



## TENTH SESSION OF THE STATISTICS DIVISION

Montréal, 23 to 27 November 2009

### Agenda Item 4: Airport traffic data

#### BEST PRACTICES IN ESTIMATING MISSING DATA

(Presented by ACI)

##### SUMMARY

This IP examines a number of issues related to the estimation of missing data, specifically data on airport economics. It focuses on the desirability of extrapolating from a sample to achieve regional and global totals. Best practices are discussed in some detail: some of these practices were learned in the proverbial school of hard knocks, a demanding institution, to be sure.

### 1. INTRODUCTION

1.1 A lesson learned. In 1995, Airports Council International (ACI) published its first *Airport Economics Survey*. This initial effort was a six-page document which provided data from 220 airports, roughly one/seventh of all ACI member airports. The sample included a number of large airports, but data from some regions was spotty and the data was simply not representative enough of ACI's membership to permit extrapolation to derive regional or global totals. ACI thus made a decision to publish totals for the sample airports only: included were important parameters such as airport employment, revenues and capital expenditure presented by region.

1.2 ACI routinely sends documents such as the *Airport Economics Survey* to its list of press contacts. There was modest press coverage of the 1995 document. One trade publication in North America mistakenly printed the data *without* mentioning it was merely tabulations from a sample. This caused considerable confusion because ACI-North America had conducted its own survey of capital expenditure (capex) and made extrapolations for the region from a larger sample. Thus there was a large discrepancy between NA's capex figure and the capex figure for North America published in the journal, which listed the *ACI Survey* as the source. ACI-NA was using its data to demonstrate that airport operators were aggressive in planning for future growth: the erroneous statistics in the trade journal created confusion and somewhat undermined that conclusion. ACI concluded neither the press nor the travelling public would appreciate the nuances involved in presenting sample data.

1.3 In a meeting with regional offices immediately after the erroneous media report, ACI resolved to build up the sample to a size for the *Economics Survey* to a point that would enable extrapolation to produce estimates of *total* global and regional statistics. The larger the sample, it was reasoned, the higher would be the level of confidence in the totals (see para 5.1). This policy decision resulted in a new, more aggressive approach to data collection. ACI's statistics department started the collection process earlier in the year, in January 1996, and sent repeated reminders to airports not complying with deadlines for submission of data.

## 2. LENGTHENING THE DATA COLLECTION CYCLE AND INCENTIVIZING PARTICIPATION

2.1 Airports' fiscal years vary: many airports had not submitted data in 1995 because their fiscal years had not yet ended, ACI learned. The data collection cycle, which had been only 3 months in duration in 1995, from transmission of the questionnaire to collation and analysis of the data, needed to be expanded to 6 to 9 months to ensure that a sufficient number of airports would report fiscal year-end data. In addition, because a large number, nearly 1,000 airports, were contributing passenger and freight data to ACI, the contacts supplying this data were encouraged to intervene in getting a higher response rate to the economics questionnaire (although in most airports, the two responsibilities were handled by different personnel, this approach had good results.) ACI also translated the questionnaire into Spanish to increase the response rate in Latin America. In 1996, the sample size doubled and ACI extrapolated the data to produce regional and global totals.

2.2 Several years later, ACI had built the *Survey* up with a number of features, including an annex on ownership of privatized airports and articles by eminent industry experts. The *Survey* had become "weighty" enough to be sold. To attract more airport participants in the sample, ACI offered participating airports a free copy. It is difficult to calculate the results of this offer, but it is worth noting that the survey sample size grew in every year between 1995 and 2008.

## 3. BUILDING MOMENTUM IN DATA COLLECTION

3.1 Over the last 14 years, ACI found that success in data collection bred more success. The 2008 survey collected data from airports representing over 70% of global passenger traffic. In some regions the sample size exceeded 80% of total regional traffic. Extrapolation to regional and global totals in this situation can be handled with a very high level of confidence. There was clearly a "tipping point", perhaps around 2003 or 2004, when the *ACI Airport Economics Survey* had acquired a reputation as an excellent resource for airport data. Some airports outside the sample may have concluded it was better not to be "on the outside looking in" and began to participate by supplying data. Reasoning that some data was better than no data, ACI also encouraged airports to skip data elements on which they had insufficient information rather than opt out of the survey altogether.

