1(a)

a. # Replace 'path\_to\_file' with the actual file path if the file is not in the working directory

vinegar\_data <- vinegar

# Boxplot to visualize acidity differences between factory locations

boxplot(vinegar\_data$pH ~ vinegar\_data$Site, data = vinegar\_data,

xlab = "Factory Location", ylab = "Acidity",

main = "Acidity Levels across Factory Locations")A chart of a row of boxes

Description automatically generated

Interpretation

1. Variations in Median and Spread:Higher acidity level observed in Sydney location , lower acidity level observed in New York
2. iqr\_values <- tapply(vinegar\_data$pH, vinegar\_data$Site,IQR)

# Calculate IQR for each factory location

iqr\_values <- tapply(vinegar\_data$pH, vinegar\_data$Site,IQR)

View(iqr\_values)

A screenshot of a computer

Description automatically generated

1. Spread :Interquartile range highest in Sydney ,followed by Birmingham suggesting higher variability in acidity levels
2. Outliers: The outliers tending to the highest acidity values are observed in Paris, while tending to the minimum acidity values are observed in Birmingham.
3. Comparison of Acidity Levels Between Locations:Sydney has consistently higher acidity values , while New York has consistently lower acidity values.

1(b)

1. Hypothesis formulation : We will go with **Alternative Hypothesis (H1):** as There is at least one pair of factory locations with significantly different mean acidity levels.
2. Perform anova test
   1. # Perform one-way ANOVA
   2. anova\_result <- aov(vinegar\_data$pH ~ vinegar\_data$Site, data = vinegar\_data)
   3. # Summary of ANOVA results
   4. summary(anova\_result)

> summary(anova\_result)

A screenshot of a computer

Description automatically generated Df Sum Sq Mean Sq F value Pr(>F)

vinegar\_data$Site 4 24.57 6.143 6.682 0.000534 \*\*\*

Residuals 31 28.50 0.919

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

**Interpretation:**As P value is less than 0.05 , we reject the null hypothesis ,indicating significant differences in mean acidity levels among factory locations.As there are differences , we go for pairwise tests

> pairwise\_result <- pairwise.t.test(vinegar\_data$pH, vinegar\_data$Site, p.adj = "none")

> pairwise\_result

Pairwise comparisons using t tests with pooled SD

data: vinegar\_data$pH and vinegar\_data$Site

Birmingham London New York Paris

London 0.10843 - - -

New York 0.07349 0.88535 - -

Paris 0.66869 0.04539 0.02842 -

Sydney 0.01358 0.00017 7.6e-05 0.03654

P value adjustment method: none

A screenshot of a computer

Description automatically generated

the value 0.04539 for the comparison between Paris and London.

The p-value (0.04539) is below the conventional significance level of 0.05.

Interpretation: There's evidence to reject the null hypothesis for the pH levels between Paris and London. It suggests a statistically significant difference in mean pH levels between these two locations.

Anova analysis

> grp = factor(vinegar\_data$Site)

> grp

[1] London London London London London London London Birmingham Birmingham Birmingham

[11] Birmingham Birmingham Birmingham Birmingham Sydney Sydney Sydney Sydney Sydney Sydney

[21] Sydney New York New York New York New York New York New York New York New York Paris

[31] Paris Paris Paris Paris Paris Paris

Levels: Birmingham London New York Paris Sydney

> y = vinegar\_data$pH

> y

[1] 3.96 4.26 2.11 3.47 3.79 2.99 2.08 3.01 2.50 5.03 3.80 5.25 4.95 4.05 6.68 4.55 5.60 4.36 6.82 5.56 4.41 3.77 3.38

[24] 4.36 3.19 1.89 4.30 2.57 1.86 3.51 5.61 5.15 4.16 3.55 3.86 4.30

> drug.lm = lm(y ~ grp)

> anova(drug.lm)

Analysis of Variance Table

Response: y

Df Sum Sq Mean Sq F value Pr(>F)

grp 4 24.571 6.1427 6.682 0.0005335 \*\*\*

Residuals 31 28.498 0.9193

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

>

> oneway.test(y ~ grp, var.equal = TRUE)

One-way analysis of means

data: y and grp

F = 6.682, num df = 4, denom df = 31, p-value = 0.0005335

A larger F-value suggests larger differences between group means relative to the variability within groups.

The p-value obtained (p-value = 0.0005335) represents the probability of observing such an extreme F-statistic if the null hypothesis (no difference in means) were true.

