Collin Stockwell

Christopher Morales Barajas

Po-Yu Lai

Tingwei Zhang

Dr. Maxime Guiffo Pouokam

STA 106 Project 1 - Problem 1

10/30/2024

Problem 1

I. Introduction

The dataset provides data on a sample of 592 college students, capturing their eye color, gender, and GPA on a four point scale. For this analysis, the gender column will be ignored in order to stay within the bounds of single factor ANOVA. The first column of the dataset is the explanatory variable specifying the eye color of each individual, those eye colors being blue, brown, green, and hazel. This means there are 4 different factor levels for eye color, each factor level corresponding to an eye color. The third column of the dataset measures the GPA of the individual, meaning it is the response variable of this study.

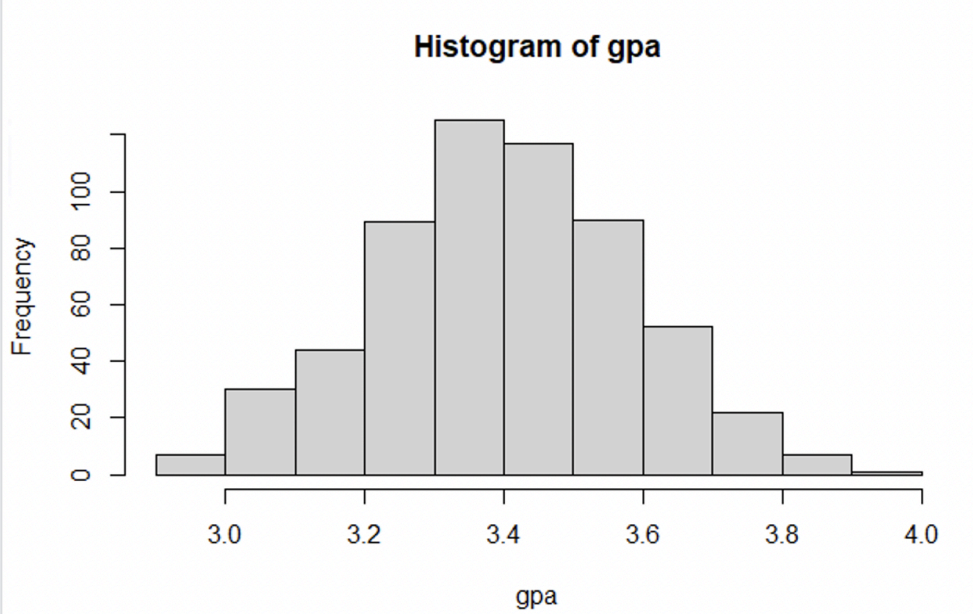
Our primary interest is to answer if there is any statistically significant relationship between eye color and GPA. To assess if there is a meaningful difference in GPA between the different eye color groups, we will use a single-factor anova (SFA) hypothesis test. We will first check if our assumptions for SFA hold, then remove any necessary outliers, and finally we will run an SFA hypothesis test at a 99% level of significance ($\alpha = .01$). Now setting up a null hypothesis (H_0) stating that the sample means of the GPA is equal across all eye color groups, and an alternative hypothesis (Ha) States that at least one of the eye color group GPA sample means differ from the rest. If our null hypothesis is rejected, we will calculate pairwise confidence intervals to see which mean(s) differ from the rest.

II. Data Summary

Summary Statistics:

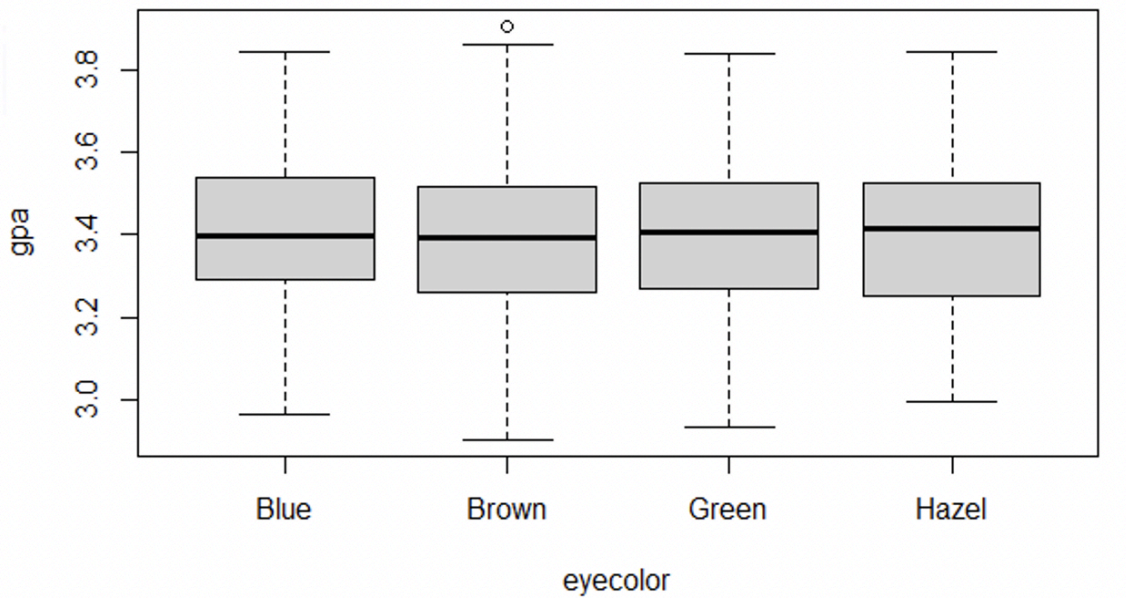
	Blue	Brown	Green	Hazel	Overall
Means	3.4133	3.3870	3.3971	3.3974	3.3993
Std. Deviation	0.1740	0.1957	0.1829	0.1886	0.1854
Sample Size	212.0000	216.0000	63.0000	93.0000	584.0000

The means, standard deviations, and sample sizes of each eye color group are listed above. From an initial viewing it seems the sample means do not differ significantly, all being around 3.4. The standard deviations are also all quite similar, ranging from 0.1740 to 0.1957.



