Warren W. Esty

Statistical analysis of hoard data in ancient numismatics

The geographic mobility of coins and the rate at which coins are lost from circulation are interesting questions which are answered utilizing statistical analyses of numerical data, especially from coin hoards. Unfortunately, hoard data are not the ideal "experimental" data treated in statistics texts. Numismatic analyses are often complicated by small sample sizes and non-randomness, which may invalidate statistical conclusions.

In statistics, in order to support a hypothesis, data are used indirectly to rule out the potential alternative explanations. Mathematics is used to describe how the data should look if some alternative explanation is true, and if the data do not look like that, the alternative is rejected, which indirectly supports the hypothesized explanation. However, it is sometimes overlooked that the mathematical deduction of "how the data should look" usually requires assumptions of large samples and randomness. If the data do not look as they should, it may be simply that these mathematical assumptions are false, not that the numismatic alternative is false. Therefore, it is difficult to prove anything statistically using hoards because the mathematical assumptions are not under "experimental" control. The following examples illustrate some of the subtle difficulties of statistical analysis of hoard data.

The importance of randomness and sample size can be seen when Roman Republican hoards are used to study die productivity and wastage (Duncan-Jones, 1999; Lockyear, 1999). Usually the term "wastage" (also known as "loss") refers to the fraction of each issue that goes out of circulation each year. However, Duncan-Jones uses it differently to refer to the decrease in the representation of issues as a fraction of the circulation pool. Even if no coins went out of circulation, new issues would expand the pool and decrease the percentage of any given old issue. With this definition, "wastage" rates would be significantly higher than the usual "loss" rates.

Suppose we wish to compute the average loss rate. First we assemble hoard data that gives the numbers of coins for dated issues (so we know the times elapsed between issues). We cannot simply compare issues sizes and expect older issues to be represented by fewer coins –the original issues were not all of the same size. If older issue A is consistently represented in hoards by half the number of coins of newer issue B, it might be because half of issue A went out of circulation in the meantime, or it might be because it was originally half the size, or some combination of the two. Even hoards with different burial dates make no contribution to distinguishing the alternatives because, in any given hoard assembled at a later date, the numbers of issues A and B are both reduced by the same fraction, leaving the relative representations of issues A and B unchanged over time.

Therefore, we need some knowledge of the relative original issue sizes. This can be provided by issues with good die counts, assuming that the issue size is proportional to the number of dies used to produce it. In some circumstances this assumption may be questionable (Buttrey, 1993 and 1994), but it seems justified in a high-production mint such as Rome. Obvious candidates for die counts are those issues with distinctive control marks such as actual Roman numerals. Suppose we have a number of well-counted issues with various dates spanning decades. Then we would have the (estimated) relative original issue sizes we need to estimate loss rates.

A numerical estimate of the loss rate uses a formula derived under an assumption of randomness of a particularly stringent sort –not only were the coins selected at random from circulating coins, but also all at one moment in time. If coins were withdrawn over a period of years, which is quite possible, the usual formula is wrong and can easily yield negative estimates of the rate (Buttrey and Buttrey, 1999, 127). Earlier issues can be added to the hoard in earlier years, as well as in later years, but later issues can only be added in later years. Therefore, earlier issues can be better represented than later issues.

Here is a much-simplified numerical example. Assume two hypothetical issues of equal size in successive years, and a 4% annual loss rate. Imagine coins withdrawn into a hoard just after the second year, say, including 50 from that year. If the 4% loss rate operated perfectly, the hoard would also include 48 (4% less than 50) from the first year. However, if, in violation of the assumption, coins were also withdrawn during the first year, the number 48 could be increased to well above 50, but the 50 from the second year could not be increased. Hoards could easily have older coins better represented! Using the usual statistical estimator of loss rates would yield a negative loss rate –erroneously, and not sensibly. However, consideration of the likely alternatives to the validity of the mathematical assumptions– not just alternatives to the numismatic assumptions –yields this simple explanation for apparently negative loss rates.

