Journal of Business and Economics, ISSN 2155-7950, USA
April 2022, Volume 13, No. 4, pp. 225-230
DOI: 10.15341/jbe(2155-7950)/04.13.2022/005

http://www.academicstar.us

Statistical Scrutiny of Mischaracterizations in Hospitality Research

Steven E. Kent
(School of Business, Molloy College, USA)

Abstract: When explaining industry trends in the hospitality sector many hotel investors and operators will use regression analysis, and correlations as evidenced by high R-squares to make the case for certain industry dynamics positively or negatively influencing industry trends. They also use subcategories of data rather than using broader information to come to industry wide conclusions. There are also academic studies within hospitality that also seem to data mine or draw insignificant conclusions. By reviewing some of these problems the quality of the research and relevance may improve.

Key words: validity, generalization, autocorrelation and statistical significance

JEL codes: A, M

In virtually every study in academia, business or research, the goal seems to be to make the analysis reliable, valid, and to use the insight to generalize to a broader population. But if researchers are truly altruistic then how can we have conflicting messages from similar studies? It seems that nearly every morning we wake up with one study that conflicts with another. From whether we should drink alcohol or coffee, eat low fat or high carbohydrate food or whether we should use a hotel's own website to get the best rates or access an online travel agent. One of the reasons for differences in finding is in fact the reliability, validity, and generalizability of the analysis from the start.

Reliability broadly means that the data or instrument being used is stable and consistent. The scores collected from a study conducted multiple and different times should be nearly the same. Furthermore, the scores should be consistent in that if an individual answers a question one way during a test they should answer a related question in a similar way when asked again (Creswell, 2015).

Validity is showing "sound" evidence that the interpretation of the test scores in fact measures the proposed use. In the past validity was sometimes measured and defined based on construct, criterion referenced and content. However, Creswell points out that "validity is seen now as a single unitary concept rather than three types" (Creswell, 2015, p. 158).

The concept of generalizability is that you can make predictions based on past observations. If an event or data has happened in the past then it will likely happen in the future. Conceptually, if a researcher were to collect enough data from the past and analyze it they should be able to support a hypothesis or premise to predict the outcome in the future, if all things are equal with some level of accuracy (August 4, 2017, Generalizability)[1]. However, this issue of generalizability, especially as it relates to social sciences or human reaction, may be the most difficult to feel confident in. For example, on Wall Street many investors will say, "The trend is my friend", meaning that if

Steven E. Kent, Ph.D., Assistant Professor, School of Business, Molloy College; research areas: economics, finance, hospitality and tourism. E-mail: stevenerickent@gmail.com & skent@molloy.edu.
[1] Accessed from https://www.iwh.on.ca/wrmb/generalizability.

they are seeing stocks or profits moving up in the past then it is likely to happen in the future. In fact, millions of dollars are spent every day trying to generalize historic data into the future of the markets, but because investors are humans and don't always act reliably, this generalization can be misleading and it leads to trading losses not gains.

Reliability, validity and generalizability are all linked, can overlap but can also be mutually exclusive. Validity might be viewed as overarching when you are trying to assess whether a certain instrument, data set or information is appropriate (Creswell, 2015). Reliability seems to be more closely aligned with whether the data or information is consistent. The challenge as it relates to reliability is that sometimes data shows greater variability so they do not always align. For example, hotel room rates can show significant variability day to day but does that make them less reliable? The researcher assessing reliability then needs to make a judgment that just because a room in New York City costs $400 on Tuesday night and $200 on a Sunday night that this data is still reliable because this relationship of higher prices on one day and lower prices on another is in fact consistent. As to generalizability, one of the biggest challenges in hospitality research is using historical data sets as the time-period of analysis may make it very hard to generalize. For example, if a researcher studied hotel trends from 2007 to 2009 (the period of the Great Recession) and then generalized these finding to forecast 2017 the results may not be valid or reliable as this period during the economic downturn was so unique.

One study that illustrates the challenges of reliability, validity and generalization, "Cultural Preferences in Hotel Guestroom Lighting Design" (Park, Pae, Joo & Meneely, 2010). This study analyzed the cultural preferences for lighting in hotel rooms. Specifically, the researchers wanted to see if Americans and South Koreans had differing views towards light color and intensities. They projected four different room designs and then asked 87 North American students and 88 South Korean students to select their favorites. The result was the North Americans preferred warm/dim lighting while Koreans preferred warm/bright lighting. The authors showed that there was statistical significance between the two groups preferences.

While this study may have been helpful in that illustrates, that hoteliers need to examine "the one size fits all" and "cookie cutter" mentality for their properties. It had several issues as it relates to reliability, validity and to generalizability. For example, they only tested this group of students once so it is not clear how truly reliable these results are. Maybe all the students studied during that time were out partying the night before and would have a different view about lighting if they had been in the library during the evenings?

