DATA ANALYSIS : FREQUENCY DISTRIBUTION, HYPOTHESIS TESTING, AND pdf

1: MBA Here I come !: frequency distribution, cross tabulation, and hypothesis testing

fieldwork (data collection), data preparation and analysis strategy, data analysis (frequency distribution, hypothesis testing and cross-tabulation), data analysis (hypothesis testing related to differences), data analysis (correlation and regression), report preparation and presentation.

Construct a pie chart Solution Summary The solution gives detailed steps on performing the data analysis including building a pie chart, histogram and calculating mean, median, variance and then conducting hypothesis testing. All are done in Excel but report in Word. Use graphs, regression analysis, and hypothesis tests to determine among other things if men spend more money purchasing vehicles than women See the attached data. Below is a breakdown of what exactly to use: The data will be fictious. Gather or create data. If you "invent" the data, please footnote that fact and in your paper describe a method that would have been used to gather the data. If you find data, then document your source and indicate the sampling method and whether your sample was random or not. Either way, your data sets will need to be related to your topic and included in your appendix. You will need a minimum of three data sets with at least 30 values in each, but you may use more as needed to complete the requirements. You will need several different sets of data, all numeric. There are several options for this data. This could be a trend line with one data set on a time interval axis. This could be "before" and "after" data. This data will be used in the hypothesis test of independent means to determine if the means are equal. For the bivariate data: Determine which is the independent and which is the dependent variable. Create a scatter diagram. Determine the coefficient of correlation and interpret it. Determine the equation of the regression line and explain. Interpret each value in the regression equation and explain how the equation and line are related. Graph the regression line on a scatter diagram. Make a prediction for some value within the x-range and explain the meaning and the reliability of the prediction. From your data set determine the point with the largest residual and explain its interpretation. For the two or more sets of data from different populations: Remember to show all steps including the statements of the hypotheses, conditions, computational results, decisions, and interpretation. These should be in paragraph form, not numbered. Remember to show all steps. For the paired data: Perform a hypothesis test for the difference of two means. Remember to show all steps and to explain and give an interpretation of the result.

DATA ANALYSIS : FREQUENCY DISTRIBUTION, HYPOTHESIS TESTING, AND pdf

2: Introductory Statistics: Concepts, Models, and Applications by David Stockburger

From frequency distributions to cross tabulation to hypothesis testing to more sophisticated types of analysis, Minitab has what you need to analyze survey data and make sound conclusions about markets, customers, or whatever you're trying to assess.