In this case, the p-value is very low (less than the conventional significance level of 0.05), suggesting strong evidence against the null hypothesis.

The obtained p-value of 0.0005335 is less than 0.05 (or any chosen significance level), indicating that there are significant differences in mean acidity levels among the different site locations.

Based on this one-way ANOVA test, there is evidence to suggest that at least one group has a significantly different mean acidity level compared to the others.

1( c )

**> shapiro.test(residuals)**

**Shapiro-Wilk normality test**

**data: residuals**

**W = 0.94045, p-value = 0.05246**

# Fitting the ANOVA model

anova\_model <- aov(formula = pH ~ Site, data = vinegar\_data)

# Extracting residuals from the ANOVA model

residuals <- residuals(anova\_model)

# Checking assumptions - Normality and Homogeneity of Variance

par(mfrow = c(1, 2))

# Residuals vs. Fitted Values (Homogeneity of Variance)

plot(anova\_model, which = 1)

# Normal Q-Q plot (Normality)

qqnorm(residuals)

qqline(residuals)

**A graph of a normal q-q plot

Description automatically generated**

**Levene's Test for Homogeneity of Variances:**

Using Levene's Test (leveneTest()):

> levene.test(acidityLevels,site)

Modified robust Brown-Forsythe Levene-type test based on the absolute deviations from the

median

data: acidityLevels

Test Statistic = 0.28038, p-value = 0.8884

**Visual Inspection of Variance Homogeneity:**

**A screenshot of a computer

Description automatically generated**

**A graph with red lines and numbers

Description automatically generated**

**Interpretation**:

**Interpretation of Diagnostic Plots:**

1. **Residuals vs. Fitted Values Plot (Homogeneity of Variance):**
   * the random scattering of points around the horizontal line is zero.
   * No distinct patterns or trends are visible; this confirms for homogeneity of variance.
   * If a cone or fan-like shape is observed, it suggests heteroscedasticity, violating the assumption – no such pattern is observed.
2. **Normal Q-Q Plot (Normality):**
   * Points fall approximately along the diagonal line.
   * Some of the points points deviate significantly from the line, it indicates departure from normality.
3. Shapiro test for normality test:

The Shapiro-Wilk test for normality of residuals from the ANOVA model yields a p-value of approximately 0.05246. Typically, in statistical practice, the commonly used significance level for hypothesis testing is 0.05.

In this case, the obtained p-value (0.05246) is slightly higher than the typical significance level of 0.05. This suggests that there is no strong evidence to reject the null hypothesis, implying that the residuals may be normally distributed.

Levene's Test Results:

A non-significant result (p > 0.05) from Levene's test suggests that the variances among groups are approximately equal.

Residuals Plot:

The absence of systematic patterns suggests that variances are homogeneous among different groups.

1(d)

Call:

aov(formula = vinegar\_data$pH ~ vinegar\_data$Site, data = vinegar\_data)

Terms:

vinegar\_data$Site Residuals

Sum of Squares 24.57082 28.49786

Deg. of Freedom 4 31

Residual standard error: 0.9587939

Estimated effects may be unbalanced

A screenshot of a computer

Description automatically generated

Interpretation of Anova Results:

* The ANOVA results suggest that there is some variation in pH levels among the different factory locations.
* The significant sum of squares associated with 'vinegar\_data$Site' indicates that the pH levels vary significantly among the factory locations.
* The residual standard error gives an estimate of the variability of the observed pH levels around the mean predicted by the model.

Based on these results, further post-hoc tests (if required) or pairwise comparisons between factory locations can be conducted to determine specific differences in acidity levels among these locations.

A screenshot of a computer

Description automatically generated

**Interpretation of Tukey's HSD Results:**

* **p adj (Adjusted p-value):** This value determines the statistical significance of the observed differences after correcting for multiple comparisons. Smaller p-values (< 0.05) indicate significant differences.

**Pairwise Comparisons (95% Confidence Level):**

1. **London vs. Birmingham:**
   * Not statistically significant (p = 0.477).
2. **New York vs. Birmingham:**
   * Not statistically significant (p = 0.363).
3. **Paris vs. Birmingham:**
   * Not statistically significant (p = 0.992).
4. **Sydney vs. Birmingham:**
   * Marginally significant (p = 0.092).
5. **New York vs. London:**
   * Not statistically significant (p = 0.999).
6. **Paris vs. London:**
   * Not statistically significant (p = 0.252).
7. **Sydney vs. London:**
   * Statistically significant (p = 0.002).
   * There is a significant difference in acidity levels between Sydney and London.
8. **Paris vs. New York:**
   * Not statistically significant (p = 0.172).
9. **Sydney vs. New York:**
   * Statistically significant (p = 0.001).
   * There is a significant difference in acidity levels between Sydney and New York.
10. **Sydney vs. Paris:**
    * Not statistically significant (p = 0.212).