The point is that we must respect the mathematical assumptions of the model of hoard composition and not calculate statistics (estimates) when the hypotheses do not hold. Even if the basic idea of a "loss rate" being constant across issues and time is correct, we will not find the true rate if the data were not assembled as assumed in the mathematical model. Similarly, data could not disprove the idea of a constant loss rate unless they are assembled as assumed in the mathematical model.

The previous examples illustrate principles of statistics which are now outlined abstractly. Statistics, like mathematics, is an "if this, then that" subject. In the following, "H" stands for some historical or numismatic hypothesis, and "M" stands for the assumptions of the mathematical model. Statistical theorems are mathematical and read like this:

A) "If H and M are correct, then the data should look like *this*" (most of the time). We can deduce:

B) "If the data do not look like *this*, then H and M are not both correct" (most of the time).

For example, to prove that *different* types were shipped east and west, assume the same types were shipped east and west. Deduce, mathematically, what should be seen in the data if they were the same. If the data does not look as it should, consider the assumption disproved. However, be careful not to treat the numismatic hypothesis H as "the" assumption, because the mathematical component M is also critical. If "H and M" is false, it may well be that only M is false and H is still valid.

When the data don't look as they are supposed to under the numismatic hypothesis H, we should often conclude they look wrong merely because the data were not random. The numismatic hypothesis, H, might not be ruled out. For example, when statistics yield an estimated loss rate which is negative, that should not lead to rejecting the hypothesis of a constant positive loss rate, rather it undermines the mathematical assumption that the entire hoard was assembled randomly at one moment in time.

From (A), we can *not* deduce:

C) "If the data do look like *this*, then H is correct" (most of the time).

Mathematically, this would be the converse of (A), which does not follow from (A).

Instead, to support the hypothesis H, we need additional results like this:

D) "If H2 is some alternative hypothesis, then the data will look like *this different result*" (most of the time).

If the data should look substantially different under different hypotheses, then the test has high "power." To support H we want our method to have high power in order to have a good chance of ruling out all the alternatives.

Consider "loss" again. Can hoard data support a particular loss rate? Would data be able to rule out alternatives? The mathematical assumption of a constant loss rate yields a well-known mathematical model, illustrated with simplified hypothetical data here. The model supposes that each year p% (2% is illustrated here) of circulating coins go out of circulation. Suppose dies are taken as a proxy for issue size and deal with four (to make the hypothetical table short) well-counted issues spanning a period of time, the latest one being from year which we can call year 0. Hoards would have many other issues as well, but, without a solid standard with which to compare the numbers of coins, the coins of those individual issues would not help determine the loss rate.

This hypothetical table gives the date of each die-counted issue, the original number of dies which struck the issue, a number of coins in the hoard from that issue (which is, in this ideal example, perfectly proportional to the number of dies), the percentage of each issue of the total, and the cumulative percentage of the issue combined with all earlier issues. The middle column gives the fraction remaining when the 2%-per-year loss is compounded for the given number of years. In the "hoard with loss" columns the "dies" column gives a number equivalent to the original number of dies decreased by the "loss," and the number of coins in the hoard is proportional to the reduced number of dies.

Table 1. Hypothetical ideal hoards without loss and with loss.

date	hoard without loss				2%, fraction	hoard with loss			
year	dies	coins	%	cum %	remaining	"dies"	coins	%	cum %
-30	200	20	33	33	.55 (30 years)	110	11	23	23
-20	100	10	17	50	.67 (20 years)	67	7	15	38
-15	50	5	8	58	.74 (15 years)	37	4	9	47
0	250	25	42	100	1.00 (0 years)	250	25	53	100
total	600	60	100				47	100	
		"P1"	"C1"		"P2"	"C2"			

Loss rates of 0% and 2% produce distinguishable "percentage" columns (P1 differs from P2) and distinguishable "cumulative percentage" columns (C1 differs from C2). Similarly, a 4% rate also produces columns distinguishable from those, so a test might have good power and be able to rule out rates that are too high or too low in favor of the "correct" rate.

