**Hypothesis Testing**

Variables in consideration:

|  |  |
| --- | --- |
| **Variable Name** | **Description** |
| **annual\_inc** | The self-reported annual income provided by the borrower during registration. |
| **loan\_amt** | The listed amount of the loan applied for by the borrower. If at some point in time, the credit department reduces the loan amount, then it will be reflected in this value. |
| **funded\_amt** | The total amount committed to that loan at that point in time. |
| **dti** | Debt to Income ratio; A ratio calculated using the borrower’s total monthly debt payments on the total debt obligations, excluding mortgage and the requested LC loan, divided by the borrower’s self-reported monthly income |
| **loan\_status\_1** | Current status of the loan grouped as:  **default\_new**: Status is “default” or “charged off”  **current\_new**: Status is “current” or “in grace period”  **late**: Status is “late 16- 30 days” or “late 31-120 days” |
| int\_rate\_grp | Interest Rate on the loan grouped as:  **Low**: Interest rate < 10%  **Medium**: Interest rate between 10% - 18%  **High**: Interest rate > 18% |

Assumptions:

1. All hypotheses are evaluated at 95 % confidence interval.
2. Since the population SD is not known we will be using the t-statistic to test the hypotheses.
3. Since the number of samples > 30, t-statistic and the z-statistic are nearly the same. For 95 % confidence interval, the null hypothesis will fail to be rejected if: **-1.96 < t < 1.96**

Hypotheses on group means distributions for continuous variables (**annual\_inc, loan\_amt, funded\_amt & dti**) in **loan\_status\_1** (*default\_new and current new*):

1. **Null hypothesis**: The loan\_amt group means are the *same* for default\_new and current\_new levels of **loan\_status\_1**
2. **Alternate hypothesis:** The loan\_amt group means are *varying* for default\_new and current\_new levels of loan\_status\_1
3. **Test statistic:** t = -17.753
4. **p value corresponding to the test statistic:** p= 2.2 e-16
5. **Insight**: The calculated t statistic does not lie within the expected interval and p<0.05, **hence the null hypothesis is rejected**.

This means that the **difference in group means of loan amounts for default and current loans is statistically significant**. The average loan amounts are different for default and current loan categories.

1. **Null hypothesis**: The funded\_amt group means are the *same* for default\_new and current\_new levels of **loan\_status\_1**
2. **Alternate hypothesis**: The funded\_amt group means are *varying* for default\_new and current\_new levels of loan\_status\_1
3. **Test statistic:** t= -17.936
4. **p value corresponding to the test statistic:** p= 2.2 e-16
5. **Insight**: The calculated t statistic does not lie within the expected interval and p<0.05, **hence the null hypothesis is rejected**.

This means that the **difference in group means of funded amounts for default and current loans is statistically significant**. The average funded amounts are different for default and current loan categories.

1. **Null hypothesis**: The dti group means are the *same* for default\_new and current\_new levels of **loan\_status\_1**
2. **Alternate hypothesis**: The dti group means are *not the same* for default\_new and current\_new levels of loan\_status\_1
3. **Test statistic**: t = -4.0768
4. **p value corresponding to the test statistic**: p= 4.823e-05
5. **Insight:** The calculated t statistic does not lie within the expected interval and p<0.05, **hence the null hypothesis is rejected**.

This means that the **difference in group means of debt to income ratio for default and current loans is statistically significant**. The average debt to income ratios are different for default and current loan categories.

1. **Null hypothesis**: The annual\_inc group means are the *same* for default\_new and current\_new levels of **loan\_status\_1**
2. **Alternate hypothesis**: The annual\_inc group means are *not the same* for default\_new and current\_new levels of loan\_status\_1
3. **Test statistic**: t = -10.773
4. **p value corresponding to the test statistic**: p= 2.2e-16
5. **Insight**: The calculated t statistic does not lie within the expected interval and p<0.05, **hence the null hypothesis is rejected**.

This means that the **difference in group means of annual incomes for default and current loans is statistically significant**. The average loan amounts are different for default and current loan categories.

Hypotheses on group means distributions for continuous variables (**annual\_inc, loan\_amt, funded\_amt & dti**) in **int\_rate\_grp** (*high and low*):

1. **Null hypothesis**: The loan\_amt group means are the *same* for high and low levels of **int\_rate\_grp**
2. **Alternate hypothesis**: The loan\_amt group means are *not the same* for high and low levels of int\_rate\_grp
3. **Test statistic**: t=27
4. **p value corresponding to the test statistic:** p=2.2e-16
5. **Insight**: The calculated t statistic does not lie within the expected interval and p<0.05, **hence the null hypothesis is rejected**.

This means that the **difference in group means of loan amounts for high and low interest loans is statistically significant**. The average loan amounts are different for high and low interest loan categories.

1. **Null hypothesis**: The funded\_amt group means are the same for high and low levels of **int\_rate\_grp**
2. **Alternate hypothesis:** The funded\_amt group means are *not the same* for high and low levels of int\_rate\_grp
3. **Test statistic:** t=27.082
4. **p value corresponding to the test statistic:** p=2.2e-16
5. **Insight:** The calculated t statistic does not lie within the expected interval and p<0.05, **hence the null hypothesis is rejected**.

This means that the **difference in group means of funded amounts for high and low interest loans is statistically significant**. The average funded amounts are different for high and low interest loan categories.

1. **Null hypothesis**: The dti group means are the same for high and low levels of **int\_rate\_grp**
2. **Alternate hypothesis:** The dti group means are not the same for high and low levels of int\_rate\_grp
3. **Test statistic:** t=4.7989
4. **p value corresponding to the test statistic:** p=1.736e-06
5. I**nsight**: The calculated t statistic does not lie within the expected interval and p<0.05, **hence the null hypothesis is rejected**.

This means that the **difference in group means of debt to income ratio for high and low interest loans is statistically significant**. The average debt to income ratio are different for high and low interest loan categories.

1. **Null hypothesis**: The annual\_inc group means are the same for high and low levels of **int\_rate\_grp**
2. **Alternate hypothesis**: The annual\_inc group means are *not the same* for high and low levels of int\_rate\_grp
3. **Test statistic**: t=11.212
4. **p value corresponding to the test statistic**: p= 2.2e-16
5. **Insight**: The calculated t statistic does not lie within the expected interval and p<0.05, **hence the null hypothesis is rejected**.

This means that the **difference in group means of annual incomes for high and low interest loans is statistically significant**. The average annual incomes are different for high and low interest loan categories.