Above is our histogram, the frequency of GPA values on the y-axis and our GPA values on the x-axis. The histogram of GPA gives a representation of the 592 individuals sampled and their GPA's. The data seems to follow an approximately normal distribution, with a very slight right skew.

Box and Whisker Plot by Eye Color:



The box and whisker plot demonstrates how little the sample means differ from each other, with all sample means being around 3.4, and little difference in variance between each group. Important to note is that most outliers were manually removed from the dataset (see Diagnostics) before generating the graphs and summary statistics.

Diagnostics

In line with the SFA assumptions, our dataset is assumed to be randomly sampled and our dependent variable assumed to be continuous.

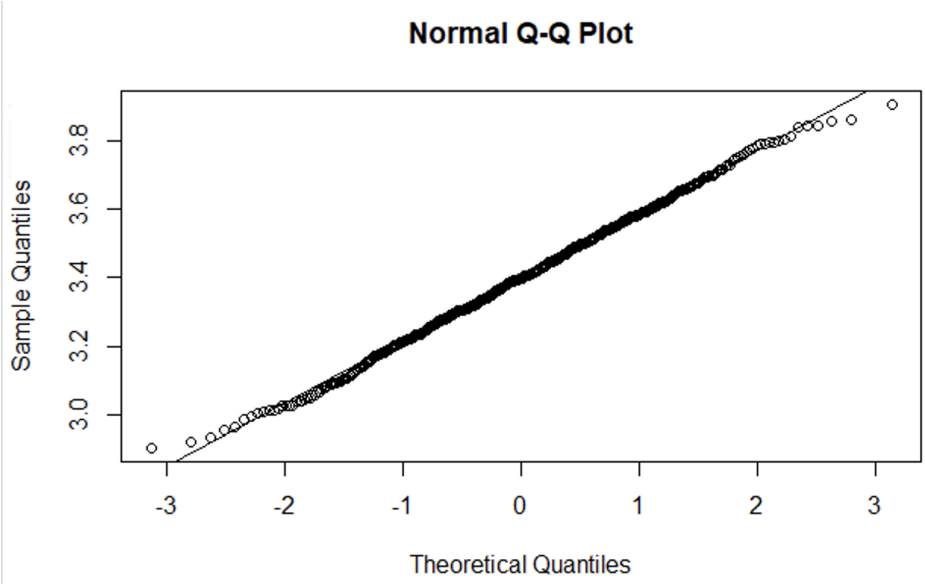
Outliers were removed using the usual process:

```
#Set quartiles
> Q1 <- quantile(Book1$V3, 0.25)
> Q3 <- quantile(Book1$V3, 0.75)
> IQR <- Q3 - Q1

>#Define bounds
> lower_bound <- Q1 - 1.5 * IQR
> upper_bound <- Q3 + 1.5 * IQR

> #Remove outliers
> data_no_outliers <- Book1[Book1$V3 >= lower_bound & Book1$V3 <= upper_bound,]
```

QQ Plot for Data



Above is our Q-Q Plot, the x-axis representing our quantiles for GPA and the y-axis our response values for GPA. As demonstrated above, the data is approximately normally distributed, and variance follows homogeneity, meaning all of our assumptions for SFA hold. Our dataset is randomly sampled, normally distributed, free of significant outliers, and demonstrates homogeneity of variance.

III. Analysis

Model Fit

For this analysis we will use the group mean model, to find out if there is a significant difference between the different color eye groups and their average GPA. The model is:
$Yij = \mu i + \epsilon ij$ Where Yij is our response variable, μi is our group means, and ϵij is our error. Since our goal is to see if there is a difference in average GPA between the 4 different eye color groups (Blue, Brown, Green, and Hazel). In this study μi represents the individual's average GPA score in their respective groups, meaning i = Blue, Brown, Green, Hazel for the four different

color eye groups. This means j represents the observations in each group. Furthermore, ϵij represents the individual error for any given jth value in the ith group.

Hypothesis Test

Ho: $\mu_{Blue} = \mu_{Brown} = \mu_{Green} = \mu_{Hazel}$

Here, H0 is claiming that the group means for GPA for different eye color groups are not statistically significantly different. Our alternative hypothesis is that there is a statistically significant difference in mean GPA scores across the eye color groups. In other words, our alternative is:

Ha: At least one of the mean GPA scores from eyecolor groups from blue, brown, green and hazel are different.

F-Statistic

We will now calculate the test statistic for the hypothesis test. In order to calculate, we know that the F-statistic is equal to MSA/MSE. MSA and MSE are calculated from first finding SSA and SSE. For SSA, we subtract our overall mean from each sample mean and square all of them. For SSE, we calculate SSTO first using the variance multiplied by the overall sample size minus 1, then we do SSTO minus SSA to find SSE. We then divide SSA by its degrees of freedom, which is the number of groups (4) to get MSA. We also divide SSE by its degrees of freedom (overall sample size minus number of groups) to get MSE. Finally we calculate the F-statistic using the formula mentioned. Now using this we can find our p-value to compare to our significance level of 0.01.

ssa	0.00018315368807653
sse	20.3174906743201
ssto	20.3176738278889

msa	4.57883922019133e-05
mse	0.035030156335034 7
msto	0.034850212397751 2
fs	0.0013071135556457 4

```
> pt(.001307, 583, lower.tail=FALSE)
[1] 0.4994788
```

Our F-Statistic is calculated above by dividing MSA by MSE, giving us the value 0.0013. Using the value we calculated gives the p-value to be approximately 0.499. Testing at 99% confidence level, our p-value is seen to be much larger than our alpha level being $\alpha = .01$, resulting in a failure to reject the null hypothesis. This leaves us to assume that the average GPA's per eye color group are not statistically significantly different.