Early use[edit] While hypothesis testing was popularized early in the 20th century, early forms were used in the s. Ronald Fisher began his life in statistics as a Bayesian Zabell , but Fisher soon grew disenchanted with the subjectivity involved namely use of the principle of indifference when determining prior probabilities , and sought to provide a more "objective" approach to inductive inference. Neyman who teamed with the younger Pearson emphasized mathematical rigor and methods to obtain more results from many samples and a wider range of distributions. Fisher popularized the "significance test". He required a null-hypothesis corresponding to a population frequency distribution and a sample. His now familiar calculations determined whether to reject the null-hypothesis or not. Significance testing did not utilize an alternative hypothesis so there was no concept of a Type II error. They initially considered two simple hypotheses both with frequency distributions. They calculated two probabilities and typically selected the hypothesis associated with the higher probability the hypothesis more likely to have generated the sample. Their method always selected a hypothesis. It also allowed the calculation of both types of error probabilities. The defining paper [34] was abstract. Mathematicians have generalized and refined the theory for decades. Neyman accepted a position in the western hemisphere, breaking his partnership with Pearson and separating disputants who had occupied the same building by much of the planetary diameter. World War II provided an intermission in the debate. Neyman wrote a well-regarded eulogy. Great conceptual differences and many caveats in addition to those mentioned above were ignored. Sometime around , [41] in an apparent effort to provide researchers with a "non-controversial" [43] way to have their cake and eat it too , the authors of statistical text books began anonymously combining these two strategies by using the p-value in place of the test statistic or data to test against the Neymanâ€"Pearson "significance level". It then became customary for the null hypothesis, which was originally some realistic research hypothesis, to be used almost solely as a strawman "nil" hypothesis one where a treatment has no effect, regardless of the context. Set up a statistical null hypothesis. The null need not be a nil hypothesis i. These define a rejection region for each hypothesis. Report the exact level of significance e. If the result is "not significant", draw no conclusions and make no decisions, but suspend judgement until further data is available. If the data falls into the rejection region of H1, accept H2; otherwise accept H1. Note that accepting a hypothesis does not mean that you believe in it, but only that you act as if it were true. Use this procedure only if little is known about the problem at hand, and only to draw provisional conclusions in the context of an attempt to understand the experimental situation. The usefulness of the procedure is limited among others to situations where you have a disjunction of hypotheses e. Early choices of null hypothesis[edit] Paul Meehl has argued that the epistemological importance of the choice of null hypothesis has gone largely unacknowledged. When the null hypothesis is predicted by theory, a more precise experiment will be a more severe test of the underlying theory. When the null hypothesis defaults to "no difference" or "no effect", a more precise experiment is a less severe test of the theory that motivated performing the experiment. Pierre Laplace compares the birthrates of boys and girls in multiple European cities. Karl Pearson develops the chi squared test to determine "whether a given form of frequency curve will effectively describe the samples drawn from a given population. He uses as an example the numbers of five and sixes in the Weldon dice throw data. Karl Pearson develops the concept of " contingency " in order to determine whether outcomes are independent of a given categorical factor. Here the null hypothesis is by default that two things are unrelated e. If the "suitcase" is actually a shielded container for the transportation of radioactive material, then a test might be used to select among three hypotheses: The test could be required for safety, with actions required in each case. The Neymanâ€"Pearson lemma of hypothesis testing says that a good criterion for the selection of hypotheses is the ratio of their probabilities a likelihood ratio. A simple method of solution is to select the hypothesis with the highest probability for the Geiger counts observed. The typical result matches intuition: Notice also that usually there are problems for proving a negative. Null hypotheses should be at least falsifiable. Neymanâ€"Pearson theory can accommodate both prior probabilities and the costs of actions resulting from decisions. The latter allows the consideration of economic issues for example as well as probabilities. A likelihood ratio remains a good criterion for selecting among hypotheses. The two forms of hypothesis testing are based on different problem formulations. In the view of Tukey [50] the former produces a conclusion on the basis of only strong evidence while the latter produces a decision on the basis of available evidence. While the two tests seem quite different both mathematically and philosophically, later developments lead to the opposite claim. Consider many tiny radioactive sources. The hypotheses become 0,1,2, There is little distinction between none or some radiation Fisher and 0 grains of radioactive sand versus all of the alternatives Neymanâ€"Pearson. The major Neymanâ€"Pearson paper of [34] also considered composite hypotheses ones whose distribution includes an unknown parameter. Neymanâ€"Pearson theory was proving the optimality of Fisherian methods from its inception. Neymanâ€"Pearson hypothesis testing is claimed as a pillar of mathematical statistics, [51] creating a new paradigm for the field. It also stimulated new applications in statistical process control , detection theory , decision theory and game theory. Both formulations have been successful, but the successes have been of a different character. The dispute over formulations is unresolved. Statisticians study Neymanâ€"Pearson theory in graduate school. Mathematicians are proud of uniting the formulations. Philosophers consider them separately. Learned opinions deem the formulations variously competitive Fisher vs Neyman , incompatible [32] or complementary. The terminology is inconsistent. Hypothesis testing can mean any mixture of two formulations that both changed with time. Any discussion of significance testing vs hypothesis testing is doubly vulnerable to confusion. Fisher thought that hypothesis testing was a useful strategy for performing industrial quality control, however, he strongly disagreed that hypothesis testing could be useful for scientists. The two methods remain philosophically distinct. The preferred answer is context dependent. Much of the criticism can be summarized by the following issues: The interpretation of a p-value is dependent upon stopping rule and definition of multiple comparison. The former often changes during the course of a study and the latter is unavoidably ambiguous. Rather than being wrong, statistical hypothesis testing is misunderstood, overused and misused. When used to detect whether a difference exists between groups, a paradox arises. As improvements are made to experimental design e. To minimize type II errors, large samples are recommended. In psychology practically all null hypotheses are claimed to be false for sufficiently large samples so " Casting doubt on the null hypothesis is thus far from directly supporting the research hypothesis. While it can provide critical information, it is inadequate as the sole tool for statistical analysis. Successfully rejecting the null hypothesis may offer no support for the research hypothesis. The continuing controversy concerns the selection of the best statistical practices for the near-term future given the often poor existing practices. Critics would prefer to ban NHST completely, forcing a complete departure from those practices, while supporters suggest a less absolute change. The American Psychological Association has strengthened its statistical reporting requirements after review, [68] medical journal publishers have recognized the obligation to publish some results that are not statistically significant to combat publication bias [69] and a journal Journal of Articles in Support of the Null Hypothesis has been created to publish such results exclusively. Major organizations have not abandoned use of significance tests although some have discussed doing so.

