Laptop Usage Pattern

A statistical Analysis

Lahari | Hima Chandh | Anuraga Chandan | Suraj Kumar | Divija

IIT HYDERABAD | 2024

Key points

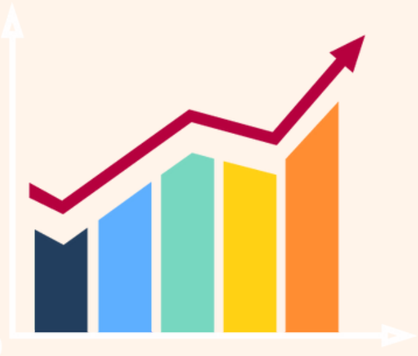
- Introduction
- Variables of Interest
- Data Visualization
- Data Analysis
    - Confidence Interval Estimation
    - Hypothesis Testing

Introduction

- Our statistics project focuses on analyzing a dataset comprising individuals computing preferences within an academic context. The dataset comprises various factors, including academic level, field of study, laptop brand, ratings on various features, primary usage, and operating system.
- Through this analysis, we aim to uncover patterns and preferences among students pursuing different academic paths, using different laptop brands, and engaging in diverse activities with their devices. By examining these factors collectively, we seek to gain insights into the relationship between academic pursuits, computing preferences, and usage patterns among students.

Data Collection

We gathered a representative sample of IITH students by soliciting voluntary participation in a survey. The survey covered diverse factors such as academic standing, field of study, laptop brand preference, rating, primary usage, and operating system, administered via email. We received a total of 245 responses from students enrolled in various programs, including Undergraduate, Postgraduate, and PhD.



Variables of Interest

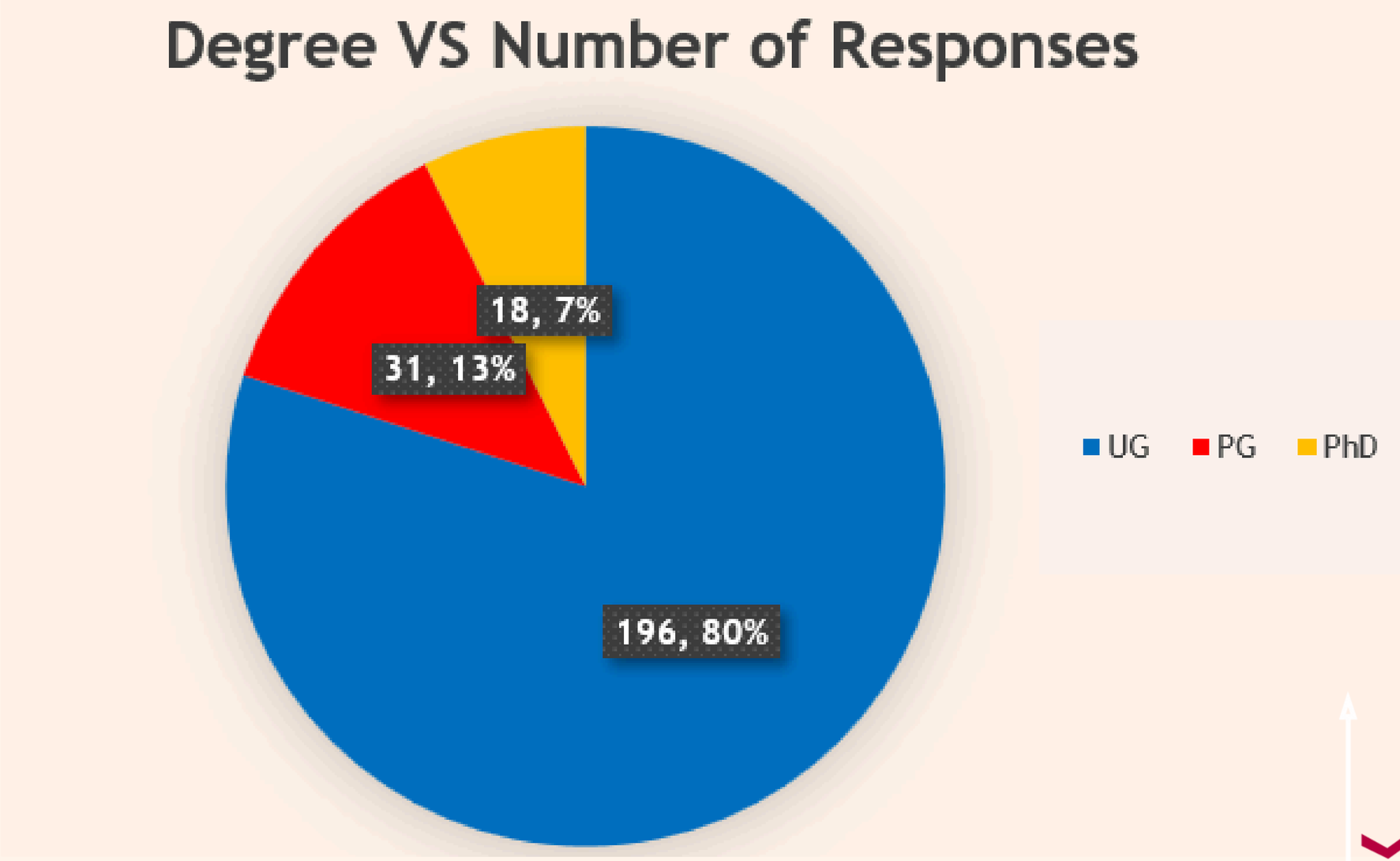
- Degree
- Laptop brand
- Hours spent on laptop per week
- Primary use of laptop
- Preferred operating system
- Importance given to following factors :Performance, Battery Life, Portability, Price, Brand, Design/Aesthetics

Data Visualization

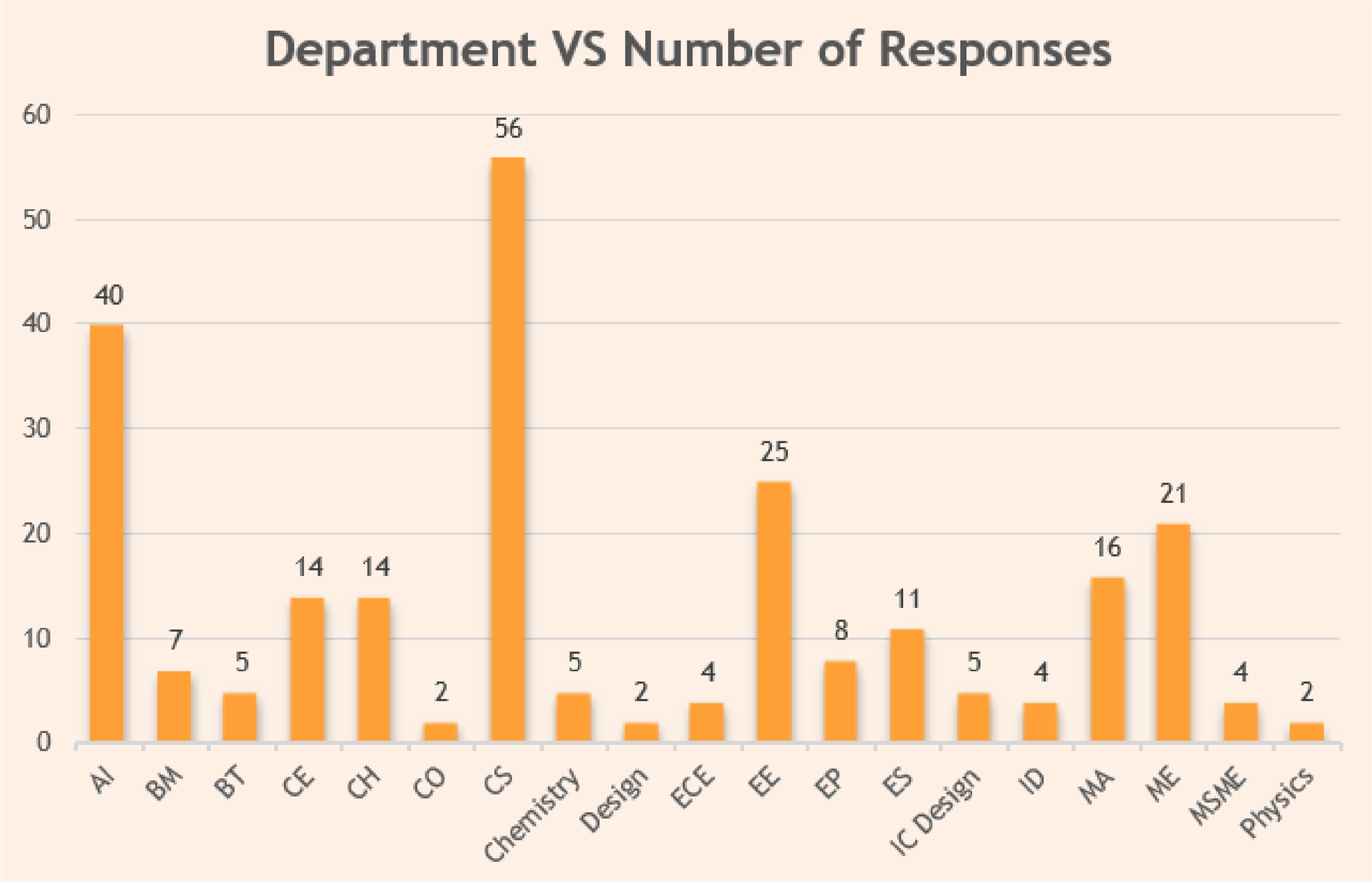
Central Tendency values of total laptop usage within a week