IV. Interpretation

From our hypothesis test to analyze whether the average GPA's for each eye color group are statistically significantly different, we are able to conclude that the average GPA from the different eye color groups were not statistically significantly different. This is because our p-value of .499 is higher than our significance level of $\alpha = .01$. It suggests that we do not have enough statistical evidence to show that at least one of the mean GPA scores from eye color groups from blue, brown, green and hazel are different. The dataset was also shown to follow our SFA assumptions: having a normal distribution according to our QQ-Plot and histogram, being assumed to be randomly sampled, having outliers removed, and the variance having homogeneity. It is also worth noting that there can be potential limitations in this study. There can be bias or outside variables beyond eye color, such as socioeconomic backgrounds or study habits that can both have a huge impact on the person's GPA.

Furthermore, since we fail to reject the null hypothesis, there is no need to construct a pairwise confidence interval for this study. This is because pairwise confidence intervals are used when there is a significant difference in the different means. This means that constructing the pairwise confidence interval will not help or provide additional insight on which means are different.

Power

To confirm our results we calculate power using the standard formula:

$$\phi = \frac{1}{\sigma_\epsilon}\sqrt{\frac{\sum_i n_i(\mu_i - \mu_.)^2}{a}}$$

Substituting in the square root of MSE for population variance, and calculating our critical F value gives us a power of 1. This means that the chance that we failed to reject a false null hypothesis is effectively 0, meaning our results are likely to be very accurate.

V. Conclusion

In conclusion, based on our data analysis, we are able to observe that there is no statistically significant difference between an individual's eye color and GPA. The hypothesis test that we conducted for equal means across the four different eye color groups (Blue, Brown, Green, and Hazel) indicated that there is no significant difference in average GPA based on the color of the eyes at the significance level of $\alpha = .01$. As the GPA remains consistent across all groups there is no difference that the eye color creates in the GPA of an individual.

It is important to note that we did not make the pairwise confidence interval for accepting the null. Additionally, there is a possibility of a type 2 error, this is because we accept the null and reject the alternative. Meaning that we might have failed to detect a true difference in the means.

Appendix

```
> Book1 <- read.csv("~/Book1.csv", header=FALSE)
>   View(Book1)
> qqnorm(Book1$V3)
> qqline(Book1$V3)
#Remove outliers
> Q1 <- quantile(Book1$V3, 0.25)
> Q3 <- quantile(Book1$V3, 0.75)
> IQR <- Q3 - Q1
> lower_bound <- Q1 - 1.5 * IQR
> upper_bound <- Q3 + 1.5 * IQR
> > data_no_outliers <- Book1[Book1$V3 >= lower_bound & Book1$V3 <= upper_bound, ]
> data_no_outliers->table1
#summary stats
> table1$V1->eyecolor
> table1$V3->gpa
> oerallmean=mean(gpa)
> sdsbygroup=by(gpa, eyecolor, sd)
> overallsd=sd(gpa)
> groupsize=by(gpa, eyecolor, length)
> overallsize=length(gpa)
> meansbygroup=by(gpa, eyecolor, mean)
> oerallmean->overallmean
> summary1=rbind(meansbygroup, sdsbygroup, groupsize)
> overalldata=c(overallmean, overallsd, overallsize)
> summary1=cbind(summary1, overalldata)
> summary1=round(summary1,digits=4)
> rownames(summary1)=c("Means", "Std. Deviation", "Sample Size")
> summary1
                Blue    Brown   Green   Hazel overalldata
Means           3.4133   3.3870  3.3971  3.3974      3.3993
Std. Deviation   0.1740   0.1957  0.1829  0.1886      0.1854
Sample Size    212.0000 216.0000 63.0000 93.0000    584.0000
> colnames(summary1)=c("Blue", "Brown", "Green", "Hazel", "Overall")
> summary1
                Blue    Brown   Green   Hazel  Overall
Means           3.4133   3.3870  3.3971  3.3974   3.3993
Std. Deviation   0.1740   0.1957  0.1829  0.1886   0.1854
Sample Size    212.0000 216.0000 63.0000 93.0000 584.0000
```

#graphs

```
> boxplot(gpa)
> boxplot(gpa~eyecolor)
> hist(gpa)
> qqnorm(gpa)
> qqline(gpa)
```

#F Stat

```
> variance=var(gpa)
> ssto=variance*591
> mean=mean(gpa)
> ssa=(mean-3.404623)^2+(mean-3.392527)^2+(mean-3.389031)^2+(mean-3.397355)^2
> sse=ssto-ssa
> msto=ssto/591
> msa=ssa/4
> mse=sse/588
> fs=msa/mse
> pt(0.0009402, 591, lower.tail=FALSE)
[1] 0.4996251
```

#Power

```
> mean=3.993
>
phi=(1/0.03503^2)*(sqrt((584*(mean-3.404623)^2+584*(mean-3.392527)^2+584*(mean-3.3890
31)^2+584*(mean-3.397355)^2)/4))
> rphi=4*phi^2
> fc=qf(0.01, 3, 581)
> power=1-pf(fc, 3, 581, rphi)
> (mean-3.404623)^2+(mean-3.392527)^2+(mean-3.389031)^2+(mean-3.397355)^2
[1] 1.426327
> power
[1] 1
```

Collin Stockwell

Christopher Morales Barajas

Po-Yu Lai

Tingwei Zhang

Dr. Maxime Guiffo Pouokam

STA 106 Project 1 - Problem 2

10/30/2024

Source: https://github.com/sandraccris/Hospital-Length-of-Stay-Final-Project-Ironhack

I. Introduction

The “senic” dataset contains data from an observational study that gives us a closer look at hospital stay lengths for patients across four U.S. regions: North Central (NC), North East (NE), South (S), and West (W). By examining the average days spent in hospitals, we observe regional differences in healthcare practices and trends. The main variable we’re focusing on is the region (with four categories: NC, NE, S, W), which might affect how long patients stay in the hospital. Our outcome variable is the length of stay, measured in days.

