|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Out variable | | | | | |
| Nominal (2 Categories as I understand) | Categorical (> 2 Categories) | Ordinal (Ordered Categories) | Quantitative  Discrete | Quantitative  Non-normal | Quantitative  Normal |
| In variable | Nominal (2 Categories as I understand) | Χ2 or Fishers | Χ2 | Χ2 trend or Mann-Whitney U-test (analogue of Student t-test, non-parametric) | Mann-Whitney U-test (analogue of Student t-test, non-parametric) | Mann-Whitney U-test (analogue of Student t-test, non-parametric) or log-rank (non-parametric for censored data) (a) | Student t-test – проверка равенства средних в двух выборках (parametric, The samples are independent, Each sample is from a normally distributed population, The population standard deviations of the groups are all equal) |
| Categorical (> 2 Categories) | Χ2 | Χ2 | Kruskall-Wallis H-test - проверка равенства медиан в нескольких выборках (extends Mann-Whitney U-test for more than two groups, analogue of ANOVA for non-normal distribution, non-parametric) (b) | Kruskall-Wallis H-test - проверка равенства медиан в нескольких выборках (extends Mann-Whitney U-test for more than two groups, analogue of ANOVA for non-normal distribution, non-parametric) (b) | Kruskall-Wallis H-test - проверка равенства медиан в нескольких выборках (extends Mann-Whitney U-test for more than two groups, analogue of ANOVA for non-normal distribution, non-parametric) (b) | ANOVA – проверка равенства средних в нескольких выборках (extends Student t-test for more than two groups, parametric, The samples are independent, Each sample is from a normally distributed population, The population standard deviations of the groups are all equal. This property is known as homoscedasticity.) (c) |
| Ordinal (Ordered Categories) | Χ2 trend or Mann-Whitney U-test (analogue of Student t-test, non-parametric) | (e) | Spearman rank (non-parametric) | Spearman rank (non-parametric) | Spearman rank (non-parametric) | Spearman rank (non-parametric) or linear regression (d) |
| Quantitative  Discrete | Logistic regression | (e) | (e) | Spearman rank (non-parametric) | Spearman rank (non-parametric) | Spearman rank (non-parametric) or linear regression (d) |
| Quantitative  Non-normal | Logistic regression | (e) | (e) | (e) | Plot data and Pearson (parametric) (? В вики написано, что для нормальных величин) or Spearman rank (non-parametric) | Plot data and Pearson (parametric) or Spearman rank (non-parametric) and linear regression |
| Quantitative  Normal | Logistic regression | (e) | (e) | (e) | Linear regression (d) | Pearson (parametric) and linear regression |

(a) If data are censored.

(b) The Kruskal-Wallis test is used for comparing ordinal or non-Normal variables for more than two groups, and is a generalization of the Mann-Whitney U test. The technique is beyond the scope of this book, but is described in more advanced books and is available in common software (Epi-Info, Minitab, SPSS).

(c) Analysis of variance is a general technique, and one version (one way analysis of variance) is used to compare Normally distributed variables for more than two groups, and is the parametric equivalent of the Kruskal-Wallis test.

(d) If the outcome variable is the dependent variable, then provided the residuals (see ) are plausibly Normal, then the distribution of the independent variable is not important.

(e) There are a number of more advanced techniques, such as Poisson regression, for dealing with these situations. However, they require certain assumptions and it is often easier to either dichotomise the outcome variable or treat it as continuous.

Categorical variables are also known as discrete or qualitative variables. Categorical variables can be further categorized as either nominal, ordinal or dichotomous.

* Nominal variables are variables that have two or more categories, but which do not have an intrinsic order. For example, a real estate agent could classify their types of property into distinct categories such as houses, condos, co-ops or bungalows. So "type of property" is a nominal variable with 4 categories called houses, condos, co-ops and bungalows. Of note, the different categories of a nominal variable can also be referred to as groups or levels of the nominal variable. Another example of a nominal variable would be classifying where people live in the USA by state. In this case there will be many more levels of the nominal variable (50 in fact).
* Dichotomous variables are nominal variables which have only two categories or levels. For example, if we were looking at gender, we would most probably categorize somebody as either "male" or "female". This is an example of a dichotomous variable (and also a nominal variable). Another example might be if we asked a person if they owned a mobile phone. Here, we may categorise mobile phone ownership as either "Yes" or "No". In the real estate agent example, if type of property had been classified as either residential or commercial then "type of property" would be a dichotomous variable.
* Ordinal variables are variables that have two or more categories just like nominal variables only the categories can also be ordered or ranked. So if you asked someone if they liked the policies of the Democratic Party and they could answer either "Not very much", "They are OK" or "Yes, a lot" then you have an ordinal variable. Why? Because you have 3 categories, namely "Not very much", "They are OK" and "Yes, a lot" and you can rank them from the most positive (Yes, a lot), to the middle response (They are OK), to the least positive (Not very much). However, whilst we can rank the levels, we cannot place a "value" to them; we cannot say that "They are OK" is twice as positive as "Not very much" for example.

Continuous variables are also known as quantitative variables. Continuous variables can be further categorized as either interval or ratio variables.

* Interval variables are variables for which their central characteristic is that they can be measured along a continuum and they have a numerical value (for example, temperature measured in degrees Celsius or Fahrenheit). So the difference between 20C and 30C is the same as 30C to 40C. However, temperature measured in degrees Celsius or Fahrenheit is NOT a ratio variable.
* Ratio variables are interval variables, but with the added condition that 0 (zero) of the measurement indicates that there is none of that variable. So, temperature measured in degrees Celsius or Fahrenheit is not a ratio variable because 0C does not mean there is no temperature. However, temperature measured in Kelvin is a ratio variable as 0 Kelvin (often called absolute zero) indicates that there is no temperature whatsoever. Other examples of ratio variables include height, mass, distance and many more. The name "ratio" reflects the fact that you can use the ratio of measurements. So, for example, a distance of ten metres is twice the distance of 5 metres.