

Secondary data analysis in tourism research

Alexander Bauer

Statistical Consulting Unit StaBLab, LMU Munich, Germany

Research seminar | March 28, 2019

Outline

1. Statistical principles
2. Sample study 1: Terrorism and tourism in Israel
3. Sample study 2: Travel constraints
4. Sample study 3: TourIST
5. Other stuff: Visualization & Spatial aggregation levels
6. Open space for discussion

Outline

1. Statistical principles
2. Sample study 1: Terrorism and tourism in Israel
3. Sample study 2: Travel constraints
4. Sample study 3: TourIST
5. Other stuff: Visualization & Spatial aggregation levels
6. Open space for discussion

① Statistical principles

Statistical analyses is an interplay between statistical methods and domain knowledge

⇒ A statistical model is most often no blackbox. Think about what information you use, what the statistical method aims for and if the findings fit your expert knowledge

Steps of a statistical analysis

1. Formulate your hypothesis / problem
(... and write it down!)
2. Figure out what data source is most appropriate
(cross-sectional vs longitudinal; individual vs aggregated level)
3. Choose an appropriate statistical method
4. Analyse and evaluate the method
5. Interpretat and properly communicate the results

① Statistical principles

Statistical analyses is an interplay between statistical methods and domain knowledge

⇒ A statistical model is most often no blackbox. Think about what information you use, what the statistical method aims for and if the findings fit your expert knowledge

Steps of a statistical analysis

1. Formulate your hypothesis / problem
(... and write it down!)
2. Figure out what data source is most appropriate
(cross-sectional vs longitudinal; individual vs aggregated level)
3. Choose an appropriate statistical method
4. Analyse and evaluate the method
5. Interpretat and properly communicate the results

① Statistical principles

Use **secondary data** wherever possible.

Oftentimes money is spent more sensibly on buying in existing data resources than on conducting a new study yourself.

Figure out your research questions to figure out appropriate data sources:

- What are your dependent, what are your independent variables?

Or is there no dependent variable?

- What's your observational unit?

Is it an individual tourist for studying his motivation?

Is it a month in which you observe the aggregated tourist count?

- What's your target population?

Regional constraints? Time constraints?

- It is necessary to account for control variables?

E.g., when analyzing the association of general travel intention and actual travel behavior, you should control for income etc.

① Statistical principles

Use **secondary data** wherever possible.

Oftentimes money is spent more sensibly on buying in existing data resources than on conducting a new study yourself.

Figure out your research questions to figure out appropriate data sources:

- What are your dependent, what are your independent variables?

Or is there no dependent variable?

- What's your observational unit?

Is it an individual tourist for studying his motivation?

Is it a month in which you observe the aggregated tourist count?

- What's your target population?

Regional constraints? Time constraints?

- It is necessary to account for control variables?

E.g., when analyzing the association of general travel intention and actual travel behavior, you should control for income etc.

Relevance vs significance

- Practitioners often aim for **statistically significant** results
(based on hypothesis tests of regression models)
 - Statistical significance as an established “objective” criterion if a finding can be generalized (e.g. to the whole population)
 - However, significance doesn't tell you if the finding is **substantially relevant**
- ⇒ Keep the focus on judging the relevance of some correlation / difference / effect in your domain.
- Significance is useful, but only of secondary interest!

Know your limits

Talking to a statistician is never bad...

- ... when formulating a research question
(and discussing its practicability)
- ... when figuring out an appropriate data source
- ... when setting up a questionnaire
(to prevent a survey that doesn't really fit your problem)
- ... when choosing an appropriate statistical model

Outline

1. Statistical principles
2. Sample study 1: Terrorism and tourism in Israel
3. Sample study 2: Travel constraints
4. Sample study 3: TourIST
5. Other stuff: Visualization & Spatial aggregation levels
6. Open space for discussion

Terrorism and tourism in Israel: Analysis of the temporal scale

Research questions

1. How big is the impact of terrorism on tourism in Israel?
2. Clarify the short-term and long-term effects of terrorism

Data base

- Combination of different data sources
- Monthly data from 2000 to 2014
- Dependent variable: Monthly number of tourist arrivals in Israel
- Independent variable: Monthly number of deaths caused by terrorism
- Control variables: economic situation, holidays, etc.

② Terrorism and tourism in Israel

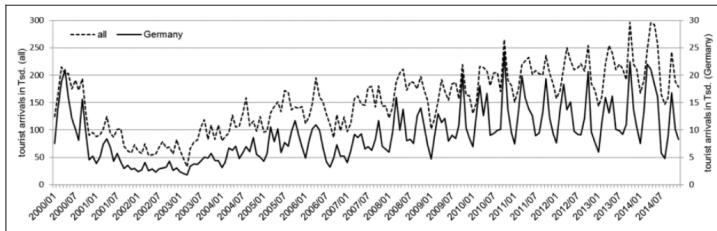


Figure 1. Tourism arrivals (total, Germany) in Israel (2000–2014) (based on CBS, 2000–2014). CBS: Central Bureau of Statistics Israel.

