Information for Decision Support, Information for Performance Evaluation – But Don't Mix Your Drinks

J. E. Everett, M. Kamperman, T. J. Howard
The University of Western Australia, Perth, Australia

jeverett@ecel.uwa.edu.au

Abstract

When the same information is used to support decisions and to evaluate performance, a falsely optimistic view of performance may result. An example occurs in quality control during ship loading for iron ore export. Ore quality depends upon consistent composition. Ore is sampled periodically during reclamation from stockpiles. The ship loader was moved from between source stockpiles when the sample assays differed from target composition. Each ship loader move incurred costly delays.

We found the apparent variations in composition could be largely ascribed to measurement error, and that intended correction during ship loading might even be harming quality.

The policy was changed, to load ships from a single stockpile without interruption. Sample assays were used to evaluate performance, not to drive decisions. Data are analyzed from 466 shiploads, spanning the years before and after the change of policy, to compare quality performance, as measured by the exporter and by the customer.

Keywords: MIS, DSS, Mining, Quality Control, Performance Evaluation.

Introduction

Management decisions require information. Evaluation of performance requires information. But when the same information is used to support decisions and to evaluate performance, a falsely optimistic view of performance may result. The paradox is illustrated by examples showing the effects can be substantial. Separation of the decision support and performance evaluation uses of information is proposed as an important principle in the study of Informing Systems.

As a simple example, consider a shooter aiming at a target. If the shooter adjusts after each shot to correct for the distance from target of the previous shot, then the error may grow without limit, if the random error is comparable in magnitude to the gun's systematic bias.

The principle has been applied to ship loading and quality control for iron ore export. Quality of exported ore depends upon consistent composition, not only in iron content, but also in percentages of silica, alumina and phosphorus. Ore is railed from the mine to the port and stored in large stockpiles, then reclaimed to load ships for export (Everett, 1996, 1997, 2001, Everett, Howard & Kamperman, 2001). The ore is sampled periodically while being reclaimed. Comparison of sample assays with the target composition was previously used to decide when to move the ship loader from one stockpile to another. Each ship loader move incurred costly delays, and a great deal of decision-making effort from the Process Control Officer (PCO). It was worrisome that customers reported greater composition variability than that estimated from the ship loading samples.

Investigation suggested random sampling errors were of comparable magnitude to the apparent composition discrepancies that drove adjustments of the source stockpiles during ship loading. If this were so, then it follows that the resulting ore variability as measured at the port was an underestimate of the true variability.

The policy was then changed, to load ships from a single stockpile without interruption. Sample assays were used to evaluate performance, not to drive decisions. It was expected that variability measured during loading would increase, but variability measured by customers should not. Quality should be maintained, while ore handling costs would be considerably reduced.

As a further benefit, separating decision support from performance evaluation would enable us to obtain separate estimates of the ore variability and of the assay errors for the exporter and for the customers, from the exporter and customer assay data. We should also be able to find which customers generate assays significantly different in level or variability from those of the exporter.

Aiming at a Target

If you fire a gun at a target and miss by an amount "t", you might be tempted to adjust your aim by an amount "–t" for the next shot. If the discrepancy were entirely caused by the gun's deviation, then that correction would be perfect. If the discrepancy were entirely caused by random error, then your next shot would probably be worse than your first. If you continued the same adjustment policy, your error would increase without limit as a "drunken walk", with amplitude proportionate to the square root of the number of shots, as shown in Figure 1. The blue dots show the results for a hundred random shots, each with normally distributed error, of unit amplitude. The red dots show the same situation if the shooter adjusts each time by the amount of the previous shot's error.

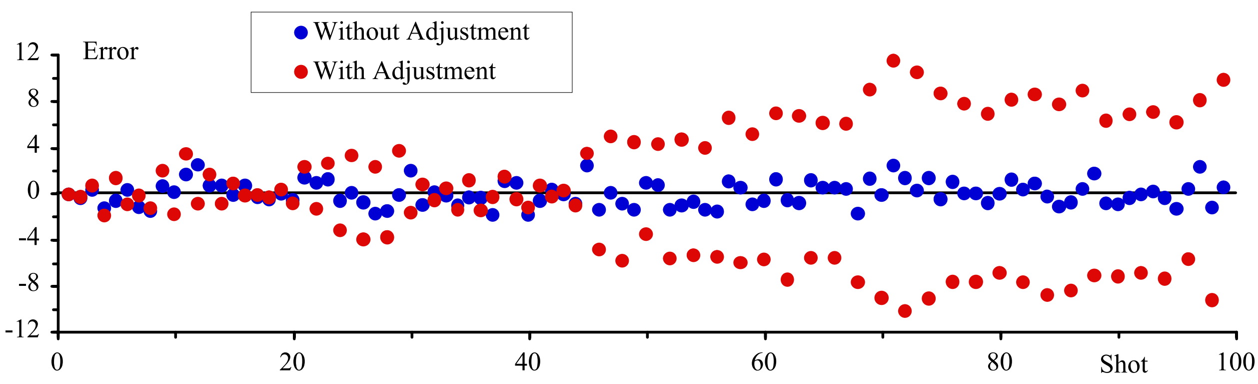

Figure 1: Aiming at a Target (Zero Real Deviation, Unit Standard Error)

Optimum Adjustment

Generally, the discrepancy of a shot will be the sum of the gun's real deviation (or consistent bias) and a random error. Assume that the gun's real deviation is unknown, but comes from a normally distributed population with zero mean and amplitude "τ". Similarly, the random error is assumed normally distributed with zero mean and amplitude "α".

The discrepancy for the original shot is the sum of the deviation and the random error:

$$d_0 = \tau\varepsilon_t + \alpha\varepsilon_{a0} \qquad \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

If you correct by a proportion "k" of the original discrepancy, then the next discrepancy is:

$$d_1 = -k(\tau\varepsilon_t + \alpha\varepsilon_{a0}) + \tau\varepsilon_t + \alpha\varepsilon_{a1} \qquad \ldots\ldots\ldots\ldots\ldots\ldots (2)$$

So the "standard discrepancy" σ_1 is given by:

$$\sigma_1^2 = E(d_1^2) = (1-k)^2\tau^2 + (1+k^2)\alpha^2 \qquad \text{……………………. (3)}$$

$E(d_1^2)$ is minimised if

$$k = \tau^2/(\alpha^2+\tau^2) \qquad \text{……………………. (4)}$$

If α and τ are equal, k should therefore be 0.5 (reducing the standard error by only 13.4%). The optimum k is always less than 1.0, if $\alpha>0$ (i.e. if there is any random error).

