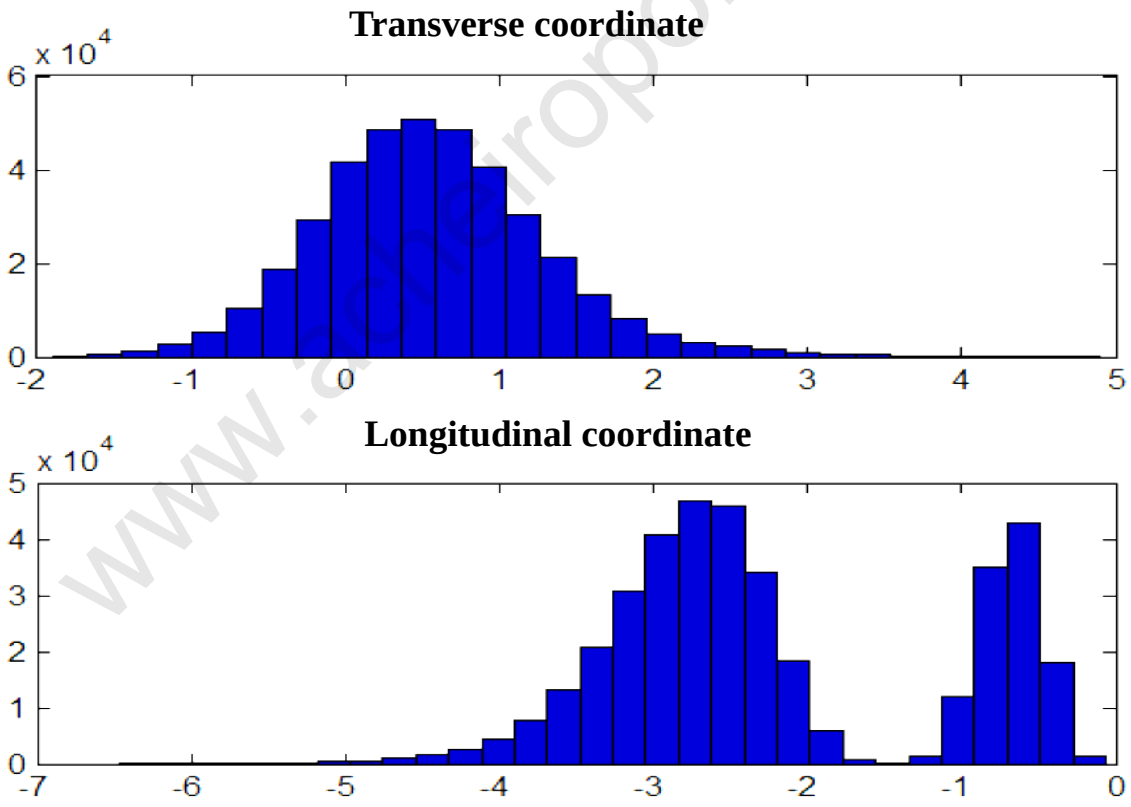

Figure 4. Two variable regression. Histograms of values of t-statistics from 387,072 possible configurations. Upper panel x_2 (transverse coordinate), lower panel x_1 (longitudinal coordinate).