**Key Findings**:

Some pairwise comparisons, such as Sydney-London and Sydney-New York, show substantial estimated mean differences with adjusted p-values (< 0.05), suggesting significant differences in pH levels between these pairs of locations.

Conversely, comparisons like New York-Birmingham and Paris-Birmingham have wide confidence intervals that include zero, indicating non-significant differences in mean pH levels between these locations.

**Conclusion**:

Significant differences in pH levels are observed between specific pairs of factory locations, while other comparisons do not demonstrate significant differences.

Consider the adjusted p-values and confidence intervals to assess the significance of the differences after adjusting for multiple comparisons..

Based on the Tukey's HSD test, significant differences in acidity levels were observed between Sydney and both London and New York.

Other comparisons did not reveal statistically significant differences in acidity levels among the factory locations.

These results suggest that the acidity levels of balsamic vinegar between Sydney and both London and New York are significantly different. However, for other pairwise comparisons, no significant differences in acidity were found between the mentioned locations.

1(e)

> p\_values < - posthoc$`vinegar\_data$Site`[, "p adj"]

London-Birmingham New York-Birmingham Paris-Birmingham Sydney-Birmingham

FALSE FALSE FALSE FALSE

New York-London Paris-London Sydney-London Paris-New York

FALSE FALSE FALSE FALSE

Sydney-New York Sydney-Paris

FALSE FALSE

> print(p\_values)

[1] 0.4765222 0.3632112 0.9923747 0.0916315 0.9998934 0.2519132 0.0015038

[8] 0.1723996 0.0006834 0.2118486

> holm\_corrected <- p.adjust(posthoc$`vinegar\_data$Site`[, "p adj"], method = "holm")

> holm\_corrected

London-Birmingham New York-Birmingham Paris-Birmingham Sydney-Birmingham

1.000000000 1.000000000 1.000000000 0.733052161

New York-London Paris-London Sydney-London Paris-New York

1.000000000 1.000000000 0.013533869 1.000000000

Sydney-New York Sydney-Paris

0.006833638 1.000000000

>

> # Apply Bonferroni correction

> bonferroni\_corrected <- p.adjust(posthoc$`vinegar\_data$Site`[, "p adj"], method = "bonferroni")

>

> # Display the corrected p-values

> bonferroni\_corrected

London-Birmingham New York-Birmingham Paris-Birmingham Sydney-Birmingham

1.000000000 1.000000000 1.000000000 0.916315201

New York-London Paris-London Sydney-London Paris-New York

1.000000000 1.000000000 0.015037632 1.000000000

Sydney-New York Sydney-Paris

0.006833638 1.000000000

>

> correction\_results = data.frame(original\_values = p\_values, Holm\_correction = holm\_corrected, bonferroni\_corrected = bonferroni\_corrected)

>

> print(correction\_results)

original\_values Holm\_correction bonferroni\_corrected

London-Birmingham 0.4765222 1.000000000 1.000000000

New York-Birmingham 0.3632112 1.000000000 1.000000000

Paris-Birmingham 0.9923747 1.000000000 1.000000000

Sydney-Birmingham 0.0916315 0.733052161 0.916315201

New York-London 0.9998934 1.000000000 1.000000000

Paris-London 0.2519132 1.000000000 1.000000000

Sydney-London 0.0015038 0.013533869 0.015037632

Paris-New York 0.1723996 1.000000000 1.000000000

Sydney-New York 0.0006834 0.006833638 0.006833638

Sydney-Paris 0.2118486 1.000000000 1.000000000

>

A screenshot of a computer code

Description automatically generated

Observations:

Holm's correction adjusts the p-values to control for multiple comparisons while maintaining a familywise error rate.

The adjusted p-values are generally higher after correction, which may indicate a stricter criterion for identifying significant differences.

Two comparisons (Sydney-London and Sydney-New York) show adjusted p-values below 0.05 after Holm's correction, suggesting significant differences in pH levels between these pairs of locations.