Figure 2 shows the standard discrepancy σ_1 as a function of k, when τ and α are equal. If the random error is as large as the bias, then full adjustment provides no improvement over zero adjustment.

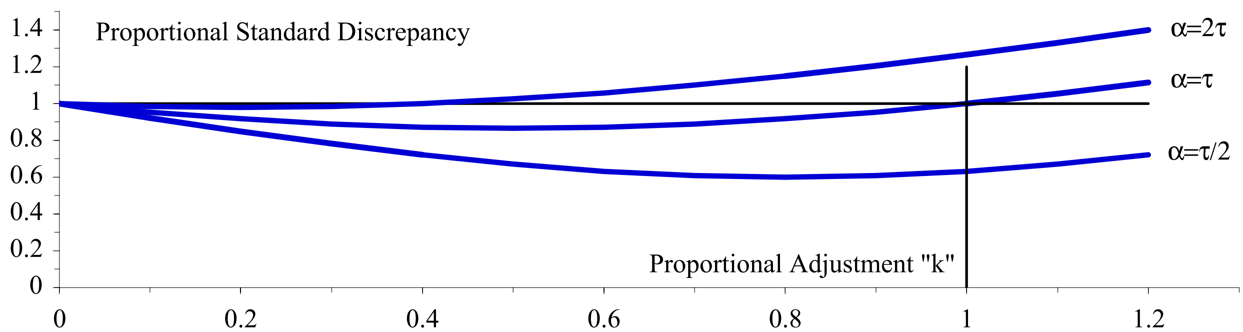

Figure 2: Standard Error Remaining after Proportional Adjustment

Loading Ships with Iron Ore

So what has all this got to do with loading iron ore onto ships?

Iron ore is railed from the mine to port, crushed, and then stacked to build stockpiles, typically of 100 to 200 kilotonne capacity. The ore is handled and the stockpiles are built with the objective of achieving as uniform ore composition as possible. This uniformity of composition is required not only in iron, but also in the contaminants such as silica, alumina and phosphorus.

Ships arrive in port to be loaded with iron ore cargoes of anything from about 50 to 120 kilotonnes. The cargoes are reclaimed from one or more stockpiles. While the ore is being loaded it is sampled and assayed at intervals of 8 or 9 kilotonnes. When the ship arrives at its destination, the customer unloads the ore. The customer then assays it again. The ore is fed into blast furnaces, whose smooth operation depends upon the ore having consistent composition, in iron and in each of the contaminant minerals.

Records are therefore available for each ship, for the assay values as measured by the exporter and by the customer. Until the end of 2000, the Process Control Officer (PCO) oversaw the loading of ships. The PCO monitored the assay values from material being sampled as a ship was being loaded. When assays were found to be drifting from target, the PCO ordered a change of source stockpile with the intention of bringing the shipload back to target composition.

It was then realised that assay measurement errors could be of a magnitude comparable to the real variation in ore composition.

Assay samples are extracted by periodically slicing off some of the ore from the conveyor belt. Assay errors had been estimated by splitting these samples into two sub-samples. The sub-samples were prepared and analysed separately, and the answers compared, to provide an estimate of the error. It was realised that this procedure underestimated the total assay error, since it considered only the preparation and assay error, and ignored the error inherent in removing the sample from the conveyor belt (see Pitard, 1993).

This suspicion was reinforced by an alternative estimation procedure. We compared the aggregate assays from stacking each stockpile with the assays later obtained when the stockpile was reclaimed. From the difference between the stacking composition and the reclaim composition, and from the number of build assays and reclaim assays, we could estimate the total sampling error. It was found to be considerably larger than the split sample error, and was comparable to the expected variation in ore composition. However, it was acknowledged that this stack/reclaim method could overestimate the sampling error for reclaim assays, because it gave a weighted average estimate of the stacking and reclaiming errors. When iron ore is being stacked to a stockpile, it is of more varied composition than when it is reclaimed. Ore is stacked along the length of the pile, and reclaimed across its width, to achieve efficient blending. Consequently, the assay errors during reclaiming are expected to be less then the assay errors during stacking.

Despite this reservation, the evidence suggested that assay errors during reclaiming were comparable to the real variation in ore composition. If so, then the situation was analogous to the problem of aiming at a target, as discussed above. Adjustment for apparent departures from target composition, if arising from measurement error, could result in more variability in ship composition rather than less. It could also cause the exporter's recorded ship variability to be less than the real variability, and thus explain why the customers were recording a greater variability when they unloaded the cargoes than the exporter was when loading the cargoes.

Using the sample assays, obtained during loading, as a decision support tool was not only suspected of spoiling quality, the reverse of intention. It also rendered the data invalid for performance evaluation, because it removed the independence between measurement error and real composition variability, as we shall see in the analysis below. Further, if we could remove the delays from changing source stockpiles, we could greatly reduce operational and demurrage costs.

