National University of the Altiplano
Faculty of Statistical Engineering and Computer Science
Theacher: Fred Torres Cruz
Author : Roberth Carlos GONZALES MAURICIO.

Commissioned Work - N° 002
loan analysis by gender.R

```
data <- read.csv("loans.csv")

female_loans <- data$female
male_loans <- data$male

mean_female <- mean(female_loans)
mean_male <- mean(male_loans)

# SD
sd_female <- sd(female_loans)
sd_male <- sd(male_loans)

# CV
cv_female <- (sd_female / mean_female) * 100
cv_male <- (sd_male / mean_male) * 100

# results
cat("Mean Female:", mean_female)
cat("Mean Male:", mean_male)
cat("SD Female:", sd_female)
cat("SD Male:", sd_male)
cat("CV Female:", round(cv_female, 0), "%")
cat("CV Male:", round(cv_male, 0), "%")

# bar plots
barplot(female_loans, main = "F", col = "pink", names.arg = 1:length(female_loans))
barplot(male_loans, main = "M", col = "lightblue", names.arg = 1:length(male_loans))

# Comparison
barplot(c(mean_female, mean_male),
        names.arg = c("Female", "Male"),
        col = c("pink", "lightblue"),
        main = "Comparison of Means by Gender",
        ylab = "Average Loan Amount")

# Boxplot
```

```
values <- c(female_loans, male_loans)
groups <- c(rep("Female", length(female_loans)), rep("Male", length(male_loans)))
boxplot(values ~ groups,
        col = c("pink", "lightblue"),
        main = "Loan Distribution by Gender",
        ylab = "Loan Amount")

# t-test
print(t.test(female_loans, male_loans))
```

Description and Justification of the Variables

1. Variable: `female`

Description: This variable represents the loan amounts granted to females. The values of this variable correspond to the amount of money loaned in each observation recorded for women.

Justification: This variable was selected to analyze the loan distribution among women, in order to assess whether there is any significant difference compared to the loans granted to men. Including this variable allows for a gender-based comparison and helps understand if loans vary between the two groups.

2. Variable: `male`

Description: This variable represents the loan amounts granted to males. Like the `Female` variable, the values of this variable reflect the amount of money loaned to men in each recorded observation.

Statistical Results

- **Mean (Female):** 1409.143
- **Mean (Male):** 1452.357
- **Standard Deviation (Female):** 889.9336
- **Standard Deviation (Male):** 662.5917
- **Coefficient of Variation (Female):** 63%
- **Coefficient of Variation (Male):** 46%

The analysis of descriptive statistics highlights differences in the loan amounts between female and male groups.

- The **mean loan amount** for females is approximately **1409.143**, while for males, it is slightly higher at **1452.357**. This suggests a marginal tendency for males to receive slightly larger loans on average within this dataset.

When examining the variability of loan amounts:

- The **standard deviation** for females (**SD = 889.9336**) is significantly larger than that for males (**SD = 662.5917**). This indicates that the loan amounts for females are more widely dispersed around their mean, whereas for males, the amounts are more closely clustered around their mean.

To understand the relative variability:

- The **coefficient of variation (CV)** for females is **63%**, reflecting a high degree of relative dispersion.
- In contrast, the **CV for males** is **46%**, indicating lower relative variability in their loan amounts.

Chart Female

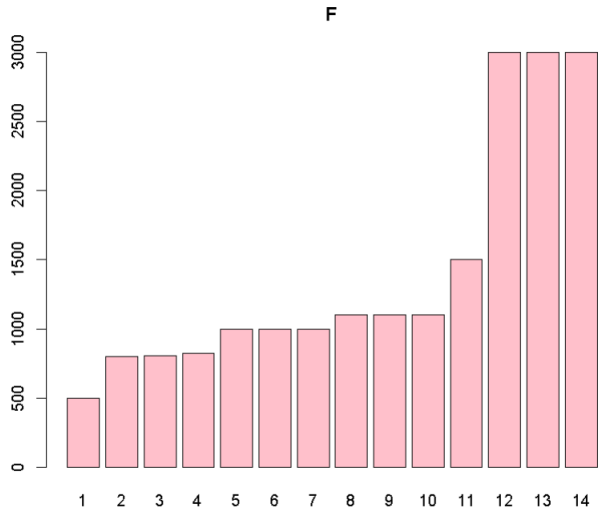


Chart F (Female)

It showed a distribution where the majority of observations were concentrated at relatively low values up to category 11, followed by a steep and significant increase in categories 12, 13, and 14, reaching the highest values. This suggested greater variability in loan amounts for women, with a notable proportion receiving significantly larger loans.

