Analyzing Radiocarbon data using Burr statistics
©2011 Remi Van Haelst
====================================
Today, Arizona uses statistics, developed by Bevington et al, and published in 1962 & 1982. "Data reduction and error analysis" Edition Mc. Graw-Hill)
Marion Scott (University of Glasgow) (1), analyzing the 1990 "International Collaborative Program" concluded :"It seems reasonable to consider that a laboratory performs adequately if it has no systematic bias and assesses its Internal and External variability adequately. IEM & EEM should not significantly different from 1."
Well, the Burr, *et al*. paper (2) does precisely this – it combines the Internal and External measured variability into one variability statistic and in this way produces an independent statistical measure of the overall variability of one laboratory's experimental results.

Important.
In reality Arizona, dated the Shroud in four sessions, during which were made two independent measurements. Only recently, Prof. Jull (3) recognized that indeed Arizona made eight measurements.
Also, the Nature paper was authored by Damon (Arizona) this combination made at the request of the British Museum was not noted.
The same may said about the strange differences in measurements given in Nature and the date given by the Italian experts Riggi and Testore.

Session			Nature Table 1
A	606-+41	574-+45	591-+30
B	753-+51	632-+49	606-+41
C	676-+59	540-+57	690-+35
D	701-+47	401-+47	701-+33

Here's the application of Eq (3) in the Burr, *et al*. article is used to evaluate to the ORIGINAL EIGHT Arizona radiocarbon data:

Radiocarbon measurements with 8 AZ observations

Arizona

	mean	Se		Wi		Σ Wi * mean		Chi sq
	606	51		0.000384		0.232987		0.82231
	574	52		0.000370		0.212278		2.26430
	753	51		0.000384		0.289504		3.90276
	632	49		0.000416		0.263224		0.17074
	676	59		0.000287		0.194197		0.16208
	540	57		0.000308		0.166205		3.87796
	701	47		0.000453		0.317338		1.07597
	701	47		0.000453		0.317338		1.07597
							χ^2 =	***13.3521***
arith mean	**647.875**		**sum =**	**0.0030557**	**sum =**	**1.9930715**	**factor**	**1.9074401**
scatter σ	**72.296**							
σ^2	**5226.696**				**wtd mean**		**652.247**	**Se**
					σ^2		**624.224**	**25**
					# obs		**8**	

In this table, Wi is = 1/Chi^2 and chi-sq is the standard calculation of: $(x_i$ – wtd mean$)^2$/Chi^2. The factor is the Chi^2 value divided by n-1. When multiplied by 1/Chi^2 Wi we get the total variance of the weighted mean the square root of which is the standard error of the weighted mean.

Using this produces a weighted mean for Arizona of **652.2 RCYBP** and a standard error of the mean of **+/-25**. This measurement incorporates both the within-measurement variability and the between-measurement variability observed in the Arizona data. The within-measurement variability was called the quoted error in the *Nature* article (5) and the quoted error does not capture all of the variability in the experiment. That is readily acknowledged in the Burr paper!

The best estimator of the total variability in these experiments is developed by employing Burr's Eq (3) – which, by the way is not a new development but has been employed for more than a decade.

If we calculate the same statistics for each of the other laboratories, we develop the following table:

	RCYBP weighted mean	std err
Arizona	652	25
Zurich	674	19
Oxford	749	18
Grand mean	*701*	
Grand std err	*30*	
χ^2 =	12.9402	
p-value	0.0015	

Obviously, with a simple Chi^2 calculation of these values we determine a p-value of 0.0015 which clearly means the data are heterogeneous. We can examine this data more closely with one-way ANOVA and find the following:

Evaluation of weighted means

ANOVA Table

Source	DF	SS	MS	F-val
Samples	2	20573.2	10286.6	**17.0229**
Error	13	7855.62	604.278	
Total	15	28428.8		

Testing the Equality of All Means

Classical F-Test

P-value under the equal variances assumption: **0.0002**

Generalized F-Test

P-value without the equal variances assumption: **0.01645**

Once we've done that, using contrast analysis, we can identify the primary cause of the non-homogeneity observed:

Planned Comparison				
Source	**DF**	**SS**	**MS**	**F-val**
Contrasts	1	20494.9	20494.9	33.9163
Error	13	7855.62	604.278	

Classical F-Test

P-value under the equal variances assumption: 0.000059

Generalized F-Test

P-value without the equal variances assumption: 0.0174

Post Hoc Comparison: Scheffe Test

Source	**DF**	**SS**	**MS**	**F-val**
AZ - OX	2	20498	10249	**16.96**
Error	13	7856	604.3	

