R Applications

Tan Sein Jone

University of British Columbia

June 27, 2024

Table of contents

Table of Contents

Hypothesis Testing

- Hypothesis testing is a statistical method that is used in making statistical decisions using experimental data.
- A hypothesis test evaluates two mutually exclusive statements about a population to determine which statement is best supported by the sample data.
- These two statements are called the null hypothesis and the alternative hypothesis.

Hypothesis Testing

- The null hypothesis is the statement being tested. Usually, the null hypothesis states that there is no effect or no difference.
- The alternative hypothesis is the statement that is accepted if the sample data provide enough evidence that the null hypothesis is false.
- The alternative hypothesis states that there is an effect or a difference.

Hypothesis Testing

- The hypothesis test is conducted by comparing the value of the test statistic to a critical value.
- The critical value is a value that determines whether the null hypothesis can be rejected.
- If the test statistic is more extreme than the critical value, then the null hypothesis is rejected.

Hypothesis Testing

- The p-value is the probability of observing a test statistic as extreme as the one computed from the sample data, assuming that the null hypothesis is true.
- If the p-value is less than the significance level, then the null hypothesis is rejected.
- The significance level is the probability of rejecting the null hypothesis when it is true.

Hypothesis Testing

- The p-value is a measure of the strength of the evidence against the null hypothesis.
- The smaller the p-value, the stronger the evidence against the null hypothesis.
- The p-value is compared to the significance level to determine whether the null hypothesis should be rejected.

Hypothesis Testing in R

- In R, the `t.test()` function is used to perform hypothesis tests.
- The `t.test()` function takes in the sample data and the null hypothesis as arguments.
- The function returns the test statistic, the p-value, and the confidence interval.

Hypothesis Testing in R

- The `t.test()` function can be used to perform one-sample t-tests, two-sample t-tests, and paired t-tests.
- The `t.test()` function can also be used to perform one-sample z-tests and two-sample z-tests.
- The `t.test()` function can be used to perform hypothesis tests for means, proportions, and variances.

Hypothesis Testing in R

```
# One-sample t-test
t.test(x, mu = 0)

# Two-sample t-test
t.test(x, y)

# Paired t-test
t.test(x, y, paired = TRUE)
```

Table of Contents

Regression Analysis

- Regression analysis is a statistical method that is used to model the relationship between a dependent variable and one or more independent variables.
- The goal of regression analysis is to estimate the parameters of the regression model that best fit the data.
- The regression model is a mathematical equation that describes the relationship between the dependent variable and the independent variables.

Regression Analysis

- There are many types of regression models, such as linear regression, logistic regression, and polynomial regression.
- Linear regression is a regression model that assumes a linear relationship between the dependent variable and the independent variables.
- Logistic regression is a regression model that is used when the dependent variable is binary.

Regression Analysis

- Polynomial regression is a regression model that is used when the relationship between the dependent variable and the independent variables is not linear.
- The regression model is estimated using the method of least squares, which minimizes the sum of the squared differences between the observed values and the predicted values.
- The estimated parameters of the regression model are the coefficients of the independent variables.

Regression Analysis

- The goodness of fit of the regression model is measured using the coefficient of determination, which is the proportion of the variance in the dependent variable that is explained by the independent variables.
- The coefficient of determination ranges from 0 to 1, with higher values indicating a better fit.
- The significance of the regression model is tested using the F-test, which tests whether the regression model is a better fit than a model with no independent variables.

Regression Analysis in R

- In R, the `lm()` function is used to fit linear regression models.
- The `lm()` function takes in the formula for the regression model and the data as arguments.
- The formula specifies the dependent variable and the independent variables in the regression model.

Regression Analysis in R

- The `summary()` function is used to display the results of the regression analysis.
- The `summary()` function displays the estimated coefficients, the standard errors, the t-values, and the p-values of the regression model.
- The `summary()` function also displays the coefficient of determination and the results of the F-test.

Regression Analysis in R

```
# Fit linear regression model
model <- lm(y ~ x1 + x2, data = data)

# Display results
summary(model)
```

Table of Contents

Time Series Analysis

- Time series analysis is a statistical method that is used to model and forecast time series data.
- Time series data is a sequence of observations that are recorded at regular intervals over time.
- Time series analysis is used in many fields, such as economics, finance, and engineering, to analyze and predict trends and patterns in the data.

Time Series Analysis

- Time series analysis is used to model the underlying structure of the time series data and to make forecasts of future values.
- The goal of time series analysis is to identify the patterns and trends in the data and to make predictions based on these patterns and trends.
- Time series analysis is used to analyze and forecast time series data using statistical models, such as autoregressive integrated moving average (ARIMA) models and exponential smoothing models.

Time Series Analysis

- ARIMA models are a class of models that are used to model time series data that exhibit autocorrelation and seasonality.
- ARIMA models are specified by three parameters: the autoregressive order (p), the differencing order (d), and the moving average order (q).
- Exponential smoothing models are a class of models that are used to model time series data that exhibit trend and seasonality.

