# Exploring Relationships using SPSS inferential statistics (Part II)

Dwayne Devonish

# Reminder: Types of Variables

- Categorical Variables
- ✓ Based on qualitative type variables.
- ✓ Gender, Ethnicity, religious affiliation, occupation.
- ✓ Age measured as categories: What is your age: 18-26 years, 27- 35 years, Over 35 years.

- Quantitative/numerical (numerical) variables
- ✓ Based on data in the form of numbers
- ✓ Age in years (What is your age _____?)
- ✓ Tourist expenditure ($)

We dealt with inferential tests used to examine differences between groups…

Now we look at inferential tests used to examine relationships between variables.

# Inferential statistics: Examining Relationships between Variables

- Inferential statistics which examine relationships or associations between two or more variables answer research questions that are associational:

✓ Is there a relationship between number of tourist arrivals and crime rates?

✓ Is there a relationship between destination satisfaction and total expenditure?

# Inferential Statistics: Examining Relationships

- Popular inferential statistics examining relationships between variables include:

➢ Correlations

➢ Chi-square test

➢ Simple linear regression and Multiple Regression

# Pearson Chi-square test: Test of Association

# Pearson Chi-square Test

- The Pearson Chi square test is used to test whether a statistically significant relationship exists between two categorical variables (e.g. gender and type of car). It accompanies a crosstabulation between the two variables.
- Categorical independent and dependent variable needed

# Research question for Chi-square

- Is there a relationship between gender and voting intentions among employees?

✓ There are two variables: gender and voting intentions.

✓ Gender is the independent variable, and voting intentions is the dependent variable.

✓ Gender is categorical, and voting intentions (whether I vote for DLP, BLP, NDP) is also categorical.

✓ Pearson Chi-square appropriate as criteria are met.

# NOTES

- Ask these questions:

- How many variables?
- Which one is the independent variable, and which one is the dependent variable?
- What types of variables are they?
- So Chi-square appropriate?

# Chi-square test in SPSS

- Let's look at the tourism data3, and answer this question:
- Is there a relationship between the nationality of the tourist and the likelihood to return to Barbados or Does the likelihood of returning to Barbados vary by nationality?
- 2 variables – nationality and likelihood of return (revisit intentions)
- Independent variable – nationality
- Dependent variable – revisit intentions
- Nationality is categorical (U.S and Canada), and revisit intentions are categorical (yes, no, don't know).

# Chi-square Output

**Crosstab**

| | | | Nationality | | |
|---|---|---|---|---|---|
| | | | U.S | Canadian | Total |
| Would you return to Barbados for a holiday? | Yes | Count | 38 | 21 | 59 |
| | | % within Nationality | 65.5% | 36.8% | 51.3% |
| | No | Count | 20 | 19 | 39 |
| | | % within Nationality | 34.5% | 33.3% | 33.9% |
| | Don't know | Count | 0 | 17 | 17 |
| | | % within Nationality | .0% | 29.8% | 14.8% |
| Total | | Count | 58 | 57 | 115 |
| | | % within Nationality | 100.0% | 100.0% | 100.0% |

**Chi-Square Tests**

| | Value | df | Asymp. Sig. (2-sided) |
|---|---|---|---|
| Pearson Chi-Square | 21.917[a] | 2 | .000 |
| Likelihood Ratio | 28.552 | 2 | .000 |
| Linear-by-Linear Association | 18.489 | 1 | .000 |
| N of Valid Cases | 115 | | |

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 8.43.

# Sample Write-up

- Is there a relationship between nationality and revisit intentions to Barbados (likelihood of return)? -- Results can be written up as:
- “A Pearson chi-square test was conducted to examine whether there was a relationship between nationality and revisit intentions. The results revealed that there was a significant relationship between the two variables (Chi square value = 21.92, df =2, $p < .001$). A significantly larger proportion of U.S tourists (66%) reported that they would return to the destination for a holiday visit compared with only 37 percent of Canadian tourists (see Tables 1 and 2).”

# More Chi-square tests in SPSS

- Let's look again at the tourism data3, and answer this question:

➢ Is there a relationship between the nationality of the tourist and the likelihood to recommend the destination (word-of-mouth intentions) or Do word-of-mouth intentions vary by nationality?

➢ 2 variables – nationality and likelihood of recommending to friends/family (word of mouth intentions)

➢ Independent variable – nationality

➢ Dependent variable – word-of-mouth intentions

➢ Nationality is categorical (U.S and Canada), and intentions are categorical (yes, no, don't know).