COUNT	245
MEAN	28.16
MEDIAN	25
MODE	25
STANDARD DEVIATION	13.55
MINIMUM	5
FIRST QUARTILE	17
SECOND QUARTILE	25
THIRD QUARTILE	40
MAXIMUM	50

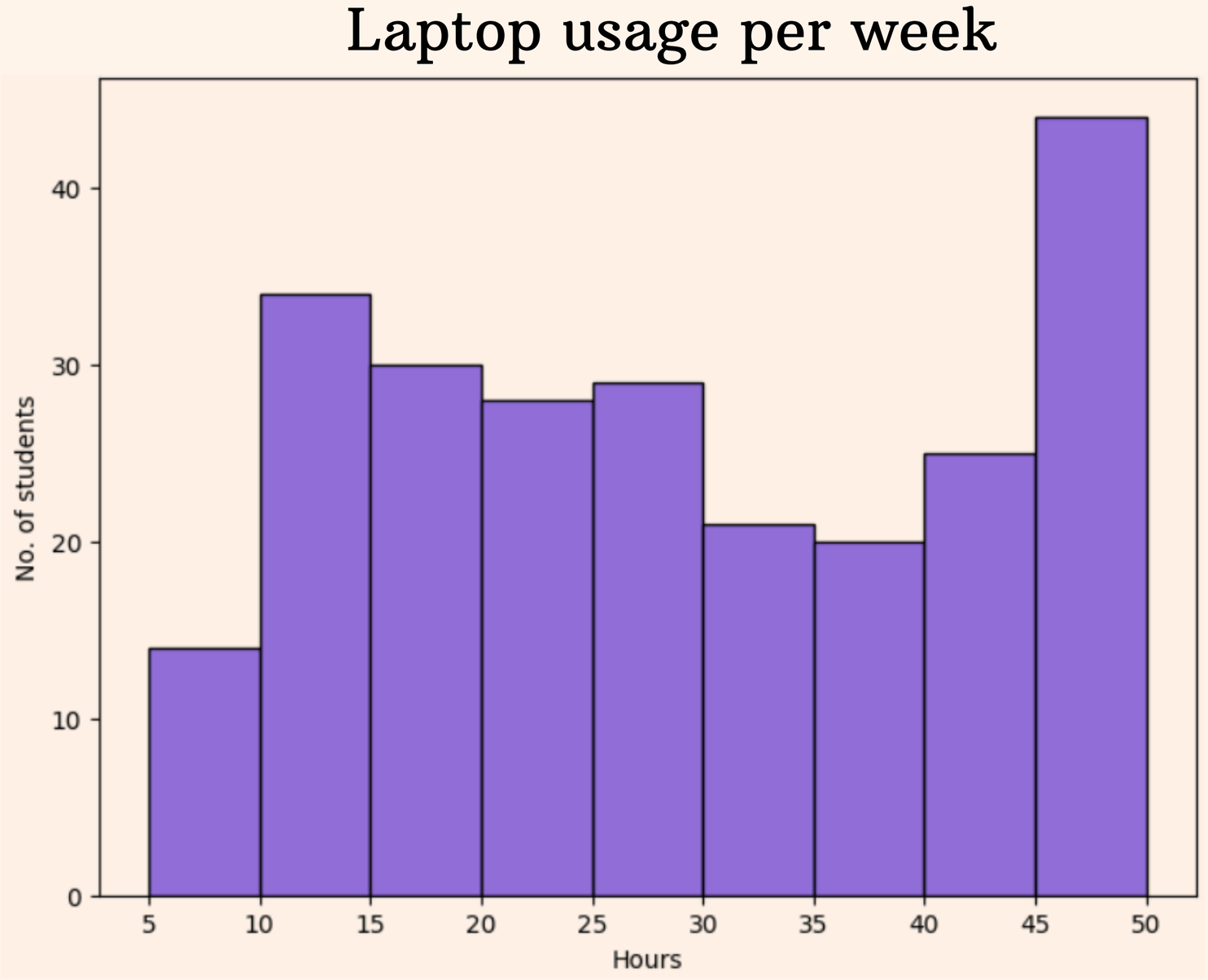
Data Visualization



Data Visualization

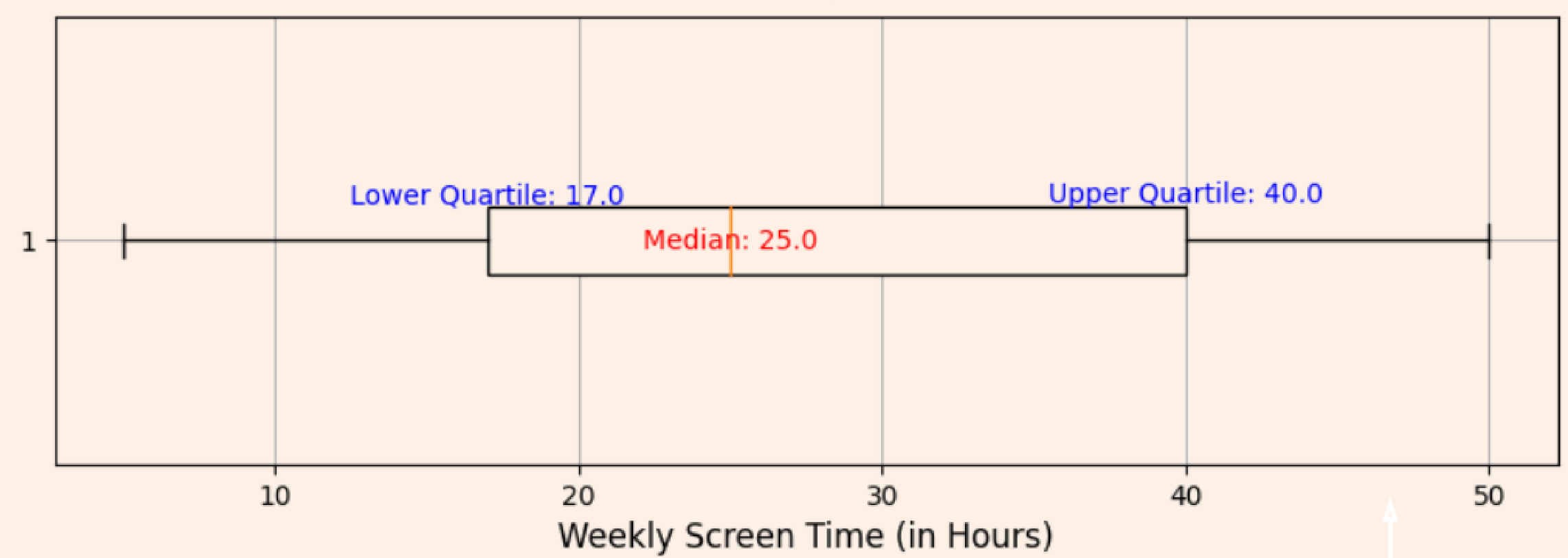


Data Visualization