As a result of these considerations, the ship loading policy was changed from the beginning of the year

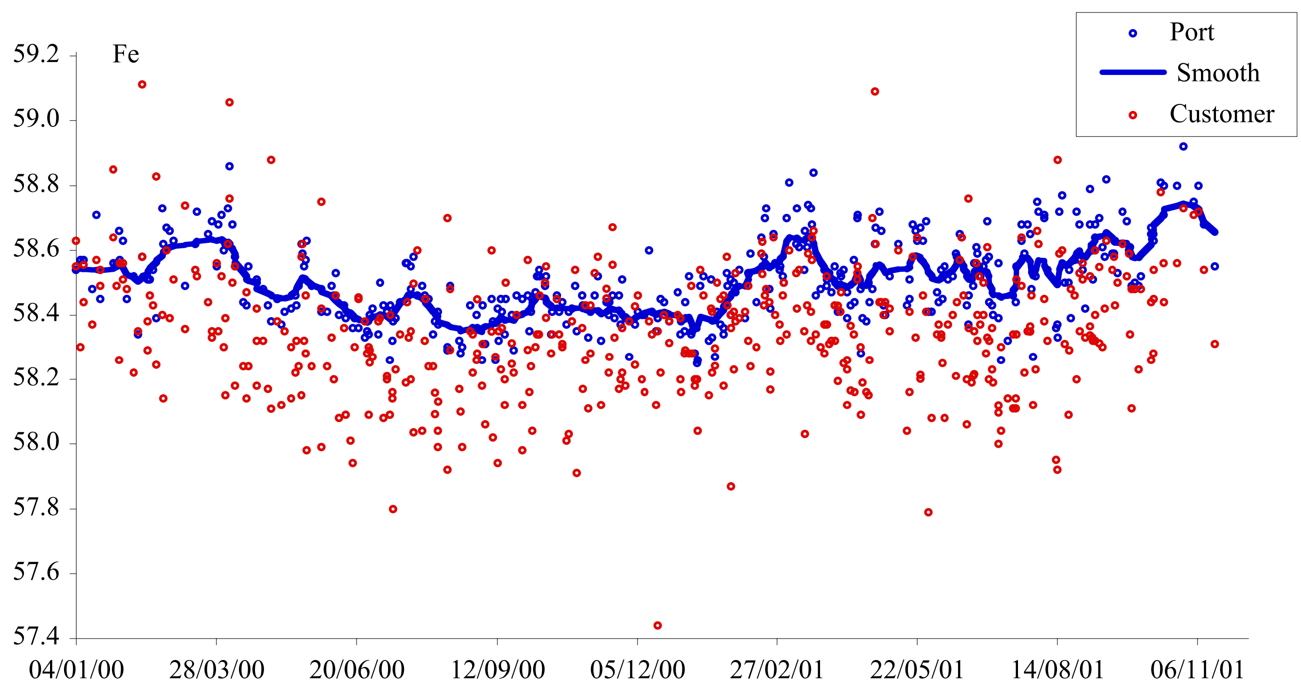

Figure 3: Port, Smoothed, and Customer Shipment Assays for Iron

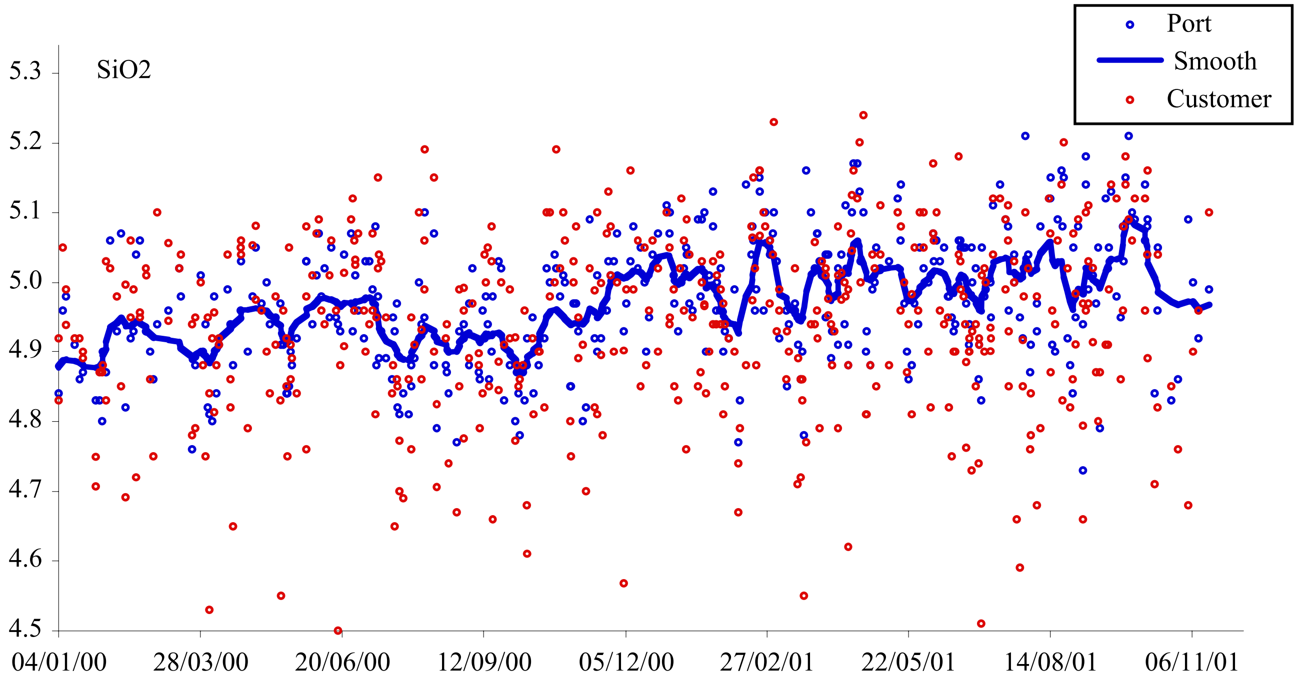

Figure 4: Port, Smoothed, and Customer Shipment Assays for Silica

2001. Ships are loaded from stockpiles to achieve target grade as planned before ship loading commences. The plan is based on assays taken while the stockpiles were being built, and is no longer revised in response to assays taken while the ship is being loaded.

Data are now available for shipments of an iron ore product, for 228 ships loaded during the year 2000 and for 240 ships loaded during the year 2001. For each shipment, records are available of the assays for each mineral, as recorded by the exporter and as recorded by the customer. Shipments were made to eight different customers.

Figures 3 and 4 show the data for the 448 shipments made during the two-year period. The figures are for iron and silica respectively. Similar data were recorded for the other two major contaminants, phosphorus and alumina. The blue circles are exporter assays made during loading, while the red circles are customer assays made during unloading of the shipments. The exporter (or port) assays have been exponentially smoothed to obtain the blue trend line.