# Chi-square Output 2

**Crosstab**

| | | | Nationality | | |
|---|---|---|---|---|---|
| | | | U.S | Canadian | Total |
| Would recommend this destination to your friends and family? | Yes | Count | 36 | 19 | 55 |
| | | % within Nationality | 62.1% | 33.3% | 47.8% |
| | No | Count | 22 | 38 | 60 |
| | | % within Nationality | 37.9% | 66.7% | 52.2% |
| Total | | Count | 58 | 57 | 115 |
| | | % within Nationality | 100.0% | 100.0% | 100.0% |

**Chi-Square Tests**

| | Value | df | Asymp. Sig. (2-sided) | Exact Sig. (2-sided) | Exact Sig. (1-sided) |
|---|---|---|---|---|---|
| Pearson Chi-Square | 9.513[b] | 1 | .002 | | |
| Continuity Correction [a] | 8.396 | 1 | .004 | | |
| Likelihood Ratio | 9.652 | 1 | .002 | | |
| Fisher's Exact Test | | | | .003 | .002 |
| Linear-by-Linear Association | 9.431 | 1 | .002 | | |
| N of Valid Cases | 115 | | | | |

a. Computed only for a 2x2 table

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 27. 26.

# Sample Write-up 2

- Is there a relationship between nationality and word-of –mouth intentions? -- Results can be written up as:
- “A Pearson chi-square test was conducted to examine whether there was a relationship between nationality and word-of-mouth intentions. The results revealed that there was a significant relationship between the two variables (Chi square value = 9.51, df =1, p = .002) U.S tourists (62%) were significantly more likely to recommend the destination to their friends/family moreso than Canadian tourists (33%) (see Tables 1 and 2).

# Do a chi-square analysis now

- Using tourism data3, answer the following questions:

➢ Is there a relationship between nationality and likelihood of returning to the hotel of stay?

# Correlations

# Correlations

- Correlation tests (Pearson correlation) are used to examine relationships between two or more quantitative/numerical variables.
- They measure the strength and direction of a relationship between variables.

# NOTES

- Ask these questions:

- How many variables?
- Which one is the independent variable, and which one is the dependent variable?
- What types of variables are they?
- So Correlation appropriate?

# Pearson Correlations

- The Pearson correlation tells you the strength and direction of a relationship between two quantitative/numerical variables. It ranges from negative (-1) to positive (+1) coefficient values.

➢ Direction:

- A negative correlation indicates that high values on one variable are associated with low values on the next. A positive correlation indicates that high values on the one variable are associated with high values of the next.

# Pearson Correlations II

- A positive correlation between height and age means that higher values on the height variable (taller persons) are associated with higher values on the age variable (older persons).
- A negative correlation means that higher values on the height variable (taller persons) are associated with lower values on age variable (younger persons).
- The p-values tells you whether the relationship or correlation between the variables are statistically significant ($p < .05$).

# Pearson Correlations III

- Strength:
- Correlations have different strengths:
- ✓ .10 to .29 = Weak correlation/relationship
- ✓ .30 to .49 = Moderate relationship/Medium correlation
- ✓ .50 and above = Strong relationship/high correlation.
- ✓ The sign of the relationship does not indicate the strength; (-).50 is the same strength as (+).50 but different direction.
- ✓ 'r' is the symbol of the correlation coefficient.

# Correlation Analysis in SPSS

- Let's look at the tourism data1, let's answer the following question:

- Is there a relationship between satisfaction with prices at the destination and shopping expenditure?
- Two variables: satisfaction with prices and shopping expenditure.
- Independent variable: satisfaction with prices
- Dependent variable: shopping expenditure
- Both variables are quantitative/numerical.

# Table 1 Correlation between Satisfaction and Expenditure

**Correlations**

| | | Satisfaction with Prices in Destination | Shopping Expenditure per person per night |
|---|---|---|---|
| Satisfaction with Prices in Destination | Pearson Correlation | 1 | .169* |
| | Sig. (2-tailed) | | .025 |
| | N | 322 | 176 |
| Shopping Expenditure per person per night | Pearson Correlation | .169* | 1 |
| | Sig. (2-tailed) | .025 | |
| | N | 176 | 176 |

*. Correlation is significant at the 0.05 level (2-tailed).

# Sample Write-up

- Is there a relationship between satisfaction with prices and shopping expenditure? – Sample write-up:
- "A Pearson correlation analysis was conducted to examine whether there is a relationship between satisfaction with prices at the destination and shopping expenditure. The results revealed a significant and positive relationship ($r = .17$, $N = 176$, $p = .03$). The correlation was weak in strength. Higher levels of satisfaction with prices were associated with higher levels of shopping expenditure (see Table 1)."

