Physical Activity in Relation to Health Outcome and Longevity in Care Home in Scotland

Exam Number: B249125

Contents

Introduction

Care homes provide accommodation, personal care and support, and/or onsite nursing care for people who can no longer live independently in their own homes according to (Goodman et al. 2016). Admission to a care home is often associated with further functional and health decline and associated adverse outcomes such as increased mortality and reduced quality of life (Luppa et al. 2010).
This report explores a data gathered from two Scottish care homes in an effort to learn more about the relationship between senior citizens' health outcomes, demographics, and physical activity. Through a thorough examination of demographic and baseline data, our goal is to identify patterns and trends that may guide public decisions to promote the increase in activity levels in order to aim for better health conditions and longevity.

In an effort to add to the growing body of knowledge on healthy aging, special attention is paid to comprehending the relationship between BMI and moderate physical activity. The results of this investigation have the potential to improve the quality of life for residents of care homes by providing practical advice.

Using Body Mass Index (BMI) is an acceptable measurement for determining the health, physical fitness, and activity level of a given population that lacks regular high-intensity physical activity (Ding and Jiang 2020).

The report will try here to first explore the data gathered and see the pattern of participants in terms of base demographic, physical level, and BMI level. If we find a relationship between physical activity type and level of moderate activity with longevity, this can encourage us to seek further promotion and encouragement to integrate such measures in care homes in Scotland.

So let's go step by step to see what this data can tell us.

Load data

Let's go on to the analysis of the given dataset in R, answering each of these questions individually. First here we are going to load the dataset and do some initial data investigation.

```r
pacman::p_load(data.table, rio, here, dplyr, epikit, janitor, lubridate, ggplot2,
               crosstable, stringr, gtsummary, flextable, Hmisc, scales, incidence,
               tidyverse, kableExtra, knitr, flextable, tidyr, fancyhdr)
carehome_data <- import(here("Assessment/Report/carehomedata_assessment2024.csv"))
carehome_data <- carehome_data %>% clean_names()
summary(carehome_data)
```

```
##  participant_id      sex               age_at_recording  age_at_death
##  Min.   :  1    Length:341         Min.   :66.53    Min.   :67.39
##  1st Qu.: 86    Class :character   1st Qu.:76.41    1st Qu.:77.77
##  Median :171    Mode  :character   Median :79.95    Median :81.56
##  Mean   :171                       Mean   :79.85    Mean   :81.52
##  3rd Qu.:256                       3rd Qu.:83.45    3rd Qu.:85.43
##  Max.   :341                       Max.   :93.49    Max.   :96.87
##  physical_activity  moderate_activity      bmi          carehome_id
##  Length:341         Min.   :19.53     Min.   :19.27   Min.   :1.000
```

```
##  Class :character   1st Qu.:43.07    1st Qu.:23.74   1st Qu.:1.000
##  Mode  :character   Median :49.65    Median :25.12   Median :1.000
##                     Mean   :49.73    Mean   :24.92   Mean   :1.493
##                     3rd Qu.:56.31    3rd Qu.:26.20   3rd Qu.:2.000
##                     Max.   :83.04    Max.   :29.81   Max.   :2.000
```

After the dataset was successfully imported into our R environment, a comprehensive preliminary assessment is to be carried out carried out to guarantee data quality and preparation for analysis. The participants' demographic and baseline characteristics will be examined as the next step.

In order to determine the age distribution, gender ratio, BMI ranges, and degree of moderate physical activity among people, this step will involve producing descriptive data. The objective is to lay the groundwork for a deeper analysis of the population being studied, with a focus on the connections between physical activity patterns, health outcomes, care home conditions, and lifespan. This methodical methodology guarantees that the conclusions drawn later are based on a solid examination of the core characteristics of the dataset.

Demographics and Baseline Characteristics (Question no. 1)

In order to understand the data, some descriptive statistics is done to learn about the demographics and baseline statistics.

```r
descriptive_stats <- carehome_data %>%
  summarise(
    min_participant_id = min(participant_id),
    max_participant_id = max(participant_id),
    mean_age_at_recording = mean(age_at_recording, na.rm = TRUE),
    sd_age_at_recording = sd(age_at_recording, na.rm = TRUE),
    min_age_at_recording = min(age_at_recording, na.rm = TRUE),
    max_age_at_recording = max(age_at_recording, na.rm = TRUE),
    mean_age_at_death = mean(age_at_death, na.rm = TRUE),
    sd_age_at_death = sd(age_at_death, na.rm = TRUE),
    min_age_at_death = min(age_at_death, na.rm = TRUE),
    max_age_at_death = max(age_at_death, na.rm = TRUE),
    mean_moderate_activity = mean(moderate_activity, na.rm = TRUE),
    sd_moderate_activity = sd(moderate_activity, na.rm = TRUE),
    mean_bmi = mean(bmi, na.rm = TRUE),
    sd_bmi = sd(bmi, na.rm = TRUE),
    min_bmi = min(bmi, na.rm = TRUE),
    max_bmi = max(bmi, na.rm = TRUE)
  )

descriptive_stats_long <- descriptive_stats %>%
  pivot_longer(cols = everything(), names_to = "Measure", values_to = "Value")
descriptive_stats_long
```