The data enable us to estimate the ore variability and measurement errors, made by the exporter and by the customers, for the 2001 data. Because of the confusion between decision support and performance evaluation arising from the earlier procedure, it is not so straightforward to carry out the analysis for 2000. The data also permit an examination as to whether the change of policy has had a significant affect on quality. It has been found that the change of procedure, no longer changing source stockpile in response to apparent disparities from target, has generated considerable operational savings, including reduction in machine time and reduction in demurrage costs, without any appreciable loss in quality as measured by customers.

It was also of interest to examine for each customer whether the assay levels or the assay variability differ significantly from the exporter's records. Such comparisons clearly have value in any negotiation regarding product payment and product quality.

Analysis of the 2001 Shipments

We will first consider the (N=240) shipments made during 2001, after the policy of responding to sampling assays during ship loading had been discontinued. For each shipment, the record gave its tonnage, and the composition as measured by the exporter's assays, and as measured by the customer's assays.

Shipments were made to J=8 different customers. "N_j", the number of shipments to customer "j", ranged from 7 to 47.

First consider a single customer "j". Using terminology analogous to the target aiming example considered above, let "τ_j" be the standard deviation of the ore's true composition, and "α_j" and "β_j" respectively be the standard error of measurements made by the exporter and customer respectively. Our estimates of τ_j, α_j, and β_j will be denoted by "$\square_j$", "$\square_j$", and "$\square_j$". Let σ_a, σ_b and σ_{ab} be the standard deviation (around trend) of ship assays made by exporter and customer, and their covariance, with s_a, s_b and s_{ab} being the estimates of these parameters derived from the data.

Let "X_{aji}" and "X_{bji}" be the shipment assay, weight w_{ji} kilotonnes, as recorded by exporter and customer for the i^{th} ship to that customer, with means "M_{aj}" and "M_{bj}".

The port data has an overall estimated trend "T(time)", which is a function of time. This trend, as shown in Figures 3 and 4, was estimated by averaging backward and forward exponential smoothing of the port data.

The random "discrepancy", "x_{aji}" or "x_{bji}" for each shipment, as measured by the exporter or customer, is the composite of the ore's true deviation from trend and its assay error, after removing any systematic bias. Let c_j be the estimated systematic bias (if any) of customer assays relative to exporter assays.

For the shipments to customer "j" we can calculate:

$$c_j = M_{bj} - M_{aj} = \sum_i w_{ji}(X_{bji}-X_{aji})/\sum_i w_{ji} \quad \text{......... (5)}$$

We can use a t-test to check if c_j is significantly different from zero. If it is, then we can conclude that the customer's assays have a significant bias relative to the exporter's assays.

$$\text{t-test} = c_j/s_{cj}, \text{ where } s_{cj}^2 = \sum_i (X_{bji} - X_{aji})^2/(N_j-1) \quad \text{......... (6)}$$

$$x_{aji} = X_{aji} - T(\text{time}) = \tau_j\varepsilon_{tji} + \alpha_j\varepsilon_{aji} \quad \text{......... (7)}$$

$$x_{bji} = X_{bi} - T(\text{time}) - c_j = \tau_j\varepsilon_{tji} + \beta_j\varepsilon_{bji} \quad \text{......... (8)}$$

Where ε_{tji}, ε_{aji}, ε_{bji} are random samplings from populations ε_{tj}, ε_{aj}, ε_{bj}, which are each normally distributed, of unit variance and zero correlation.

$$\sigma_{aj}^2 = E_i(x_{aji}^2) = \tau_j^2 + \alpha_j^2 \quad \text{......... (9)}$$

$$\sigma_{bj}^2 = E_i(x_{bji}^2) = \tau_j^2 + \beta_j^2 \quad \text{......... (10)}$$

$$\sigma_{abj} = E_i(x_{aji}x_{bji}) = \tau_j^2 \quad \text{......... (11)}$$

So we can calculate estimates for each customer set of data:

$$(\text{Ore Standard Deviation})^2 = t_j^2 = \sum_i x_{aji}x_{bji}/(N_j-1) = s_{abj} \quad \text{......... (12)}$$

$$(\text{Exporter Standard Error})^2 = a_j^2 = \sum_i x_{aji}^2/(N_j-1) - t_j^2 = s_{aj}^2 - s_{abj} \quad \text{......... (13)}$$

$$(\text{Customer Standard Error})^2 = b_j^2 = \sum_i x_{bji}^2/(N_j-1) - t_j^2 = s_{bj}^2 - s_{abj} \quad \text{......... (14)}$$

Finally, we can combine the results from all customers to find overall estimates t, a and b.:

$$t^2 = \sum_{i,j} x_{aji}x_{bi}/(N-J) = s_{ab} \quad \text{......... (15)}$$

$$a^2 = \sum_{i,j} x_{ai}^2/(N-J) - t^2 = s_a^2 - s_{ab} \quad \text{......... (16)}$$

$$b^2 = \sum_{i,j\,i} x_{bi}^2/(N\text{-}J) - t^2 = s_a^2 - s_{ab} \qquad \text{......... (17)}$$

F-tests were used to test whether the customer standard errors were significantly larger than the exporter standard errors. The F-ratio was calculated as $F = s_b^2/s_a^2 = (t^2 + b^2)/(t^2 + a^2)$.

Because each variance was the sum of the assay error variance plus the ore composition variance, which could be assumed independent, an F-ratio significantly greater than one is evidence that the customer assay errors are larger than the exporter assay errors (i.e. that $\beta > \alpha$).

Summary of the 2001 Shipments Analysis

For each element in each shipment to customer "j", we have a data record $\{w_{ji}, X_{aji}, X_{bji}\}$, where w_{ji} is the weight in kilotonnes, and X_{aji} and X_{bji} are the exporter and customer assays for the shipment.

From the set of data records for customer j, we derive estimates $\{M_{aj}, M_{bj}, c_j, s_{aj}, s_{bj}, s_{abj}\}$,

Applying a t-test, we check whether c_j differs significantly from zero. If it does, we conclude that assays carried out by customer j have a systematic bias relative to those carried out by the exporter.

