Tourism data pitfalls: a discussion paper

David Edwards

Scattered Clouds

September 2022

Contents

1 Introduction

This report has been prepared for ETOA by Scattered Clouds.

This discussion paper considers a range of factors that those making use of tourism data should always be mindful of in order to minimise the risk of reaching a misleading conclusion when interpreting statistical evidence.

The overarching advice that applies to much of the following discussion is that those who use tourism statistics should always be inquisitive about the data; how was it gathered, how is it being presented, might the entity reporting the data have a particular set of 'findings' that it wishes to promote.

As a sector tourism is fortunate in having a wealth of data that can be used to help paint a picture as to the volume and value of visitors, their behaviours and attitudes, but this abundance of data definitely comes with a health warning and the following sections endeavour to explain why.

2 Aggregation of arrivals data

The issue here is essentially one of geography. If all travellers made simple trips from origin A to destination B and then returned back to their origin, we would have nothing to worry about.

However, travel is not that simple, and this has implications when we try to compare figures collated on outbound travel from a source market with arrivals data complied at a destination.

For example, data from the US Department of Commerce tells us that in 2019 US residents made 17.3 million visits to Europe. So, were we to seek out data from each country that the US categorises in Europe and summed figures on arrivals from the US in 2019 would we get 17.3 million?

No.

There are several reasons for this. One explanation is that data will typically be based on sample surveys, and as will be discussed in a later section of this note these come with a margin of error.

But of equal importance in accounting for why aggregating arrivals data from a series of destinations and expecting this to tally with outbound figures is that visitors can travel to more than one destination on a single trip.

Clearly in terms of a departure from the US they are only adding 1 to the tally. But each time they enter a new country on their trip the National Tourist Board or whoever compiles inbound tourism data will chalk-up another arrival from the US.

Figures that profile US outbound travel to Europe are taken from the US Survey of International Air Travellers, and in 2019 this survey had a sample size for those heading to Europe of 15,429, which may sound like a hefty number, but recall that the total size of the market was 17.3 million.

Based on this sample the US data reveals that the typical traveller to Europe visited 1.5 countries on their trip, a figure that has varied hardly at all over the past two decades.

As such adding up arrivals from the US as collated for inbound data series in each country of Europe we could reasonably expect the figure for 2019 to be about 50% higher than the true number of Americans who visited Europe that year.

Clearly this only becomes an issue when we are looking at aggregated inbound data for a series of destinations, but nonetheless we might well fall into the trap of assuming that the number of Americans 'visiting Europe' was substantially higher than is the reality.

The same issue can arise when looking at geographically disaggregated data at a destination level.

Taking inbound visits to the UK as an example, figures from the International Passenger Survey available from VisitBritain indicate that in 2019 the number of 'staying visits' (that's visits lasting at least 1 night) stood at 39.418 million.

However, if we add up the estimates for staying visits in each of London, other parts of England, Scotland, Wales and Northern Ireland we reach a grand total of 43.352 million, so almost four million higher than the true figure for inbound visits to the UK.

Again, the reason for this discrepancy is that an inbound visitor will often spend nights in different parts of the UK during their trip.

The insight here is that tourism statistics users should be highly cautious of adding up inbound data relating to a series of disaggregated destinations as it will almost certainly not represent the true figure for arrivals from a particular source market to the aggregated area.

3 Simplification of trip purpose

A range of challenges presents itself when looking at trip purpose breakdowns of tourism data.

One is that different destinations may use different definitions for what counts as which type of trip, or simply not include some trip purpose categories altogether.

An oft used categorisation for trip purpose is Holiday, Business, Visits to Friends and Relatives, Study, Other.

In one sense it is possible to figure out what 'Other' represents based on it having to be anything that doesn't neatly fit within any of the other categories, but exploring (if it is known) the range of trip purposes that are destined to be counted as 'Other' can be rewarding and puzzling in equal measure, as some would argue that individual trip purpose reasons are important enough to be presented as a standalone group, while others might argue that other purposes could reasonably be bundled in with one of the headline journey purpose groupings.

Looking at data for the UK in 2019 we find that out of a total 40.9 million inbound visits 2.3 million were classified as having an 'Other' or 'Miscellaneous' trip purpose, that's almost 6% of the total.