## 4. BACK TO BASICS – THE IMPORTANCE OF CLARITY AND BREVITY IN THE QUESTIONNAIRE

4.1 The airport personnel who provide data to ACI are busy managers, often under pressure to deliver data to many competing organisations, as well as needing to generate data to support in-house decision processes. Recognizing this fact, ACI from its first *Survey* in 1995 took a decision to limit the

questionnaire to one page. While over the years, certain data elements were added to the questionnaire, this was accomplished by decreasing the font size, rather than adding a page. (This trick works!).

4.2 Another important consideration in issuing any questionnaire is complete clarity in the presentation. It is important to recall that many persons answering the questions will not be native speakers of the language(s) of the survey. After several years, ACI began issuing definitions of some of the terms used as an annex to the survey to further clarify the concepts presented. This had two positive outcomes: it led to better compliance (fewer blank responses) and it cut down on airport inquiries to ACI about the terms they found confusing.

## 5. **PROBLEMS WITH EXTRAPOLATION OF DATA**

5.1 While a large sample normally results in more accurate extrapolation, if key airports in a region are missing from the sample, distortions can occur. An example in ACI's PAXFLASH monthly statistics would be if Atlanta or Chicago (ORD) were missing from the sample: omitting these airports, the two busiest in the world, would distort the North America regional totals. Indeed, sometimes it is worth waiting the extra day or two for data from a key large airport, even though timely data is the objective of PAXFLASH. With economic data, the problem can be more pronounced. For example, if a large airport with a costly expansion underway is missing from the sample, capital expenditure may be understated for the region.

## 6. **SPOTTING BIZARRE DATA**

6.1 People, even statisticians and economists, are capable of mistakes. One of the most common is a 'single-digit' mistake, which is normally an extra zero at the end of a statistic. These large errors are readily spotted and ACI is vigilant in weeding them out. (Perhaps more worrisome are the smaller errors, which are not so easily caught.)

6.2 In any annual survey, the staff doing the analysis climbs a learning curve and becomes adept at spotting statistical anomalies. For example, there are certain ratios, such as number of airport employees to passengers, which are fairly stable over time in each region. Wild deviations from the ratio range, say for example an airport serving 1 million passengers have 10,000 employees, are immediately apparent to the analyst.

6.3 Bizarre data can also come from poor questionnaires, or from asking questions which might be better left off the survey. One year, ACI tried to gather data on the valuation of airport assets. Due to differing accounting practices (including widely varying treatment of depreciation) across nations and regions, the results were deemed "impossible to interpret". Asset valuation at two airports of the roughly the same size in two different nations could be as much as 3,000% different! This data element was left out of all subsequent year's questionnaires.

## 7. FILLING IN THE BLANKS, ESTIMATING MISSING DATA

7.1 There are plenty of options when one needs to estimate data from an airport which has not returned a questionnaire. (But some of these are time-consuming and a drain on scarce resources):

- a) **Look at recent data from the same airport.** Airports' economic statistics would be unlikely to vary hugely in a one-year period unless there were some external shock (tsunami, SARS) in the region. A quick look at time-series data from previous years can provide a fairly accurate assessment of the missing year's data. This method, of course, only works when an airport has participated in the survey regularly in the past;
- b) **Alternative sources.** ACI is not the only source of economic data on airports, although it is considered by many to be the most authoritative and accurate. Over the years, ACI has found *Momberger's Airport Information* to be an excellent source on capital expenditure and *Airline Business* to have solid data on the top one hundred airports' operating results and profits, for example;
- c) **Google it!** Many airports have good websites, loaded with information, including financial reports. This approach, however, is quite time-consuming: no two airport websites are alike and it can take a long time to find the data. And even when the data is available it often will not be in the same format ACI needs. Capex for instance, may not be provided by year, but over the entire period of an airport master plan, which could have a 15 to 20 year horizon and
- d) **Phone the airport.** Another labor-intensive approach to finding missing data. The problem here is two-fold: finding the right contact and; once found, actually connecting with the person for a meaningful exchange of views.

## 8. CONCLUSION

8.1 As described above, there are many ways to estimate missing data. But ACI wants to stress two factors for success in minimizing missing data. The first is to collect data through a brief, logical, clear questionnaire. Secondly, if the airports in the sample appreciate the value of the final product of a survey, and the analysis accompanying the data is timely and profound, the response rate is likely to be much higher. If airport operators view ACI's annual *Survey* as an essential reference document and actually benchmark themselves against the data, one could safely assume that they will be enthusiastic about their participation in the *Survey*.

— END —