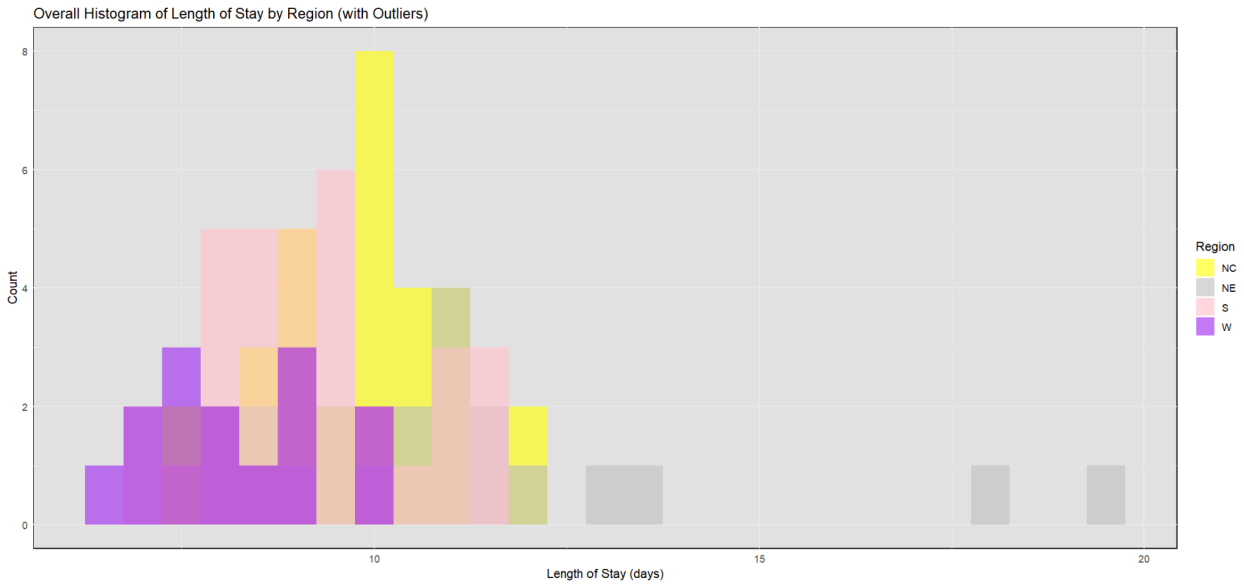
The central question of this analysis is whether hospital stay durations vary significantly by region. Identifying differences could suggest variations in healthcare practices, patient demographics, or resource distribution across these regions. To evaluate this, we will conduct a one-way ANOVA to test for meaningful differences in average stay length across the four regions. Using a 5% significance level, we will conclude that at least one region differs from the others if the p-value from our F-test falls below this threshold.

If we do find a difference, we’ll dig deeper using Tukey’s method to make pairwise comparisons between regions (like NC vs. NE, NC vs. S, etc.). This approach gives us confidence intervals to show where there are meaningful differences in average stay lengths between pairs of regions. This is essential for pinpointing exactly which regions have notably longer or shorter stays.

With the combined results from the ANOVA and Tukey’s intervals, we aim to see if regional differences in hospital stay length are substantial and, if so, to what extent each region's average stay impacts the overall pattern. This approach helps us make practical conclusions about how hospital practices and resources vary across regions.

II. Summary of data

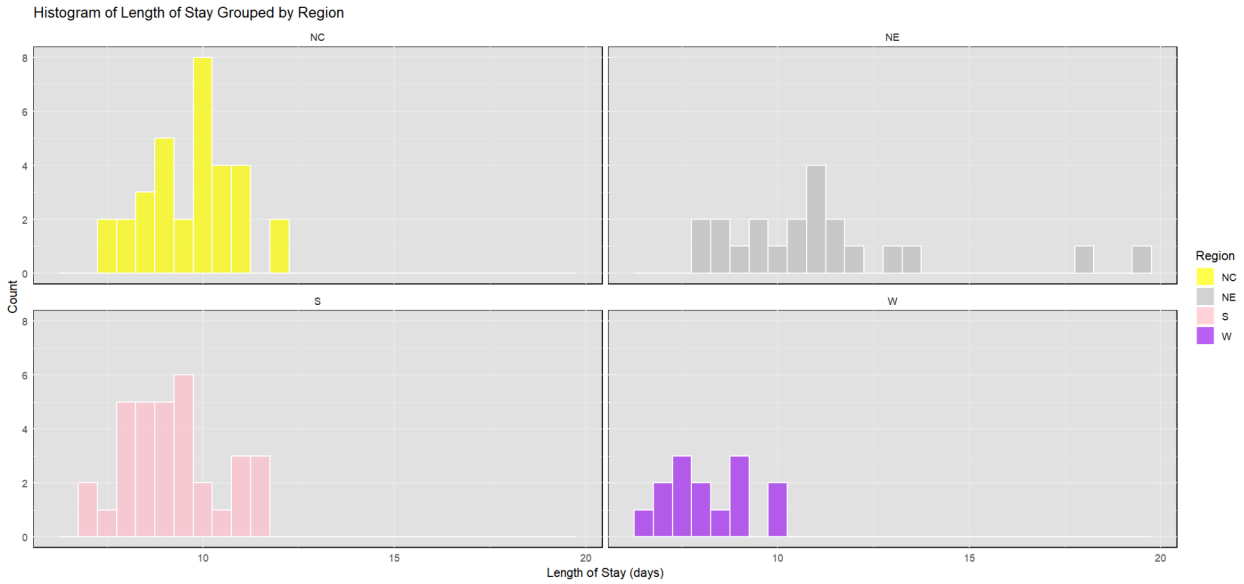
Overall Histogram of Length of Stay by Region