Measure	Value
min_participant_id	1.000000
max_participant_id	341.000000
mean_age_at_recording	79.846328
sd_age_at_recording	5.176227
min_age_at_recording	66.532592
max_age_at_recording	93.491975
mean_age_at_death	81.516802
sd_age_at_death	5.498229
min_age_at_death	67.393800
max_age_at_death	96.871826
mean_moderate_activity	49.732026
sd_moderate_activity	9.928704
mean_bmi	24.924828
sd_bmi	1.802885
min_bmi	19.273478
max_bmi	29.809399

The exploration here aims to see every variable of the provided data and explore what it can tell us to have a better idea of this study population and how they are doing at the two care homes.

Highlights from the above table

- There are 341 people in the dataset, and their IDs range from 1 to 341.
- The age range of the participants during recording is 66.53 to 93.49 years, with an average of 79.85 years and a standard deviation of 5.18 years.
- The average age of death is determined to be 81.52 years, with a standard deviation of 5.50 years and a range of 67.39 to 96.87 years.
- The average percentage of time that individuals spent engaging in moderate physical activity was 49.73%, with a standard deviation of 9.93%.
- The BMI measurement ranges from 19.27 to 29.81, with a standard deviation of 1.80 and an average of 24.92. Participants are split almost evenly between the two care home facilities.

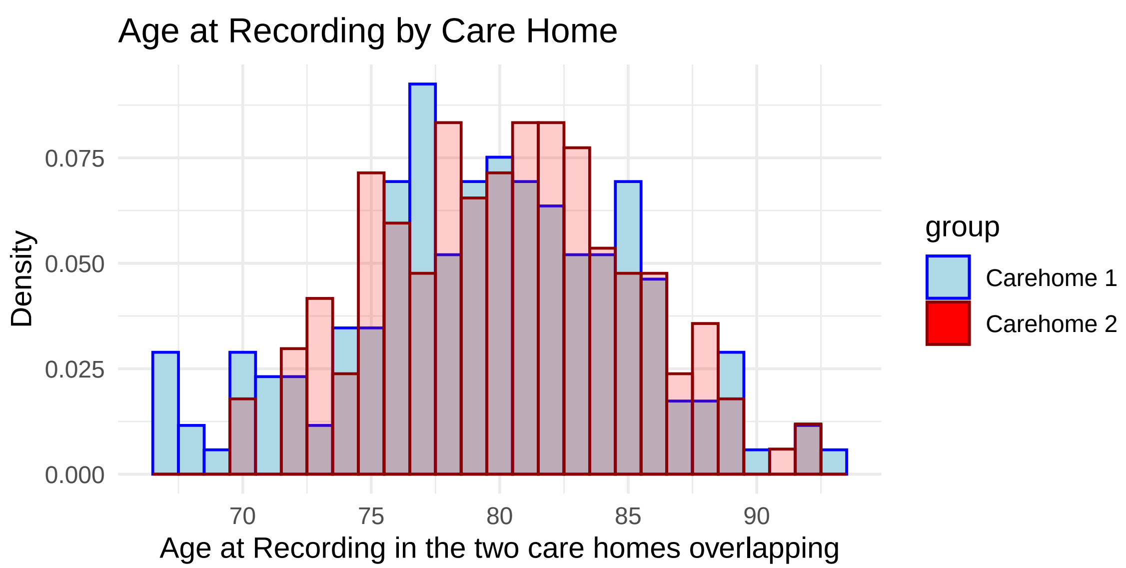
Visualizaiton of the demographics and baseline characteristics

In order to visualize the data, and see the current baseline comparing the two care homes against each other, we can check the following in details:

Age at recoding

```
carehome_data$group <- factor(carehome_data$carehome_id, levels = c(1, 2),
                              labels = c("Carehome 1", "Carehome 2"))
ggplot(carehome_data, aes(x = age_at_recording, y = ..density..,
```