# Correlation Analysis in SPSS

- Let's look at the tourism data1, let's answer the following question:

➢ Is there a relationship between the number of accompanying children and total expenditure?

➢ Two variables: number of children and total expenditure.

➢ Independent variable: number of children

➢ Dependent variable: total expenditure

➢ Both variables are quantitative/numerical.

# Table 1 Correlation between Children and Total Expenditure

**Correlations**

| | | How many children do you have? | Total Expenditure per person per night |
|---|---|---|---|
| How many children do you have? | Pearson Correlation | 1 | .186* |
| | Sig. (2-tailed) | | .020 |
| | N | 322 | 156 |
| Total Expenditure per person per night | Pearson Correlation | .186* | 1 |
| | Sig. (2-tailed) | .020 | |
| | N | 156 | 156 |

*. Correlation is significant at the 0.05 level (2-tailed).

# Sample Write-up

- Is there a relationship between number of children and total expenditure? – Sample write-up:
- "A Pearson correlation analysis was conducted to examine whether there is a relationship between number of children and total expenditure. The results revealed a significant and positive relationship ($r = .19$, $N = 156$, $p = .02$). The correlation was weak in strength. Higher numbers of children were associated with higher levels of total expenditure (see Table 1)."

# Do this correlation:

- Using tourism data1, answer the following:

✓ Is there a relationship between the number of visits to the destination and overall satisfaction.

# Regression Analyses

# Regression

- Regression analyses are used to examine the effect of different (predictor/independent) variables on a single outcome (dependent) variable.
- The use of the term “prediction” is central to regression analyses. One examines whether if one variable predicts (explains/impacts) another variable.

# Regression

- The independent or predictor variables must be either dichotomous (categorical variable with only 2 categories/groups) or quantitative/numerical variables.
- The dependent variable must be a quantitative/numerical variable.
- Simple Linear regression examines the relationship between one predictor variable and one outcome variable. Produces the same results as a bivariate Pearson correlation.
- Let's see a previous example on tourism data1, - the number of children and total expenditure example.

# Multiple Regression

- Multiple regression is a more popular extension of linear regression.
- Multiple regression examines the effects of the multiple predictors or independent variables on a single outcome variable.
- Same output as the linear regression analysis, and interpretation.

# Multiple Regression in SPSS

- Let's look at tourism data1.sav:

➢ Do nationality, satisfaction with restaurants, and satisfaction with prices in the destination predict (impact) overall destination satisfaction

➢ There are three general tables that must be interpreted in the write-up of the regression analysis.

# First Table

**Model Summary**

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|---|---|---|---|---|
| 1 | .365[a] | .133 | .121 | .902 |

a. Predictors: (Constant), Satisfaction with Restaurants, Nationality of Tourist, Satisfaction with Prices in Destination

The information that needs to taken from this table is the R-square (.133). The R-square is the proportion of variation in the dependent variable (overall satisfaction) that is explained by the three independent variables. It is expressed as a percentage. So 13.3 percent of the variation in overall satisfaction can be explained by three independent variables in the model.

# Second Table

**ANOVA[b]**

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| 1 | Regression | 27.068 | 3 | 9.023 | 11.092 | .000[a] |
| | Residual | 175.709 | 216 | .813 | | |
| | Total | 202.777 | 219 | | | |

a. Predictors: (Constant), Satisfaction with Restaurants, Nationality of Tourist, Satisfaction with Prices in Destination

b. Dependent Variable: Overall satisfaction with the destination

The table shows whether the proportion of variance explained in the first table is significant. It also tells whether the overall effect of the three independent variables on overall satisfaction is significant. The sig. (or p-value) is .000 which is below the .05 level; hence, we conclude that the overall model is statistically significant, or that the variables have a significant combined effect on the dependent variable.

# Third Table: Look at the effects of individual independents.

**Coefficients [a]**

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | 1.898 | .354 | | 5.355 | .000 |
| | Nationality of Tourist | .533 | .131 | .259 | 4.052 | .000 |
| | Satisfaction with Prices in Destination | .008 | .039 | .013 | .204 | .838 |
| | Satisfaction with Restaurants | .273 | .068 | .253 | 3.987 | .000 |

a. Dependent Variable: Overall satisfaction with the destination

Look at the sig. (*p*-values) first. We can see that nationality (.000) and satisfaction with restaurants (.000) are significant predictors (or significantly related to) of overall satisfaction. The standardised beta tell us the strength and direction of the relationships (interpreted like correlation coefficients). Satisfaction with restaurants is positively related to overall satisfaction (.26). High levels of this satisfaction correspond to higher overall satisfaction.