Other comparisons have adjusted p-values greater than 0.05, indicating non-significant differences after Holm's correction.

Conclusion:

Holm's correction has made the criteria for significance more stringent, resulting in fewer significant differences between factory locations in terms of pH levels.

Consider these adjusted p-values to interpret the significance of differences after applying the correction. In this case, Sydney-London and Sydney-New York comparisons remain statistically significant after Holm's correction.

**Bonferroni Correction**:

**Observations**:

Bonferroni correction is a conservative method that adjusts the p-values to control for multiple comparisons, maintaining a stricter significance threshold.

After the Bonferroni correction, most adjusted p-values remain above 0.05, suggesting non-significant differences between the majority of the pairs of factory locations regarding pH levels.

Only one comparison (Sydney-London) shows an adjusted p-value below 0.05 after Bonferroni correction, indicating a significant difference in pH levels between these two locations.

Conclusion:

Bonferroni correction tends to be more conservative, requiring a lower threshold for significance compared to Holm's correction.

In this case, after Bonferroni correction, only the Sydney-London comparison remains statistically significant, while other comparisons show non-significant differences in pH levels between factory locations.

**Summary:**

* In both correction methods, some pairwise comparisons resulted in adjusted p-values lower than 0.05, indicating statistical significance even after multiple comparison adjustments.
* For instance, 'Sydney-London' and 'Sydney-New York' comparisons have adjusted p-values below 0.05, indicating significant differences in acidity levels between these pairs after correction.

1(f)

The experiment displays statistically significant acidity differences between certain factory locations, particularly evident in pairs like Sydney-London (p = 0.0015) and Sydney-New York (p = 0.00068) based on both ANOVA (p < 0.05) and Tukey HSD tests. Enhancements could involve larger sample sizes per location to increase statistical robustness and reliability. Exploring additional factors, such as production methods or environmental variables, may elucidate influential factors impacting acidity variation. Implementing complementary analyses or controlling potential confounders could yield a more comprehensive understanding of acidity disparities among factory locations.

2(a)

# Boxplot of Venom Yield by Body Class and Expression

boxplot(`Yield (mg)` ~ `Body Class` \* `Expression`, data = VenomYield,

xlab = "Body Class and Expression", ylab = "Venom Yield (mg)",

main = "Venom Yield Distribution by Body Class and Expression")

A graph of a graph showing a number of different sizes

Description automatically generated with medium confidence

Interpretation:

1.Venom yield Median values are observed higher for Body Class ‘Large’ and Expression ‘High’.

2.For Body Class ‘Small and Expression ‘High’ , there is a variability of Venom Yield values given by IQR(Inter quartile ranges)

3.For

Body Class ‘Small and Expression ‘Low’ smaller median values for Venom yield are observed as well as min values

4.The outliers which were at 1.5 + IQR are highest for Body Class ‘Large’ and Expression ‘High’.

5.The outliers ,both 1.5+IQR,1.5 -IQR were uniform for Body Class ‘Small and Expression ‘Low’

2(b)

model <- aov(`Yield (mg)` ~ `Body Class` \* `Expression`, data = VenomYield)

# Perform the two-way ANOVA

summary(model)

> model <- aov(`Yield (mg)` ~ `Body Class` \* `Expression`, data = VenomYield)

> summary(model)

Df Sum Sq Mean Sq F value Pr(>F)

`Body Class` 1 68190 68190 49.160 3.17e-08 \*\*\*

Expression 1 53329 53329 38.446 3.75e-07 \*\*\*

`Body Class`:Expression 1 976 976 0.704 0.407

Residuals 36 49936 1387

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Interpretation:

'Body Class' shows a significant effect on 'Yield (mg)' with a very low p-value (p < 0.001). This implies that there is a significant difference in venom yield between the 'Large' and 'Small' body classes of spiders.

'Expression' also exhibits a significant effect on 'Yield (mg)' with a very low p-value (p < 0.001). This suggests that differences in gene expression levels ('High' and 'Low') significantly influence venom yield.

However, the interaction between 'Body Class' and 'Expression' does not appear to have a significant effect on 'Yield (mg)' as indicated by a non-significant p-value (p = 0.407). This means that the impact of one factor on venom yield does not depend significantly on the levels of the other factor.

In summary, both 'Body Class' and 'Expression' have a significant individual effect on venom yield, but their interaction does not significantly influence the venom yield of the spiders in this study.

The residual term represents unexplained variability or random error in the model. The residual mean square (1387) indicates the average variability of data points around the fitted values.

