التحليل الإحصائي المتقدم
باستخدام SPSS

اعداد

الأستاذ الدكتور مهند العكش

استشاري التحليل الإحصائي التطبيقي

Ph.D., Louisiana State University-USA

Master of Applied Statistics, Louisiana State University-USA

ANOVA with Repeated Measures using SPSS Statistics

- Repeated measures are used for the measurement of exactly the <u>same</u> response variable on exactly the <u>same</u> piece of experimental material in time or space.

Application Example:

- In a memory recall experiment.

  (Three methods of printing material were randomized to 9 subjects. Memory recall was recorded over 4 days).

- FILE NAME "memory recall".

WHAT TO REPORT:

- A repeated measures ANOVA with a Greenhouse-Geisser correction determined that mean memory recall differed statistically significantly between time points ($F_{(3, 18)} = 23.794$, $p < 0.0001$). Post hoc tests using the Bonferroni correction revealed that there was a slight non-significant reduction in memory recall from day1 to day2 ($p=1.0$). However, memory recall had been reduced to 10.1078 ± 1.5391, which was statistically significantly different to day2 ($p < .006$).

Practice Example:

- In a pest control experiment and as target pests live on the bottom of plant leaves, three types of pest control sprayers were tested. The effect of sprayers as a function of location in 3 heights was evaluated using 4 replications? **File Name "Pest Control"**

Chi-Square Goodness-of-Fit Test in SPSS Statistics

- It is used to determine whether the distribution of cases in a single categorical variable follows a known or hypothesized distribution.

Application Example

- In a car makers data, researchers want to test the ratio of car origins used in USA (American origin, European, or Japanese). Data File:" Cars.sav"

WHAT TO REPORT:

- The table above, **Test Statistics**, provides the actual result of the chi-square goodness-of-fit test. We can see from this table that our test statistic is statistically significant: $\chi^2(2) = 154.8$, $p < .0001$. Therefore, we can reject the null hypothesis and conclude that there are statistically significant differences in the preference of the car origins used in USA, with more people preferring the "American" ($N = 253$) compared to either the " European " ($N = 73$) or the " Japanese " ($N = 79$).

Practice example:

- In "F2_AFLP" file test whether any of the present molecular markers does follow the Mendelian Inheritance Rule of segregation?

Chi-Square Test for Association using SPSS Statistics

- It is used to discover if there is a relationship between two categorical variables.

Application Example:

- In the "car.sav" data, we are interested in whether the three country of origin have different responses to cylinder numbers.

- (The null hypothesis is that vehicle country of origin does not affect response to cylinder number)

WHAT TO REPORT:

- You want to report the Chi-Square value, df, and p, parenthetically, following a statement that describes the patterns in the data. For example: vehicles with different country of origin did differ in their response to number of cylinders (chi-square test of independence, Chi-Square =185.794, df=8, p<0.001).

Practice example:

- In the "Technology.sav" file you need to test the association between:

- Q2 &Q5

- Q2&Q5&Q26

Linear Regression Analysis (simple & Multiple) using SPSS Statistics

- It is used when we want to predict the value of a variable based on the value of another variable(s). The variable we want to predict is called the dependent variable. The variable we are using to predict the other variable's value is called the independent variable. When we have single independent variable it called simple regression, however, when we have two or more independent variables, rather than just one, it called multiple regression.

Application Example

- In a car makers data, researchers want to test the relationship between "Time to accelerate" as independent variable and "Vehicle weight" as independent variable. Data File:" Cars.sav"

WHAT TO REPORT:

- Following a statement that describes the patterns in the data, you should parenthetically report the F, df, and p from the ANOVA, as well as the R^2 value. For example: Vehicle weight can significantly predict Time to acceleration (F=84.159, df=1, p<0.001; R^2=0.172).

Practice Example:

- Regress "Time" on "weight", "engine", "horse", "cylinder" and "year".

Logistic Regression using SPSS Statistics

- It is useful for situations in which you want to predict the presence or absence of a variable based on values of a set of predictor variables. It is similar to a linear regression model but is suited to models where the dependent variable is dichotomous. Logistic regression coefficients can be used to estimate odds ratios for each of the independent variables in the model.

- Logistic regression has been especially popular with medical research in which the dependent variable is whether or not a patient has a disease.

Application Example

- Consider a study on cancer remission (Lee; [1974]). The data consist of patient characteristics and whether or not cancer remission occurred. The variable *remiss* is the cancer remission indicator variable with a value of 1 for remission and a value of 0 for nonremission. The other six variables are the risk factors thought to be related to cancer remission. Data File:" cancer remission".

WHAT TO REPORT:

- A logistic regression was performed to ascertain the effects of six risk factors on the likelihood that participants have cancer remission. The logistic regression model was statistically not significant, $\chi 2(7) = 4.989$, $p < .672$. The model explained 52% (Nagelkerke R2) of the variance in heart disease and correctly classified 74.1% of cases. Males were 7.02 times more likely to exhibit heart disease than females.

Practice example:

- What lifestyle characteristics are risk factors for coronary heart disease (CHD)? Given a sample of patients measured on smoking status, diet, exercise, alcohol use, and CHD status, you could build a model using the four lifestyle variables to predict the presence or absence of CHD in a sample of patients. The model can then be used to derive estimates of the odds ratios for each factor to tell you, for example, how much more likely smokers are to develop CHD than nonsmokers. The goal of this study is explore the relationship between age and the presence or absence of CHD in this study population. File name "Heart disease"

Principal Components Analysis (PCA) using SPSS Statistics

- Principal components analysis (PCA, for short) is a variable-reduction technique that shares many similarities to exploratory factor analysis.

- Its aim is to reduce a larger set of variables into a smaller set of 'articifial' variables, called 'principal components', which account for most of the variance in the original variables.

- Term often used interchangeably with "factor analysis", however, there are slight differences

Application Example

- Socioeconomic data provided by Harman (1976). The five variables represent total population (*Population*), median school years (*School*), total employment (*Employment*), miscellaneous professional services (*Services*), and median house value (*HouseValue*). Each observation represents one of twelve census tracts in the Los Angeles Standard Metropolitan Statistical Area.

- Data File: "Socioeconomic.xls"

WHAT TO REPORT:

- the first two principal components provide an adequate summary of the data for most purposes.

Practice example:

- Data file: "places.xls"