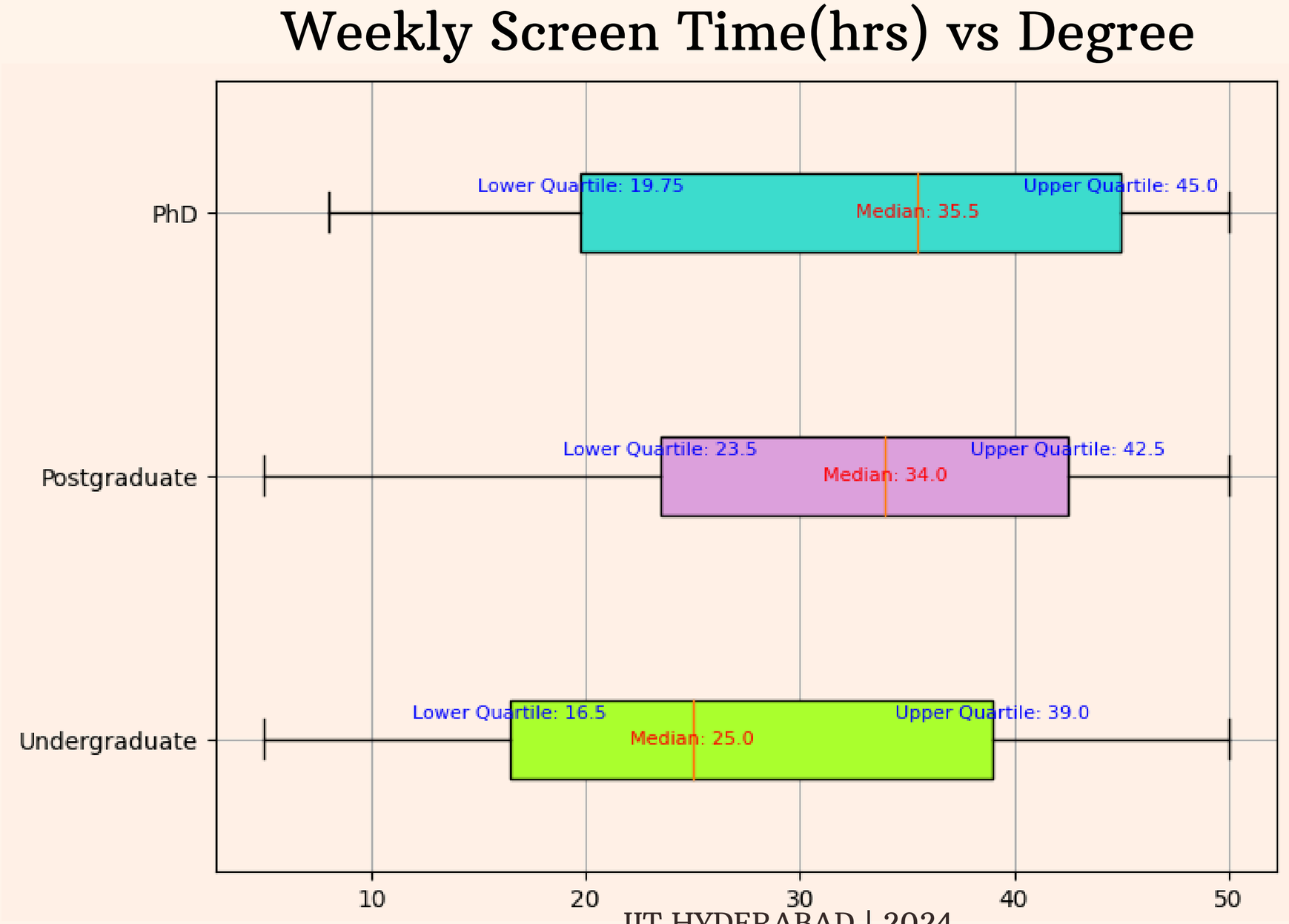


Data Visualization

Box Plot on weekly screen time

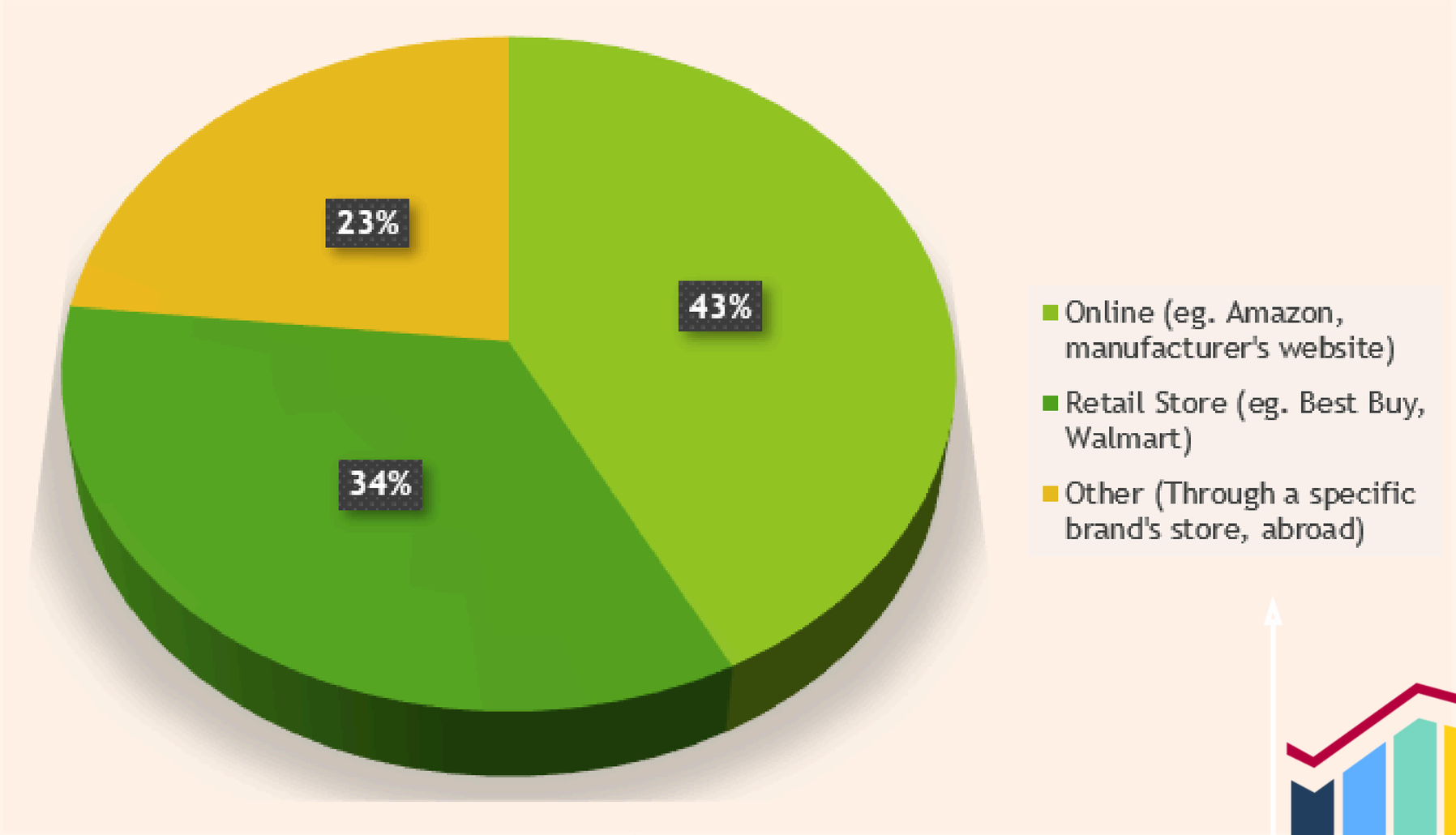


Data Visualization



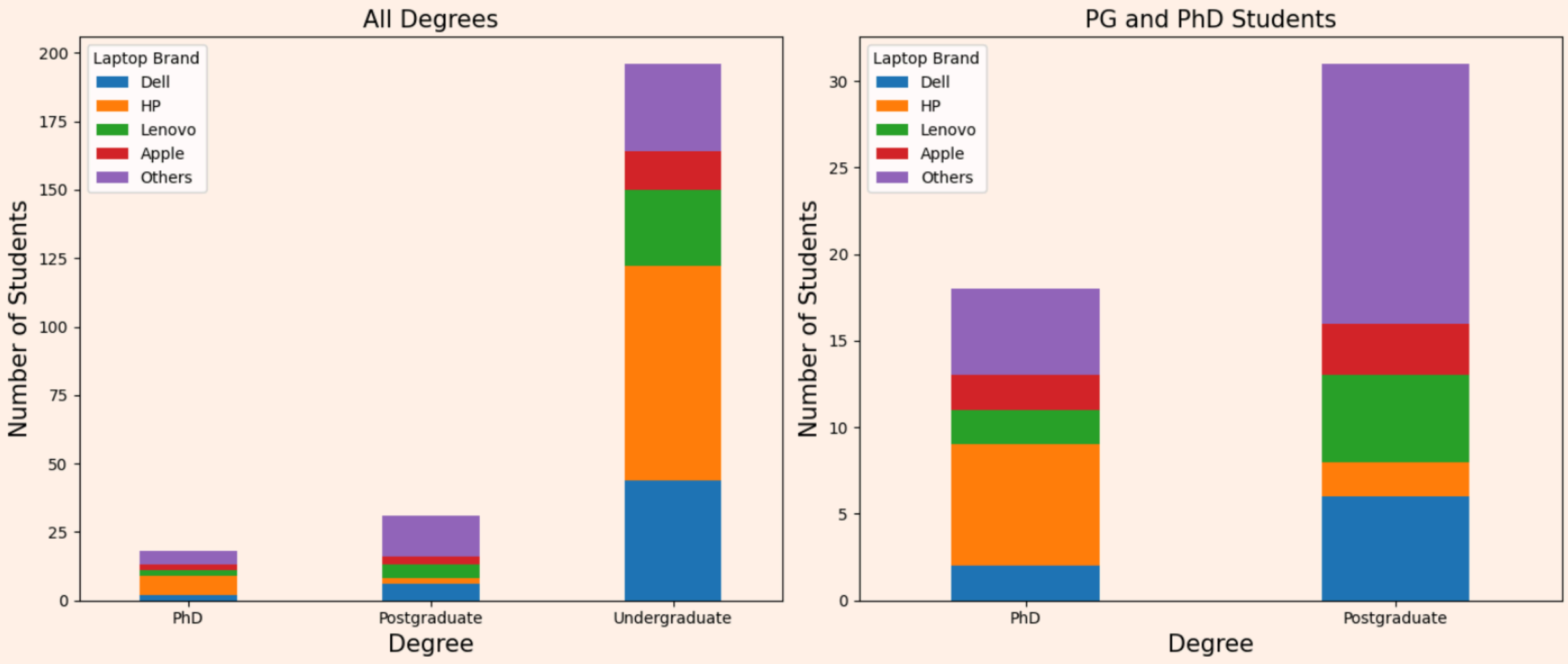
Data Visualization

Laptop purchase analysis



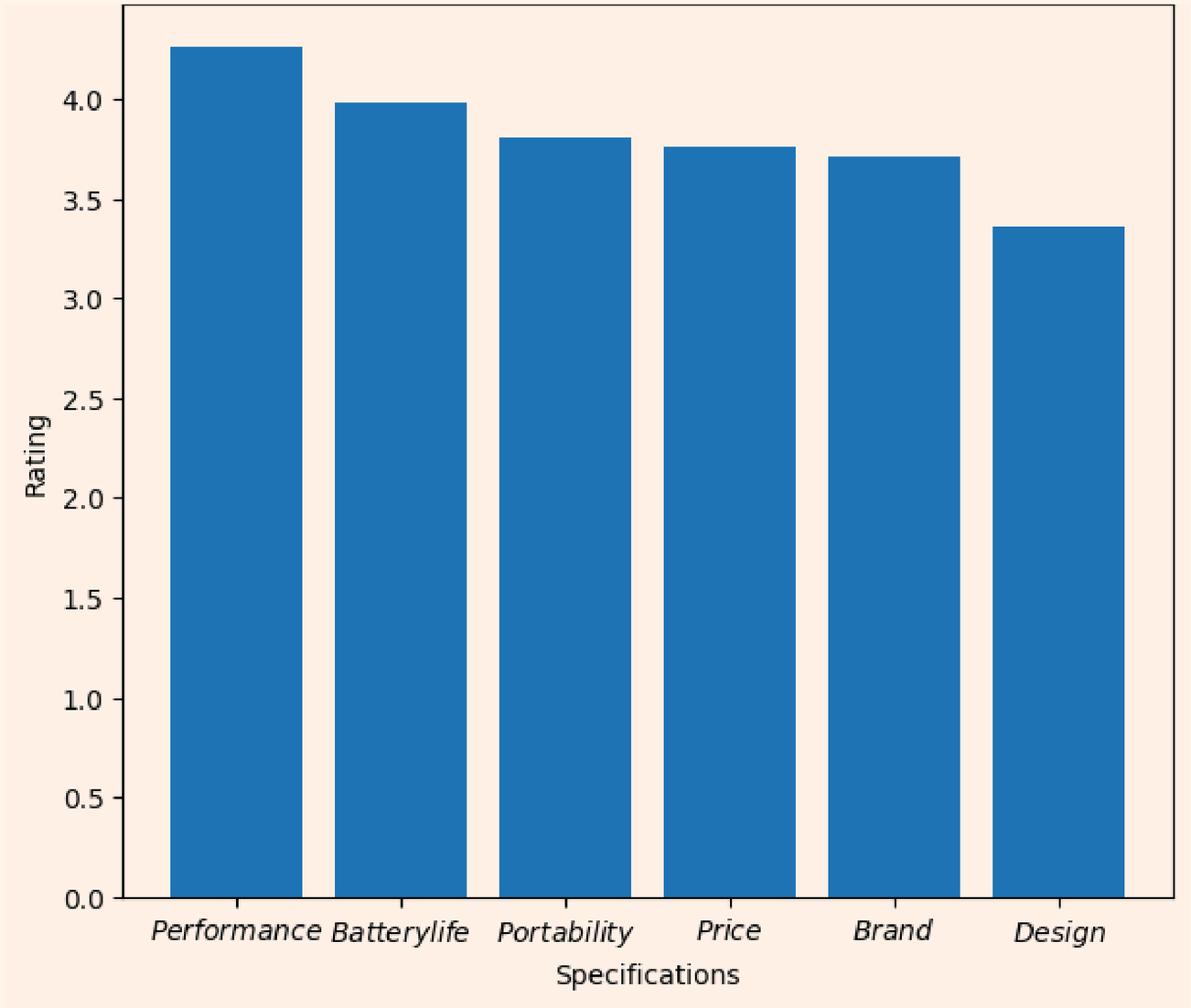
