Takeover - Solution

We will mainly discuss the problem within the following context:

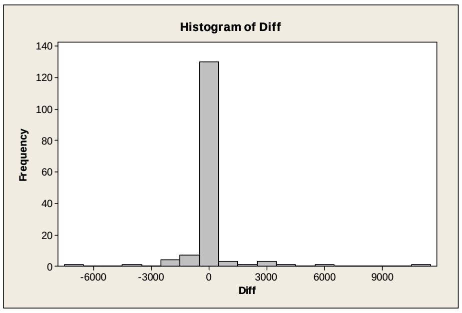
Context 1:
Buyer and seller had at the outset come to terms with that the updating procedures were satisfactory, and that the system prices should be the basis for the valuation. Sampling with subsequent data analysis was nevertheless done on the buyers own initiative, and claims were raised with hindsight.

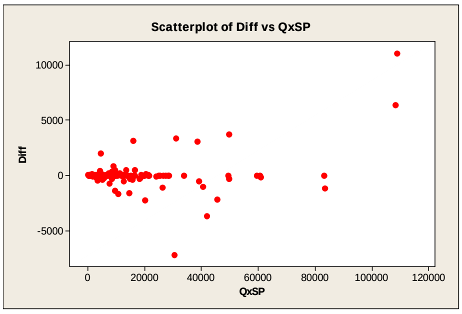
From the variables Q=Quantity, SP=SysPrice and IP=InvPrice, we compute the system value QxSP and the invoice value QxIP and their difference Diff=QxSP-QxIP for each of the n=153 sampled inventory items. Selected descriptive statistics are

Descriptive Statistics: QxSP; QxIP; Diff

Variable	Total Count	Mean	SE Mean	StDev	Sum	Minimum	Maximum
QxSP	153	13715	1530	18930	2098468	35	108876
QxIP	153	13664	1493	18470	2090584	23	101957
Diff	153	52	114	1408	7884	-7202	11035

We note from the graph below that most differences are zero or close to zero, and that there are extreme outliers both on the positive and the negative side. If we plot the differences against the system value we see that two extreme positive differences are associated with inventory item of high value, and otherwise there is just a weak tendency for the differences to vary more as the value of the inventory item increases.





From this it seems that it is not likely to be any systematic overstatement of the values by the system prices. If we try to estimate of the total overstatement, there are several ways to do it:

1. The difference method
2. The ratio method

The difference method just takes the mean difference, and scale it up by the total number of inventory items N=5501. This gives an estimated total amount of 5501 x 52=286 052. Scaling up the standard error of the mean accordingly we get 5501 x 114= 627 114 (see formula later). With errors margins based on plus/minus two standard errors, which gives an approximate 95% guarantee we have (rounded to nearest thousand)

$$286\,000 \pm 1\,254\,000$$

saying that the estimated overpay just as well could be due to sampling error.

Remark. Formally the 95% guarantee relies on normally distributed observations. We have seen from the graph that this is not so. However, the mean of many independent observations will have distribution closer to normal even if the parent distribution is non-normal.

The alternative Ratio method compares the total overstatement in the sample with the system value of the sample, and scales this up by the total system value of the population, which is known. We have the estimate

$$\begin{aligned}\text{Total overstatement} &= \frac{\text{Total overstatement in sample}}{\text{Total system value of sample}} \; x \; \text{Total system value} \\ &= \frac{7\,884}{2\,098\,468}\, x\; 72\,649\,991 \\ &= 272\,963\end{aligned}$$

The corresponding standard error for the ratio estimate requires some theory (see technical note below). Using errors margins based on plus/minus two standard errors we report (rounded)

$$273\,000 \pm 1\,231\,000$$

The result is close to the one above, and with the same general conclusion.

Main conclusion (context 1):

There is no compelling statistical evidence that the inventory is overpriced by the system prices. The error margins are so large that even underpricing may be possible (but not as likely). Admittedly there are some large deviations between the two prices, but they are likely to be balanced off, and do not justify the effort in the context of this takeover to find the specific reason.

*We have (**i**n context 1) agreed upon the system prices as basis, and they have to be trusted until it is proved beyond reasonable doubt that they cannot be. The sample taken does not give any support for such doubt. We cannot practice a decision making process where we risk to reject accepted and acceptable procedures based on unjust statistical arguments.*

Remark. In the context of continuing the system, it is clearly of interest to find the reason for the large outliers, and possibly take some corrective actions on the system.

From the data we have estimated the standard deviation of the differences in worth to be about S=1400. To find the sample size n which gives error margins about +/- 500 000 we have to solve the equation (see technical note below)

$$2 \cdot 5\,501 \cdot \frac{1\,400}{\sqrt{n}} = 500\,000$$

giving n=949. Since this is a fairly high compared with the population size (more than 10%) we multiply by the computed finite correction factor 0.853 (see technical note), giving n=809.

<u>Technical note</u>: For both methods the error margins are of form

$$\pm k \cdot N \cdot \frac{S}{\sqrt{n}}$$

where

N = total number of inventory items (in the calculations used 5501)
n = number of inventory items in the sample = 153
S = computes standard deviation from the sample (S_D =1408, S_R =1384)
k = safety factor (here used 2 giving an approximate 95% guarantee)

These error margins are based on normal approximation, which may be justified when the sample is small in comparison with population, say less than 10%. Otherwise theory suggests the error margins to be reduced by following multiplicative finite correction factor:

$$\sqrt{1-\frac{n}{N}}$$

The formula for the error margin may be used to determine the size of the sample n to achieve a desired error margin. If n turns out large in comparison with the population size N, it should be reduced by the multiplicative factor

$$\frac{1}{1+\frac{n}{N}}$$

For the difference method may take S or $S/\sqrt{n}$ directly from the descriptive statistics. Or we could have used the analysis option for confidence intervals for difference in means for paired samples. For the ratio method the corresponding S have to be computed according to the (easily programmable) formula

$$S_R^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - R \cdot X_i)^2 \quad \text{where } R = \frac{\overline{Y}}{\overline{X}}$$

In general the ratio method works best when the deviations between the two amounts are likely to be increasing with (and proportional to) the size of invoice amount, while the difference method works best when the deviations are independent of the size. In this data the large deviations typically occur for the large recorded amount, so we believe the ratio method to be the best.