DATA ANALYSIS : FREQUENCY DISTRIBUTION, HYPOTHESIS TESTING, AND pdf

3: Statistical hypothesis testing - Wikipedia

1. A study was conducted on the distribution and packaging of ounce packages of plain m&mÂ® candies. A random sample of 78 bags was collected and the data reported in the attached Excel file.

Basic analysis invariably involves some hypothesis testing. Select an appropriate statistical technique and the corresponding test statistic. Choose the level of significance. Determine the sample size and collect the data. Calculate the value of the test statistic. Make the statistical decision to reject or not reject the null hypothesis. Express the statistical decision in terms of the marketing research problem. Accepting the alternative hypothesis will lead to changes in opinions or actions. The appropriate way to formulate the hypotheses is: On the other hand, if Ho is not rejected, then the new service should not be introduced unless additional evidence is obtained. The power of a statistical test is discussed further in step 3. Select an Appropriate Test To test the null hypothesis, it is necessary to select an appropriate statistical technique. This statistic would be computed as follows: Choose Level of Significance. Two types of errors can occur. As a compromise, a is often set at 0. The test statistic z can be calculated as follows: The shaded area between and 1. Steps 6 and 7: Compare the Probability or Critical Value and Make the Decision The probability associated with the calculated or observed value of the test statistic is 0. This is the probability of getting a p value of 0. Hence, the null hypothesis is rejected. This sign shift can be easily seen: As can be seen from Figure In tests of differences. Tests of differences could relate to distributions, means.

DATA ANALYSIS : FREQUENCY DISTRIBUTION, HYPOTHESIS TESTING, AND pdf

4: Data Analysis, Graphs and Hypothesis Testing

A symmetric bell-shaped distribution that is useful for small sample testing, when the mean and the variance of population are unknown. t test A parametric test based on t distribution to determine the statistical significance of the difference between a sample mean and a population parameter.

