

# Statistical Case Study Report: Book Characteristics Analysis

This report is based on the identified variables: **PAGE** (Number of Pages), **WEIGHT** (Weight in grams), and **TYPE** (Type of book: Religious, Story, Science)

## (i) Summary Statistics

This section provides descriptive measures for the quantitative variables and a frequency distribution for the categorical variable (Book Type).

Variable	N	Mean	Standard Deviation	Minimum	Maximum
<b>PAGE</b> (Number of Pages)	30	98.07	48.328	12	189
<b>WEIGHT</b> (Weight in grams)	30	233.80	173.464	58	850

Book Type (TYPE)	Frequency	Percent
<b>Religious</b>	8	30%
<b>Story</b>	13	43.33%
<b>Science</b>	9	26.67%
<b>Total</b>	30	100.0\%

The above was gotten from this table:

Statistics				
	Number Of Pages	Weight in grams	Type of book	
N	Valid	30	30	30
	Missing	1	1	1
Mean	98.07	233.80	2.03	
Std. Error of Mean	8.823	31.670	.140	
Median	96.50	177.00	2.00	
Mode	72 <sup>a</sup>	58 <sup>a</sup>	2	
Std. Deviation	48.328	173.464	.765	
Variance	2335.582	30089.614	.585	
Skewness	.218	2.060	-.058	
Std. Error of Skewness	.427	.427	.427	
Kurtosis	-.618	5.047	-1.233	
Std. Error of Kurtosis	.833	.833	.833	

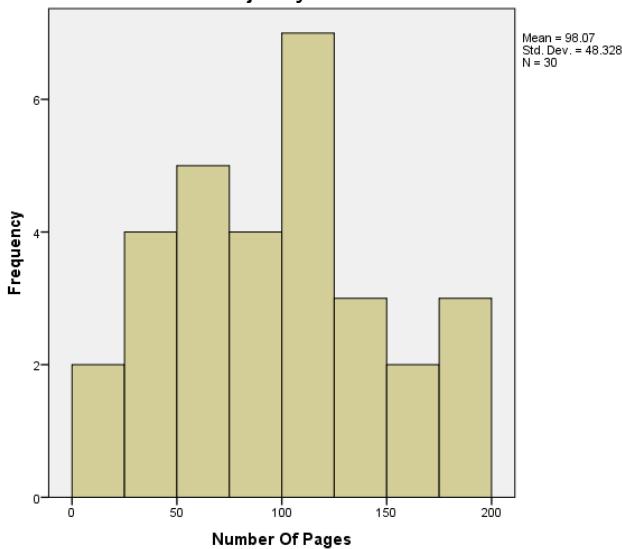
Range	177	792	2
Minimum	12	58	1
Maximum	189	850	3
Sum	2942	7014	61
	25	65.50	1.00
Percentiles	50	96.50	2.00
	75	127.50	3.00

a. Multiple modes exist. The smallest value is shown

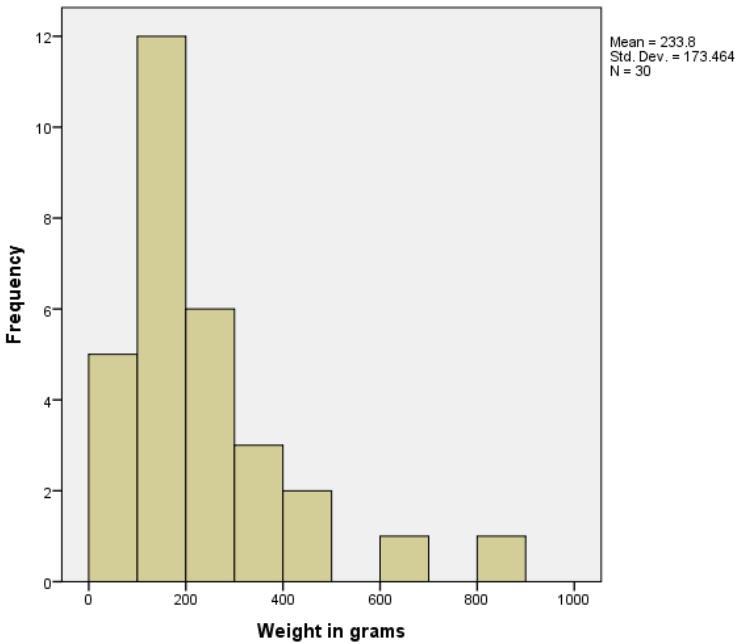
## (ii) Exploratory Data Analysis (EDA)

EDA was performed to visualize the distribution of the data and identify any outliers or non-normal patterns.