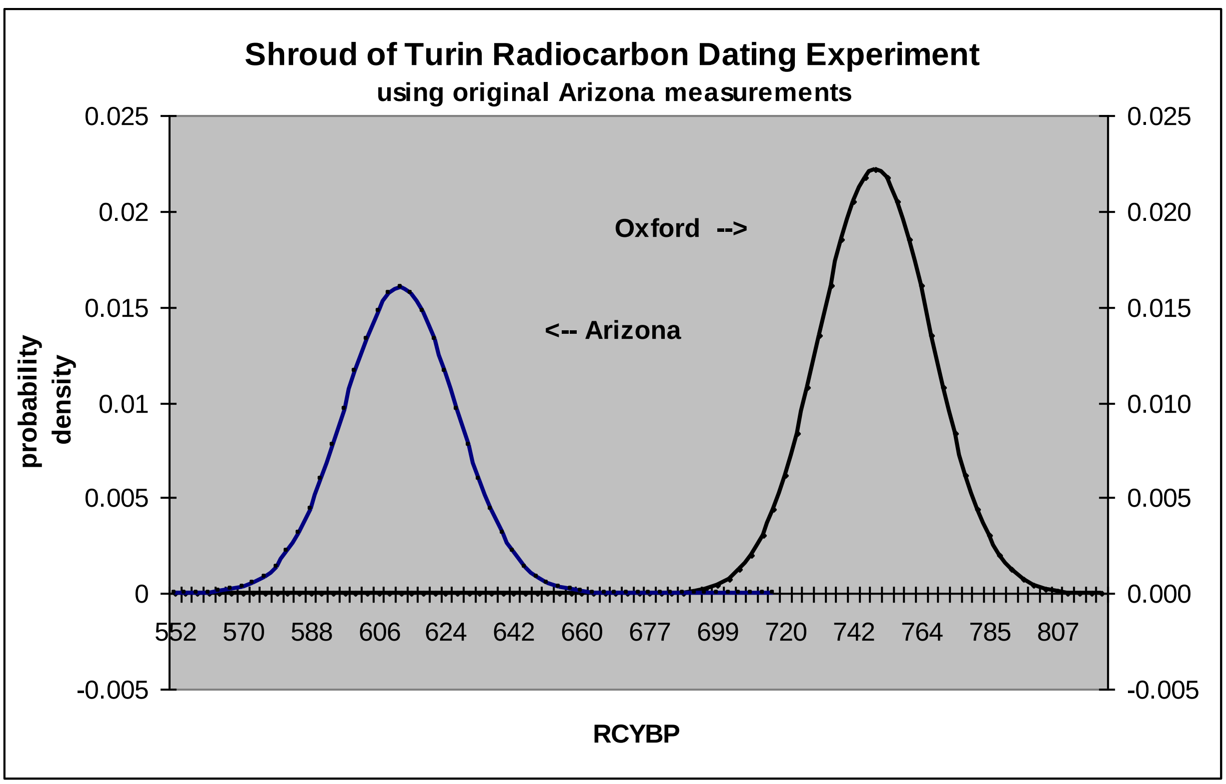
Classical F-Test

P-value under the equal variances assumption: **0.000238**

Generalized F-Test

P-value without the equal variances assumption: **0.03515**

In this post-hoc contrast, we find that the hypothesis that the weighted mean measurements for Arizona and Oxford come from the same population <u>is rejected,</u> regardless of what assumptions we make about equal variances in the two samples. This can be illustrated with a graphic display of the likely distribution of the individual measurements associated with the two laboratories:



We can also evaluate the Arizona and Zurich data in the same way:

Multiple Comparisons

Laboratory	**Contrast**
Arizona	-0.5
Zurich	0.5
Oxford	0

Estimate of linear combination of means (Std Error): 10.83

Planned Comparison

Source	**DF**	**SS**	**MS**	**F-val**
Contrasts	1	1443.56	1443.56	2.38889
Error	13	7855.62	604.278	

Classical F-Test

P-value under the equal variances assumption: 0.1462

Generalized F-Test

P-value without the equal variances assumption: 0.1850

Post Hoc Comparison: Scheffe Test

Source	**DF**	**SS**	**MS**	**F-val**
Contrasts	2	1444	721.8	1.194
Error	13	7856	604.3	

Classical F-Test

P-value under the equal variances assumption: 0.334

Generalized F-Test

P-value without the equal variances assumption: 0.3901

In this post-hoc evaluation, the hypothesis that the Arizona and Zurich means are the same is accepted. Thus, the primary source of the differences noted derives from the Arizona and Oxford mean measurements and, as a result, these measurements should not be combined.

Conclusion:
========
One should examine all possible reasons for the very large scatter of results.
A warning against contamination is given in the Burr et al paper.
The possibility of contamination is also confirmed in a Radiocarbon paper co-authored by Prof. Ramsey and Prof Hedges of Oxford. (4)
In the ICProgamme (1) 23 labs out of 38 failed to meet the 3 basic criteria.
This may be a solution for the many differences between archeological and radiocarbon dates.
Following the Burr et al paper, in the case of such dubious results, one should tune up the AMS facility.

Note:
This ANOVA statistical analysis, based on the Burr paper, confirms the conclusions I reached, using the ANOVA method given in "Perry's Chemical Engineers Handbook" (Fourth Edition McGraw-Hill).
I used this method for a lecture on statistics, given in Rome 1993.
The Scheffe contrast analysis noted in this paper confirms the simple Wilcoxon test, which states the Oxford and Arizona samples should NOT be combined.

Remi Van Haelst.

The author likes to thank Bryan Walsh for his precious help in applying one line ANOVA analyses. Using on-line ANOVA excludes any bias.

References
1 : Marion Scott et al. "Report International Collaborative Program"
2: "Error analysis at the NSF Arizona facility"
Burr et al. in Science Direct 2007 www.sciencedirect.com
3 : « Radiocarbon Dating the Shroud of Turin » Damon et al
Nature Volume 337 N° 6200 pp. 601- 615
4 : Private E-mail correspondence.