Business administration and marketing class notes - these notes are taken using voice recognition software and may have grammar errors. Saturday, June 09, frequency distribution, cross tabulation, and hypothesis testing Chapter 15 Basic data analysis provides viable insights and guides the rest of the data analysis as well as the interpretation of the results. A frequency distribution should be obtained for each variable in the data. This analysis produces a table of frequency counts, percentages, and cumulative percentages for all the values associated with that variable. It indicates the extent of out of range, missing, or extreme values. The mean, mode, and median of a frequency distribution are measures of central tendency. The variability of the distribution is described by the range, the variants or standard deviation, coefficient of variation, and interquartile range. Skewness and kurtosis provide an idea of the shape of the distribution. Cross tabulations are tables that reflect the joint distribution of two or more variables. In cross tabulation, the percentages can be computed either column wise, based on column totals, or row wise, based on row totals. The general rule is to compete the percentages in the direction of the independent variable, across the dependent variable. The Chi Square statistic provides a test of the statistical significance of the observed association in a cross tabulation. Parametric and non-parametric tests are available for testing hypothesis related to differences. And the parametric case, the t test is used to examine hypotheses related to the population mean. Different forms of the t test are suitable for testing hypotheses based on one sample, two independent samples, or paired samples. In the nonparametric case, popular one sample tests include the Kolmogorov-Smirnov, chi-square, runs test, and the binomial test. For two independent nonparametric samples, the Mann-Whitney U test, median test and the Kolmogorov-Smirnov test can be used. For paired samples, the Wlicoxon matched-pairs signed-ranks test and assign tests are useful for examining hypotheses related to measures of location. If the null hypothesis is not rejected, no changes will be made alternative hypothesis -- a statement that some difference or effect is expected. Excepting the alternative hypothesis will lead to changes in opinions or actions one tailed test -- a test of the null hypothesis where the alternative hypothesis is expressed directionally two tailed test -- a test of the null hypothesis where the alternative hypothesis is not expressed directionally test statistic -- a measure of how close the sample has come to the null hypothesis. It often follows a well-known distribution, such as the normal, t, or chi- squared distribution type I error -- also known as Alpha error, occurs when a sample results lead to the rejection of a null hypothesis that is in fact true level of significance -- the probability of making a type 1 error type II error -- also known as beta error, occurs when the sample results lead to the non-rejection of a null hypothesis that is in fact false power of a test -- the probability of rejecting the null hypothesis when it is in fact false and should be rejected Cross tabulation -- a statistical technique that describes two or more variables simultaneously and results in tables that reflect the joint distribution of two or more variables that have a limited number of categories or distinct values contingency table -- a cross tabulation table. It contains a cell for every combination of categories of the two variables chi-square statistic -- the statistic used to test the statistical significance of the observed association and cross tabulation. It assists us in determining whether a systematic association exists between the two variables chi-square distribution -- a skewed distribution and shape depends solely on the number of degrees of freedom. Lambda also varies between zero and one symmetric lambda -- the symmetric lambda does not make an assumption about which variable is dependent. It measures the overall improvement when production is done in both directions tau b -- test statistic that measures the association between two ordinal-level variables. It makes adjustment for ties and is most appropriate when the table of variables is square tau c -- test statistic that measures the association between two ordinal-level variables. It makes adjustment for ties and is most appropriate when the table of variables is not square but a rectangle Gamma -- test statistic that measures the association between two ordinal-level variables. It does not make an adjustment for ties parametric tests -- hypothesis testing procedures that assume that the variables of interest are measured on at least an interval scale non-parametric tests -- hypothesis testing procedures that assume that the variables are measured on a nominal or ordinal scale t test -- a univariate hypothesis test using the t distribution, which is used in the standard deviation is unknown and the sample size is small t statistic -- a statistic that assumes that the variable has a symmetric bell shaped distribution in the mean is known or assumed to be known and the population variants is estimated from the sample t distribution -- symmetric bell shaped distribution that is useful for small sample testing z test -- a univariate hypothesis test using the standard normal distribution independent samples -- to samples that are not experimentally related. The measurement of one sample has no effect on the values of the second sample f test -- a statistical test of the equality of the variances of two populations f statistic -- the f statistic is computed as the ratio of two sample variances f distribution -- a frequency distribution that depends on two sets of degrees of freedom -- the degrees of freedom in the numerator and the degrees of freedom in the denominator paired samples -- and hypothesis testing, the observations are paired so that two sets of observations relate to the same respondents paired samples t test -- a test for differences in the means of paired samples Kolmogorov-Smirnov one-sample test - A one sample nonparametric goodness of fit test that compares the cumulative distribution function for a variable with a specified distribution runs test -- a test of randomness for a dichotomous variable binomial test -- a goodness of fit statistical test for dichotomous variables. It tests the goodness of fit of the observed number of observations in each category to the number expected under a specified binomial distribution Mann-Whitney U test -- a statistical test for the variable measured on an ordinal scale comparing the difference in the location of two populations based on observations from two independent samples two-sample median test -- non-parametric test statistic that determines whether two groups are drawn from populations with the same median. This test is not as powerful as the Mann- Whitney U Kolmogorov-Smirnov two-sample test -- nonparametric test statistic that determines whether to his divisions are the same. It takes into account any differences in the two distributions including median, dispersion, and skewness Wilcoxon matched-pairs signed-ranks test -- a nonparametric test that analyzes the differences between the paired observations, taking into account the magnitude of the differences sign test -- a nonparametric test for examining differences in the location of two populations, based on paired observations, that compares only the signs of the differences between pairs of variables without taking into account the magnitude of the differences Posted by.

DATA ANALYSIS : FREQUENCY DISTRIBUTION, HYPOTHESIS TESTING, AND pdf

5: Statistics | science | www.enganchecubano.com

frequency distribution, cross tabulation, and hypothesis testing Chapter 15 Basic data analysis provides viable insights and guides the rest of the data analysis as well as the interpretation of the results.