For customer j we obtain t_j, a_j, and b_j, estimates of the ore standard deviation, and the exporter and customer standard errors. Combining the data records provides overall estimates t, a, and b. F-tests on the variance ratios test for significant difference in standard error between exporter and customers.

Analysis of the 2000 Shipments

Up to the end of 2000, the PCO adjusted source stockpiles in an attempt to compensate for discrepancies from target apparent from assays taken during loading.

With this policy, if the exporter (or port) assay error is comparable to the ore variability, we had a situation analogous to the target-aiming model depicted in Figures 1 and 2. The correction for apparent discrepancies may well have done more harm than good to the shipment quality, quite apart from its delays to the loading process.

Using the same terminology as for the 2001 data, we add the proviso that, under the earlier policy, the material loaded was chosen so as to correct for an unknown proportion "k" of the apparent discrepancy.

Equations (5) and (6), as developed for the 2001 data above, remain unaltered. So we can examine differences in assay means as for the 2001 data.

Comparison of assay variability is more difficult, because we can no longer treat ore deviations and exporter assay error as independent. When we allow for the correction for apparent discrepancies during loading, equation (7) becomes:

$$x_{aji} = \tau_j\varepsilon_{tji} + \alpha_j\varepsilon_{aji} - k(\tau_j\varepsilon_{tji} + \alpha_j\varepsilon_{aji}) = (1\text{-}k)\tau_j\varepsilon_{tji} - k\alpha_j\varepsilon_{aji} + \alpha_j\varepsilon_{aji} \qquad \text{......... (18)}$$

In equation (18), the real ore deviation is $(1\text{-}k)\tau_j\varepsilon_{tji} - k\alpha_j\varepsilon_{aji}$, and the exporter assay error is $\alpha_j\varepsilon_{aji}$.

Thus the real ore deviation and the exporter assay error are now negatively correlated. The real ore standard deviation (as shipped) is given by:

$$(\text{Ore Standard Deviation})^2 = \underset{\cdot}{\tau}_j^2 = (1\text{-}k)^2\tau_j^2 + k^2\alpha_j^2 \qquad \text{......... (19)}$$

If k approaches one, then the apparent variability as measured at the port would approach zero, but the real variability would approach α_j, which would be a net increase in real variability: $\underset{\cdot}{\tau}_j > \tau_j$ if $\alpha_j > \tau_j$.

This is analogous to the situation discussed in the section on "Aiming at a Target". The apparent variability of equation (18), as measured by the exporter, will be decreased by the factor (1-k). For $\underset{\cdot}{\tau}_j$ to be less than τ_j requires $k < 2/(1+\alpha_j^2/\tau_j^2 n_i)$. To minimise $\underset{\cdot}{\tau}_j$ requires $k = 1/(1+\alpha_j^2/\tau_j^2)$.

Equations (8) to (11) must be revised to:

$$x_{bji} = (1-k)\tau_j\varepsilon_{tji} - k\alpha_j\varepsilon_{aji} + \beta_j\varepsilon_{bji} \quad \ldots\ldots\ldots (20)$$

$$\sigma_{aj}^2 = E(x_{aji}^2) = (1-k)^2(\tau_j^2 + \alpha_j^2) = \underset{\cdot}{\tau}_j^2 + \alpha_j^2 - 2k\alpha_j^2 \quad \ldots\ldots\ldots (21)$$

$$\sigma_{bj}^2 = E(x_{bji}^2) = (1-k)^2\tau_j^2 + k^2\alpha_j^2 + \beta_j^2 = \underset{\cdot}{\tau}_j^2 + \beta_j^2 \quad \ldots\ldots\ldots (22)$$

$$\sigma_{abj} = E(x_{aji}x_{bji}) = (1-k)^2\tau_j^2 - k(1-k)\alpha_j^2 = \underset{\cdot}{\tau}_j^2 - k\alpha_j^2 \quad \ldots\ldots\ldots (23)$$

The three equations (21, 22 and 23) now contain four unknowns (τ_j, α_j, β_j and k), so cannot be solved without further evidence. This indeterminacy arises because the real ore deviation and the exporter assay error cease to be independent when the exporter assay is used to change the ore selection. However the following general observations are relevant:

- Customer assay variance is still the sum of the real ore variance and the customer error variance.
- Exporter assay variance is now less than the sum of the real ore variance and the exporter error variance, by the amount $2k\alpha_j^2$.
- Covariance is now less than the real ore variance, by the amount $k\alpha_j^2$.

For the 2001 data, we used the F-ratio, calculated as $F = s_b^2/s_a^2 = (t^2 + b^2)/(t^2 + a^2)$, to test whether the customer assay errors were larger than the exporter assay errors. The adjustment procedure used in the year 2000 means that the exporter assay error cannot be assumed independent of the ore variation, so we cannot use this test on the variance ratio to find if the customer errors are larger than the port errors. Although we can say $s_b^2 = \underset{\cdot}{t}^2 + b^2$, we now have $s_a^2 = \underset{\cdot}{t}^2 + a^2 - 2ka^2 \neq \underset{\cdot}{t}^2 + a^2$.

Results for the Assay Means

Table 1 shows the mean port and customer assays results from the analysis of the 2000 and 2001 shipping data. For each customer, and for each mineral (iron, phosphorus, silica and alumina) the table shows the number of shipments, the mean assay as reported by the exporter and by the customer, and the difference between the means, which is tested for significance by a t-test. The results are summarised for the 228 shipments in 2000, the 240 shipments in 2001, and for the total 468 shipments.

In the t-tests of Table 1, many comparisons are being made simultaneously, so a Bonferroni correction to the significance levels has been applied (Bonferroni, 1936; Miller, 1981). If we use raw alpha values, twenty simultaneous independent tests on purely random data would yield an average of one test spuriously significant at the 5% level. To compensate, the Bonferroni significances are calculated as (1-(1-α)^n), where n is the number of simultaneous independent tests. The difficulty remains as to how many independent tests are being made simultaneously, since the data are partially correlated. Much ink has been spilt arguing the question (see, for example, Perneger, 1998). In this example, an n of 8 (the number of customers) has been used, to calculate the reported Bonferroni significances for the individual customers. No Bonferroni correction has been applied to the “All Ships” summary statistics.