Data Visualization

Stacked bar Graph



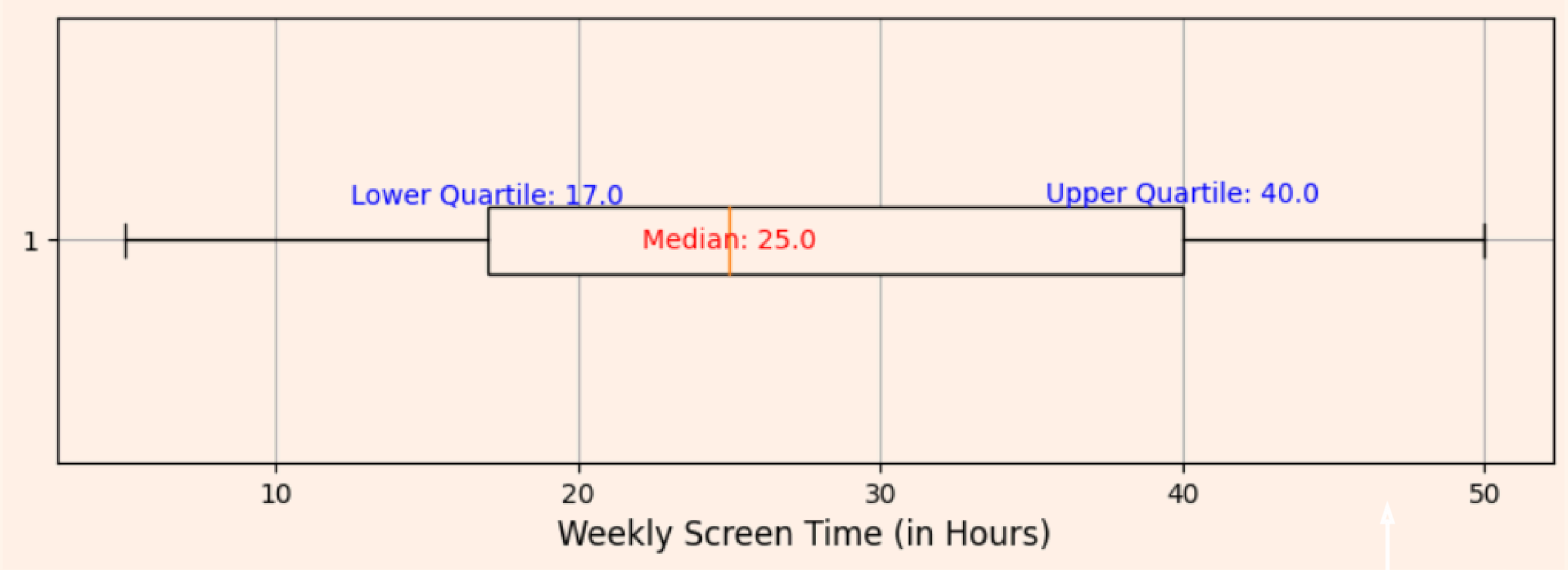
Data Visualization

Average rating of specifications of HP laptop



Analysis 1 : Screen Time

Considering screen time per week overall responses



Confidence Interval Estimation

i. Mean (μ):