This article has been cited by other articles in PMC. The assumption of normality needs to be checked for many statistical procedures, namely parametric tests, because their validity depends on it. The aim of this commentary is to overview checking for normality in statistical analysis using SPSS. Normality, Statistical Analysis 1. Many of the statistical procedures including correlation, regression, t tests, and analysis of variance, namely parametric tests, are based on the assumption that the data follows a normal distribution or a Gaussian distribution after Johann Karl Gauss, â€" ; that is, it is assumed that the populations from which the samples are taken are normally distributed 2 - 5. The assumption of normality is especially critical when constructing reference intervals for variables 6. Normality and other assumptions should be taken seriously, for when these assumptions do not hold, it is impossible to draw accurate and reliable conclusions about reality 2 , 7. If we have samples consisting of hundreds of observations, we can ignore the distribution of the data 3. Although true normality is considered to be a myth 8 , we can look for normality visually by using normal plots 2 , 3 or by significance tests, that is, comparing the sample distribution to a normal one 2 , 3. It is important to ascertain whether data show a serious deviation from normality 8. The purpose of this report is to overview the procedures for checking normality in statistical analysis using SPSS. Visual Methods Visual inspection of the distribution may be used for assessing normality, although this approach is usually unreliable and does not guarantee that the distribution is normal 2 , 3 , 7. However, when data are presented visually, readers of an article can judge the distribution assumption by themselves 9. The frequency distribution histogram , stem-and-leaf plot, boxplot, P-P plot probability-probability plot , and Q-Q plot quantile-quantile plot are used for checking normality visually 2. The frequency distribution that plots the observed values against their frequency, provides both a visual judgment about whether the distribution is bell shaped and insights about gaps in the data and outliers outlying values The stem-and-leaf plot is a method similar to the histogram, although it retains information about the actual data values 8. The P-P plot plots the cumulative probability of a variable against the cumulative probability of a particular distribution e. After data are ranked and sorted, the corresponding z-score is calculated for each rank as follows: This is the expected value that the score should have in a normal distribution. The scores are then themselves converted to z-scores. The actual z-scores are plotted against the expected z-scores. If the data are normally distributed, the result would be a straight diagonal line 2. A Q-Q plot is very similar to the P-P plot except that it plots the quantiles values that split a data set into equal portions of the data set instead of every individual score in the data. Moreover, the Q-Q plots are easier to interpret in case of large sample sizes 2. The boxplot shows the median as a horizontal line inside the box and the interquartile range range between the 25 th to 75 th percentiles as the length of the box. The whiskers line extending from the top and bottom of the box represent the minimum and maximum values when they are within 1. Scores greater than 1. A boxplot that is symmetric with the median line at approximately the center of the box and with symmetric whiskers that are slightly longer than the subsections of the center box suggests that the data may have come from a normal distribution 8. Normality Tests The normality tests are supplementary to the graphical assessment of normality 8. For small sample sizes, normality tests have little power to reject the null hypothesis and therefore small samples most often pass normality tests 7. For large sample sizes, significant results would be derived even in the case of a small deviation from normality 2 , 7 , although this small deviation will not affect the results of a parametric test 7. The K-S test is an empirical distribution function EDF in which the theoretical cumulative distribution function of the test distribution is contrasted with the EDF of the data 7. A limitation of the K-S test is its high sensitivity to extreme values; the Lilliefors correction renders this test less conservative It has been reported that the K-S test has low power and it should not be seriously considered for testing normality Moreover, it is not recommended when parameters are estimated from the data, regardless of sample size The Shapiro-Wilk test is based on the correlation between the data and the corresponding normal scores 10 and provides better power than the K-S test even after the Lilliefors correction Power is the most frequent measure of the value of a test for normalityâ€"the ability to detect whether a sample comes from a non-normal distribution Some researchers recommend the Shapiro-Wilk test as the best choice for testing the normality of data SPSS provides the K-S with Lilliefors correction and the Shapiro-Wilk normality tests and recommends these tests only for a sample size of less than 50 8. In Figure , both frequency distributions and P-P plots show that serum magnesium data follow a normal distribution while serum TSH levels do not. It is clear that for serum magnesium concentrations, both tests have a p-value greater than 0. Lack of symmetry skewness and pointiness kurtosis are two main ways in which a distribution can deviate from normal. The values for these parameters should be zero in a normal distribution. These values can be converted to a z-score as follows:

DATA ANALYSIS : FREQUENCY DISTRIBUTION, HYPOTHESIS TESTING, AND pdf

6: Introduction to Hypothesis Testing Marketing Research Assignment Help and Homework Help

Depending on the actual distribution of the data, you may be able to approximate the Poisson distribution with a Normal distribution, and you could use a t-test, analysis of variance (ANOVA), or linear regression.