There is a consistent tendency for customers to report significantly lower iron and silica, and significantly higher phosphorus and alumina, than does the exporter. This systematic bias is particularly noticeable for iron, which may relate to the fact that payment is directly related to the customer’s iron assay.

		Load at Port				Customer Unload				Customer - Port				Signif. Customer - Port			
Customer	Ships	Fe	P	SiO_2	Al_2O_3	Fe	P	SiO_2	Al_2O_3	Fe	P	SiO_2	Al_2O_3	Fe	P	SiO_2	Al_2O_3
Yr 2000																	
1	19	58.45	.042	4.95	1.29	58.31	.041	4.80	1.28	-.14	-.001	-.15	-.01	4%		3%	
2	17	58.43	.042	4.95	1.29	58.28	.044	4.96	1.35	-.14	.001	.01	.06	2%			4%
3	45	58.46	.042	4.93	1.28	58.28	.043	4.92	1.29	-.18	.001	-.01	.01	0%			
4	10	58.50	.041	4.94	1.28	58.48	.041	5.02	1.27	-.02	.000	.09	-.01				
5	2	58.37	.044	4.98	1.35	57.96	.045	5.01	1.44	-.42	.001	.04	.09				
6	8	58.49	.042	4.94	1.27	58.06	.043	5.01	1.26	-.43	.002	.07	.00				
7	49	58.48	.041	4.93	1.29	58.32	.042	4.97	1.32	-.16	.001	.04	.04	0%		0%	0%
8	32	58.41	.043	4.95	1.28	58.26	.043	4.93	1.30	-.15	.001	-.02	.02	0%			
9	46	58.47	.042	4.93	1.30	58.48	.043	4.88	1.31	.01	.001	-.05	.02			3%	
All Ships	228	58.45	.042	4.94	1.28	58.31	.043	4.93	1.30	-.15	.001	-.01	.02	0%	0%		0%
Yr 2001																	
1	25	58.57	.041	5.01	1.29	58.41	.041	4.84	1.30	-.16	.001	-.17	.01	0%		0%	
2	21	58.57	.040	5.02	1.29	58.52	.043	4.97	1.38	-.05	.002	-.04	.10				0%
3	42	58.55	.041	5.00	1.28	58.36	.042	4.95	1.30	-.18	.001	-.05	.02	0%	0%	1%	
4	7	58.55	.041	4.98	1.30	58.46	.042	5.06	1.27	-.09	.002	.07	-.03				
6	8	58.54	.041	4.96	1.29	58.15	.043	4.96	1.30	-.39	.002	.00	.00				
7	55	58.55	.041	5.00	1.29	58.37	.042	5.00	1.33	-.18	.001	-.01	.04	0%	0%		0%
8	47	58.53	.041	5.01	1.29	58.39	.041	4.95	1.32	-.14	.001	-.06	.03	0%		2%	1%
9	35	58.53	.040	5.00	1.30	58.35	.040	4.97	1.31	-.18	.000	-.03	.01	0%			
All Ships	240	58.54	.041	5.01	1.29	58.39	.042	4.95	1.32	-.16	.001	-.05	.03	0%	0%	0%	0%
Two Yrs																	
1	44	58.52	.041	4.98	1.29	58.37	.041	4.82	1.29	-.15	.000	-.16	.00	0%		0%	
2	38	58.52	.041	4.99	1.29	58.43	.043	4.97	1.37	-.09	.002	-.02	.08		0%		0%
3	87	58.50	.041	4.96	1.28	58.32	.042	4.93	1.30	-.18	.001	-.03	.02	0%	0%	1%	0%
4	17	58.53	.041	4.96	1.29	58.47	.042	5.04	1.27	-.05	.001	.08	-.02				
5	2	58.37	.044	4.98	1.35	57.96	.045	5.01	1.44	-.42	.001	.04	.09				
6	16	58.52	.041	4.95	1.28	58.11	.043	4.98	1.28	-.41	.002	.03	.00	3%	3%		
7	104	58.51	.041	4.97	1.29	58.35	.042	4.98	1.33	-.17	.001	.02	.04	0%	0%		0%
8	79	58.48	.041	4.99	1.29	58.34	.042	4.95	1.31	-.14	.001	-.04	.02	0%		3%	0%
9	81	58.50	.041	4.96	1.30	58.41	.042	4.92	1.31	-.08	.001	-.04	.01	1%		0%	
All Ships	468	58.50	.041	4.98	1.29	58.35	.042	4.94	1.31	-.15	.001	-.03	.02	0%	0%	0%	0%

Table 1: Mean Assays and Port to Customer Differences

As we have seen in the theoretical discussion, the comparison of means is not affected by the earlier adjustment policy, so the conclusions regarding systematic bias are as valid for the year 2000 data as for the year 2001 data.

Results for the Assay Variances

Table 2 shows the standard deviations of port and customer assays for year 2000 and year 2001 shipments. The F-statistic (the ratio of customer to port variances) has been calculated for each customer, and for the year's total shipments.

In each year, the variance of the customer assays significantly exceeded the variance of the port assays, for each mineral.

For the year 2001, equations (9) and (10) show that the port and customer assay variances are $(\tau^2+\alpha^2)$ and $(\tau^2+\beta^2)$ respectively. The estimate of the customer variance $(\tau^2+\beta^2)$ is significantly greater that the estimate of the port variance $(\tau^2+\alpha^2)$. Therefore we have significant evidence that the customer error (β) is greater than the port standard error (α).