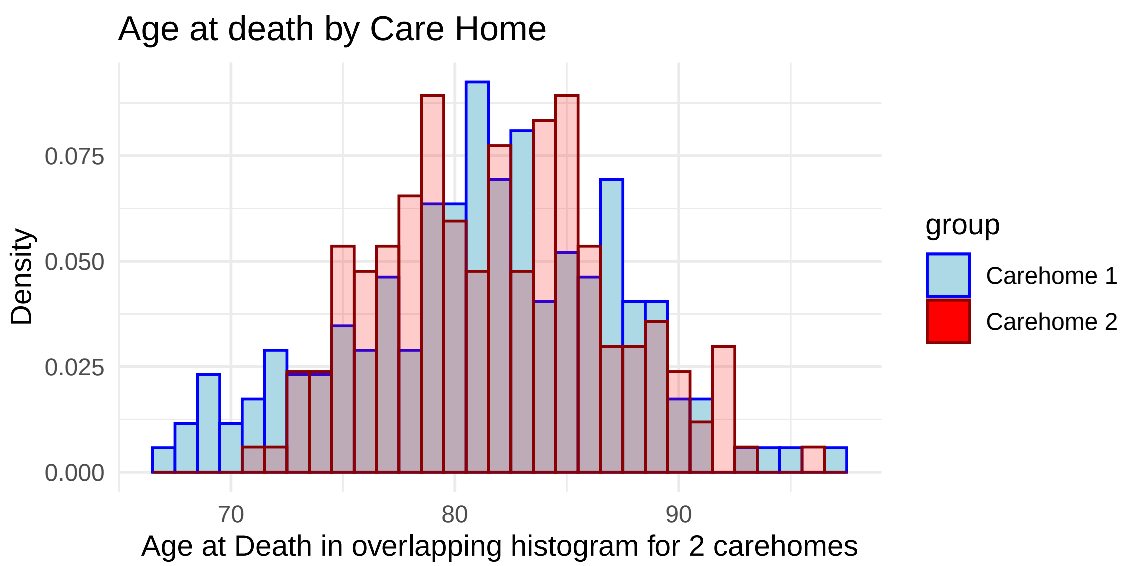
```
                          fill = group, color = group)) +
  geom_histogram(data = subset(carehome_data, carehome_id == 1),
                 binwidth = 1, alpha = 1.5) +
  geom_histogram(data = subset(carehome_data, carehome_id == 2),
                 binwidth = 1, alpha = 0.2) +
  scale_fill_manual(values = c("Carehome 1" = "lightblue", "Carehome 2" = "red")) +
  scale_color_manual(values = c("Carehome 1" = "blue", "Carehome 2" = "darkred")) +
  theme_minimal() +
  labs(title = "Age at Recording by Care Home",
       x = "Age at Recording in the two care homes overlapping", y = "Density")
```



For the participants age, in numerical terms, one might say, The majority of our participants are **clustered around their late 70s to early 80s**, with fewer individuals below 70 or above 90.

Age at death

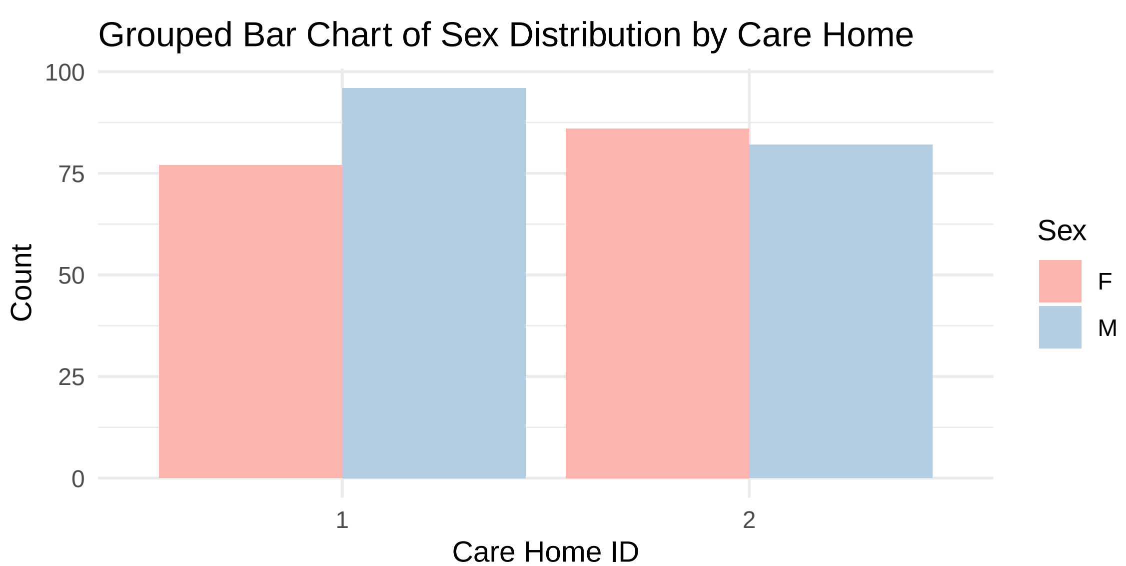
```
carehome_data$group <- factor(carehome_data$carehome_id, levels = c(1, 2), labels = c("Carehome 1", "Carehome 2"))

ggplot(carehome_data, aes(x = age_at_death, y = ..density.., fill = group, color = group)) +
  geom_histogram(data = subset(carehome_data, carehome_id == 1),
                 binwidth = 1, alpha = 1.5) +
  geom_histogram(data = subset(carehome_data, carehome_id == 2),
                 binwidth = 1, alpha = 0.2) +
  scale_fill_manual(values = c("Carehome 1" = "lightblue", "Carehome 2" = "red")) +
  scale_color_manual(values = c("Carehome 1" = "blue", "Carehome 2" = "darkred")) +
  theme_minimal() +
  labs(title = "Age at death by Care Home",
       x = "Age at Death in overlapping histogram for 2 carehomes", y = "Density")
```



The average age of death is determined to be 81.52 years, with a standard deviation of 5.50 years and a range of 67.39 to 96.87 years.