$$\bar{X} = 28.1591$$
$$S = 13.5542$$
$$95\%\ CI \Rightarrow \alpha = 0.05$$
$$n = 245$$
$$t_{\alpha/2,\ n-1} = 1.9697$$
$$CI \equiv \left[\bar{X} - t_{\alpha/2,\ n-1}\left(\frac{S}{\sqrt{n}}\right), \bar{X} + t_{\alpha/2,\ n-1}\left(\frac{S}{\sqrt{n}}\right)\right]$$
$$\Rightarrow CI \equiv [26.4535, 29.8647]$$

$\therefore$ *The Confidence Interval for mean μ of the weekly screen time is*

$$[26.4535, 29.8647]$$

Confidence Interval Estimation

ii. Variance (σ):

$$S = 13.5542 \Rightarrow S^2 = 183.7163$$

$$\alpha = 0.05$$

$$n = 245$$

$$a = \chi^2_{1-\alpha/2,n-1} = \chi^2_{0.975,244} = 202.6272$$

$$b = \chi^2_{\alpha/2,n-1} = \chi^2_{0.025,244} = 289.1591$$

$$CI \equiv \left[\frac{(n-1)S^2}{b}, \frac{(n-1)S^2}{a}\right]$$

$$\Rightarrow CI \equiv [155.0246, 221.2278]$$

$\therefore$ *The Confidence Interval for the variance* σ^2 *of the weekly screen time is* $[155.0246, 221.2278]$ *and for standard deviation* σ *is* $[12.45, 14.8737]$.

Hypothesis Testing

- Is the mean of weekly screen time by student is more than 25hrs?

Hypothesis:

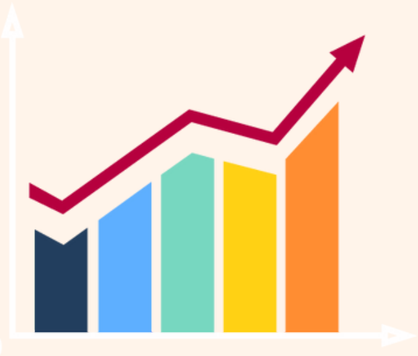
$$\mu_0 = 25$$

$$H_0: \mu \leq \mu_0 \qquad H_a: \mu > \mu_0$$

$$\bar{X} = 28.1591$$

$$S = 13.5542$$

$$n = 245 \Rightarrow df = n - 1 = 244$$



Hypothesis Testing

Test Statistic:

$$t^* = \frac{\bar{X} - \mu_0}{S/\sqrt{(n)}}$$

$$\Rightarrow t^* = \frac{28.1591 - 25}{13.5542/\sqrt{244}} = 3.6408$$

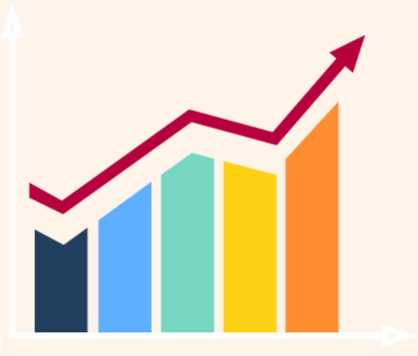
Rejection-Region Approach:

$$\alpha = 0.01$$

Reject if $t^* \geq t_{\alpha,n-1}$

$$t_{\alpha,n-1} = 2.3417$$

$\therefore$ Reject H_0



Hypothesis Testing

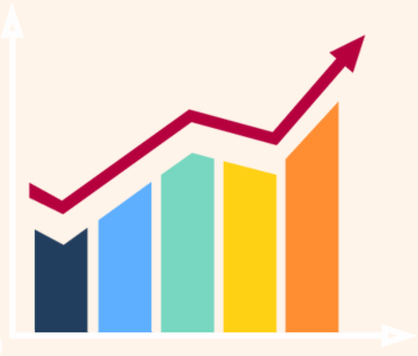
p-Value approach:

$$P(t \geq t^*) = 0.0003$$

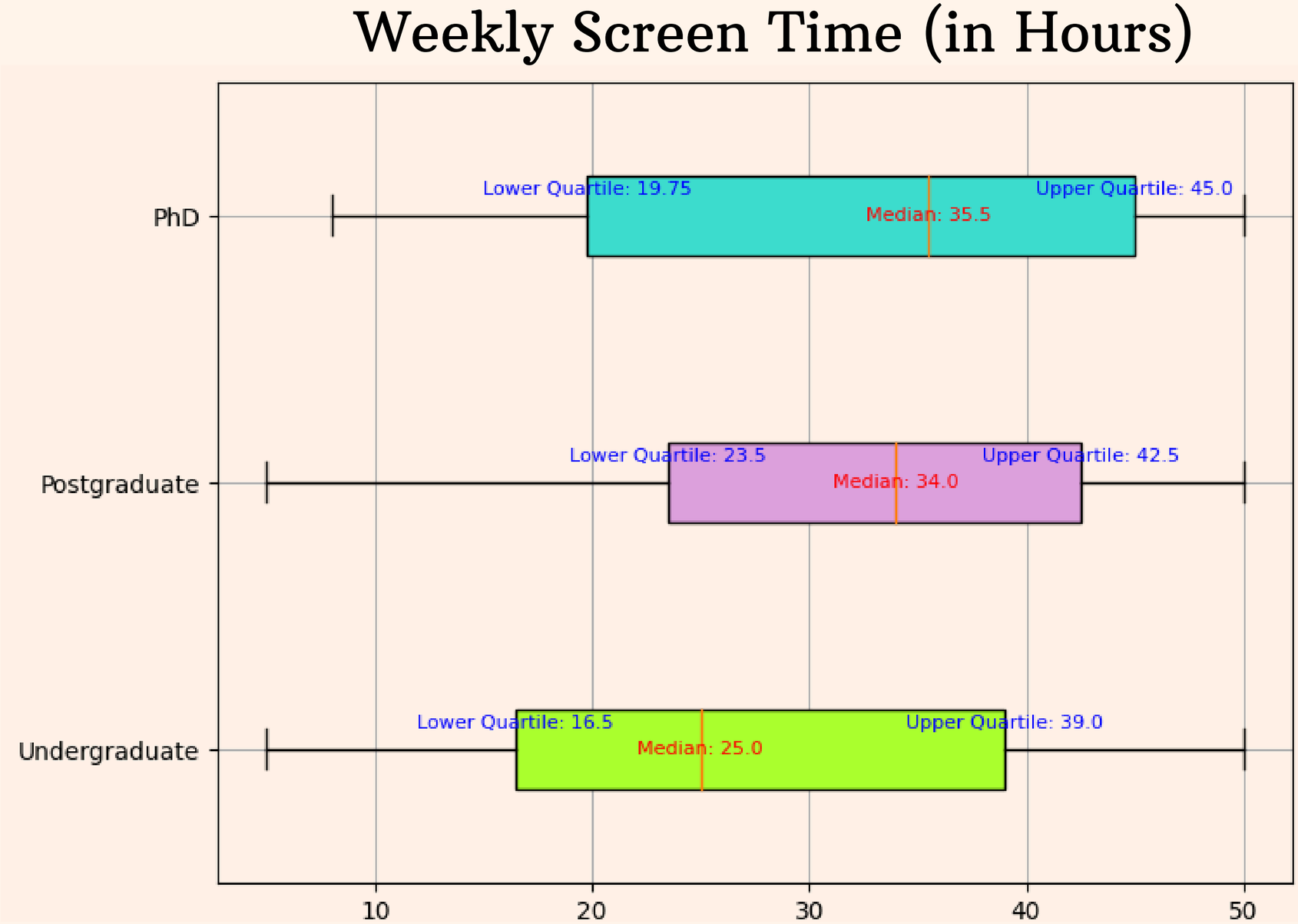
$$0.0003 \leq 0.01$$

$$p \leq \alpha$$

$$\Rightarrow \text{Reject } H_0$$



Considering 2 samples : UG and PG



Confidence Interval Estimation

For difference of means of screen time of UG ($\mu 1$)and PG ($\mu 2$)