Retrieve Value Given a set of specific cases, find attributes of those cases. What is the value of aggregation function F over a given set S of data cases? What is the sorted order of a set S of data cases according to their value of attribute A? What is the range of values of attribute A in a set S of data cases? What is the distribution of values of attribute A in a set S of data cases? What is the correlation between attributes X and Y over a given set S of data cases? Barriers to effective analysis[edit] Barriers to effective analysis may exist among the analysts performing the data analysis or among the audience. Distinguishing fact from opinion, cognitive biases, and innumeracy are all challenges to sound data analysis. Confusing fact and opinion[edit] You are entitled to your own opinion, but you are not entitled to your own facts. Daniel Patrick Moynihan Effective analysis requires obtaining relevant facts to answer questions, support a conclusion or formal opinion , or test hypotheses. Facts by definition are irrefutable, meaning that any person involved in the analysis should be able to agree upon them. This makes it a fact. Whether persons agree or disagree with the CBO is their own opinion. As another example, the auditor of a public company must arrive at a formal opinion on whether financial statements of publicly traded corporations are "fairly stated, in all material respects. When making the leap from facts to opinions, there is always the possibility that the opinion is erroneous. Cognitive biases[edit] There are a variety of cognitive biases that can adversely affect analysis. In addition, individuals may discredit information that does not support their views. Analysts may be trained specifically to be aware of these biases and how to overcome them. In his book Psychology of Intelligence Analysis, retired CIA analyst Richards Heuer wrote that analysts should clearly delineate their assumptions and chains of inference and specify the degree and source of the uncertainty involved in the conclusions. He emphasized procedures to help surface and debate alternative points of view. However, audiences may not have such literacy with numbers or numeracy ; they are said to be innumerate. Persons communicating the data may also be attempting to mislead or misinform, deliberately using bad numerical techniques. More important may be the number relative to another number, such as the size of government revenue or spending relative to the size of the economy GDP or the amount of cost relative to revenue in corporate financial statements. This numerical technique is referred to as normalization [7] or common-sizing. There are many such techniques employed by analysts, whether adjusting for inflation i. Analysts apply a variety of techniques to address the various quantitative messages described in the section above. Analysts may also analyze data under different assumptions or scenarios. For example, when analysts perform financial statement analysis , they will often recast the financial statements under different assumptions to help arrive at an estimate of future cash flow, which they then discount to present value based on some interest rate, to determine the valuation of the company or its stock. Smart buildings[edit] A data analytics approach can be used in order to predict energy consumption in buildings. Analytics and business intelligence[edit] Main article: Analytics Analytics is the "extensive use of data, statistical and quantitative analysis, explanatory and predictive models, and fact-based management to drive decisions and actions. Initial data analysis[edit] The most important distinction between the initial data analysis phase and the main analysis phase, is that during initial data analysis one refrains from any analysis that is aimed at answering the original research question. The initial data analysis phase is guided by the following four questions: Data quality can be assessed in several ways, using different types of analysis: Test for common-method variance. The choice of analyses to assess the data quality during the initial data analysis phase depends on the analyses that will be conducted in the main analysis phase. One should check whether structure of measurement instruments corresponds to structure reported in the literature. There are two ways to assess measurement: If the study did not need or use a randomization procedure, one should check the success of the non-random sampling, for instance by checking whether all subgroups of the population of interest are represented in sample. Other possible data distortions that should be checked are: It is especially important to exactly determine the structure of the sample and specifically the size of the subgroups when subgroup analyses will be performed during the main analysis phase. The characteristics of the data sample can be assessed by looking at: Basic statistics of important variables Scatter plots Cross-tabulations [31] Final stage of the initial data analysis[edit] During the final stage, the findings of the initial data analysis are documented, and necessary, preferable, and possible corrective actions are taken. Also, the original plan for the main data analyses can and should be specified in more detail or rewritten. In order to do this, several decisions about the main data analyses can and should be made: In the case of non- normals: In the case of missing data: In the case of outliers: In case items do not fit the scale: In the case of too small subgroups: In case the randomization procedure seems to be defective:

DATA ANALYSIS : FREQUENCY DISTRIBUTION, HYPOTHESIS TESTING, AND pdf

7: Normality Tests for Statistical Analysis: A Guide for Non-Statisticians

Data Analysis. Frequency Distribution In a frequency distribution, one variable is considered at a time. A frequency distribution for a variable produces a.