Gender distribution

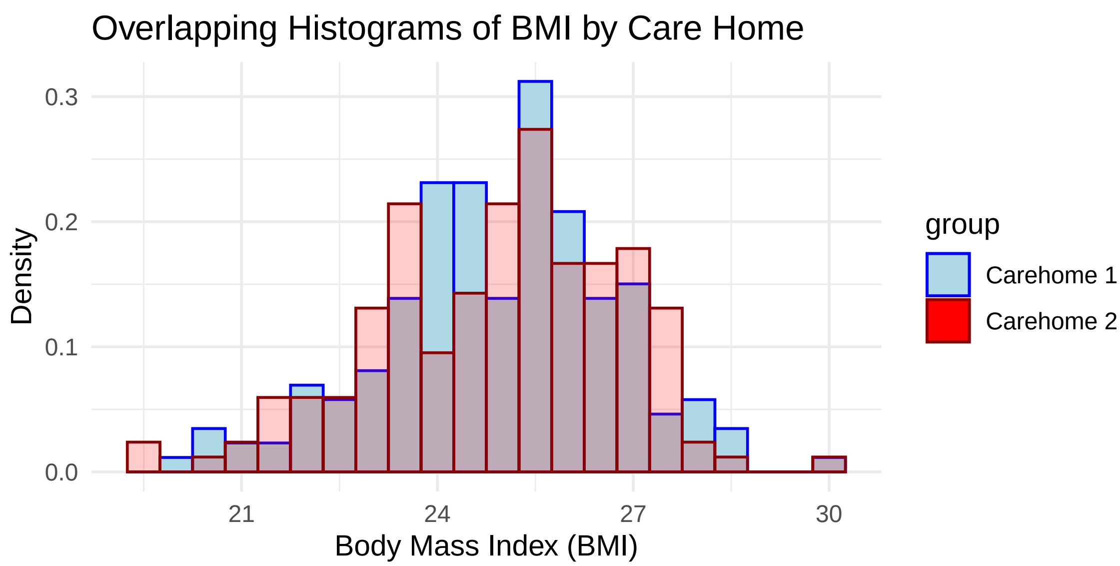
```r
# Grouped bar chart for sex distribution within each care home
ggplot(carehome_data, aes(x = as.factor(carehome_id), fill = sex)) +
  geom_bar(position = "dodge") +
  scale_fill_brewer(palette = "Pastel1", name = "Sex") +
  theme_minimal() +
  labs(title = "Grouped Bar Chart of Sex Distribution by Care Home",
       x = "Care Home ID",
       y = "Count")
```



The sex distribution are very close especially in care home 2 and with sligh increase in males in the care home 1.

BMI distibution between the two care homes

```
carehome_data$group <- factor(carehome_data$carehome_id, levels = c(1, 2), labels = c("Carehome 1", "Carehome 2"))

ggplot(carehome_data, aes(x = bmi, y = ..density.., fill = group, color = group)) +
  geom_histogram(data = subset(carehome_data, carehome_id == 1),
                 binwidth = 0.5, alpha = 1.5) +
  geom_histogram(data = subset(carehome_data, carehome_id == 2),
                 binwidth = 0.5, alpha = 0.2) +
  scale_fill_manual(values = c("Carehome 1" = "lightblue", "Carehome 2" = "red")) +
  scale_color_manual(values = c("Carehome 1" = "blue", "Carehome 2" = "darkred")) +
  theme_minimal() +
  labs(title = "Overlapping Histograms of BMI by Care Home",
       x = "Body Mass Index (BMI)", y = "Density")
```