Mean of UGs' screen time: $\overline{X_1} = 27.1122$

Mean of PGs' screen time: $\overline{X_2} = 32.13$

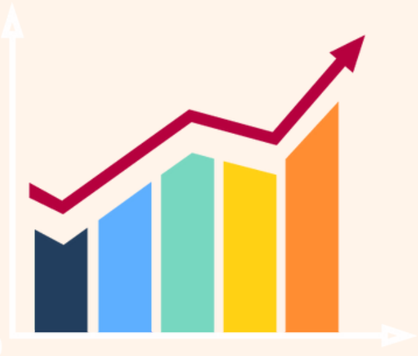
Number of UGs in the sample: $n = 196$

number of PGs in the sample: $m = 31$

$$S_1^2 = 13.2594$$

$$S_2^2 = 13.6326$$

$$\overline{X_1} - \overline{X_2} = -5.0178$$



Confidence Interval Estimation

$$\frac{{S_1}^2}{{S_2}^2} < 4 \Rightarrow \text{Two sample pooled interval}$$

$${S_p}^2 = \text{pooled sample variance} = \frac{(n-1)S_x^2+(m-1)S_y^2}{n+m-2}$$

$${S_p}^2 = \frac{(195 \times 13.2594) + (30 \times 13.6326)}{196 + 31 - 2} = 13.30916$$

$$S_p = 3.6482$$

$$t_{\alpha/2,m+n-2} = t_{0.025,225} = 1.97056$$

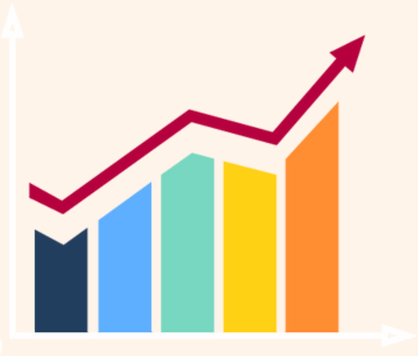
Confidence Interval Estimation

$$CI \equiv \left[\{\overline{X_1} - \overline{X_2}\} - t_{\alpha/2,n+m-2} \times S_p \sqrt{\frac{1}{n} + \frac{1}{m}}, \{\overline{X_1} - \overline{X_2}\} + t_{\alpha/2,n+m-2} \times S_p \sqrt{\frac{1}{n} + \frac{1}{m}} \right]$$

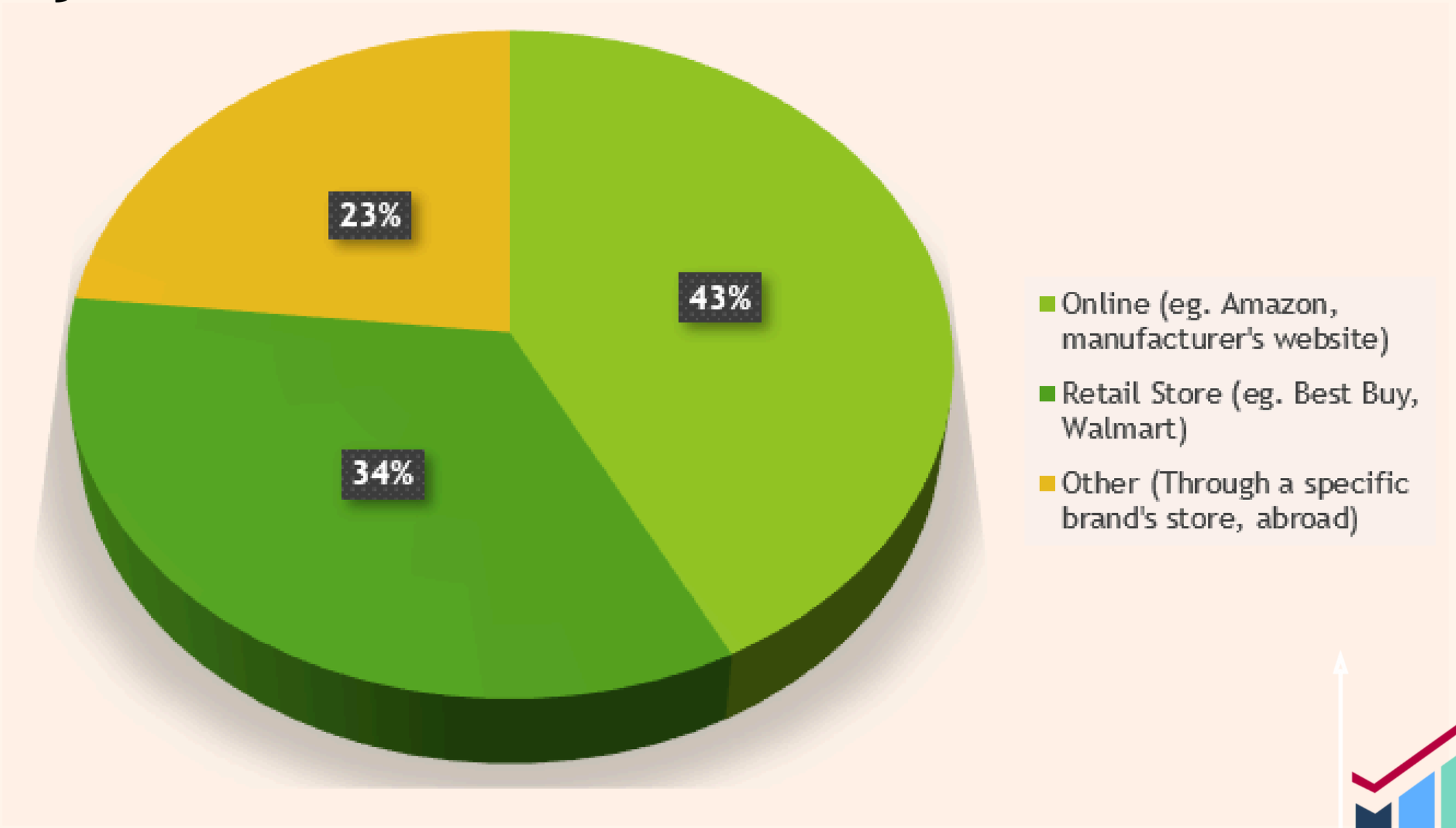
$$\left[-5.0178 - 1.97056 \times 3.6482 \sqrt{\frac{1}{196} + \frac{1}{31}}, -5.0178 + 1.97056 \times 3.6482 \sqrt{\frac{1}{196} + \frac{1}{31}} \right]$$

$$CI \equiv [-6.40734, -3.6282]$$

We see that the interval is negative, so we can say that mean screen time of PGs is higher than UGs by $x \in [3.6282, 6.40734]$ with a 95% confidence level.



Analysis - 2 : Mode of Purchase



Confidence Interval Estimation

For proportion of students using online mode of purchase with 95% confidence level

Total = 245

Students who chose online mode: 104

Confidence level = 95%

$$\alpha = 0.05$$

$$\hat{p} = \frac{104}{245} = 0.424$$

$$z_{\alpha/2} = 1.96$$

$$CI \equiv \left[\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

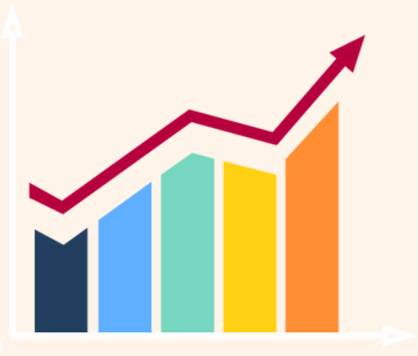
Confidence Interval Estimation

For proportion of students using online mode of purchase with 95% confidence

$$\left[0.424 - 1.96\sqrt{\frac{0.424(0.576)}{245}}, 0.424 + 1.96\sqrt{\frac{0.424(0.576)}{245}}\right]$$

$$CI \equiv [0.3621, 0.4859]$$

$\therefore$ the confidence interval for the proportion of students choosing online mode of purchase is $[0.3621, 0.4859]$ with a 95% confidence level.