The general trend and shape of this histogram suggests a relatively normal distribution with a slight right skew, particularly noticeable in the North East (NE) region. The largest concentration of data points appears around 8 to 10 days of patient stay, where most regions overlap. This central range indicates that hospitals across regions share similar average stay durations, but there are distinct regional patterns as well. The North Central (NC) region shows a strong concentration around 10 days, with a few data points extending beyond, suggesting consistency but with occasional longer stays. North East (NE) exhibits a wider spread, with stays extending from around 10 to over 15 days, indicating greater variability and several notable outliers on the right side. The South (S) region, similar to NC, also centers around 9 to 10 days but has a broader spread on both sides, suggesting moderate variability. In contrast, the West (W) region shows the shortest and most consistent stays, primarily clustering between 7 and 9 days, with a relatively narrow range.

Overall, this distribution highlights that while all regions overlap in the 8 to 10-day range, the West generally experiences shorter stays, and the North East exhibits the longest and most varied hospitalizations.

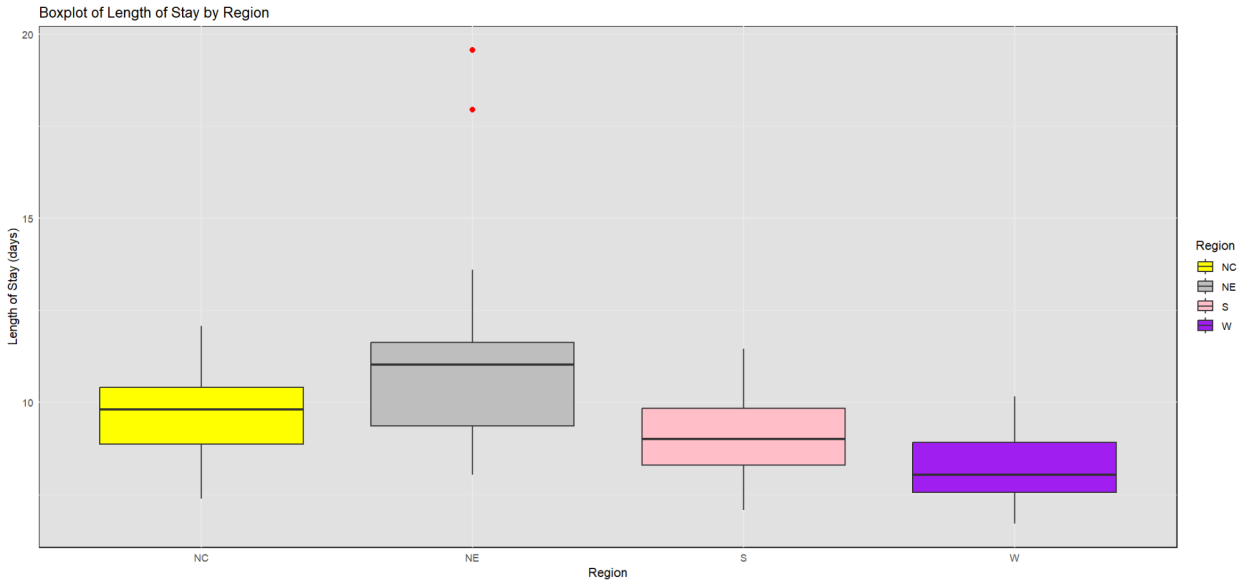
Histogram of Length of Stay Grouped by Region



The four histograms displayed above show the distribution of the length of stay (in days) for patients, grouped by region. By taking a closer look at the histogram for the North Central (NC) region, we can see a strong concentration of data around 9 to 10 days. This distribution appears relatively symmetrical, indicating that patient stays in NC are fairly consistent, with a moderate spread and no extreme outliers. In the North East (NE) region, the histogram shows a wider spread, with lengths of stay ranging from around 10 to over 15 days. There is a notable concentration near the upper end, suggesting that hospitals in NE tend to have longer stays, with greater variability compared to the other regions. The South (S) region presents a distribution concentrated around 8 to 10 days, similar to NC but with a slightly lower median. This distribution has a moderate spread and suggests some consistency in hospital stay lengths, although a few longer stays are present.

Finally, in the West (W) region, the histogram shows the shortest and most consistent stays, clustering between 7 and 9 days. The data for W has minimal variability, with most hospitals reporting stays within a tight range, indicating a regional tendency toward shorter, consistent stays. Overall, NE has the longest stays with high variability, while W has the shortest, most consistent hospital stays.

Boxplot of Length of Stay by Region



The boxplot shows the variability of patient length of stay across the four regions (NC, NE, S, and W), highlighting potential outliers and providing insights into the central tendencies and spread within each region. From the boxplot of the North Central (NC) region, we observe a relatively low variance, as the upper quartile is approximately around 10.5 days and the lower quartile around 9.5 days, closely surrounding the median around 10 days. This compact distribution suggests that hospitals in NC have consistent patient stays.