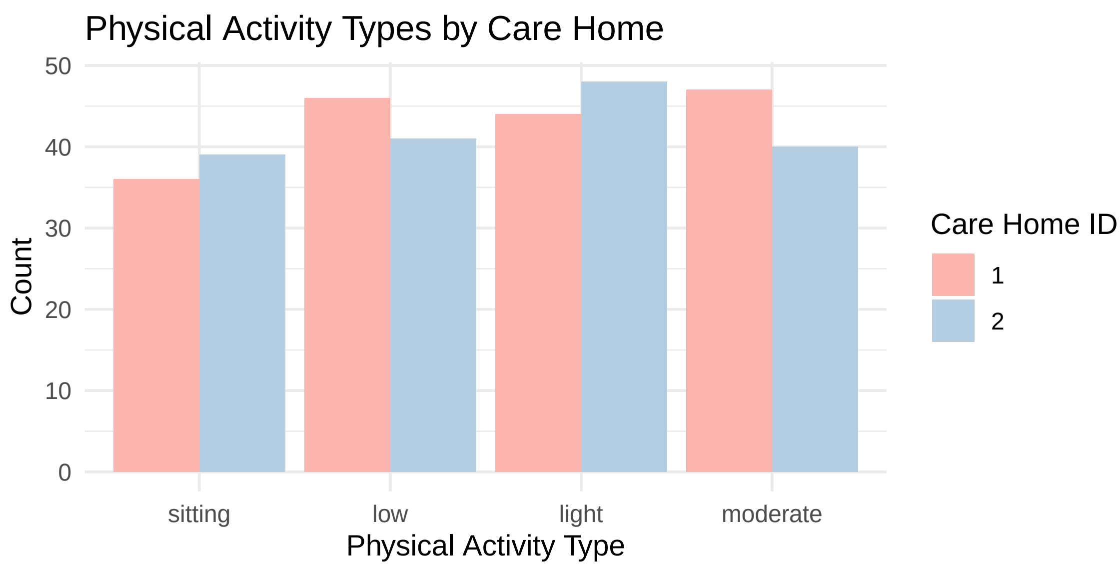


As mentioned before, the participants' Body Mass Index (BMI) ranges from a minimum of 19.27 to a maximum of 29.81 showing a wide range of body weight statuses from normal to overweight. This would reflect somehow the health status and be later used as a measure of higher activity levels especially if there is positive relationship detected between BMI and activity level, which we will examin later in this study.

As the histogram shows there is big overlap between the two care homes in the ranges with a normally distributed shape for the BMI in each one of them.

Activity type distribution between the two care homes

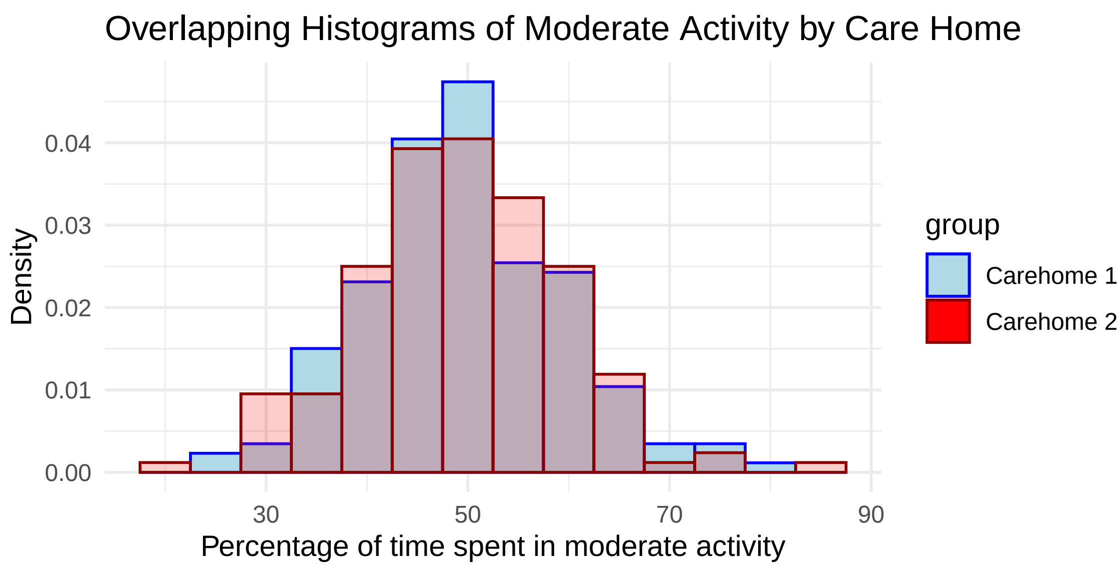
```
ggplot(carehome_data, aes(x = factor(physical_activity, levels = c("sitting", "low", "light", "moderate")), fill =
  geom_bar(position = "dodge") +
  scale_fill_brewer(palette = "Pastel1", name = "Care Home ID") +
  theme_minimal() +
  labs(title = "Physical Activity Types by Care Home",
       x = "Physical Activity Type",
       y = "Count")
```



When comparing the two care facilities as shown above, there is also an almost equal distribution of the four categories of physical activity, which range from low, light, moderate, and seated to practically extremely close levels.

Distribution of moderate activity comparing the two care homes (Percentage of time each individual spent undertaking activity classified as moderate)

```
# histograms with density on the y-axis
carehome_data$group <- factor(carehome_data$carehome_id, levels = c(1, 2), labels = c("Carehome 1", "Carehome 2"))

ggplot(carehome_data, aes(x = moderate_activity, y = ..density.., fill = group, color = group)) +
  geom_histogram(data = subset(carehome_data, carehome_id == 1),
                 binwidth = 5, alpha = 1.5) +
  geom_histogram(data = subset(carehome_data, carehome_id == 2),
                 binwidth = 5, alpha = 0.2) +
  scale_fill_manual(values = c("Carehome 1" = "lightblue", "Carehome 2" = "red")) +
  scale_color_manual(values = c("Carehome 1" = "blue", "Carehome 2" = "darkred")) +
  theme_minimal() +
  labs(title = "Overlapping Histograms of Moderate Activity by Care Home",
       x = "Percentage of time spent in moderate activity", y = "Density")
```



On average, participants across both care homes engaged in moderate activity around 49.73% of the time indicating a generally active lifestyle among the residents.