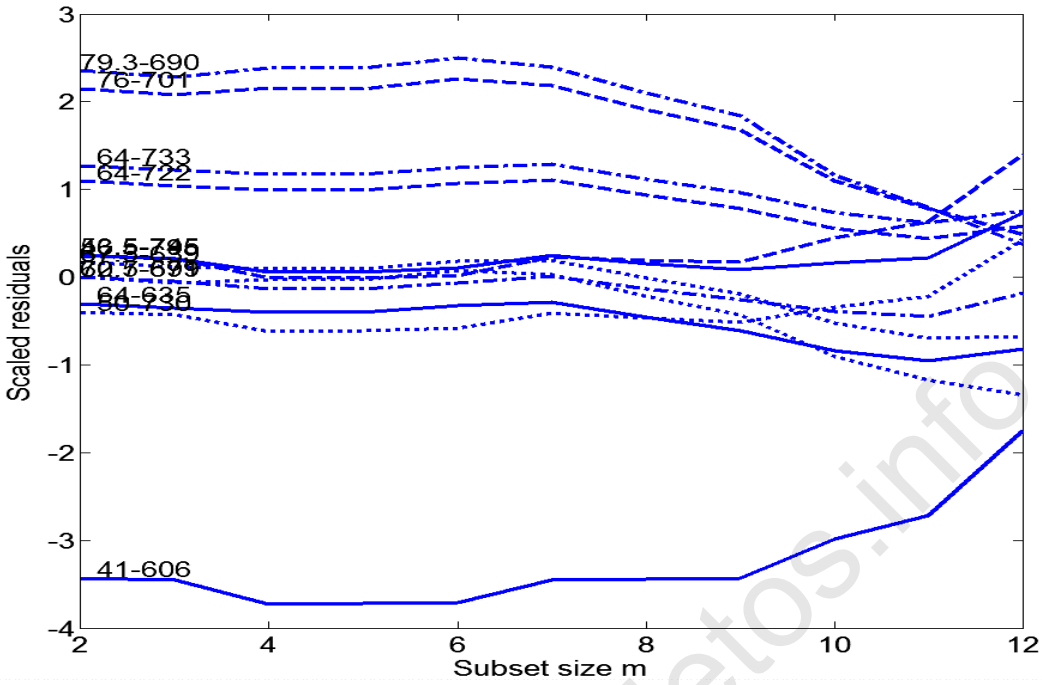

Figure 5. Analysis of residuals for one typical configuration when x1=41, y1=591. Forward plot of scaled residuals showing that this assignment produces an outlier.

7. DISCUSSION

The arguments in favour of the authenticity of the TS are rehearsed in other papers in this volume. For example, the formation mechanism of the body images has not yet been scientifically explained. One so far unexplained feature is that the body image is extremely superficial in the sense that only the external layer of the topmost linen fibre is coloured [8]. See also [9] and [10].

At a more mundane level, we note that the weights used in Section 3, taken from Ref. 1, were obtained from up to 8 repeat determinations. Burr et al. [11] describe the process of analysis used at Arizona. As always, in any data analysis, it is a help in understanding and modeling the truth of a situation to work with the original data, rather than data which have already been summarized, even if only lightly.

REFERENCES

1. Damon P.E., Donahue D.J., Gore B.H., Hatheway A.L., Jull A.J.T., Linick T.W., Sercel P.J., Toolin L.J., Bronk C.R., Hall E.T., Hedges R.E.M., Housley R., Law I.A., Perry C., Bonani G., Trumbore S., Wölfli W., Ambers J.C., Bowman S.G.E., Leese M.N., Tite M.S.: Nature, **337**, 611-615 (1989).

2. Riani M., Atkinson A.C., Fanti G., Crosilla F.: "*Carbon Dating of the Shroud of Turin: Partially Labelled Regressors and the Design of Experiments*" see: www.lse.ac.uk/collections/statistics/research/RAFC04May2010.pdf

3. Atkinson, A.C. and M. Riani: *Robust Diagnostic Regression Analysis*, (New York: Springer–Verlag) 2000.

4. Christen, A.: Applied Statistics, **43**, 489-503 (1994).

5. Abraham, B. and Box, G.E.P. Applied Statistics, **27**, 131-138 (1978).

6. Walsh B. *The 1988 Shroud of Turin radiocarbon tests reconsidered.* Proceedings of the 1999 Shroud of Turin International Research Conference Richmond, Virginia USA, pp. 326–342. B. Walsh Ed., Glen Allen VA: Magisterium Press (1999).

7. Ballabio G. *New statistical analysis of the radiocarbon dating of the Shroud of Turin* (unpublished manuscript). See http://www.shroud.com/pdfs/doclist.pdf.

8. Fanti G., Botella J.A. Di Lazzaro P., Heimburger T., Schneider R., Svensson N. "Journal of Imaging Science and Technology **54**, 040201-(8) (2010).

9. Fanti, G. Basso R., "The Turin Shroud, Optical Research in the Past Present and Future", Publisher Nova Science Pub Inc., 2008.

10. Fanti G.: "La Sindone, una sfida alla scienza moderna", Aracne ed., Roma, Italy, 2008.

11. Burr G.S., Donahue D.J., Tang Y., Beck W.J., McHargue L., Biddulph D., Cruz R. and Jull A.J.T.: Nucl. Instr. and Methods in Physics B **259**, 149-153 (2007).