Chart Male

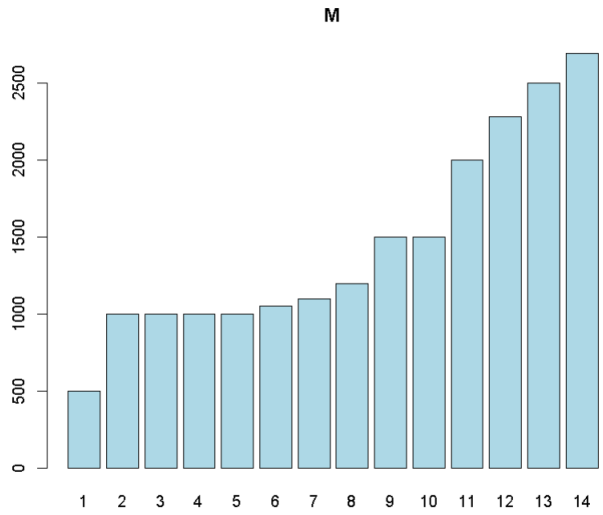
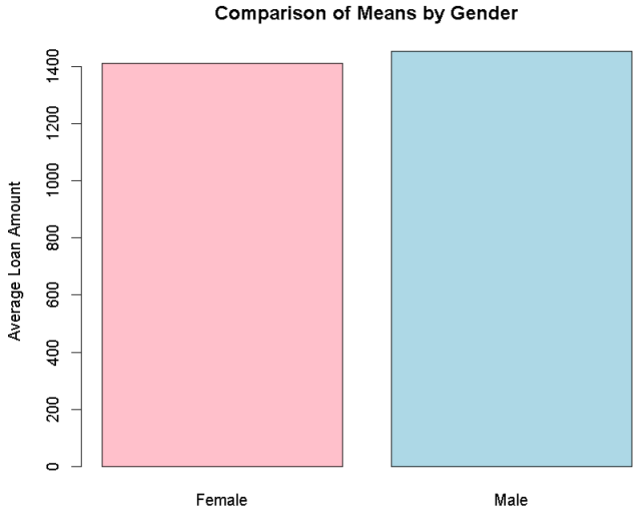


Chart M (Male)

This showed a distribution with a more gradual and sustained increase in loan amounts across categories. While the highest values were also found in the lower categories, the increase was less pronounced than in the female group. This suggested a smaller relative dispersion in loan amounts for men.

comparison chart



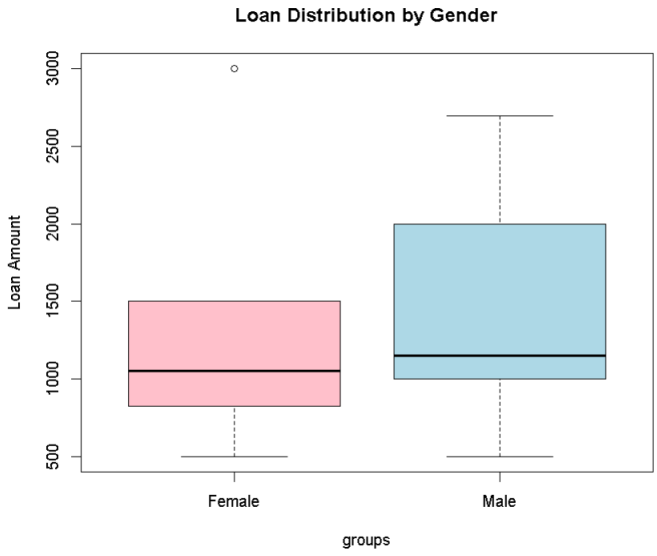
Bar Chart Analysis: Average Loan Amount by Gender

The bar chart compares the average loan amount between the female and male groups.

- **Horizontal Axis:** Represents the gender categories: "Female" and "Male."
- **Vertical Axis:** Represents the "Average Loan Amount."