The difference in moderate activity levels between the two care homes is relatively small. Care home 1 in blue may show higher activity especially at different points which suggests slightly more active or engaged resident population when it comes to moderate physical activities.

Association between the care home and type of physical activity and also level of moderate activity level (Question no. 2)

To answer the question regarding the association between care home and the type of physical activity undertaken, and if a longer amount of moderate activity is observed depending on the care home. A statistical test called chi-square of independence can be used for the first part to analyze the association between categorical variables (care home and type of physical activity).

For the second part, another statistical test called t-test can be used to compare the means of moderate activity between the two care homes if the data meets the assumptions for these tests.

First Part; chi-square

The results of the chi-square test will show if there's a statistically significant association between care home and physical activity type. A significant result suggests that the type of activity depends on the care home.

```
# Create a contingency table for care home and activity type
contingency_table <- table(carehome_data$carehome_id, carehome_data$physical_activity)
contingency_table
```

```
##
##      light low moderate sitting
```

```
##    1    44  46        47       36
##    2    48  41        40       39
```

```
# Perform chi-square test
chi_square_result <- chisq.test(contingency_table)

# Print the results
print(chi_square_result)
```

```
##
##  Pearson's Chi-squared test
##
## data:  contingency_table
## X-squared = 1.0714, df = 3, p-value = 0.784
```

The purpose of the Pearson's Chi-squared test is to determine whether there is a relationship between two categorical variables—in this case, the residents' types of physical activity and the care facility. A chi-square statistic value of 1.0714 was obtained from the test. Although this figure may seem abstract on its own, it is a component of the p-value calculation, which indicates the likelihood that any apparent correlation between these variables was the result of chance.

The p-value, which came out at 0.784, is an important component of this puzzle. We can interpret our results by comparing this value to a standard threshold (alpha level) of 0.05. Our p-value, which is much larger than 0.05, indicates that there may be random variation rather than a particular pattern or trend to be the cause of the variations in physical activity kinds that we observe amongst care homes.

Put another way, this high p-value suggests that there is insufficient data to conclude that a person's kind of physical activity is significantly influenced by the sort of care home they live in.

Second part

The second section of the analysis compares the average amount of time spent in each of the two care homes using a t-test to see whether there are any differences. The mean moderate activity levels are compared using the t-test, which assumes, until otherwise demonstrated, that there is no significant difference.

```
carehome1_data <- filter(carehome_data, carehome_id == 1)
carehome2_data <- filter(carehome_data, carehome_id == 2)

t_test_result <- t.test(carehome1_data$moderate_activity, carehome2_data$moderate_activity, var.equal = TRUE)
print(t_test_result)
```

```
##
##  Two Sample t-test
##
## data:  carehome1_data$moderate_activity and carehome2_data$moderate_activity
## t = 0.17578, df = 339, p-value = 0.8606
## alternative hypothesis: true difference in means is not equal to 0
```

```
## 95 percent confidence interval:
##  -1.929110  2.307742
## sample estimates:
## mean of x mean of y
##  49.82530  49.63598
```

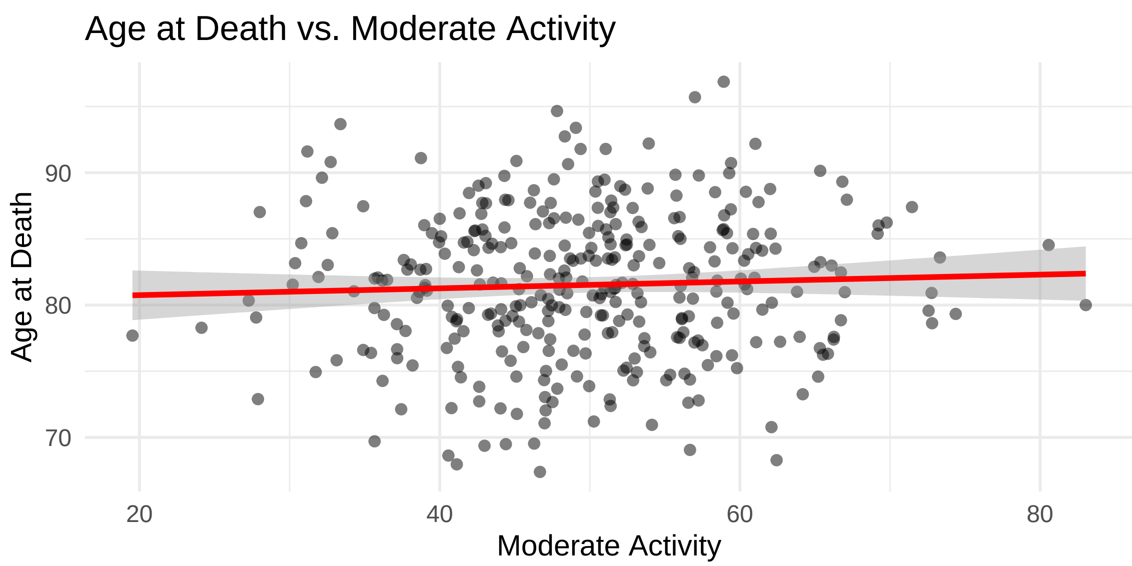
In order to determine whether there is a discernible difference, we compared the average levels of physical activity between two care facilities using a statistical technique known as a t-test. The test yielded a p-value of 0.8606, significantly higher than the typical 0.05 cut-off threshold. Given the large p-value, it is likely due to chance that the two care facilities' levels of physical activity differ, and it is not significant enough to conclude that one encourages more moderate activity than the other.