Hypothesis Testing

The proportion of students preferring online mode for purchasing laptop less than HALF ?

$H_0: p \geq 0.5$ $H_a: p < 0.5$

Total = 245

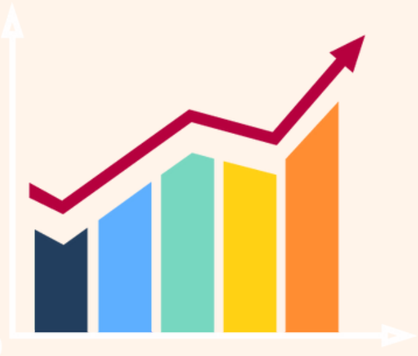
Students who chose online mode: 104

significance level $\alpha = 0.01$

$$\hat{p} = \frac{104}{245} = 0.424$$

$$z^* = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.424 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{245}}} = -2.37$$

critical value $z_\alpha = 2.326$



Hypothesis Testing

Rejection region Approach:

$$\text{critical value } z_\alpha = 2.326$$

$$\text{Reject } H_0 \text{ if } z^* \leq -z_\alpha$$

$$-2.37 < -2.326 \Rightarrow z^* < -z_\alpha$$

$$\therefore \text{ we reject } H_0$$

Analysis - 3 : Operating Systems

The Proportions of Operating Systems vs Degree

Consider the sample with 31 members from PG and 18 members from PhD

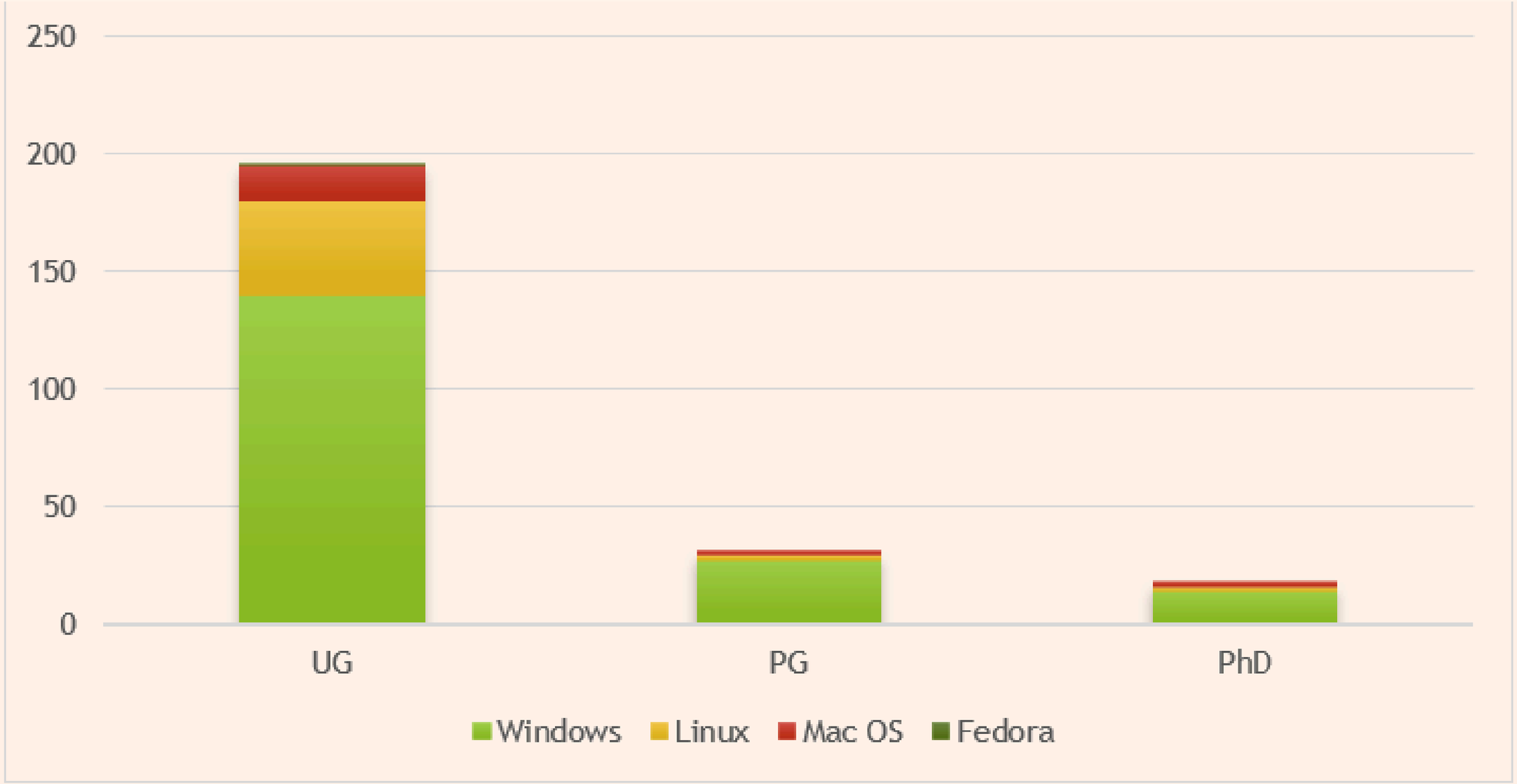
$p_1 \rightarrow$ Proportion of Windows users in PG $\quad n_1 =$ number of PG students $= 31$

$p_2 \rightarrow$ Proportion of Windows users in PhD $\quad n_2 =$ number of PhD students $= 18$

	Windows	Non-Windows
PG ($n_1 = 31$)	27	4
PhD ($n_2 = 18$)	14	4

Analysis - 3 : Operating Systems

Stacked Bar Graph of Operating Systems vs Degeree



Confidence Interval Estimation

Confidence interval estimation of difference in proportions of Windows users in PG and PhD

Confidence Level = 95%

$$\alpha = 0.05$$

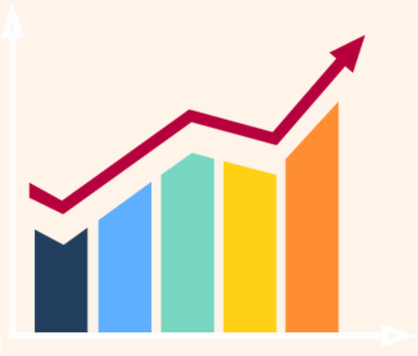
$$CI \equiv \left[(\widehat{p_1} - \widehat{p_2}) - z_{\alpha/2} \times \sqrt{\frac{\widehat{p_1}(1-\widehat{p_1})}{n_1} + \frac{\widehat{p_2}(1-\widehat{p_2})}{n_2}}, \right.$$
$$\left. (\widehat{p_1} - \widehat{p_2}) + z_{\alpha/2} \times \sqrt{\frac{\widehat{p_1}(1-\widehat{p_1})}{n_1} + \frac{\widehat{p_2}(1-\widehat{p_2})}{n_2}} \right]$$

$$\widehat{p_1} = \frac{27}{31} = 0.870$$

$$\widehat{p_2} = \frac{14}{18} = 0.777$$

$$z_{\alpha/2} = 1.96$$

$\therefore$ The Confidence Interval for $\widehat{p_1} - \widehat{p_2}$ is $[-0.132, 0.318]$



Hypothesis Testing

Whether the proportions of Windows users in PG is equal to that of PhD ?

$$\widehat{p_1} = \frac{27}{31} = 0.870$$

$$\widehat{p_2} = \frac{14}{18} = 0.777$$

$$z_{\alpha/2} = 1.96$$

Hypothesis:

$$H_0: p_1 - p_2 = 0$$

$$H_a: p_1 - p_2 \neq 0$$

Test Statistic:

$$z^* = \frac{(\widehat{p_1} - \widehat{p_2}) - p_0}{\sqrt{\frac{\widehat{p_1}(1-\widehat{p_1})}{n_1} + \frac{\widehat{p_2}(1-\widehat{p_2})}{n_2}}} = \frac{0.093}{0.115} = 0.808$$