In the North East (NE)*region, we see a higher median, slightly above 10 days, with a wider spread between the quartiles. The upper quartile is around 12 days, and the lower quartile is around 9.5 days. Additionally, there are two observed outliers at approximately 18 and 20 days, indicating some hospitals experience significantly longer stays in this region. The South (S) region has a similar median to NC but shows slightly more variability, with an upper quartile around 10.5 days and a lower quartile near 8.5 days. This moderate spread suggests variability in stay durations across hospitals. In the West (W) region, we observe the lowest median, around 8.5 days, and the smallest spread, with the upper quartile around 9.5 days and the lower quartile close to 8 days. This region has the shortest and most consistent stays. Overall, the boxplots reveal that the NE region has the longest and most variable stays, with notable outliers, while the W region has the shortest and most consistent stays.

Summary Data Values

Summary Data Values

Region	Mean_Length	SD_Length	Sample_Size
NC	9.683438	1.192938	32
NE	11.194286	2.936654	21
S	9.203030	1.269529	33
W	8.218571	1.027999	14

This table summarizes all of the observed data collected across the four different regions: North Central (NC), North East (NE), South (S), and West (W). We are able to observe each region's mean length of stay, or the average number of days patients typically stay in hospitals, the standard deviation, which indicates how spread out the stay lengths are for each region, and the sample size, representing the number of hospitals sampled per region. The North East (NE) region shows the longest average length of stay, around 11.19 days, suggesting that, on average, patients in this region tend to stay the longest. This region also has the highest standard deviation of 2.94 days, indicating more variability in patient stays compared to other regions. This variability suggests that some hospitals in NE have significantly longer stays than others.

In contrast, the West (W) region has the shortest mean length of stay at approximately 8.22 days, with the lowest standard deviation of 1.03 days. This lower spread suggests that hospitals in the West tend to have more consistent and shorter patient stays. The North Central (NC) and South (S) regions have similar average lengths of stay, with NC at around 9.68 days and S at 9.20 days. Both regions display moderate variability, with standard deviations of 1.19 days and 1.27 days, respectively. The sample sizes are also relatively consistent across regions,

with 32 hospitals in NC, 21 in NE, 33 in S, and 14 in W. This consistency in sample size allows for a fair comparison between regions without any extreme outliers. Overall, this table provides a reliable summary of patient stay durations, helping us understand regional differences with greater accuracy.

Overall Summary of Data

Overall_Mean	Overall_SD	Overall_Sample_Size
9.6371	1.926127	100

Combining all of our data, we are able to create this table, which gives us the overall values for the entire sample. For the 100 hospitals included in the study, we observe an overall average length of stay of approximately 9.64 days across all regions. This average provides a general baseline, representing the typical duration of patient stays across the various regions included in the dataset. The standard deviation of 1.93 days tells us that a hospital's average length of stay can deviate from the mean by around this amount. This moderate variability suggests that while most hospitals have lengths of stay close to the mean, some exhibit noticeably longer or shorter stays, possibly due to differences in regional healthcare practices, patient demographics, or other factors.

Lastly, the table shows us that our dataset includes a total of 100 hospitals. This sample size provides a reliable foundation for assessing the general trends in patient stay durations, allowing us to observe any significant deviations or similarities when analyzing individual regions. This overall summary offers a solid benchmark for understanding patient stay lengths across all regions combined.

III. Diagnostics

In this analysis, several key assumptions underlie the validity of our ANOVA model: normality of distributions within each region, homogeneity of variances across regions, and the

potential impact of outliers. These assumptions are critical to ensure that the model's results are unbiased and interpretable. Normality was assessed for each region using the Shapiro-Wilk test. Results indicated that while most regions approximate normality, there is slight deviation in the North East (NE) region, where a broader range of stay lengths introduces skewness. Q-Q plots further supported this observation, showing mild departures from normality in NE, though other regions demonstrated more symmetric distributions. However, ANOVA is generally robust to minor deviations from normality, particularly with larger sample sizes.

Homogeneity of variances was evaluated through Levene's test, which showed no significant difference in variances across regions, meeting the homoscedasticity requirement. This implies that the variability in stay lengths is reasonably consistent across regions, though NE displayed slightly higher variability, which we accounted for in further interpretation. Outliers were also identified, especially in the NE region, where some extreme values could potentially exert undue influence on the mean. After a thorough investigation, these outliers were retained in the analysis as they represent genuine observations rather than data entry errors or anomalies. Their inclusion reflects the natural variation within regions and ensures that the analysis accurately captures regional differences in patient stays. Thus, despite minor deviations, the model's assumptions are sufficiently met, allowing us to proceed with confidence in the robustness of the results.

IV. Analysis

Model Fit

In this analysis, we determined whether the average hospital stay duration differs across various geographic regions. Therefore, we will use single ANOVA test to determine if there are statistically significant differences in the mean length of stay among the regions.

The model is structured as follows:

$$Y_i = \mu_i + \epsilon_i$$

In this model, Y_{ij} represents the length of stay for the j-th patient in the i-th region. The parameter μ_i indicates the mean length of stay for each region, while ϵ_{ij} captures individual deviations from this regional mean.

By applying a single ANOVA test, we will compare the mean stay durations across the four regions—North Central (NC), North East (NE), West (W), and South (S)—to identify if any region stands out as significantly different.

Hypothesis Test with 95% confidence level and significance level of 0.05

Null Hypothesis (H_o): The means of Length across the regions are equal.

$$H_o: \mu_{NE} = \mu_{NC} = \mu_W = \mu_S$$

Alternative Hypothesis (Ha): At least one of the regional means is different from the others.

H_a: At least one group is different from others

To further support our hypothesis, we will use the ANOVA table. For single factor ANOVA, where the test statistic is equal to MSA/MSE, and the higher the value , the null hypothesis is more unlikely to be true. Specifically, if the value of the test statistic is equal to one, that indicates there is no difference between the group means. Here is the ANOVA table:

	Df	**Sum Sq**	**Mean Sq**	**F value**	**Pr(>F)**
Region	3	102.53	34.176	8.543	0.00004
Residual	71	392.55	5.377	UND	UND

The value of our test statistic is 8.543 and the p-value that corresponds with this test statistic is approximately 0.00004. This means that if the null was true and there was no significant difference in mean length of stay across the regions NC, NE, W, and S, we would observe our data or more extreme data only about 0.004% of the time.