Additionally, we examined a measure known as a 95% confidence interval, which spans from roughly -1.93 to 2.31. This range includes zero, which in this case indicates that there may not be much of a difference in activity levels. This interval essentially confirms our first conclusion, which was that there is little data to suggest that a resident's choice of care home influences their level of moderate activity. This conclusion clarifies that the tenants' activity levels in these residences may not be determined by their surroundings.

Physical activity and longevity (Question no. 3)

This quesiton explores whether the data provide any evidence that those who are more physically active live longer. It requires a formal recommendation on whether an intervention should be provided based on these results. This question aims to explore the relationship between physical activity and longevity, and it's essential to use appropriate statistical methods to analyze this relationship and provide clear understandable recommendations.

```r
ggplot(carehome_data, aes(x = moderate_activity, y = age_at_death)) +
  geom_point(alpha = 0.5) +  # Plot the individual data points
  geom_smooth(method = "lm", color = "red") +  # Add a linear regression line
  theme_minimal() +
  labs(title = "Age at Death vs. Moderate Activity",
       x = "Moderate Activity",
       y = "Age at Death")
```




```
# Perform linear regression
model <- lm(age_at_death ~ moderate_activity, data = carehome_data)

summary(model)
```

```
##
## Call:
## lm(formula = age_at_death ~ moderate_activity, data = carehome_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.0442  -3.7428   0.1046   3.9076  15.1179
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       80.23381    1.52355  52.662   <2e-16 ***
## moderate_activity  0.02580    0.03004   0.859    0.391
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.5 on 339 degrees of freedom
## Multiple R-squared:  0.00217,    Adjusted R-squared:  -0.0007732
## F-statistic: 0.7373 on 1 and 339 DF,  p-value: 0.3911
```

To find any possible proof that more physical activity could result in a longer life, statistical analysis is used to examine the relationship between physical activity and lifespan. This investigation is essential when deciding whether to suggest programs meant to increase levels of physical activity in order to improve longevity.

We plotted the ages of participants at death against their levels of moderate exercise using a graphical approach. To find any trends, this image was used with a linear regression analysis.

The study did find a modest increase in lifespan—roughly 0.026 years for every unit increase in activity—for moderate activity, with a coefficient of 0.02580. Notably, this finding lacked statistical significance (p-value = 0.391), and the model accounted for only a tiny fraction of the variance in lifespan (Multiple R-squared: 0.00217). These results imply that factors beyond our current scope may play more pivotal roles in influencing longevity.

These results suggest that although physical activity is clearly good for health, there is no statistically significant correlation between it and longer life expectancy. Therefore, taking into account the wider range of factors impacting lifetime, any proposal for an intervention aimed at improving longevity that just focuses on increasing moderate physical activity should be further researched and see maybe additional or different factors.

BMI Relationship with Moderate Activity Level (Question no. 4)

Question four explores the hypothesis that BMI decreases as the proportion of moderate activity increases. It involves using appropriate visualizations and a linear regression model to test this hypothesis and quantify the relationship, providing insights into the impact of moderate activity on BMI.