Rejection Region Approach:

$$0.808 < 1.96$$

$$\Rightarrow |z^*| < z_{\alpha/2}$$

$\therefore$ We fail to reject H_0

Analysis - 4 : Performance

Comparison of avg performances of HP and DELL

$$\mu_1 = \text{Average performance of HP}$$

$$\mu_2 = \text{Average performance of Dell}$$

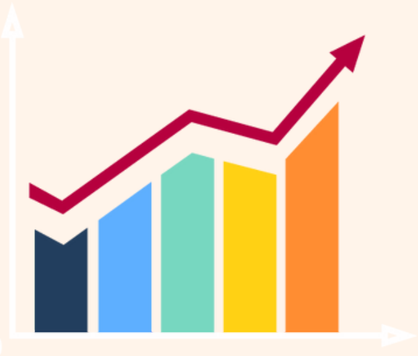
$$\overline{X_1} = 4.264$$

$$\overline{X_2} = 4.596$$

$$S_1^2 = 1.139$$

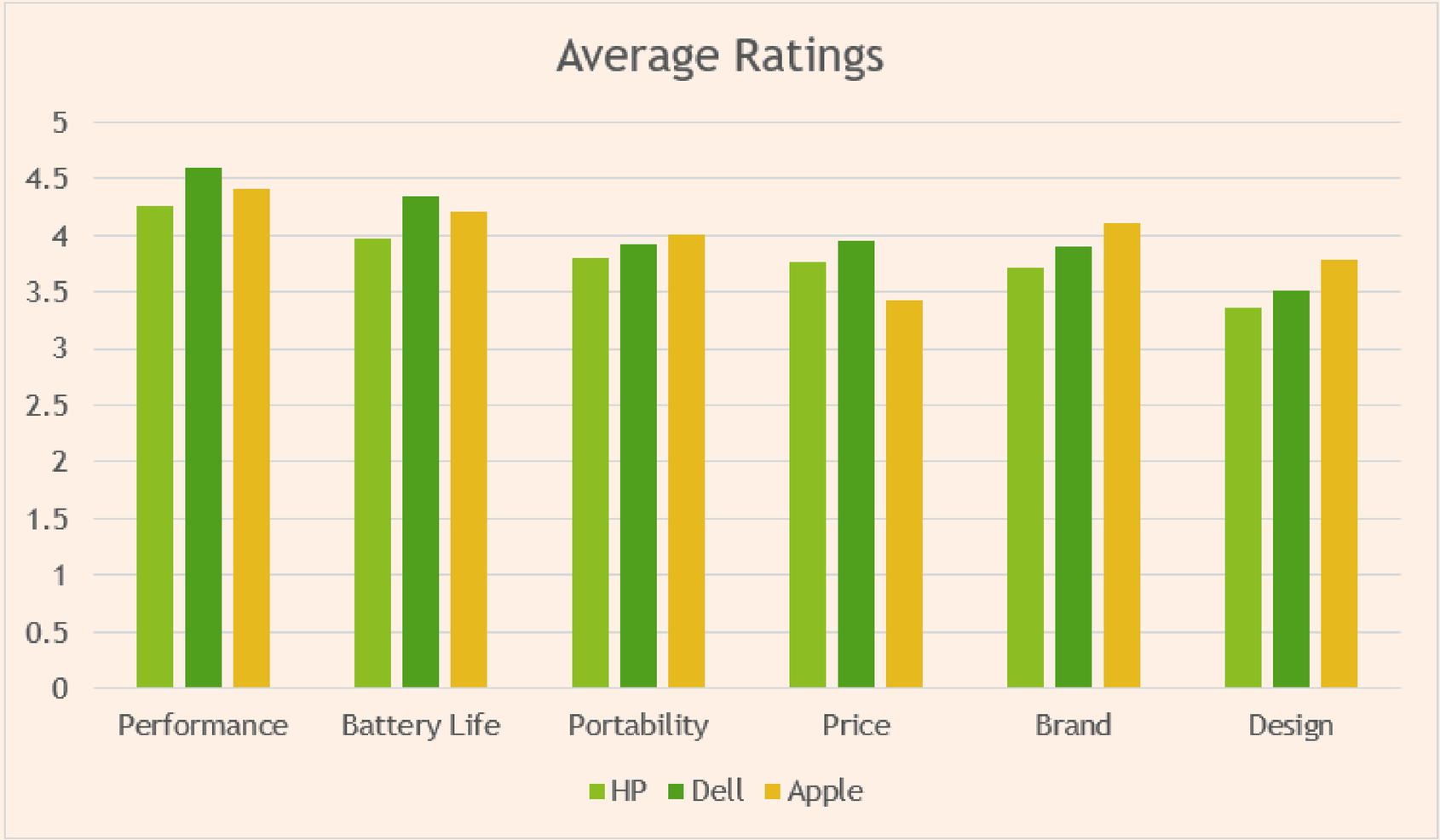
$$S_2^2 = 0.4023$$

$$n_1 = 87, n_2 = 52$$



Analysis - 4 : Performance

Stacked Bar Graph of Operating Systems vs Degeree



Hypothesis Testing

Whether the difference b/w avg performances of HP and DELL is atmost 0.3?

Hypothesis:

$$H_0: \mu_1 - \mu_2 \geq 0.3 \qquad H_a: \mu_1 - \mu_2 < 0.3$$

Test Statistic:

$$t^* = \frac{(\overline{X_1} - \overline{X_2}) - 0.3}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = -0.032$$

$$S_p = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}} = 0.9299$$

Hypothesis Testing

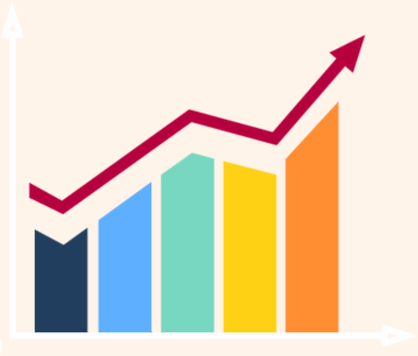
$$df = n_1 + n_2 - 2 = 137$$
$$t_{\alpha,df} = 2.3538$$

Rejection Region Approach:

Reject H_0 if $t^* \leq -t_{\alpha,df}$

$$-0.032 > -2.3538 \Rightarrow t^* > -t_{\alpha,df}$$

$\therefore$ We fail to reject H_0

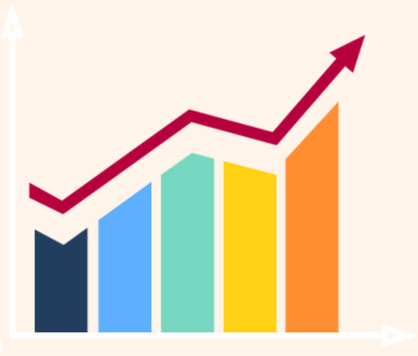


Analysis - 5 : Laptop Brand

Contingency Table of Degree and Laptop Brand

Original Frequencies

DEGREE	Brand			total
	ASUS	LENOVO	OTHER	
UG	20	28	148	196
PG	10	5	16	31
PhD	4	2	12	18
total	34	35	176	245



Analysis - 5 : Laptop Brand

$$E_{ij} = \frac{(\text{Row Total}_i) \times (\text{Column Total}_j)}{\text{Grand Total}}$$

Expected Frequencies

DEGREE	Brand			total
	ASUS	LENOVO	OTHER	
UG	27.2	28	140.8	196
PG	4.3	4.4	22.3	31
PhD	2.5	2.6	12.9	18
total	34	35	176	245

Chi square test of Independence

H_0: There is no association between degree and laptop brand.

H_a: There is an association between degree and laptop brand.

Test Statistic:

$$\chi^{2^*} = \sum_{i=1}^{r \times c} \frac{(O_i - E_i)^2}{E_i}$$

where, r is number of rows in the Contingency Table.

c is number of columns in the Contingency Table.

$$\chi^{2^*} = 12.7927$$

$$\alpha = 0.01$$

Chi square test of Independence

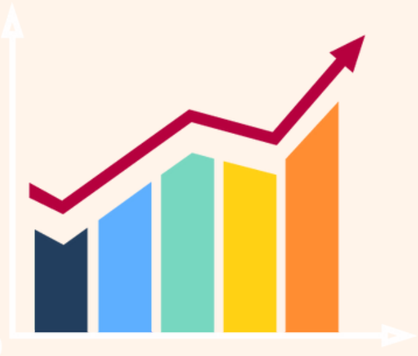
$$H_0: \chi^{2^*} \sim \chi^2_{(r-1)(c-1)}$$

$$\chi^2_{\alpha,4} = 13.2767$$

$$12.7927 < 13.2767 \Rightarrow \chi^{2^*} < \chi^2_{\alpha,04}$$

$\therefore$ We fail to reject H_0

There is no significant evidence to suggest an association between degree and laptop brand.



Thank you