Descriptive statistics Descriptive statistics are tabular, graphical, and numerical summaries of data. The purpose of descriptive statistics is to facilitate the presentation and interpretation of data. Most of the statistical presentations appearing in newspapers and magazines are descriptive in nature. Univariate methods of descriptive statistics use data to enhance the understanding of a single variable; multivariate methods focus on using statistics to understand the relationships among two or more variables. To illustrate methods of descriptive statistics, the previous example in which data were collected on the age, gender, marital status, and annual income of individuals will be examined. Tabular methods The most commonly used tabular summary of data for a single variable is a frequency distribution. A frequency distribution shows the number of data values in each of several nonoverlapping classes. Another tabular summary, called a relative frequency distribution, shows the fraction, or percentage , of data values in each class. The most common tabular summary of data for two variables is a cross tabulation, a two-variable analogue of a frequency distribution. For a qualitative variable, a frequency distribution shows the number of data values in each qualitative category. For instance, the variable gender has two categories: Thus, a frequency distribution for gender would have two nonoverlapping classes to show the number of males and females. A relative frequency distribution for this variable would show the fraction of individuals that are male and the fraction of individuals that are female. Constructing a frequency distribution for a quantitative variable requires more care in defining the classes and the division points between adjacent classes. For instance, if the age data of the example above ranged from 22 to 78 years, the following six nonoverlapping classes could be used: A frequency distribution would show the number of data values in each of these classes, and a relative frequency distribution would show the fraction of data values in each. A cross tabulation is a two-way table with the rows of the table representing the classes of one variable and the columns of the table representing the classes of another variable. To construct a cross tabulation using the variables gender and age, gender could be shown with two rows, male and female, and age could be shown with six columns corresponding to the age classes 20â€”29, 30â€”39, 40â€”49, 50â€”59, 60â€”69, and 70â€” The entry in each cell of the table would specify the number of data values with the gender given by the row heading and the age given by the column heading. Such a cross tabulation could be helpful in understanding the relationship between gender and age. Graphical methods A number of graphical methods are available for describing data. A bar graph is a graphical device for depicting qualitative data that have been summarized in a frequency distribution. Labels for the categories of the qualitative variable are shown on the horizontal axis of the graph. A bar above each label is constructed such that the height of each bar is proportional to the number of data values in the category. A bar graph of the marital status for the individuals in the above example is shown in Figure 1. There are 4 bars in the graph, one for each class. A pie chart is another graphical device for summarizing qualitative data. The size of each slice of the pie is proportional to the number of data values in the corresponding class. A pie chart for the marital status of the individuals is shown in Figure 2. A pie chart for the marital status of individuals. A histogram is the most common graphical presentation of quantitative data that have been summarized in a frequency distribution. The values of the quantitative variable are shown on the horizontal axis. A rectangle is drawn above each class such that the base of the rectangle is equal to the width of the class interval and its height is proportional to the number of data values in the class. Page 1 of 8.

8: hypothesis testing - Statistical Test for frequency distribution - Cross Validated

Conduct a descriptive analysis on any two interval/ratio variables you wish using the www.enganchecubano.com and the www.enganchecubano.com file. Explain the output. a.) www.enganchecubano.com (File).

