# Part A: Data Analysis

## Summary

This report shows an analysis done to determine if the presence of a sales charge in mutual funds will cause a difference on the mutual fund returns. A T-test was used to compare the categorical group against the various returns. The results collected are displayed in tables for comparison and an analysis was done.

## Description of Problem

Based on a data sample of 158 mutual funds, the project aims to determine if the **presence of sales charges (fees) in funds** affects the following numerical variables:

1. **2001 Return:** Twelve-month return in 2001
2. **Three Year Return:** Annualized return from 1999-2001
3. **Five Year Return**: Annualized return from 1997-2001

## Description of Data

The following box plots below show the difference in the means between having no sales charges and having sales charges tested on different returns.

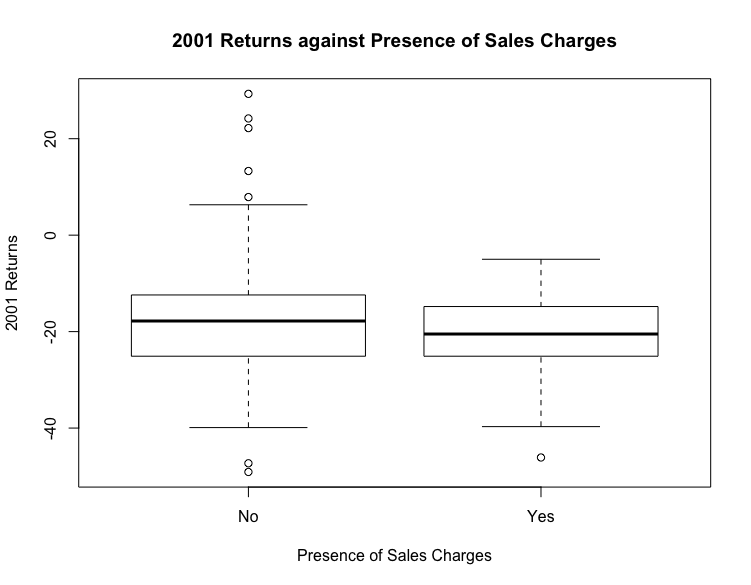


Figure 3.1 Box plot of 2001 Returns against Presence of Sales Charges

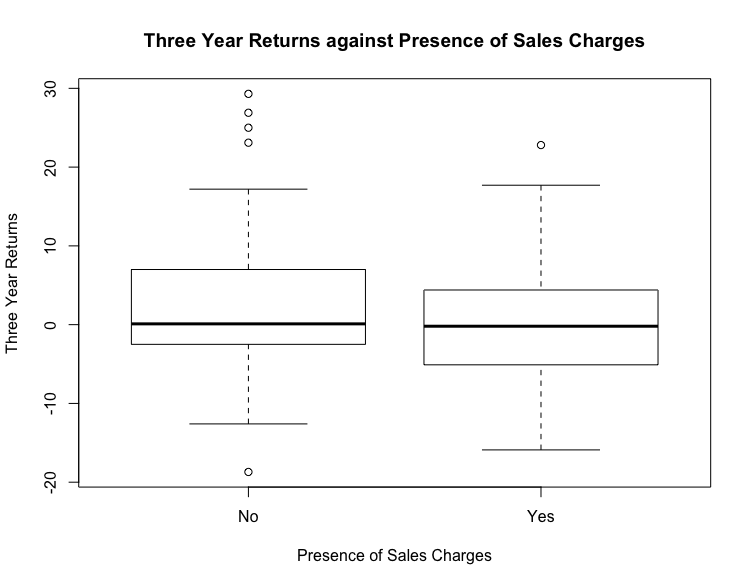


Figure 3.2 Box plot of Three Year Returns against Presence of Sales Charges

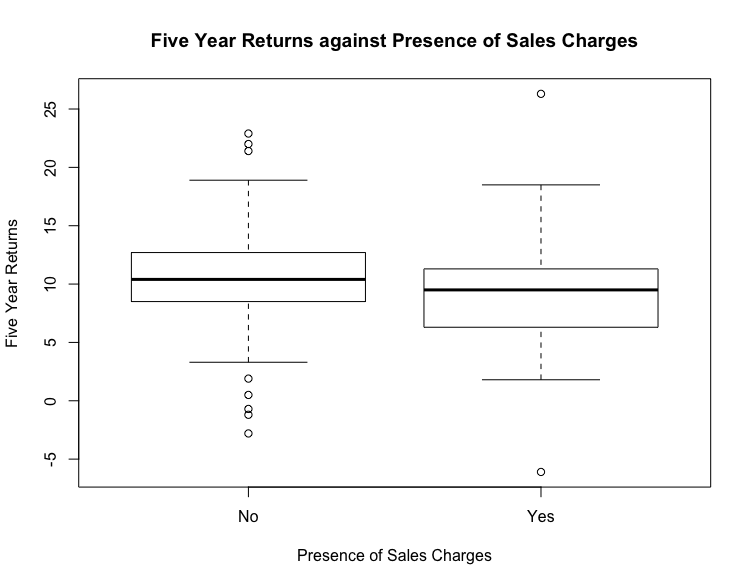


Figure 3.3 Box plot of Five Year Returns against Presence of Sales Charges

Based on the box plots above, it can be shown that the means between having or not having sales charges against different returns are different. However, further tests need to be done to ensure the claim in which the means are different are valid. This test will be discussed in the next section.

## Discussion of Statistic Analysis Method Used

A T-test is used for this experiment.

### Hypothesis Setup

The hypotheses are set up as such:

**Null Hypothesis:** The means between having sales charges and no sales charges have no significant difference between each other.

**Alternative Hypothesis:** There is a significant difference between the means of having sales charges and no sales charges.

### Assumptions Made

For the T-test to be valid, the following assumptions are explained below.

#### Continuous Dependent Variable

Since the 2001 Returns, Three-Year Returns, Five-Year Returns are all measured variables, all three of them can be considered as continuous variables.

#### Bivariate Independent Variable

Since the independent variable, *fees*, has exactly two categories (No and Yes), *fees* is said to be bivariate.

#### Normality of Variables

The numerical variables must follow a normal distribution for the T-test to be valid.