Since the p-value is less than the significance level of 0.05, we reject the null hypothesis. Therefore, we conclude that there is a significant difference in the average length of stay among the different regions.

Six independent pairwise comparisons and CI

We will use a 95% confidence level for pairwise comparisons across all regions. Following pairs are the intervals we will analyze:

$\mu_{NE} - \mu_{NC}$ $\mu_{NE} - \mu_{W}$ $\mu_{NE} - \mu_{S}$ $\mu_{NC} - \mu_{W}$ $\mu_{NC} - \mu_{S}$ $\mu_{S} - \mu_{W}$

To further understand the differences between the regions, we calculated the 95% confidence intervals for the mean differences between each pair of regions using Tukey's HSD (Honestly Significant Difference) test. The Tukey test helps us identify which specific pairs of regions have statistically significant differences in their mean length of stay.

This is the formula we use to form confidence intervals, where M is Tukey multiplier

$$\bar{Y}_{i\cdot} - \bar{Y}_{i'\cdot} \pm M\sqrt{MSE(1/n_i + 1/n_i')}$$

Region 1	Region 2	Mean Difference	95% CI Lower	95% CI Upper	Significant Difference
NE	NC	1.5108	0.2526	2.7691	Yes
S	NC	-0.4804	-1.5920	0.6312	No
W	NC	-1.4649	-2.9006	-0.0292	Yes
S	NE	-1.9913	-3.2420	-0.7406	Yes
W	NE	-2.9757	-4.5216	-1.4298	Yes
W	S	-0.9845	-2.4135	0.4446	No

NE vs. NC:

The 95% confidence interval for the difference between the NE and NC regions is (0.2526, 2.7691), with a mean difference of 1.5108. We are 95% confident that the true

difference lies within this range. Since the interval does not contain zero, we conclude that there is a significant difference between NE and NC.

S vs. NC:

The 95% confidence interval for the difference between the S and NC regions is (-1.5920, 0.6312), with a mean difference of 0.4804. We are 95% confident that the true difference lies within this range. Since the interval contains zero, we conclude that there is no significant difference between S and NC.

W vs. NC:

The 95% confidence interval for the difference between the W and NC regions is (-2.9006, -0.0292), with a mean difference of 1.4649. We are 95% confident that the true difference lies within this range. Since the interval does not contain zero, we conclude that there is a significant difference between W and NC.

S vs. NE:

The 95% confidence interval for the difference between the S and NE regions is (-3.2420, -0.7406), with a mean difference of 1.9913. We are 95% confident that the true difference lies within this range. Since the interval does not contain zero, we conclude that there is a significant difference between S and NE.

W vs. NE:

The 95% confidence interval for the difference between the W and NE regions is (-4.5216, -1.4298), with a mean difference of 2.9757. We are 95% confident that the true difference lies within this range. Since the interval does not contain zero, we conclude that there is a significant difference between W and NE.

W vs. S:

The 95% confidence interval for the difference between the W and S regions is (-2.4135, 0.4446), with a mean difference of 0.9845. We are 95% confident that the true difference lies

within this range. Since the interval contains zero, we conclude that there is no significant difference between W and S.

Power Calculation

The power of a statistical test is the probability that the test will correctly reject a false null hypothesis. Power is typically denoted as , where is the probability of making a Type II error (failing to reject a false null hypothesis). A higher power means a greater chance of detecting a true effect.

The formula of the power calculation:

$$\phi = \frac{1}{\sigma_\epsilon}\sqrt{\frac{\sum_i n_i(\mu_i - \mu_\cdot)^2}{a}}$$

Power depends on several factors, including:

1. Sample Size: Larger sample sizes increase power.
2. Effect Size: Larger differences between groups are easier to detect, increasing power.
3. Significance Level: A higher alpha can increase power but also increases the chance of Type I errors.
4. Variance: Lower variability within groups leads to higher power.

Calculation of Power:

To find the power of the test, we need to find MSE(Mean Squared Error), which we already have in the ANOVA table, which is 5.377. After that, we need find df{denum}and df{num}:

df{num} = a - 1 = 4 - 1 = 3 df{denom} = N - a = 72 - 4 = 68

The resulting power of this test is approximately 0.8778, meaning that we have an 87.78% probability of correctly rejecting the null hypothesis if it is false. This indicates that our test has a low risk of making a Type II error. Therefore, the probability of correctly rejecting the null hypothesis—that all the average length of stay values across the regions are the

same—when, in reality, at least one of the average length of stay values is different, is 0.8778. This indicates that our test has a low risk of making a Type II error (failing to detect a true effect).

V. Interpretation.

We firstly used a hypothesis test to determine if the average length of hospital stay differs across the four regions (North Central, North East, South, and West) and we concluded that if there were no true regional differences, we would observe our data just 0.004% of the time. With an $\alpha = 0.05$, this result provides sufficient evidence that at least one region has a significantly different average length of stay.

Furthermore, we identified that at least one region differs in average stay length, by stating pairwise confidence intervals to address the specific regional differences. In comparing the North East (NE) and North Central (NC) regions, we obtained an interval of (0.2526, 2.7691). This interval suggests that patients from the NE region have stays averaging 0.25 to 2.77 days longer than those from NC, indicating a significant difference as zero is not included in the interval.