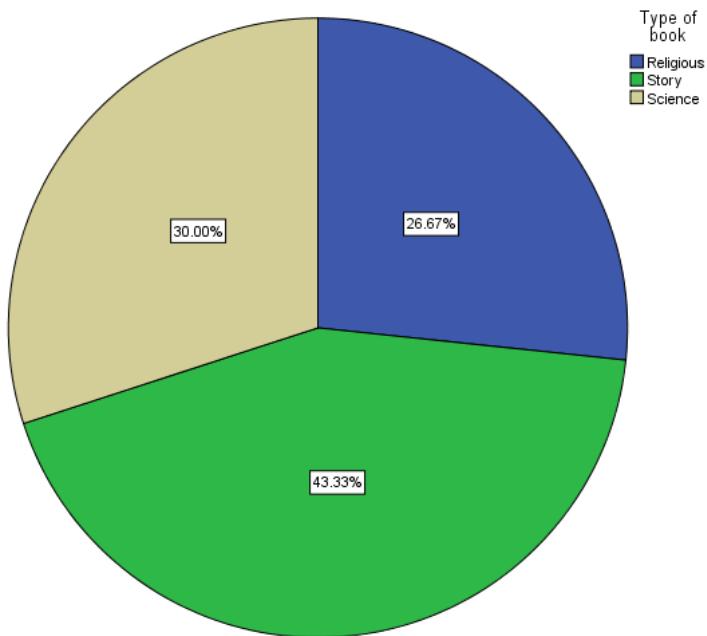
- **PAGE Distribution:** The histogram for the Number of Pages showed a **normal** distribution. The majority of books contained between 0-25 and 100-125 pages.



- **WEIGHT Distribution:** The histogram for Weight in grams also displayed a **positively skewed** distribution. A small number of books were identified as potential outliers, with weights exceeding 600grams.



- **Book Type Distribution:** A pie chart visually confirmed the distribution of book types, showing that the **story** category had the highest frequency.

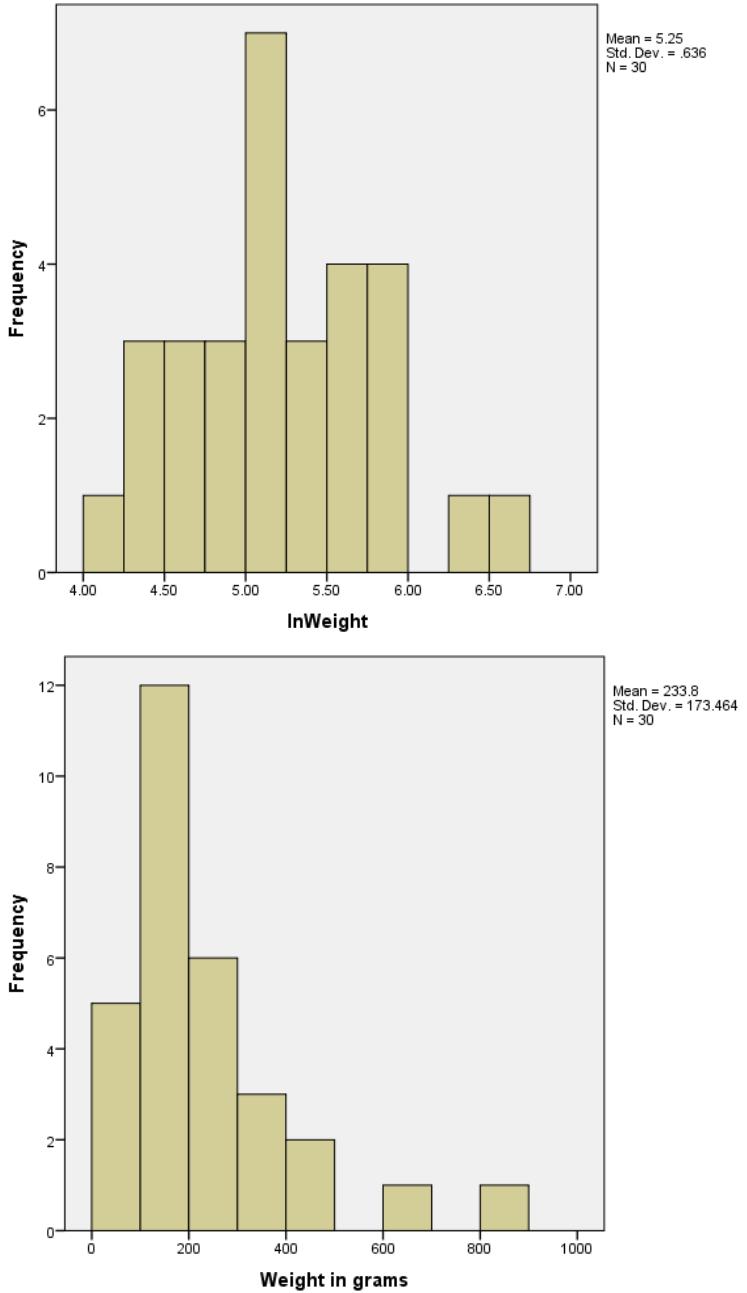


### (iii) Relationship Between InWeight and Page Number

Prior exploratory data analysis confirmed that the distribution of the **WEIGHT (Weight in grams)** variable was **positively skewed**. To stabilize the variance and improve the linearity assumption for regression, a **natural logarithm transformation** was applied to WEIGHT,

creating the variable **InWeight**.