		Load at Port Std Devn				Customer Std Devn				F-Ratio Customer/Port				Signif. Customer/Port			
Customer	Ships	Fe	P	SiO_2	Al_2O_3	Fe	P	SiO_2	Al_2O_3	Fe	P	SiO_2	Al_2O_3	Fe	P	SiO_2	Al_2O_3
Yr 2000																	
1	19	.07	.001	.06	.04	.13	.002	.12	.11	3.7	2.9	3.5	8.3	4%		4%	0%
2	17	.06	.001	.06	.03	.12	.003	.12	.04	4.4	7.4	4.6	2.2	2%	0%	2%	
3	45	.06	.001	.06	.04	.19	.002	.09	.05	11.8	3.6	2.1	2.0	0%	0%		
4	10	.06	.001	.08	.03	.15	.001	.08	.03	6.2	1.2	1.1	1.1				
5	2	.07	.002	.06	.09	.10	.002	.09	.11	1.8	1.0	2.5	1.3				
6	8	.07	.001	.05	.04	.18	.001	.09	.03	7.4	1.8	3.2	0.8				
7	49	.06	.001	.05	.03	.13	.001	.07	.04	4.9	3.3	2.0	2.6	0%	0%		1%
8	32	.06	.001	.05	.03	.14	.002	.14	.05	4.8	7.1	7.1	2.4	0%	0%	0%	
9	46	.09	.001	.07	.04	.21	.002	.11	.08	5.9	4.3	2.2	4.2	0%	0%	3%	0%
All Ships	228	.07	.001	.06	.03	.17	.002	.11	.06	6.2	4.1	2.9	3.3	0%	0%	0%	0%
Yr 2001																	
1	25	.10	.001	.09	.03	.13	.001	.11	.07	1.7	2.6	1.7	4.5				0%
2	21	.08	.001	.08	.03	.20	.003	.12	.04	5.5	16.9	2.1	2.0	0%	0%		
3	42	.08	.001	.09	.04	.14	.002	.10	.05	3.5	3.1	1.4	1.7	0%	0%		
4	7	.08	.000	.06	.05	.17	.001	.09	.07	4.5	2.6	2.4	1.7				
6	8	.09	.001	.07	.04	.19	.001	.11	.04	4.4	1.0	2.6	1.3				
7	55	.10	.001	.06	.04	.17	.002	.09	.05	2.9	3.3	2.1	1.8	0%	0%	3%	
8	47	.09	.001	.07	.03	.16	.003	.14	.05	3.0	16.1	3.5	3.1	0%	0%	0%	0%
9	35	.10	.001	.08	.03	.16	.001	.09	.06	2.4	1.6	1.2	2.6	5%			3%
All Ships	240	.09	.001	.08	.04	.16	.002	.11	.05	3.0	5.3	2.0	2.3	0%	0%	0%	0%
F-Ratio 2001/2000																	
Customer 1		2.0	1.0	1.8	0.7	1.0	0.9	0.9	0.4								
2		2.3	0.5	1.9	1.1	2.8	1.1	0.9	1.0								
3		1.8	1.0	1.9	1.0	0.5	0.8	1.2	0.8								
4		1.7	0.2	0.5	2.8	1.2	0.4	1.2	4.4								
6		2.0	1.7	1.6	1.2	1.2	0.9	1.3	2.0								
7		2.7	1.5	1.5	2.0	1.6	1.5	1.5	1.5								
8		2.2	1.1	1.9	1.0	1.3	2.6	0.9	1.3								
9		1.3	1.3	1.2	0.8	0.5	0.5	0.7	0.5								
All Ships		1.9	1.0	1.5	1.1	0.9	1.3	1.0	0.7								
Signif. 2001/2000																	
Customer 1																	
2																	
3																	
4																	
6																	
7		0%			5%												
8							2%										
9																	
All Ships		0%		0%			2%		1%								

Table 2: Port and Customer Standard Deviations Compared

For the year 2000, where the port material was being adjusted in response to port assays, the port and customer variances are represented by the more complex equations (21) and (22) so cannot be directly used to compare the port and customer standard errors.

Table 2 also shows the results of F-tests checking whether the variances in 2001 are significantly changed from 2000. The customer variances have not changed consistently: the customer phosphorus and alumina variances have gone up and down by 30% respectively, while the iron and silica variances are virtually unchanged. By contrast, the port variances for iron and silica (the major control variables under the old policy) have gone up by a highly significant 90% and 50% respectively. This result is consistent with the conclusion that the year 2000 variances in port assay were artificially reduced by the compensation policy, which was largely driven by measurement error.

		Ore SD (Around Trend)				Port Ship Error				Customer Ship Error			
Customer	Ships	Fe	P	SiO_2	Al_2O_3	Fe	P	SiO_2	Al_2O_3	Fe	P	SiO_2	Al_2O_3
Yr 2001													
1	25	0.100	.0008	0.064	0.025		.0005	0.058	0.019	0.082	.0013	0.095	0.062
2	21	0.045	.0005	0.045	0.017	0.071	.0004	0.065	0.024	0.192	.0027	0.107	0.037
3	42	0.057	.0009	0.076	0.025	0.050	.0004	0.039	0.025	0.132	.0015	0.069	0.038
4	7	0.082		0.053	0.055		.0005	0.018		0.150	.0008	0.069	0.038
6	8	0.061	.0010	0.081	0.038	0.068	.0005		0.012	0.183	.0004	0.070	0.023
7	55	0.086	.0008	0.060	0.037	0.047	.0004	0.017	0.013	0.143	.0014	0.067	0.039
8	47	0.089	.0004	0.075	0.029	0.031	.0007		0.009	0.135	.0033	0.113	0.045
9	35	0.062	.0008	0.076	0.034	0.080	.0008	0.034	0.009	0.144	.0012	0.053	0.045
All Ships	240	0.077	.0007	0.068	0.031	0.051	.0006	0.033	0.016	0.141	.0020	0.084	0.043

Table 3: Estimation of Ore Standard Deviation, Port and Customer Errors, for 2001 Ships

Covariance Analysis

For the year 2001 data, we can use equations (12) to (14) to obtain t_j as estimates of the ore standard deviation, τ_j, and a_j and b_j as estimates of the port and customer standard errors, α_j and β_j.

Table 3 shows the results of this calculation for the year 2001 data. The blank spaces are where the statistics were not calculable (square roots of negative numbers). Results for individual customers are indicative only, but we can place some reliance on the overall figures, based on the 240 shipments.