However, there are some significant statistical difficulties. We have already discussed how the data might not be random. Limiting consideration to issues dated well before the close of the hoard probably would mitigate that problem. However, sample size is also likely to be a significant concern. It takes a very large hoard to produce numbers of coins-per-issue as large as these, and even these dip to as low as 4. The total number of coins, 47, in the "hoard with loss" column is misleadingly low because it not the size of the whole hoard, but only the total of those four illustrative issues. Republican hoards may be composed from well over 150 circulating issues (the Cosa hoard, Buttrey, 1980, 81), so a hoard with numbers as large as those in the table would probably number over 1000 coins. Smaller hoards, as most are, would produce issue-samples so small as to violate the usual statistical admonition to use categories with at least five members. Random variation in small numbers may play a big role in making individual issues look over –or under-represented. That is a reason for using cumulative numbers instead of raw numbers per issue, especially when hoards are not so huge. We can hope the fluctuations in several small numbers may tend to "average out" and yield representative cumulative percentages.

Lockyear (1999) uses Roman Republican hoard data as above. He does not explicitly disregard the most recent years to try to avoid the hoard assembly problem, but he considers randomness and rightly notes that the fact that hoards are often similar suggests they are close enough to random samples, except for very large hoards. He notes that rather than use one large hoard with 1000 or 2000 coins, it is better to use several hoards with 100-300 coins. Then they are too small to analyze individual issue-sizes, but large enough to analyze cumulative data. He uses Crawford's die-number estimates, which are known to be too low, but because he uses them only for relative (as opposed to absolute) issue sizes that is not a problem if they are all low proportionally, which they probably are. He makes the power of the estimator clear by considering numerous alternative rates. Because different hoards from different regions have similar cumulative-percentage profiles that resemble the profile from a particular mathematical model, the hypothesis of a constant loss rate is supported.

Determining a loss rate was an analytical success because there was a mathematical model sufficiently –developed to yield results of the forms (A), (B), and (D) above. Without such results, it is not possible to find statistical analysis credible.

Suppose we want to know if different regions of the Roman Empire received the same types of coins, or if different regions were shipped different issues (Duncan-Jones, 1996). Could we detect differences by comparing hoards from various regions? The first problem is, again, sample size. Under the empire there were large numbers of different issues so most hoards will have most issues represented by too few coins to compare numbers and still meet the usual statistical admonition to have at least 5 counts in each category. So we might, again, try using cumulative percentages as data. But this introduces a new difficulty: The power goes down relative to an obvious alternative.

Here is a simplified explanation. Suppose coins were issued in equal amounts for 100 years, with coins from the 50 odd years going to the east and coins in even years going to the west. Then each cumulative curve would go up 1/50 = 2% every second year, and the cumulative percentages would never differ by more than 2% even though the issues sent to the two regions were completely different! The test has very low power. Cumulative data will not necessarily show individual numbers are different.

So, if we can not accumulate all the data, perhaps we might accumulate small categories into fewer, larger categories. But how? Consider grouping all issues of a single year into a one category (Duncan-Jones, 1996, in which dated denarius issues after AD 148 make identification by year possible). If this makes the numbers large enough to find a numerical difference between East and West, we must still seek and rule out alternative explanations. Furthermore, if there is truly a difference in the issues sent East and West, will we necessarily see it in the type of data we gather?
It is easy to imagine that, among the numerous issues each year, some were sent entirely to the East and some entirely to the West. If this continued year after year, there could have been a nearly complete difference in types sent to the East and West, yet the yearly fraction sent each region could be almost identical. Grouping by years would not be able to discern this type of difference, even though it is dramatic. The test has low power.

On the other hand, if the yearly fractions are statistically significantly different, some explanation is required. Perhaps it is that the fractions really were different, but perhaps it reflects lack of randomness or that the sample is too small to justifiably assert statistical significance. This problem can usually be identified by inspecting the original data, but sometimes the problem is concealed when the author uses percentages and fails to present the raw numbers. Presuming the amount of data is not too large, readers and future researchers wishing to re-analyze the data will be thankful for the original numbers. Duncan-Jones (1999) utilizes only 23 denarius issues in 3 hoards, for 69 raw numbers, few enough to give explicitly, but all we see are percentages, some hovering below 1%. This makes it hard to tell whether inappropriate mathematics might be a factor. Similarly, a study of coin mobility (Duncan-Jones, 1996) considers issues from 14 years in 10 hoards from east, west, and center. That is only 140 data points, which could and should be shown in a table of the raw data. Duncan-Jones admits that many of the raw numbers are too small to be confident using the usual ways of comparing such data, so the numbers are reworked three ways. One way converts them to rank-orders, which are then converted to a type of "distance" between hoards, which is presented as meaningful. But, is it? We are presented no mathematical model of the data-generation process that would lead us to believe the test has good power. With this type of "distance," about which the reader knows nothing, we cannot even be sure that basically similar hoards would be "close," or that hoards with significantly different representations of the issues would "distant." And, most significantly, if the numbers are too small for the usual, well-understood, techniques, why should we expect this, or any other, methodology to produce credible results?