This article highlights several basic tools in Minitab that will help you interpret your survey data accurately. Surveys are an important tool for market research. Businesses use them to systematically and objectively gather information from respondents to discover what people want and identify market needs. Surveys also have many applications beyond market research, and students in statistics and many other types of classes can learn a great deal from gathering and analyzing survey data. How much data do you need? You also can use Minitab to ensure that you are selecting a truly random sample of participants. By hand-picking them yourself, you may introduce bias into your results even though you are trying to pick households at random. Minitab also can help you prepare your data for analysis. For example, one of your survey questions asked people to rank a product on a 7-point scale, and you want to classify responses 6 and 7 as positive, 3 to 5 as neutral, and 1 and 2 as negative. The Data menu also contains a wide array of tools that you can use to sort, clean, and otherwise prepare your raw data for analysis. A good first step in your analysis is to conveniently summarize the data by counting the responses for each level of a given variable. These counts, or frequencies, are called the frequency distribution and are commonly accompanied by the percentages and cumulative percentages as well. A frequency distribution can quickly reveal: Suppose a pet adoption and rescue agency wants to find out whether dogs or cats are more popular in a certain location. To answer this question, we survey a random sample of local pet owners to find out if dogs are more popular than cats, or vice versa. We also can summarize the data using descriptive statistics. This can be very helpful when looking at continuous variables that might have a broad range, such as the age when people got their first dog or cat. Cross Tabulation A frequency distribution can tell you about a single variable, but it does not provide information about how two or more variables relate to one another. To understand the association between multiple variables, we can use cross tabulation. Are men more likely to want a dog than a cat compared to women, or vice versa? To summarize data from both variables at the same time, we need to construct a cross-tabulation table, also known as a contingency table. This table lets us evaluate the counts and percents, just like a frequency distribution. But while a frequency distribution provides information for each level of one variable, cross tabulation shows results for all level combinations of both variables. Based on these percentages, we can conclude that males are more likely to own a dog while females are more likely to own a cat. But what if there is third variable to consider, such as marital status? We could then create a similar cross tabulation, but break it down into two tables: Hypothesis Testing Frequency distributions and cross tabulation are great starting points for survey analysis, but they may not be sufficient for a comprehensive analysis. To get a fuller understanding of your data, we need to include hypothesis testing. For our pet survey, we want to make sure the difference we see between gender and pet preference is due to a true association, and not random chance. A hypothesis test can tell us if the difference we see in the percentages is statistically significant, and whether the pet preference and gender variables are independent or not. To evaluate the statistical significance of cross tabulation results, we use a hypothesis test called the chi-square test. Minitab makes it easy to evaluate several variations of the Chi-Square test. Therefore, we can reject the null hypothesis that the variables are independent and conclude that a statistically significant relationship exists between gender and pet preference. Based on these results, we would focus on marketing a local dog rescue drive to male residents and a local cat rescue in a way that appeals to female residents. A chi-square test was the tool we needed in this case, but there are other hypothesis tests commonly used for survey data, including t-tests and proportion tests. These types of tests can be used to compare averages or proportions to a target value, or to compare averages or proportions to each other. These hypothesis tests can answer questions such as: Is one brand of cola preferred over its competitor? Is there a difference in the average rating given to a cell phone company by teenagers compared to parents? For more information and additional examples detailing how to use these and other useful tools, Minitab offers an extensive Help system and free Technical Support. See a sample issue.

9: Data analysis - Wikipedia

frequency data in Table From the histogram, one could examine whether the observed distribution is consistent with an expected or assumed distribution, such as the normal distribution.

DATA ANALYSIS : FREQUENCY DISTRIBUTION, HYPOTHESIS TESTING, AND pdf

Association law handbook Lower extremity arterial disease coexisting with coronary artery disease : management and prognosis Denni The Son of the Bondwoman Osha 1926 All My Stars and Animals Pl sql basic interview questions A mister Blenkinsop, a diarist. The wonderful clouds Bagatelles for a Masochist Sticker Safari/Snakes Ultimate eu test book American and Asian Slavery Writing the New Economy Implementing adaptive governance 1997 ford contour repair manual Love sonnets of an office boy Lifes schooling E. B. Van Arsdale. Beethoven virus violin sheet Cults, too good to be true Theory and research on leadership in organizations Gary Yukl, David D. Van Fleet 2000 World Market Forecasts for Imported Unworked, Unset Diamonds Dc adventures rpg 12th science question paper maharashtra board 2013 Preschool classroom managment plan created in alabama Dyslexia and reading Cathy Svensson. The reign of terror (Nov. 1939) Interaction of Catholic bioethics and secular society Death note another note los angeles bb case The old fireside. Science in a Technical World James A. Garfield by Allan Peskin Formation of econometrics Mirza ghalib history in urdu Organizational and managerial issues in logistics How to start and operate your own design firm Performance measurement Malcolm Horner Ice, Rock, and Beauty Power and sample size determination Cooking with Very Little Energy To repair the world