As to validity does the results from an analysis students' impressions really illustrate the proposed use? The authors likely used students out of convenience/availability to assess room lighting, but if the core goal was to assess how consumer feel, it would have made more sense to test the rooms with business travelers who are the target customer. Finally, it seems hard to generalize a finding based on less than 100 students in each region.

The notion of statistical significance being viewed as arbitrary is not usually noted by the layman as the reader of a study must usually dig deeper into the method and results rather than reading the headlines or the summary to find these values. When a researcher says, something is statistically significant it is implied that the findings are not due to chance. In statistics, it means that if the null hypothesis is true than there would be a low probability that the findings in the study would be so large or substantial (Sauro, 2014). Another way of stating this concept is "A result is said to be significant if it is very unlikely to occur when the null hypothesis is true. That is the result is sufficient to reject the null hypothesis" (Gravetter & Wallnau, 2013, p. 251)

Beyond the issue of sampling error, which would indicate that the researcher has simply examined the wrong subset or there maybe random noise or fluctuations, the other issue concerning statistical significance is that this is still probability not certainty. When somebody bases a decision based on statistics they are inherently saying that

they can live with a 10%, 5% 1%, etc. chance that the decision is wrong because it is virtually impossible to guarantee 100% certainty based on a sample or subset (Sauro, 2014)

The reader needs to decide whether they feel a 5% risk (1 out of 20) or a 1% risk (1 out of a 100) is worth making the decision based on the analysis. So, while it is true a statistician can arbitrarily pick a level of statistical significance, the practitioner may not feel as comfortable. Furthermore, the willingness to accept the arbitrary statistical significance may depend on the circumstances. For example, a person in need of hydration in a desert may be willing to live with 10% statistical significance (1 out of 10) that the water they are about to drink is contaminated. However, a patient undergoing experimental elective heart surgery may not be willing to proceed with a 1% statistical significance of procedural success (1 out of 100).

Additional information beyond the statistical significance is also important to assessing whether an analysis is robust and not arbitrary. For example, statistical significance does not give any insight into treatment effect size, magnitude of change and the direction of results (Page, 2014). Furthermore, the statistical significance does show you whether the results were influenced by the number or variability of the subjects tested or magnitude of the effect. Finally, it may not demonstrate whether results achieve statistically significant differences is influenced by factors such as the number and variability of subjects, as well as the magnitude of effect. Most importantly, most inferential statistical analyses assume that the sample being tested is representative of the population that is normally distributed (Page, 2014)

An example of additional information or support to build their case that their study is not only statistically significant but also demonstrates impact researchers should provide the effect size or magnitude of change. For example, a medical study may be able to show that one drug versus another drug results in a statistically significant difference in cancer treatments. However, while the difference maybe significant between the two drugs, it may only extend life of a Stage 4 cancer patient *on average* by two weeks which may not be impactful enough for a consumer to purchase a drug. Similarly if the variance of the cancer patients in the study is extraordinary wide, so some outlier patients live for three years while other have no effect this may show that the possibility of a significant outcome is compelling.

An example of a paper in hospitality literature that illustrates how the writers use statistical significance, that may need to be looked at more critically, is “How Fast Do New Hotels Ramp Up Performance?” (Enz, Peiro-Signes, & Segarra-Ona, 2013). The authors used statistical analysis to see how fast hotels ramp up their occupancy, average daily rates and RevPAR (revenue per available room) relative a group of peer hotels. The authors initial hypothesis was that while many startup businesses are risky the propensity for failures in hotels is relatively modest. They prove this by showing that with 1.75 years’ new hotels get to occupancy levels similar to the peer group. They achieve similar revenue per available room 2.5 years after opening compared to the peer set and that branded hotels achieve comparable RevPAR by 2.25 years while unbranded hotels took longer to ramp.

The authors use a data set of nearly 5,000 hotels and compare it to the broader industry. By using such a large subset of hotels in the sample the statistical significance is likely to be high. This is core to the study or power in statistics as Gravetter and Wallnau, write “The power of a statistical test is the probability that the test will correctly reject a false null hypothesis. That is, power is the probability that the test will identify a treatment effect of one really exists” (p. 265). They go on to note “One factor that has a huge influence on power is the size of the sample” (p. 268) and indicate if researchers want to increase the power they should increase sample size.

One of the criticisms of the analysis by Enz et al. (2015) is that they make broad conclusions but do not (1) specifically state how they got to such statistically significant results, and (2) they seem to “bury” the statistical

significance in the tables as seen in Figure 1. If the analysis was so robust and statistically significant they would trumpet it more loudly throughout the paper, but instead it can only be observed in the tables.