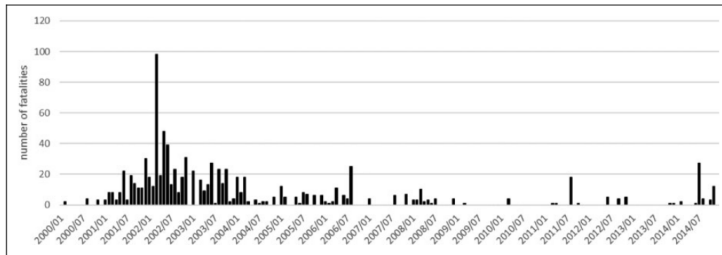


Figure 2. Terrorism in Israel (2000–2014).

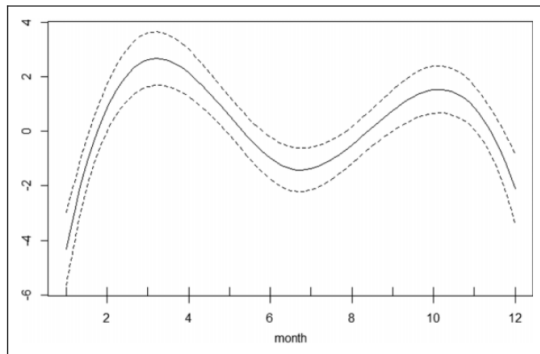
② Terrorism and tourism in Israel

Table 2. Results of research model with interaction effect.

Parameter	Estimate	p-value
Intercept	5.94	0.7456
High_Season	6.94	<0.0001***
Terr_death_log2_lag0	-0.34	0.0241*
Terr_death_log2_lag0 × High_Season	0.56	0.05
Terr_death_log2_lag1	-0.23	0.1350
Terr_death_log2_lag1 × High_Season	-0.24	0.38
Terr_death_log2_lag2to6months	-0.38	0.1310
Terr_death_log2_lag2to6months × High_Season	-1.71	<0.0001***
Long-term_Trend	0.42	<0.0001***
Easter	-0.23	0.7154

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

② Terrorism and tourism in Israel



② Terrorism and tourism in Israel

Statistical analysis

- Time series data
 - ⇒ Neighbored monthly observations are correlated
- Regression model with nonlinear effects
- Adjusting for natural seasonality of tourist arrivals
(both for general seasons and religious holidays)
- Instant and lagged effects of terrorism
- Evaluation of differences between high season and off-season

Outline

1. Statistical principles
2. Sample study 1: Terrorism and tourism in Israel
3. Sample study 2: Travel constraints
4. Sample study 3: TourIST
5. Other stuff: Visualization & Spatial aggregation levels
6. Open space for discussion

What stops the world travel champions from travelling?

The impact of travel constraints on German residents' travel frequencies

Research questions

1. How do travel constraints affect travel frequency?
2. How does travel motivation affect travel frequency?
3. Are effects different, comparing actual travel frequency to travel intention?

Data base

- representative individual level data for German residents
- 7 000 (face-to-face) to 12 000 (incl. online) respondents per year (1995–2019)
- information about travel intention, motivation and behavior

③ Travel constraints

Inhalte der Reiseanalyse

Reiseintensität

Reisemotive

Urlabsaktivitäten letzte 3 Jahre

Urlabsziele letzte 3 Jahre

Reiseabsichten

Informationssuche

Urlabsformenpotenziale

Soziodemografie inkl. Sinus-Milieus®



Reiseziel

Reiseausgaben

Reisedauer

Reiseart

Organisationsform

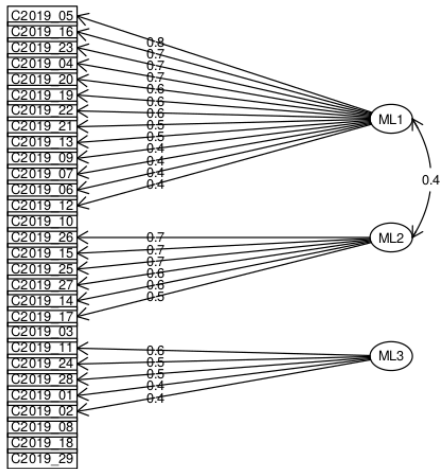
Unterkunft

Verkehrsmittel

... und vieles mehr!