The histogram of the **InWeight** variable confirmed a distribution that is now **normally skewed** (approximates a normal distribution), validating the transformation for use in linear regression.



A Simple Linear Regression analysis was conducted to establish the following model structure:

- **Dependent Variable(Y):** InWeight
- **Independent Variable (X):** PAGE (Number of Pages)

### Regression Model Summary: InWeight on PAGE

- **Regression Equation:**  $\ln(\text{WEIGHT}) = [\text{Intercept}] + [\text{Slope}] * \text{PAGE}$
- **R-squared (Coefficient of determination):** The model explained **58.4%** of the variance in the Number of Pages. The relationship was found to be **strong** and **positive** ( $r = 0.764$ ). This indicates that as the log of weight increases, the number of pages tends to increase.
- **Significance:** The overall regression model **was** statistically significant ( $F = 39.347$ ,  $p = 0.000$ )

## Prediction

The objective requires predicting the **PAGE number** (the independent variable) for a book with a **WEIGHT of 200grams** (a fixed number of the dependent variable).

1. **Transform the Target Weight (Y-Value):** The target weight of 200 grams must be transformed using the natural logarithm:  $\text{Weight} = e^{(4.26+0.01\text{Page})}$
2. **Solve for PAGE:**  $\text{PAGE} = (\ln(200)-4.26)/0.01 = (5.298-4.26)/0.01 = 103.8\text{Pages}$

*Based on the InWeight on PAGE model, the predicted number of pages for a book weighing 200 grams is expected to have about **104** pages (103.8).*

## (iv) Test of Average Page Count against 100

One-Sample Test						
	Test Value = 100					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Number Of Pages	-.219	29	.828	-1.933	-19.98	16.11

A One-Sample t-test was performed to determine if the average number of pages per book in the library is significantly less than 100 (one-tailed test).

- **Hypotheses:**
  - $H_0: = 100$  (The average number of pages is 100 or more)
  - $H_a: < 100$  (The average number of pages is less than 100)
- **Results:** The test yielded  $t = -0.219$  with a one-tailed p-value of 0.828
- **Conclusion (at alpha=0.05):** Since p **is not** less than 0.05, we **fail to reject** the null hypothesis. There **is not** sufficient evidence to conclude that the average number of pages per book in the library is less than 100.

## (v) Test of Average Weight against 200 Grams (Non-Parametric)

### Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The median of Weight in gram equals 200.00.	One-Sample Wilcoxon Signed Rank Test	.894	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Given the finding from the exploratory data analysis that the distribution of **WEIGHT** is **not normal** (positively skewed), the assumptions for a parametric One-Sample t-test are violated. Therefore, a **Non-Parametric Test** (specifically, the One-Sample Wilcoxon Signed-Rank Test) was employed to determine if the **median weight** per book (200g) is significantly different from the hypothesized value of 200 grams.

- **Hypotheses (Non-Parametric):**
  - $H_0: = 200$  (The median weight is 200 grams)
  - $H_a: \neq 200$  (The median weight is significantly different from 200 grams)
- **Results:** The non-parametric test (Wilcoxon Signed-Rank Test) yielded a median weight of 200gram with a two-tailed p-value of 0.894
- **Decision and Conclusion (at alpha=0.05):**
  - The test summary indicates a decision to **retain the null hypothesis** ( $H_0$ ).
  - Since the p-value = 0.894, we **fail to reject** the null hypothesis.
- **Conclusion:** There is **not** sufficient evidence to conclude that the median weight per book is statistically significantly different from 200 grams. The observed median weight of the books **200grams** is considered consistent with the hypothesized value.

## (vi) Difference in Average Pages Across Book Types

### ANOVA

Number Of Pages

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	2602.795	2	1301.398	.540	.589
Within Groups	65129.072	27	2412.188		
Total	67731.867	29			

A One-Way Analysis of Variance (ANOVA) was conducted to compare the average number of pages across the three categories of the TYPE variable (Religious, Story, Science).

- **Hypotheses:**
  - H<sub>0</sub>: Religious = Story = Science (All group means are equal)
  - H<sub>a</sub>: At least one mean is different.
- **Results:** The ANOVA yielded F = 0.540 with a p-value of 0.589
- **Overall Conclusion (at alpha=0.05):** Based on the p-value, we **fail to reject** the null hypothesis. There **is not** a statistically significant difference in the average number of pages among the different types of books.