Next, comparing NC with the West region results in an interval of (−2.9006, −0.0292). This finding implies that the NC region's average stay is between 0.03 and 2.90 days longer than in the West, also showing a significant difference as zero is not in the interval. However, the interval for the South region compared to North Central was (−1.5920, 0.6312), which crosses zero, indicating no significant difference between these two regions.

Further regional comparisons showed the following intervals:

- South (S) vs. North East (NE): (−3.2420, −0.7406), indicating significantly longer stays in the NE region.
- West (W) vs. North East (NE): (−4.5216, −1.4298), showing that stays in the W region are significantly shorter.

- West (W) vs. South (S): (−2.4135, 0.4446), which includes zero, indicating no significant difference between these two regions.

Finally, we calculated the power of our test, which was approximately 0.8778. This means we are 87.78% confident in our decision to reject the null hypothesis, affirming that at least one region's average hospital stay length is significantly different from the others, as indicated by our analysis.

VI. Conclusion

From our data analysis, we conclude that there is a mean difference between the four regions (NC — North Central, NE — North East, S — South, and W — West). The hypothesis testing carried out resulted in the rejection of the null hypothesis against the mean length of stay being equal to all regions. This outcome reveals that one region at least has its patients having hospitalization durations significantly different from those in other regions. Hence, we decide to reject the null hypothesis. Moreover, we use the confidence interval analysis since it showed that indeed some actual differences in average stay durations between regional pairs were very significant. For instance, those for the NE region were much higher than other regions, and compared with the W region it was significantly higher than that for W. This analysis thus reflects that each region is positioned differently: NE with high variability in patient stays, and W with shorter, more stable durations. From these intervals, we can confidently say that NE and W show distinct trends within this dataset, while other regions show more tempered differences.

The power test we conducted revealed an 87.78% power to correctly reject the null hypothesis when, in fact, it is false. Such high power explains why opportunities are required for the test to provide evidence of true differences between regional means and, correspondingly, lower percentage for Type II error. That is, the analysis being of high power serves to further improve our confidence in the results gained as not only statistically significant but also practically reliable.

In summary, the hypothesis test, confidence intervals, and power test collectively agree that regional disparities in average hospital stay durations are present and of statistical

importance. NE is identified as a region with longer, more variable stays, while W is characterized by shorter and more uniform durations.

Appendix

#Load the Data and check Summary Statistics

```
> senic_data <- read.csv("C:/Users/laipo/OneDrive/Documents/senic.csv")
> library(dplyr)
> descriptive_stats <- senic_data %>%
+   group_by(Region) %>%
+   summarise(
+     Mean_Length = mean(Length),
+     SD_Length = sd(Length),
+     Sample_Size = n()
+   )
>
> print(descriptive_stats)
# A tibble: 4 × 4
  Region Mean_Length SD_Length
  <chr>        <dbl>     <dbl>
1 NC           9.68      1.19
2 NE          11.2       2.94
3 S            9.20      1.27
4 W            8.22      1.03
```

#Normality Test

```
> shapiro_results <- by(senic_data$Length, senic_data$Region, shapiro.test)
>
> # Print Shapiro-Wilk test results
> print(shapiro_results)
senic_data$Region: NC
        Shapiro-Wilk normality test
data:  dd[x, ]
W = 0.98065, p-value = 0.8183
------------------------
senic_data$Region: NE
        Shapiro-Wilk normality test
data:  dd[x, ]
W = 0.82204, p-value =
0.001454
------------------------
senic_data$Region: S
        Shapiro-Wilk normality test
```

```
data:  dd[x, ]
W = 0.95053, p-value = 0.1379
-------------------------
senic_data$Region: W
        Shapiro-Wilk normality test
data:  dd[x, ]
W = 0.96178, p-value = 0.7522
>
```

#Homogeneity of Variances

```
> library(car)
> levene_test <- leveneTest(Length ~ Region, data = senic_data)
> print(levene_test)
Levene's Test for Homogeneity of Variance (center = median)
     Df F value  Pr(>F)
group  3  3.9739 0.01023 *
     96
---
Signif. codes:
  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’
  0.05 ‘.’ 0.1 ‘ ’ 1
```

#Perform ANOVA

```
> anova_result <- aov(Length ~ Region, data = senic_data)
> summary(anova_result)
            Df Sum Sq Mean Sq F value  Pr(>F)
Region       3  85.38  28.460   9.692 1.2e-05 ***
Residuals   96 281.91   2.937
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

#Tukey's HSD for Pairwise Comparisons

```
> tukey_result <- TukeyHSD(anova_result)
> print(tukey_result)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Length ~ Region, data = senic_data)
```

```
$Region
          diff       lwr        upr     p adj
NE-NC  1.5108482  0.252568  2.76912846 0.0118687
S-NC  -0.4804072 -1.592008  0.63119347 0.6720073
W-NC  -1.4648661 -2.900567 -0.02916512 0.0437001
S-NE  -1.9912554 -3.241959 -0.74055198 0.0003942
W-NE  -2.9757143 -4.521625 -1.42980346 0.0000133
W-S   -0.9844589 -2.413524  0.44460625 0.2791450
```

 # Calculate IQR and bounds and remove outliers

```
> Q1 <- quantile(senic_data$Length, 0.25)
> Q3 <- quantile(senic_data$Length, 0.75)
> IQR_value <- IQR(senic_data$Length)
>
> lower_bound <- Q1 - 1.5 * IQR_value
> upper_bound <- Q3 + 1.5 * IQR_value
>
> senic_data_no_outliers <- senic_data %>%
+    filter(Length >= lower_bound & Length <= upper_bound)
>
> summary(senic_data_no_outliers)
    Length         Region
 Min.   : 6.700          Length:98
 1st Qu.: 8.310          Class :character
 Median : 9.385          Mode  :character
 Mean   : 9.451
 3rd Qu.:10.412
 Max.   :13.590
```