It is my supposition that while the analysis did show statistical significance there were several assumptions and techniques that require further analysis. First, was the basic postulation that a hotel would want to achieve similar occupancy to its peers. Nearly all hotels want to do the opposite; choosing to ramp up more slowly (soft openings, reducing initial occupancy) so their staff can get up to speed and they don't disappoint customers. The initial weak performance might be by design, not a function of operating results.

Second, the authors conclude with the finding that "that brand hotels, and particularly brand-managed new hotels, were able to produce market-comparable occupancy and RevPAR rates two years faster than new independents, hotels" (Enz et al., 2013, p. 148). This would suggest that brands are better than independents but having a brand costs money and the authors do not address profits but rather focus on revenues. A concern is that because the statistical significance seems so high, in part due to the sample size, it may result in misleading validity and conclusions.

Exhibit 2:
New Entrant Performance Ramp-up Compared with Incumbents in the First Two Years.

		Occupancy		ADR		RevPAR	
Time Period	Sample Size	Median (Wilcoxon Signed Test Z-Statistic)	% Positive (Binomial Sign Test Z-Statistic)	Median (Wilcoxon Signed Test Z-Statistic)	% Positive (Binomial Sign Test Z-Statistic)	Median (Wilcoxon Signed Test Z-Statistic)	% Positive (Binomial Sign Test Z-statistic)
Opening year							
1st quarter	3,127	−31.47 (−46.95)***	5.4 (−49.86)***	2.75 (−15.7)***	62.65 (−14.13)***	−28.44 (−46.66)***	5.44 (−49.82)***
2nd quarter	3,472	−14.42 (−40.18)***	17.83 (−37.9)***	1.67 (−16.1)***	61.46 (−13.49)***	−13.01 (−39.59)***	18.58 (−37.01)***
3rd quarter	3,472	−7.83 (−27.2)***	29.81 (−23.8)***	0.97 (−11.69)***	57.46 (−8.78)***	−6.96 (-25.84)***	30.99 (−22.41)***
4th quarter	3,487	−4.46 (−17.46)***	37.05 (−15.27)***	1.17 (−10.06)***	57.64 (−9.01)***	−3.55 (−14.87)***	38.83 (−13.18)***
Second year							
1st quarter	3,489	−1.96 (−8.56)***	44.08 (−6.98)***	1.34 (−11.15)***	57.52 (−8.87)***	−1.41 (-5.38)***	46.26 (−4.4)***
2nd quarter	3,486	−0.42 (−2.42)*	48.97 (−1.2)	1.9 (−13.43)***	59.32 (−10.99)***	0.26 (−1.22)	50.69 (−0.8)
3rd quarter	3,482	0.33 (−1.48)	50.86 (−1)	2.02 (−13.78)***	59.85 (−11.61)***	1.04 (−4.82)***	53.22 (−3.78)***
4th quarter	3,483	1.37 (−5.65)***	53.95 (−4.64)***	2.25 (−14.66)***	60.29 (−12.13)***	2.49 (−9.36)***	55.96 (−7.01)***

Note. ADR = average daily rate; RevPAR = revenue per available room.
*Significantly different from zero (50% in the case of percent positive) at the .05 level for the two-tailed test.
**Significantly different from zero (50% in the case of percent positive) at the .01 level for the two-tailed test.
***Significantly different from zero (50% in the case of percent positive) at the .001 level for the two-tailed test.

Figure 1 New Entrant Performance Ramp Up Compared with Incumbents in the First Two Years

One of the mistakes in research is interpreting that correlation implies a cause and effect between two variables. There are lots of reports of relationships between one variable and another variable but this does not mean causation. For example, between the years 1860 and 1940, as the number of Methodist ministers living in New England increased and so did the importation of rum. Causation would suggest that ministers moving to New England were bringing caseloads of liquor with them but in fact it was that the population was growing in the region (Borwein & Rose, 2017). Just because one variable moves in lock step with another variable, indicating high correlation, does not necessarily mean that it suggests causation. Gravetter and Wallnau indicate that "to establish a cause and effect relationship it is necessary to conduct a true experiment in which one variable is manipulated by a researcher and other variables are rigorously controlled" (p. 522).

However, it is not always possible to create such a controlled experiment, especially in social sciences. Nonetheless, when reading hospitality literature, it sometimes appears that the researchers are coming to conclusions that are based on correlation or causation. The findings may have more to do with goodness of fit, rather than exogenous factors that may be influencing the findings.

For example, in the article "Fear and Fear and Managing in Las Vegas" (Eisendrath, Bernhard, Lucas & Murphy, 2008) the authors looked at the rapid recovery of Las Vegas gaming trends after the 9/11 attacks. Specifically, they noted that slot revenues (used as a proxy for gaming trends) declined for only 5 months and then showed a rapid recovery. The authors looked at other terrorist driven exogenous events and referenced work that classified these types of attacks as permanent and abrupt, temporary and abrupt, permanent and gradual and temporary and gradual. Using 179 data points (monthly slot coin in from January 1990 to November 2004) they adjusted for the upward trend, seasonality and a surge in capacity, the predictions they were still able to show "a good fit". From Feb 2002 to February 2004 the slot trends came in toward the bottom end of predicted range but they did not stay depressed, as operators had predicted.