Time Series Analysis

- Exponential smoothing models are specified by three parameters: the level parameter (alpha), the trend parameter (beta), and the seasonality parameter (gamma).
- The parameters of the time series models are estimated using the method of maximum likelihood, which maximizes the likelihood of the observed data given the model.
- The goodness of fit of the time series model is measured using the mean squared error (MSE) and the Akaike information criterion (AIC).

Time Series Analysis in R

- In R, the `arima()` function is used to fit ARIMA models to time series data.
- The `arima()` function takes in the time series data and the parameters of the ARIMA model as arguments.
- The parameters of the ARIMA model are specified by the order argument, which is a vector of the autoregressive order, the differencing order, and the moving average order.

Time Series Analysis in R

- In R, the `ets()` function is used to fit exponential smoothing models to time series data.
- The `ets()` function takes in the time series data and the parameters of the exponential smoothing model as arguments.
- The parameters of the exponential smoothing model are specified by the model argument, which is a string that specifies the type of exponential smoothing model to fit.

Time Series Analysis in R

```
# Load packages
library(forecast)

# Generate time series data
data <- ts(data, start = 1, end = n, frequency = f)

# Fit ARIMA model
model <- arima(data, order = c(p, d, q))

# Fit exponential smoothing model
model <- ets(data, model = "ZZZ")
```

Table of Contents

Instrumental Variables

- Instrumental variables are used in econometrics to estimate the causal effect of an independent variable on a dependent variable.
- Instrumental variables are used when the independent variable is correlated with the error term in the regression model.
- Instrumental variables are used to identify the causal effect of the independent variable by removing the correlation between the independent variable and the error term.

Instrumental Variables

- Instrumental variables are variables that are correlated with the independent variable but are uncorrelated with the error term.
- Instrumental variables are used to estimate the causal effect of the independent variable by using the variation in the instrumental variables to identify the causal effect.
- Instrumental variables are used in regression analysis to estimate the parameters of the regression model that best fit the data.

Instrumental Variables in R

- In R, the `ivreg()` function is used to estimate instrumental variables regression models.
- The `ivreg()` function takes in the formula for the regression model, the data, and the instrumental variables as arguments.
- The formula specifies the dependent variable, the independent variables, and the instrumental variables in the regression model.

Instrumental Variables in R

- The `ivreg()` function estimates the parameters of the regression model using the method of instrumental variables.
- The `ivreg()` function returns the estimated coefficients, the standard errors, the t-values, and the p-values of the regression model.
- The `ivreg()` function is used to estimate the causal effect of the independent variable on the dependent variable by removing the correlation between the independent variable and the error term.

Instrumental Variables in R

```r
# Generate some data
set.seed(123)
n <- 100
x <- rnorm(n)
z <- rnorm(n)
y <- 1 + 2 * x + 3 * z + rnorm(n)

# Fit instrumental variables regression model
model <- ivreg(y ~ x | z)
summary(model)
```

Table of Contents

Presenting Results

- When presenting the results of a statistical analysis, it is important to clearly communicate the findings to the audience.
- The results should be presented in a clear and concise manner that is easy to understand.
- The results should be presented in a format that is appropriate for the audience and the purpose of the analysis.

Stargazer

- Stargazer is an R package that is used to create tables of regression results.
- Stargazer is used to create tables of regression results that can be easily formatted and exported to different file formats.
- Stargazer is used to create tables of regression results that can be included in reports, presentations, and publications.

Stargazer

- Stargazer is used to create tables of regression results that include the estimated coefficients, the standard errors, the t-values, and the p-values of the regression model.
- Stargazer is used to create tables of regression results that can be customized with different formatting options, such as bolding the coefficients and adding stars to indicate significance.
- Stargazer is used to create tables of regression results that can be exported to different file formats, such as HTML, LaTeX, and Word.

Stargazer

```
# Load packages
library(stargazer)

# Generate some data
set.seed(123)
n <- 100
x1 <- rnorm(n)
x2 <- rnorm(n)
y <- 1 + 2 * x1 + 3 * x2 + rnorm(n)

# Fit linear regression model
model <- lm(y ~ x1 + x2, data = data)

# Create table of regression results
stargazer(model, type = "html", out = "results.html")
```

Plotting Linear Regression

- When presenting the results of a linear regression analysis, it is often helpful to include a plot of the data and the regression line.
- The plot of the data and the regression line can help to visualize the relationship between the dependent variable and the independent variables.
- The plot of the data and the regression line can help to identify patterns and trends in the data and to assess the goodness of fit of the regression model.

Plotting Linear Regression

- In R, the `plot()` function is used to create a scatter plot of the data.
- The `abline()` function is used to add a regression line to the plot.
- The `points()` function is used to add points to the plot.

Plotting Linear Regression

```
# Generate some data
set.seed(123)
n <- 100
x <- rnorm(n)
y <- 1 + 2 * x + rnorm(n)

# Create scatter plot of data
plot(x, y)

# Add regression line to plot
abline(model)

# Add points to plot
points(x, y)
```