The list of trip purposes that fall into this category for inbound data to the UK includes:

- Looking for work
- Au pair
- Shopping
- Join/accompany
- Medical treatment
- Overnight transit
- To look for a place to live
- Selling or buying a house (not for business)
- To get married (even if travelling on a package)
- To bring a child to school
- To renew a visa, permit
- To fulfil conditions for retaining residence permit
- To take an examination
- Watching any event for pleasure (including a sporting event)
- Work experience
- To receive an honour
- To do research for own interest, not as part of a profession/job
- To live in a religious community
- To take part in a competition – not amateur sports players/arts performers
- Voluntary work
- Working holiday (if both work/business and holiday are considered main reasons)
- Flight only trips
- Attending graduation

As can be seen this is both a long and rather varied list.

Those trips that are classified as Business, Holiday or VFR may of course in reality represent a blend of activities, with a Business visit encompassing a few hours or even a day or two of holiday-like activities, a visit to family featuring some days out that would be akin to a holiday and so on.

Perhaps one of the biggest misunderstandings can relate to trips classified as 'Business'. There is a tendency to assume that these are all in scope for tourist boards to market the idea of extending the trip for leisure, or that all

Business trips involve attending meetings or conferences, typically referred to as the MICE segment, whereas very many do not.

Data for inbound visits to the UK in 2019 reveals that MICE accounted for 1.8 million inbound Business visits (with the 'M' in MICE being Meetings of 21+ people).

Moving down the meeting size spectrum those with 6-20 people accounted for 1.6 million Business visits while those featuring 1-5 people accounted for 1.9 million inbound Business visits.

Those visiting to attend training events or development represented just over 0.5 million inbound Business visits while the group nearly aways not thought of as contributing to the overall tally for Business are lorry drivers, and in 2019 their number stood at 1.9 million, or 21% of all Business visits, although their spending represented just 3% of the total.

Another term that should always engender a question or two about what is or what is not meant is when a table or report refers to 'Leisure' as a purpose. It is often unclear as to whether VFR is included as arguably some VFR trips may well be purely for leisure, whereas others, for example attending a funeral or visiting an ill relative are rather more non-discretionary than discretionary in the sense of a leisure trip.

4 Unpicking trends in visitor spend

One of the commonest questions that tourism stats will be used to proffer an answer to is how much the sector is worth, whether this amount is going up or down, and how it compares with equivalent figures for other destinations.

Before thinking about comparing figures for one destination with another let us first consider the pitfalls that exist when looking at trends for a single destination.

It is all too easy to look up figures on the value of tourism each year, populate a spreadsheet with this information and draw a trend line and / or calculate how much the amount has changed.

It is essential to always figure out whether, if the data for several years is taken from a single report, figures are reported in outturn prices or constant prices. If the user is having to source figures for separate years from separate reports they will almost certainly be in outturn prices for that year.

This matters because of inflation. Inflation erodes the value of an amount by lessening its purchasing power. A bank balance of £1,000 in 1972 was worth a lot more than a bank balance of £1,000 in 2022.

The risk is that if we do not adjust value figures for inflation then we might reach the wrong conclusion as to whether or not the worth of tourism is increasing. If what is driving the value higher is underlying inflation the nominal total may be higher than last year, but in real terms it could be unchanged, or even lower.

To illustrate this point back in 1980 inbound visitors spent £2.961bn in the UK whereas in 2019 they spent £28.448bn, an uplift of 861%. However, if we adjust for inflation, the 1980 figure presented at 2019 prices is £12.792 bn, meaning that in real terms inbound visitor spending has increased by a rather more modest 122% over the forty-year period.

When comparing spending figures between different countries we need not only to take account of inflation, which will likely vary from one to the other, but in many cases also consider how to accommodate figures being reported in different currencies.

Clearly £1,000 is not the same as US$1,000 or as CHF1,000, so if we want to make a like-for-like comparison of say the amount spent per inbound visit, then we need to convert to a common currency.

This then raises the issue of deciding on the appropriate exchange rate to use. These of course vary from day to day, month to month and year to year. If we know that our data on spend per visit relates to an annual average then the right course of action is to use the annual average exchange rate during that period rather than simply looking up today's rate.

Imagine that during the 2016 we know that the average amount spent by an inbound visitor to the UK was £750 and that for an inbound visitor to the US it was $1,000. Were we to convert the $1,000 to sterling using the average exchange rate during January 2016 ($1.44 = £1) then we'd conclude that the sterling equivalent of $1,000 was £695. However, if we used the average exchange rate during December 2016 ($1.25 = £1) then our $1,000 converts to £801. So using the January rate we would have proclaimed that inbound visitors typically spent less per visit than those to the UK, whereas using the December rate we would have boldly claimed that the exact opposite were true.

5 Sampling methodologies

Whether small ad hoc studies or estimates published by national statistics offices, very often tourism data is generated from a sample survey.