A Shapiro-Wilk test for normality was conducted on the variables. The test was conducted at the 1% significance level. This means that any variable that produces a p-value smaller than 0.01 follows a normal distribution.

The results of the test are shown in the table below.

|  |  |  |
| --- | --- | --- |
| **Test Variable** | **p-value** | **Follow Normal Distribution?** |
| 2001 Return | 5.946e-6 | Yes |
| Three Year Return | 2.254e-5 | Yes |
| Five Year Return | 0.006959 | Yes |

Table 4.1 Shapiro-Wilk Test for Normality on numerical variables.

Based on the test results, all test variables produced a p-value smaller than 0.01. Therefore it can be said that the three variables follow a normal distribution.

#### Equal Variances

The population variances must be equal across the numerical variables for the group levels.

Levene’s Test for Equal Variances was used. The test is conducted at the 5% significance level. The null hypothesis for the test states that variances between two variable pairs are equal.

The p-values obtained from the test are shown in the table below.

|  |  |  |
| --- | --- | --- |
| **Test Variable** | **p-Value** | **Violation of Equal Variance Assumption** |
| 2001 Returns | 2.51e-05 | Yes |
| Three Year Returns | 0.001966 | Yes |
| Five Year Returns | 0.0004714 | Yes |

Table 4.2 Levene’s Test for Equal Variances.

All test variables produced p-values less than 0.05, which means the null hypotheses for these variables are rejected at the 5% significance level, showing that there is evidence of unequal variances for each variable within the asset type groups. Therefore, Welch Correction must be used to test the variables.

## Interpretation of Test Results

The T-test results are shown in the table below.

|  |  |  |  |
| --- | --- | --- | --- |
| **Test Variable** | **Mean** | | **p-Value** |
| **No** | **Yes** |
| 2001 Returns | -17.60 | -21.21 | 0.004329 |
| Three Year Returns | 2.243 | 0.7930 | 0.2693 |
| Five Year Returns | 10.55 | 9.439 | 0.1694 |

Table 5 T-test for Fees against returns.

Based on the results in Table 5, the p-values generated for the different return categories against the presence of sales charges all fall below 0.05. Therefore, for each test variable paired with presence of sales charges, the null hypothesis is rejected. Therefore, it can be said that there is evidence that there a significant different in the means between each asset type for the 2001 Returns, Three Year Returns, Five Year Returns, Best Quarter and Worst Quarter.

## Conclusion