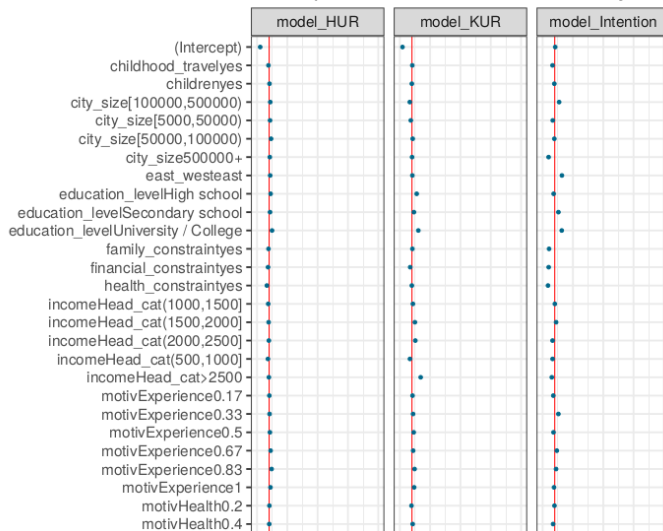
③ Travel constraints

Factor Analysis



③ Travel constraints

effects on exponential scale with restricted y axis



Statistical analysis

- Development of a wholistic model including travel motivations, constraints and sociodemography
- Generalized regression (binary and count data)
- Factor analysis
- Joint work of Geography and Statistics to ensure a suitable and interpretable model

Outline

1. Statistical principles
2. Sample study 1: Terrorism and tourism in Israel
3. Sample study 2: Travel constraints
4. Sample study 3: TourIST
5. Other stuff: Visualization & Spatial aggregation levels
6. Open space for discussion

Touristic behavior in a spatio-temporal context: Statistical analyses for the identification and development of behavior patterns

Research questions

1. Detection of typical tourist groups
2. Analysis of spatial differences and temporal development

Data base

Same as in sample study 2

Statistical analysis

- Cross-sectional data over many years
 - ⇒ No track over personal developments, only of overall developments in the population
- Cluster analysis
 - (purely based on tourist properties, excluding sociodemography)
- (Nonlinear) Regression with clusters as dependent variable
 - ⇒ "What traits are associated with ones tourist type"
 - and "How do the typical tourist types change over time?"

Outline

1. Statistical principles
2. Sample study 1: Terrorism and tourism in Israel
3. Sample study 2: Travel constraints
4. Sample study 3: TourIST
5. Other stuff: Visualization & Spatial aggregation levels
6. Open space for discussion

**Getting a good feeling for your data (based on descriptives)
is always the way to start a proper data analysis!**

Data visualization tools

- **SPSS** for basic stuff
- **Tableau** for basic stuff *plus* better maps (see [here](#))
and the possibility to create interactive graphics and dashboards

It's free for academics: <https://www.tableau.com/academic>

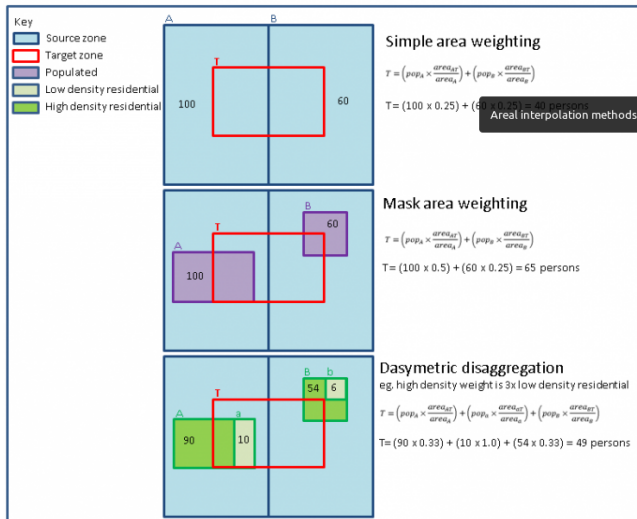
⑤ Spatial aggregation levels

General issue

- Combining data from different sources is often hard as of different spatial aggregation levels
- Solution to this problem is very application dependent
- Most common problem: Need to “downscale” highly aggregated data
⇒ Solution: **Spatial disaggregation**

5 Spatial aggregation levels

Spatial disaggregation



Source: http://www.integrated-assessment.eu/eu/guidebook/spatial_disaggregation.html

Outline

1. Statistical principles
2. Sample study 1: Terrorism and tourism in Israel
3. Sample study 2: Travel constraints
4. Sample study 3: TourIST
5. Other stuff: Visualization & Spatial aggregation levels
6. Open space for discussion

⑥ Open space

Insert questions here