Surveys can be a crucial vehicle for providing insight, but they can equally convey findings that mislead, whether through unforeseen design flaws or due to deliberate sleight of hand when it comes to question composition or selective reporting.

Back in 1936 ahead of the US presidential election the Literary Digest undertook a poll. Its findings predicted Alfred Landon would win 57% of the vote and Franklin D Roosevelt 43% based on a hefty sample of 2.4 million voters.

The reason you will never have heard of President Landon is that because on election day 62% voted for Roosevelt against 38% for Landon, an outcome George Gallup was able to accurately predict from a sample of just 50,000.

We will return to why the Literary Digest got it so spectacularly wrong in due course, but first it is important to describe the many and varied types of 'survey error'.

First up is 'sampling error'. The clue is in the name, surveys are based on a sample, they are not a census of the entire population. As there is likely to be variability rather than homogeneity within the entire population in relation to whatever it is we are trying to measure, we cannot expect to be fully accurate in our assessment when just looking at the characteristics of a proportion of that population.

If our study is well designed and we were to keep repeating the survey the 'result' would be close to the actual figure more often than not, but from time to time the result would be an outlier, simply because we are basing our results on a sample.

Turning next to 'coverage error' - one of the most common reasons why survey results might not be as conclusive as we expect. In essence the issue here is that the sample is not fully representative of the population about which we are trying to make discoveries.

Imagine we are planning to survey passengers departing from an airport and to keep costs down we decide not to conduct interviews in the evening or at weekends. This may lead to results that fail to reflect the true nature of the population of departing passengers if those travelling late in the day or at weekends have characteristics that are different to passengers travelling during 'working hours' between Monday and Friday. Poor coverage will lead to poor findings which may lead to poor decisions.

'Measurement error' is a form of survey error that has become more of an issue in recent decades, as the internet age and readily available survey software has meant too many have opted to "do a survey" without devoting sufficient care and attention to questionnaire design.

Poorly constructed questions can arise because the author has made them too long-winded, imprecise, included multiple clauses, failed to offer a complete set of feasible answers, or if they are being mischievous, chosen to devise a leading question in order to conjure up a desired answer.

Examples of poor or leading questions "Had you not seen the advert would you have not visited?", "Do you agree the countryside here is exceptionally beautiful?" and "How would you rate the staff at reception and in the restaurant".

Finally there is 'non-response error', often referred to as bias. This arises when some groups within the sample are more likely than others to complete the survey. An example from the world of tourism might be an airport

survey conducted in a single language when it is known that a significant proportion of travellers are not fluent in that language.

Sometimes we can correct for non-response error by weighting our findings, for example if we know the true age distribution of the population and know that of our sample, we can apply a set of factors to each 'observation' to deliver a more representative set of findings.

The Literary Digest had mailed out its poll to 10 million people, but these were all individuals whose name appeared in telephone directories or who were known to subscribe to magazines. Remember this was 1936, an era when access to a telephone was far from universal. Given that anyone without a telephone or who failed to subscribe to a magazine had no ability to participate the survey findings were only going to be accurate if the voting intentions of telephone owners were pretty much identical to those without telephone access – a classic case of coverage error.

Furthermore, while 2.4 million did respond to the survey it had been sent to ten million, meaning that the response rate was 24% (high by today's online standards). The Literary Digest made the heroic assumption that the 76% not responding were no different to those who did, when there was a very real risk of non-response error.

So when looking at tourism statistics that are based on a sample survey it is certainly helpful to have an idea of the sample size relative to the size of the population, but our willingness to trust the survey findings ought not be dominated by whether the sample size sounds large or small. The design and execution of the survey matter just as much, as does who it is that is sponsoring the survey and whether they are seeking to promote a particular viewpoint.

Traditionally one of the ways in which the accuracy of a sample survey can be judged is to report confidence intervals – that is to say a range of values so defined that there is a specified probability that the parameter we are trying to estimate lies within it.

It is comparatively rare for confidence intervals to be shown when the findings of a survey are being presented. There are common sense reasons for this being the case as most users will be looking for 'an estimate' rather than the range within which an estimate sits, especially as that range will vary according to the level of probability that we assign to the confidence interval.

It is however good practice to draw users' attention to the overall confidence interval within which estimates typically sit.

For the key metrics relating to inbound tourism to the UK the Office for National Statistics does publish 95% confidence intervals, which indicate that in 2019 the estimate for total visits had a confidence interval of +/-2.5%, for total visitor nights it was +/-3.4% and for visitor spending +/-3.9%.