It appears that the port assay errors are comparable, or a little smaller than, the ore standard deviation. The customer assay errors are consistently larger than the ore standard deviation.

We cannot carry out the same analysis for the year 2000 data, because the adjustment factor k makes equations (21) to (23) indeterminate. However, it is not unreasonable to assume the same assay error estimate "a" for the year 2000 as for the year 2001. Combining equations (21) and (23) (for the whole year's data) gives us the estimator:

$$s_a^2 - s_{ab} = (1-k)a^2 \qquad \text{......... (24)}$$

So, assuming the error a is consistent between the two years:

$$k = 1 - (s_a^2 - s_{ab})_{2000}/a^2 \approx 1 - (s_a^2 - s_{ab})_{2000}/(s_a^2 - s_{ab})_{2001} \qquad \text{......... (25)}$$

Hence we can also estimate the real ore variance and the customer error variance:

$$\square^2 \approx (s_{ab})_{2000} + ka^2 \qquad \text{......... (26)}$$

$$b^2 \approx (s_b^2)_{2000} - \square^2 \qquad \text{......... (27)}$$

The results of these calculations are shown in Table 4. Using the assumption that the port error variance is the same for the year 2000 as for 2001, we obtain estimates for the "correction" applied to each mineral as a proportion of the apparent discrepancy from target, as measured at the port. The correction factor esti-

Year	Statistic	Interpretation	Fe	P	SiO_2	Al_2O_3
2001	s_a^2	Port Variance	8.3E-03	8.4E-07	5.6E-03	1.2E-03
	s_b^2	Customer Variance	2.9E-02	4.8E-06	1.4E-02	3.3E-03
	s_{ab}	Covariance	5.8E-03	5.5E-07	4.5E-03	9.5E-04
	t	Ore Std Devn	0.076	0.0007	0.067	0.031
	a	Port Std Error	0.050	0.0005	0.034	0.016
	b	Customer Std Error	0.141	0.0020	0.084	0.043
2000	s_a^2	Port Variance	4.3E-03	8.3E-07	3.7E-03	1.1E-03
	s_b^2	Customer Variance	3.6E-02	3.7E-06	1.4E-02	4.2E-03
	s_{ab}	Covariance	3.0E-03	5.5E-07	2.8E-03	1.1E-03
	a	Port Std Error	0.050	0.0005	0.034	0.016
	k	'Correction' Propn	48%	5%	20%	91%
	t'	'Real' Ore Std Devn	0.065	0.0008	0.055	0.036
	b	Customer Std Error	0.179	0.0018	0.104	0.053

Table 4: Estimation of Statistics, for 2000 Ships

mate turns out to be in the feasible range 0% to 100% for each mineral, ranging from 91% for alumina down to 5% for phosphorus. This range is broadly compatible with the effort the PCO applied in trying to correct each mineral in response to the port assays. The calculations suggest that the "real" ore standard deviation and the customer standard errors in 2000 were little different from those in 2001. Consequently, nothing has been lost, and much operational efficiency gained, by abandoning the attempts at correction in response to port assays during loading.

Conclusion

This study has illustrated some of the problems and confusions that may arise when the same data are used for operational decision support and control as well as for performance evaluation. The main danger is that adjustment in response to discrepancies from target may actually spoil performance if the discrepancies are largely due to measurement error

It would be an overstatement to claim that the same data should never be used for operational decision support and control as well as for performance evaluation. As we have seen in the discussion of aiming at a target, a proportional response to error-bearing signal may improve performance, but full response is likely to cause hunting and increased discrepancy from target. It is also important to recognise that the use of the data to adjust operations destroys the independence between discrepancy and measurement error, making analysis more complex and lessening the information that can be extracted from the data, as has been seen in our analysis of data from 2000, as compared with the simpler fuller analysis possible for the 2001 data, which had not been used for operational adjustment.

The ship loading example has also provided an example where abandoning a complex adjustment policy has led to considerable operational savings with little or no decrease in quality.

A side benefit of the change of policy and consequent analysis is that we are now able to identify which customers have assays differing significantly in level or variability from those of the exporter. Since such assays figure largely in negotiations determining price and quality, this information has considerable potential value to the exporter.

References

Bonferroni, C.E. (1936) Teoria statistica delle classi e calcolo delle probabilità. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze,* 8, 3-62.

Everett, J.E. (1996). Iron ore handling procedures enhance export quality. *Interfaces*, 26(6), 82-94.

Everett, J.E. (1997). Simulation to reduce variability in iron ore stockpiles. *Mathematics and Computers in Simulation*, 43, 563-568.

Everett, J.E. (2001). Iron ore production scheduling to improve product quality. *European Journal of Operational Research*, 129, 355-361.

Everett, J.E., Howard T. & Kamperman, M. (2001). The development of an MIS for iron ore mining operations. *Proceedings of the Informing Science 2001 Conference, 337-345*. Kracow, Poland; IS2001.

Perneger, T.V. (1998). What's wrong with Bonferroni adjustments? *British Medical Journal*, 316, 1236-1238.

Pitard, F.F. (1993) *Pierre Gy's Sampling Theory and Sampling Practice*. 2nd ed.., Boca Raton, FL, CRC.

Miller, R.G. (1981) *Simultaneous statistical inference*. 2nd ed.. Springer Verlag, 6-8.

Biographies

Jim Everett is Professor of Information Management in the School of Business at the University of Western Australia. Jim's publications and research interests are outlined on his website http://www.imm.ecel.uwa.edu.au/je.

Marcel Kamperman and **Terry Howard** are Research Fellows in the School of Business at the University of Western Australia. Both are currently managers in the iron ore industry, with extensive experience in process control and quality management.

Terry is a qualified metallurgist with a Masters Degree in the field of secondary metallurgy. He has 33 years experience in the Australian minerals processing industry, in both primary and secondary processing fields.

Marcel is a qualified geologist and chemist with a Ph.D. in the field of mineral chemistry. He has 17 years experience in the Australian minerals industry, in exploration, mining and process control.