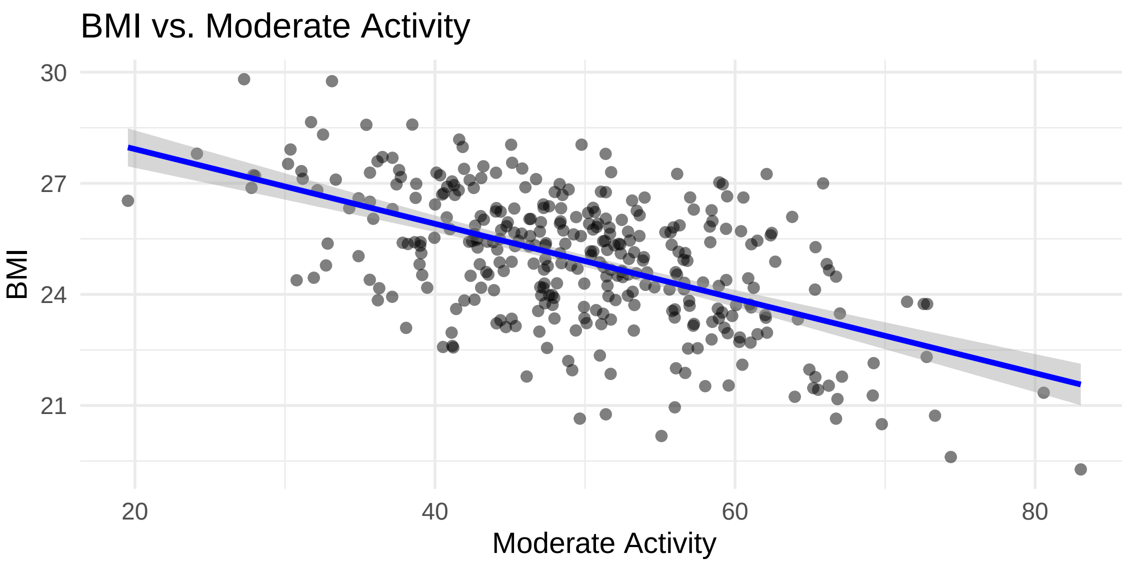
To perform the linear regression analysis exploring the relationship between BMI and moderate activity, and to visualize this relationship, you can use the following R code:

```r
model2 <- lm(bmi ~ moderate_activity, data = carehome_data)

# Display the summary of the regression model
summary(model2)
```

```
##
## Call:
## lm(formula = bmi ~ moderate_activity, data = carehome_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.2880 -0.9991  0.0876  1.0663  3.6989
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       29.938631   0.415956   71.98   <2e-16 ***
## moderate_activity -0.100816   0.008203  -12.29   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.502 on 339 degrees of freedom
## Multiple R-squared:  0.3083, Adjusted R-squared:  0.3062
## F-statistic: 151.1 on 1 and 339 DF,  p-value: < 2.2e-16
```

```r
# Plot BMI vs. Moderate Activity with regression line
ggplot(carehome_data, aes(x = moderate_activity, y = bmi)) +
```

```
geom_point(alpha = 0.5) +  # Plot individual data points
geom_smooth(method = "lm", color = "blue") +  # Add linear regression line
theme_minimal() +
labs(title = "BMI vs. Moderate Activity",
     x = "Moderate Activity",
     y = "BMI")
```



The linear regression analysis results indicate a significant relationship between moderate activity and BMI. The coefficient for moderate activity is -0.100816, with a highly significant p-value (<2e-16), suggesting that for every unit increase in moderate activity, BMI decreases by approximately 0.101 units. The negative sign of the coefficient confirms that the relationship is inverse, aligning with the hypothesis that increased moderate activity is associated with lower BMI.

The intercept, 29.938631, represents the estimated BMI when moderate activity is zero. The t-value for the moderate activity coefficient, -12.29, further emphasizes its statistical significance.

The model's residual standard error is 1.502, indicating the average distance of the data points from the fitted regression line. The R-squared value of 0.3083 suggests that approximately 30.83% of the variability in BMI can be explained by the model, which is a moderate amount of explanatory power.

Overall, the analysis provides strong evidence supporting the hypothesis that higher levels of moderate activity are associated with lower BMI values, making a compelling case for promoting moderate physical activity as part of weight management strategies.

As we referred earlier as Body Mass Index (BMI) can be an indication for better health outcome as per the reference (Ding and Jiang 2020), and we can find higher moderate activity is directly related with lower BMI.

Conclusion

We acknowledge the well-established benefits of physical activity on overall health, which may indirectly influence longevity through factors like improved cardiovascular health and weight management. Despite the lack of a significant direct link between physical activity levels, including moderate activity, and increased longevity, our study's conclusion takes this relationship into account. Findings also show an intriguing correlation between BMI and physical activity, indicating that regular physical activity may help sustain a lower BMI, which is linked to increased longevity.

The complex relationship that exists between longevity, BMI, and physical activity highlights the diverse range of factors that affect health and lifespan. It suggests that although moderate exercise by itself might not be able to prolong life, it is still very important for preserving a healthy body mass index and improving general wellbeing.

In order to fully understand the complex web of influences on longevity, future study should adopt a more holistic approach and take a wider range of factors into account, such as physical activity and BMI, among others. This thorough comprehension will be essential for creating recommendations and actions in health care that are more successful.

References

Ding, Chen, and Yumei Jiang. 2020. "The Relationship Between Body Mass Index and Physical Fitness Among Chinese University Students: Results of a Longitudinal Study." *Healthcare* 8 (4): 570. https://doi.org/10.3390/healthcare8040570.

Goodman, Claire, Tom Dening, Adam L. Gordon, Susan L. Davies, Julienne Meyer, Finbarr C. Martin, John R. F. Gladman, et al. 2016. "Effective Health Care for Older People Living and Dying in Care Homes: A Realist Review." *BMC Health Services Research* 16 (1): 269. https://doi.org/10.1186/s12913-016-1493-4.

Luppa, M., T. Luck, S. Weyerer, H.-H. Konig, E. Brahler, and S. G. Riedel-Heller. 2010. "Prediction of Institutionalization in the Elderly. A Systematic Review." *Age and Ageing* 39 (1): 31–38. https://doi.org/10.1093/ageing/afp202.