# Third Table: Look at the effects of individual independents.

**Coefficients [a]**

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | 1.898 | .354 | | 5.355 | .000 |
| | Nationality of Tourist | .533 | .131 | .259 | 4.052 | .000 |
| | Satisfaction with Prices in Destination | .008 | .039 | .013 | .204 | .838 |
| | Satisfaction with Restaurants | .273 | .068 | .253 | 3.987 | .000 |

a. Dependent Variable: Overall satisfaction with the destination

Nationality is a dichotomous variable where 1 = European, and 2 = U.S. The positive coefficient (correlation) for nationality suggests that high value on this variable (which is 2 = U.S) corresponds to higher scores on the dependent variable (i.e. high levels of overall satisfaction). It is interpreted as U.S tourists (=2) reported higher levels of overall satisfaction compared with European tourists (=1).

# Sample Write-up

- “Multiple regression was conducted to examine whether nationality, satisfaction with restaurants, and satisfaction with prices impact on overall satisfaction with destination. The overall model explained 13.3 percent of variance in productivity, which was revealed to be statistically significant, $F(3,216) = 11.09$, $p < .001$. An inspection of individual predictors revealed that satisfaction with restaurants (Beta = .25, $p < .001$) and nationality (Beta = .26, $p < .001$) are significant predictors of overall satisfaction with destination. Higher levels of satisfaction with restaurants are associated with higher levels of overall satisfaction with the destination, and U.S tourists reported significantly more overall satisfaction than did European tourists.

# Multiple Regression in SPSS (2)

- Let's look at tourism data1:

➢ Do satisfaction with accommodation, and with local transport, and number of children affect overall satisfaction?

➢ There are three general tables that must be interpreted in the write-up of the regression analysis.

# First Table

**Model Summary**

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|---|---|---|---|---|
| 1 | .538[a] | .289 | .277 | .846 |

a. Predictors: (Constant), How many children do you have?, Satisfaction with Accommodation, Satisfaction with Local transport

The R-square is the proportion of variation in the dependent variable (overall satisfaction) that is explained by the three independent variables. It is expressed as a percentage. So 28.9 percent of the variation in overall satisfaction can be explained by three independent variables (no. of children, satisfaction with restaurants, and with transport) in the model.

# Second Table

**ANOVA[b]**

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| 1 | Regression | 52.655 | 3 | 17.552 | 24.551 | .000[a] |
| | Residual | 129.399 | 181 | .715 | | |
| | Total | 182.054 | 184 | | | |

a. Predictors: (Constant), How many children do you have?, Satisfaction with Accommodation, Satisfaction with Local transport

b. Dependent Variable: Overall satisfaction with the destination

As stated before, this table shows that the overall model explains a significant proportion of variance (see Table 1 on prior slide), or that the overall model is statistically significant – all three independent variables have a significant combined effect on overall satisfaction, $F(3, 181) = 24.55$, $p < .001$.

# Third Table cont'd

**Coefficients[a]**

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | 1.401 | .304 | | 4.604 | .000 |
| | Satisfaction with Local transport | .242 | .062 | .289 | 3.890 | .000 |
| | Satisfaction with Accommodation | .387 | .089 | .321 | 4.336 | .000 |
| | How many children do you have? | .116 | .074 | .099 | 1.579 | .116 |

a. Dependent Variable: Overall satisfaction with the destination

The table also shows that satisfaction with local transport (Beta = .289, $p < .001$) and satisfaction with accommodation (Beta = .32, $p < .001$) were positively correlated with overall satisfaction, suggesting that higher levels of satisfaction with the two categories are associated with higher levels of overall satisfaction. Number of children is not a significant predictor of overall satisfaction ($p = .12$).

# Sample Write-up

- “Multiple regression was conducted to examine whether satisfaction with local transport, satisfaction with accommodation, number of children and impact on satisfaction. The overall model explained 28.9 percent of variance in overall satisfaction, which was revealed to be statistically significant, $F(3,181) = 24.55$, $p < .001$. An inspection of individual predictors revealed that satisfaction with local transport (Beta = .29, $p < .001$) and satisfaction with accommodation (Beta = .32, $p < .001$) are significant predictors of overall satisfaction. Higher satisfaction with accommodation, and with local transport were associated with higher levels of overall satisfaction.”

# Do this example:

- Using tourism data1, answer the following,
- Do satisfaction with accommodation, number of children, and overall satisfaction with destination impact on total expenditure?