The problem with this analysis is that it showed correlation between time from the exogenous events and slot win, suggesting that Las Vegas had recovered in a very short period. They make the statement, "that the terrorist-related research that we reviewed indicated that once terrorist activity stops tourism and travel markets tend to recover relatively quickly from their initial rapid falloff in volume. These findings appear to be consistent with our findings in Las Vegas in relation to the 9/11 attacks" (Eisendrath et al., 2008, p. 158)

Their analysis was based explicitly on slot player activity, suggesting that the correlation between slot revenues and time, were an effective measure of the recovery in Las Vegas. They also suggest that recoveries can be much faster than operators predicted (Eisendrath et al., 2008, p. 159). However, there are other indicators that also should have been evaluated, such as visitor counts, hotel occupancy, total gaming revenue to assess whether this recovery was as robust as the correlation suggested.

In fact, slots revenues were likely the wrong variable to look at as gaming revenues as a percentage of total revenue have declined from 59% in 1984 to 34% in 2016 with the biggest shifts in this spending pattern occurring before 2001 (Nevada Casinos: Departmental Revenues, 1984-2016, n.d.). Looking back overall spending took much longer to come back, than just the slot spending metric. In fact, hotel had to lower prices on rooms, food and beverage to get consumer to come back to Las Vegas after 9/11.

An example where the author focused more on causation rather than correlation was, "Is the gaming industry still recession-proof?" (Zheng T., Farrish J., Lee M., & Yu H., 2013). The author examined Iowa gaming trends during the recession of 2007 to 2009 and found that the Iowa gaming revenues during the recession were not dramatically impacted by the economic slowdown as, "the results of analyses show that: slot coin-in was not affected by the recession; table drop was slightly affected, but started to recover in late 2010; and monthly admission was not affected by the recession, and showed a significant increase after the recession" (Zheng et al., p. 1135).

The authors suggest that Iowa was insulated from the recession and there was no causality. However, the authors could have examined the supply side of the gaming equation as during the period scrutinized there were very few new casinos opened or expanded. The authors do note at the end of the paper, that Iowa did not face the same negative economic drivers (mortgage delinquency, unemployment etc.) of the 2007-2009 recession as the rest of the U. S. In contrast, other gaming markets saw weakness because of overbuilding. The other gaming markets also saw the negative impact of lower economic activity. So, while Iowa may have been insulated from the downturn relative to other markets, it may be to strong of a statement to say that gaming is recession proof.

References

Borwein J. and Rose M (2017). "Clearing up confusion between correlation and causation", available online at: http://theconversation.com/clearing-up-confusion-between-correlation-and-causation-30761.

Creswell J. W. (2015). *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research*, Pearson.

Eisendrath D., Bernhard B. J., Lucas A. F. and Murphy D. J. (2008). "Fear and managing in Las Vegas", *Cornell Hospitality Quarterly,* Vol. 49, No. 2, pp. 145-162.

Enz C. A., Peira-Signes A. and Segarra-Om (2013). "How fast do new hotels ramp up performance?", *Cornell Hospitality Quarterly*, Vol. 55, No. 2, pp. 141-151.

Gravetter F. J. and Wallnau L. B. (2013). *Statistics for the Behavioral Sciences*, Boston, MA: Cengage Learning.

Institute for Work & Health. (n.d.). Available online at: https://www.iwh.on.ca/wrmb/generalizability.

Jani D. and Han H. (2015). "Influence of environmental stimuli on hotel customer emotional loyalty response: Testing the moderating effect of the big five personality factors", *International Journal of Hospitality Management*, Vol. 44, pp. 48-57.

Nevada Casinos: Departmental Revenues, 1984-2016. (n.d.). Available online at: http://gaming.unlv.edu/reports.

Page P. (2014, October). "Beyond statistical significance: Clinical interpretation of rehabilitation research literature", available online at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4197528/.

Park Nam-Kyu, Pae Joo and Meneely Jason. (2010). "Cultural preferences in hotel guestroom lighting design", *Journal of Interior Design*, doi: 36.10.1111/j.1939-1668.2010.01046.x.

Sauro J. (2014, October 21). "What does statistically significant mean?", available online at: https://measuringu.com/statistically-significant/.

Zheng T., Farrish J., Lee M., and Yu H. (2013). "Is the gaming industry still recession-proof?", *International Journal of Contemporary Hospitality Management*, Vol. 25, No. 7, pp. 1135-1152.