SAS - Hypothesis Testing

Hypothesis testing is the use of statistics to determine the probability that a given hypothesis is true. The usual process of hypothesis testing consists of four steps as shown below.

### Step-1

Formulate the null hypothesis H0 (commonly, that the observations are the result of pure chance) and the alternative hypothesis H1 (commonly, that the observations show a real effect combined with a component of chance variation).

### Step-2

Identify a test statistic that can be used to assess the truth of the null hypothesis.

### Step-3

Compute the P-value, which is the probability that a test statistic at least as significant as the one observed would be obtained assuming that the null hypothesis were true. The smaller the P-value, the stronger the evidence against the null hypothesis.

### Step-4

Compare the p-value to an acceptable significance value alpha (sometimes called an alpha value). If p <=alpha, that the observed effect is statistically significant, the null hypothesis is ruled out, and the alternative hypothesis is valid.

SAS programming language has features to carry out various types of hypothesis testing as shown below.

|  |  |  |
| --- | --- | --- |
| **Test** | **Description** | **SAS PROC** |
| **T-Test** | A t-tests is used to test whether the mean of one variable is significantly different than a hypothesized value. We also determine whether means for two independent groups are significantly different and whether means for dependent or paired groups are significantly different. | **PROC TTEST** |
| **ANOVA** | It is also used to compare means when there is one independent categorical variable. We want to use one-way ANOVA when testing to see if the means of the interval dependent variable are different according to the independent categorical variable. | **PROC ANOVA** |
| **Chi-Square** | We use chi square goodness of fit to assess if frequencies of a categorical variable were likely to happen due to chance. Use of a chi square test is necessary whether proportions of a categorical variable are a hypothesized value. | **PROC FREQ** |
| **Linear Regression** | Simple linear regression is used when one wants to test how well a variable predicts another variable. Multiple linear regression allows one to test how well multiple variables predict a variable of interest. When using multiple linear regression, we additionally assume the predictor variables are independent. | **PROC REG** |