

46. <そのデータ本当?>

私は下水道管理支援センターに併任していますが、その併任先で毎月発行している「維持管理総合支援メールマガジン」に“ここだけの話”と題して維持管理にまつわる私の数少ない経験談を連載しています。管理支援メルマガは、このコーナーの他、中央官庁等 HP 記事のトピックスと下水道用語解説「下水道を科学する」の 3 本立てで構成しています。幸か不幸か、まだまだ読者が少ないのでこれ幸い！ 宣伝も兼ねて技術開発に関連ありそうな話題を、一部改編して転載させていただくことにします。題して「そのデータ本当?」・・・

先日、全国の下水道建設事業量を統計処理する機会がありました。原データは約5500工事の4年間の年度別事業費で、総計は約10兆円になりました。こんなものかなと思いそのまま整理しましたが、その後前年度の同様の調査データを集計したら4年間で50兆円弱という結果になりました。いくら何でもおかしいと思い、原データをチェックしたところ、百万円単位で記入すべきところを千円単位で記入しているデータが20箇所以上発見されました。ひょっとしたらと思い、前に整理したデータもチェックしたところ誤記が6箇所発見され、4年間の集計は約5兆円になりました。わずか1000件に1件程度の誤記でしたが、総計は倍半分違いました。

全国データなど大量のデータを扱う場合は、1つ1つのデータをチェックするのが億劫なので、統計処理結果から異常を判断することをよくやります。しかし、この類の異常値の混入は、アンケート調査など多数の人間がデータ作成に関与する統計データでは避けられないエラーであり、なじみのないデータを扱う場合は大きなリスクが伴います。特に、合計値や平均値ではその影響が大きく出るので注意しなければなりません。

統計処理は地道な作業ですが、それだけに得られた結果は見る側を信じ込ませる大きな力があり、データが一人歩きすることもしばしばです。最近、下

水道統計の電子版が公表されるなど、原データに直接触れられる機会も増えているようで、表計算ソフトなどを使えば原データのチェックが簡単に行えます。もし「このデータ本当？」と思われたときには、一度チェックされたら如何でしょうか？・・・

< 川口 幸男 >

※No. 52号(2006/3/3)に掲載