Still another "distance" between cumulative distribution curves can be defined and used to compare hoards (Duncan-Jones, 1996, 147, Figure 5). However the mere fact a "distance" can be defined does not mean its use is appropriate or that conclusions drawn using it are valid. For example, if more coins of the single year 149 had entered the Bristol hoard it would closely resemble the Tell Kalak hoard from which is supposedly differs greatly. Although the difference is primarily in one early year, that one difference is repeated through most of the years because the data is cumulative. That yields an high "average distance" between the two cumulative data curves, in spite of the fact that there is no justification for averaging cumulative differences. These two hoards are the "least similar" (among all ten hoards discussed), and the reader is encouraged to note that they are on opposite sides of the empire. Whether or not it is true that different parts of the empire received different types of coins and their mobility did not mix them thoroughly thereafter, this type of analysis cannot provide much statistical support for the hypothesis because we do not know what "should" happen to this novel statistical measure in the various likely cases.

It is well known that small numbers may be too small to utilize famous formulas to make estimates and deductions. Readers are often rightly skeptical about statistics, especially from small samples. Readers should also be wary of conclusions supported by statistical techniques that lack a developed mathematical theory complete enough to address not only the results expected from the hypothesized explanation, but also the results expected from the likely alternatives.

Here are a few suggestions that can make an application of statistics to numismatics more credible.

1) Give the raw data, if feasible.

2) Do not use small samples to make estimates or draw unjustified conclusions by assuming that large-sample formulas and methods will give valid results for small samples. Do not conceal when samples are small by, for example, only giving percentages.

3) If the data suggest a hypothesis is false, consider that it may be the mathematical hypotheses of randomness and large sample-size that may be false, instead of the numismatic hypothesis necessarily being false.

4) Be concerned about the power of the method. Know, mathematically, what "should" happen under the main hypothesis and should happen under alternative hypotheses. This is especially important if the data are assembled in a manner unique to numismatics so that the math is not well-known, as is the case when data from hoards are used to study loss rates, or to study the geographic distribution of coin types.

5) Be sure to consider all likely alternative hypotheses. The logic of statistics is such that data do not directly support a hypothesis –all they can do is rule out alternatives. That means that data which "look as they should" do not strongly support the hypothesis the researcher has in mind unless, under the other alternatives, the data could not look that same way. That is, the statistical test must have good power.

References

Buttrey, T. (1980): *Cosa, The Coins. Memoirs of the American Academy in Rome*, volume XXXIV, p. 1-155 and plates I-X.

Buttrey, T. (1993): Calculating ancient coin production: Facts and fantasies. *NC*, 1993, p. 335-351.

Buttrey, T. (1994): Calculating ancient coin production II: Why it cannot be done. *NC*, 1994, p. 341-352.

Buttrey, S., Buttrey, T. (1997): Calculating ancient coin production, again. *AJN* second series 9, p. 113-135.

Duncan-Jones, R. (1996): Empire-wide patterns in Roman coin hoards, *Coin Finds and Coin Use in the Roman World, The Thirteenth Oxford Symposium on Coinage and Monetary History 25-27.3.1993* [ed. King, C., Wigg, D.] (Gebr. Mann Verlag) p. 140-152.

Duncan-Jones, R. (1999): Die productivity and wastage in Roman Republican coinage, *NC*, 1999, p. 245-254.

Lockyear, K. (1999): Hoard structure and coin production in antiquity –an empirical investigation, *NC*, 1999, p. 215-243.