Evaluate the relationship/association between two variables.

|  |  |  |
| --- | --- | --- |
| **Type of variables** | **measurement** | **R function** |
| Two continuous variables | Pearson CC | Cor.test (x,y, method = "pearson") |
| One continuous and one qualitative ordinal (no matter how many levels that the ordinal feature has) | Spearman CC or  Kendall’s tau-b | Cor.test (x,y, method = "spearman")  Cor.test (x,y, method = "kendall") |
| Two qualitative ordinal variables  (no matter how many levels that the ordinal variables have) | Spearman CC or  Kendall’s tau-b | Cor.test (x,y, method = "spearman")  Cor.test (x,y, method = "kendall") |
| 我们可以看出，对于continuous and ordinal features, the calculation of relationship/association is very simple, 做Cor.test就好，回答的是两个变量是否correlated or associated to each other;当涉及到nominal variables， 我们做t-test/wilcox.test （2个 levels/groups）或者Anova/ kruskal.test（多个 levels/groups）回答的是不同level/group之间的连续变量是否显著不同。 | | |
| One continuous and one two-level nominal | Pearson CC after or  point-biserial correlation  two categories must be encoded as 0 and 1 numbers!!!! | Cor.test (x,y, method = "pearson")  Only when the method="pearson" in cor.test, it generates the same result as Biserial.cor. |
|  | Biserial.cor(x,y,level=2) from “ltm”  Package  Either x or y can be a text factor |
| 尽管以上两种方法可以用于计算One continuous and one two-level nominal variable之间的association，一般情况下，我们并不采用它们，而是采用下面的t-test/wilcox.test （2个 levels/groups）或者Anova/ kruskal.test（多个 levels/groups） | |
| For this method, the categories don’t have to be encoded as 0 and 1 (they can be text)  **当我们做wilcox.test的时候，我们不需要考虑variance 相等与否。** | t.test (Note that, unpaired two-samples t-test can be used only under certain conditions:   * when the two groups of samples (A and B), being compared, are **normally distributed**. This can be checked using [**Shapiro-Wilk test**](http://www.sthda.com/english/wiki/normality-test-in-r). * and when the **variances** of the two groups are equal. This can be checked using **F-test**.)   wilcox.test (NOT normal distribution)   1. must be with formula format 2. The factor must only have two levels |
| One continuous and one multiple-level nominal | **当我们做kruskal.test的时候，我们不需要考虑variance 相等与否。** | ANOVA (if continuous data with normal distribution)+ Tukey/Dunnett/pairwise.t.test  kruskal.test (NOT normal distribution) +pairwise.wilcox.test |
| One ordinal and one two-level nominal |  | For all these four conditions,  coin::chisq\_test can be used to fulfill this purpose depending on different setting on “scores = list “ in coin::chisq\_test  chisq\_test适合于analysis has large sample size; while Fisher’s exact test 适合于 small sample size analysis. Quantitatively, Chi-square test is not appropriate when the **expected** values in one of the cells of the contingency table is less than 5, and in this case the Fisher’s exact test is preferred .而**expected** values of the contingency table can be get by chisq.test(dat)$expected function |
| One ordinal and one multiple-level nominal |
| Two nominal variables |
| Two ordinal variables |