It can be seen that the height of the bar for the male group is slightly higher than the height of the bar for the female group. This indicates that, on average, the loan amount for men in this data set is marginally higher than the average loan amount for women. However, visually, the difference between the heights of the two bars appears to be quite small.

loan distribution by gender



Box Plot Analysis: Loan Amount Distribution by Gender

This box plot summarizes the distribution of loan amounts for the female and male groups, providing information on the median, quartiles, interquartile range (IQR), and potential outliers.

Box for the Female Group (Pink)

- The horizontal line inside the box represents the median loan amount for women, which is around 1000.
- The lower and upper borders of the box indicate the first quartile (Q1) and third quartile (Q3), respectively. Approximately 25% of female loan amounts are below Q1 (around 800), and 25% are above Q3 (around 1500).

- The height of the box represents the interquartile range (IQR = Q3 - Q1), which contains the middle 50% of the data. For women, the IQR is approximately 700.
- The vertical lines extending from the box (the "whiskers") reach the most extreme values within a range of 1.5 times the IQR from the quartiles. The lower whisker reaches approximately 500, and the upper whisker extends to approximately 2000.
- The small circle above the upper whisker indicates an outlier, representing a significantly higher loan amount than the rest of the data for the female group (approximately 3000).

Box for the Male Group (Blue)

- The median loan amount for men is slightly above 1,100.
- The first quartile (Q1) is around 1,000, and the third quartile (Q3) is close to 2,000.
- The IQR for men is approximately 1,000.
- The whiskers extend from approximately 500 to approximately 2,700.

Comparison Between Groups

- **Median:** The median loan amount is slightly higher for men than for women.
- **Dispersion:** The box for men is higher (higher IQR) than the box for women, suggesting greater dispersion in the middle 50% of male loan amounts. However, when considering the full range of the data (including the whiskers), the dispersion appears similar between the two groups.
- **Outliers:** An outlier is observed in the female group, indicating exceptionally high borrowing. No clear outliers are identified in the male group within the 1.5 times IQR criterion.
- **Symmetry:** The distribution for women appears slightly skewed toward higher values due to the outlier. The distribution for men appears more symmetrical within the IQR.

Comparison Between Groups

```
        Welch Two Sample t-test

data:  female_loans and male_loans
t = -0.14573, df = 24.025, p-value = 0.8853
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -655.1858  568.7572
sample estimates:
mean of x mean of y
 1409.143  1452.357
```

t-statistic ($t = -0.14573$)

The t-statistic obtained is $t = -0.14573$. This value measures the difference between the means of the two samples in terms of standard error. A value near zero suggests that the difference between sample means is small relative to the data variability. The negative sign indicates that the mean loan amount for females is slightly less than that for males; however, the magnitude of the difference is minimal.

Degrees of Freedom (df $= 24.025$)

The estimated degrees of freedom for this test are df = 24.025. Unlike traditional t-tests, Welch's t-test does not assume equal variances, so the degrees of freedom are calculated using the Welch-Satterthwaite approximation, which can result in a non-integer value. This value determines the shape of the t-distribution used to compute the p-value.

p-value ($p = 0.8853$)

The p-value associated with the t-statistic is $p = 0.8853$. This represents the probability of observing a difference as extreme as the one seen, or more, assuming the null hypothesis of no real difference between population means is true.

- A high p-value (commonly above the significance level $\alpha = 0.05$) suggests that the observed data is consistent with the null hypothesis.
- Here, $p = 0.8853$ is significantly greater than 0.05, meaning we **fail to reject the null hypothesis**.

Alternative Hypothesis

The alternative hypothesis posited that there is a significant difference between the mean loan amounts for females and males. However, the high p-value indicates insufficient statistical

evidence to support this claim.

95% Confidence Interval for the Difference of Means (−655.1858 to 568.7572)

The 95% confidence interval for the mean difference between the groups is −655.1858 to 568.7572. This interval provides a plausible range for the true mean difference between the populations.

- Crucially, this interval includes zero, indicating that a true difference of zero between population means is plausible.
- This reinforces the conclusion of no statistically significant difference between the groups.

Sample Mean Estimates

The sample mean loan amounts are:

- **Female group (x)**: 1409.143
- **Male group (y)**: 1452.357

While there is a numerical difference of approximately 43.214 in favor of males, the Welch's t-test indicates that this difference is not statistically significant given the within-group variability and sample sizes.

https://github.com/robert1357/entropia.git