The ONS goes as far as to publish spreadsheets on a quarterly basis showing 95% confidence intervals for a range of estimates, for example journey purpose, region of the UK visited and country of residence.

This data reveals just how cautious we need to be in reading too much into the figures. For example, in Q3 2019 the 95% confidence interval for the number of inbound Holiday visits to the UK was +/-6.5%, that for visits to Scotland was +/-12.4% and for visits to the UK from China was +/-17.0%.

Having sight of these confidence intervals is a useful tool in its own right but becomes particularly important when looking at year-on-year change. We might observe that the number of visits in the most recent quarter is 4% up on a year ago, but once we take account of the respective confidence intervals for the estimate this year and last we see that the confidence intervals overlap, meaning that the true trend might have been for a decline in the number of visits.

In summary the best advice is always to recognise that a survey will shine a light on the characteristics, attitudes and behaviours of a sample of the population, and however well designed that survey is its findings may not

accurately reflect the wider population. This can be a particular risk when results from a survey are grossed up to claim how many people in the population think or behave in a certain way, or in using results from a sample of businesses to gross up to entire sector.

6 Units of measurement

We have already referenced the importance of checking whether monetary data is being reported in constant or outturn prices, but the issue of tourism statistics being reported in units of measurement that can spread confusion does not end there.

Two destinations may report the number of inbound visits during a particular time period and initially we may think that it is straightforward to compare these two datapoints, and it might be. However, in some instances what is being reported is the number of overnight visits (visits that last at least 1 night) whereas other destinations may be including same-day visits in the total.

Some destinations have a preference to report visitor nights, so for example if that destination welcomed 100 visitors each of which stayed for 4 nights their total would be 400. When data is being reported in this form it is important to recognise that day-trip visitors (or excursionists to use the correct UNWTO terminology) are not included in the total.

This might be of particular concern when calculating average spend per night metrics, as if the total spend figure does include excursionists but we are excluding this group from the denominator then a calculated spend per night estimate will be higher than the true value for those who do stay for at least 1 night.

When reporting average length of stay figures it should be the case that we are only including overnight visitors in both the numerator and the denominator, as otherwise we are at risk of reporting an average length of stay, measured as nights per visit, that is lower than the reality.

Accommodation occupancy stats are another example of where caution is required in interpretation, both of snapshot figures and when considering trends over time.

As with so many tourism statistics what might at first appear to be an 'easy to understand' metric may have layers of nuance. Suppose that a destination reports its hotel occupancy to be 65% then the first question we ought to ask ourselves, and examine any footnotes that accompany the reported figure, is whether this is room occupancy or bedspace occupancy.

Imagine an hotel that has 100 rooms and 65 of these are occupied, then the room occupancy rate is indeed 65%. But if each room can accommodate two guests the our maximum bedspace capacity is 200. If 45 of our 65 occupied rooms have two guests but the other 20 rooms are being used by solo travellers then 110 (45x2 +20) guests are being accommodated, leading to a bedspace occupancy rate of 55% (110/200).

Once having resolved what exactly the occupancy rate is referring to we need to consider another important factor, whether or not we are calculating the occupancy rate as the number of guests (or occupied rooms) divided by the number of available rooms (or bedspaces) during the reference period under the spotlight.

The number of rooms available will vary depending on which hotels are open and whether any open establishments currently have rooms that are not being offered to the public due to refurbishment etc). If a significant part of the potential supply-side is unavailable we may misconstrue a high occupancy rate as indicating a strong performing market, whereas what might be happening is that constrained supply is artificially inflating our view of lacklustre demand.

This risk will be accentuated if a destination is gradually losing hotel rooms over time. If we just look at trends in hotel occupancy we may feel reassured that the situation is quite favourable, whereas the truth might be that demand is in retreat, only less slowly than supply, and as such the occupancy rate is disguising a tailing off of demand.

7 When to be wary

As this discussion paper has illustrated there is a lot to be mindful of when poring over tourism statistics. Diligent users should always check for footnotes or appendices that describe the methodological approach that was adopted, the sample size and headline confidence intervals for survey based estimates and what criteria was used for selecting who it was that got to answer the survey questions in the first place.

Great caution is equally sensible in comparing metrics for one destination with those of another, as differences may be as much to do with the precise method of data collation and presentation as it is with any real difference in performance between the two jurisdictions.

The availability of so many tourism statistics and surveys is a blessing rather than a curse, but the wise should always be inquisitive about the numbers to avoid being hoodwinked, accidentally or intentionally, into reaching an ill-judged interpretation.