2 (c )

# Running ANCOVA

ancova\_result <- lm(`Yield (mg)` ~ `Expression` \* `Body Length (cm)`, data = VenomYield)

summary(ancova\_result)

> summary(ancova\_result)

Call:

lm(formula = `Yield (mg)` ~ Expression \* `Body Length (cm)`,

data = VenomYield)

Residuals:

Min 1Q Median 3Q Max

-69.880 -15.209 -7.919 14.999 56.419

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 99.338 16.136 6.156 4.30e-07 \*\*\*

Expressionlow -53.046 23.213 -2.285 0.0283 \*

`Body Length (cm)` 40.678 4.908 8.287 7.28e-10 \*\*\*

Expressionlow:`Body Length (cm)` -7.467 6.990 -1.068 0.2925

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 28.25 on 36 degrees of freedom

Multiple R-squared: 0.8334, Adjusted R-squared: 0.8195

F-statistic: 60.01 on 3 and 36 DF, p-value: 4.406e-14

Here is the interpretation of the ANCOVA results:

Expression (High/Low):

The coefficient for 'Expressionlow' (-53.046) indicates that, on average, spiders with 'low' expression have 53.046 mg lower venom yield compared to those with 'high' expression, holding 'Body Length' constant.

This effect is statistically significant as indicated by the p-value (0.0283 < 0.05).

Body Length (cm):

The coefficient for 'Body Length (cm)' (40.678) signifies that for every 1 cm increase in body length, there is an average increase of 40.678 mg in venom yield, adjusting for 'Expression'.

This effect is statistically significant with a very low p-value (7.28e-10).

Interaction Term: Expressionlow x Body Length (cm):

The interaction term 'Expressionlow:Body Length (cm)' (-7.467) shows the change in the slope of 'Body Length' concerning 'Expression'.

The interaction effect is not statistically significant (p-value = 0.2925 > 0.05), indicating that the relationship between 'Body Length' and 'Yield (mg)' does not differ significantly between 'Expression' levels.

Overall Model:

The overall model is statistically significant (p-value: 4.406e-14), suggesting that the combined effects of 'Expression', 'Body Length', and their interaction can predict 'Yield (mg)' significantly well.

The Adjusted R-squared value (0.8195) indicates that around 81.95% of the variability in 'Yield (mg)' can be explained by 'Expression', 'Body Length', and their interaction in this model.

This ANCOVA indicates that both 'Expression' and 'Body Length' have significant effects on 'Yield (mg)' in funnel-web spiders. 'Expression' affects venom yield, and for every additional centimeter in 'Body Length', there is an increase in venom yield, while the interaction between 'Expression' and 'Body Length' does not significantly affect venom yield.

2(d)

Body Length seems to be a suitable covariate for the ANCOVA based on the following reasons:

Relevance to the Outcome Variable (Venom Yield): Body Length often correlates with various biological traits in organisms. In this context, the body size of spiders might affect venom production. A larger body size could accommodate larger venom glands, affecting venom yield.

Theoretical Justification: There exists a plausible theoretical relationship between body size and venom yield in spiders. Larger spiders potentially have larger venom glands or can produce more venom due to increased metabolic activity.

Statistical Significance: The ANCOVA output indicates that 'Body Length' is statistically significant (p < 0.05) in predicting 'Yield (mg)' after adjusting for 'Expression'. This suggests that changes in 'Body Length' are associated with changes in 'Yield (mg)'.

Adjusted R-squared Value: The adjusted R-squared value of the model including 'Body Length' as a covariate is relatively high (0.8195), suggesting that 'Body Length' explains a substantial portion of the variability in 'Yield (mg)'.

2(e)

he choice between performing an ANOVA (b) and ANCOVA (c) depends on the research question, the variables' relationships, and the study's objectives:

1. **ANOVA (b):**
   * **Advantages:** ANOVA helps analyze whether there are significant differences in mean venom yield among different levels of the 'Body Class' and 'Expression' factors. It examines group differences without considering other potential influential variables (covariates).
   * **Suitability:** ANOVA is appropriate when there is no need to control for or consider the influence of other continuous variables on the outcome variable (venom yield). It assumes that the effect of the factors (e.g., 'Body Class' and 'Expression') on the outcome is not confounded by other variables.
   * **Limitations:** ANOVA does not account for or control the potential impact of covariates like 'Body Length'. If 'Body Length' has a significant influence on venom yield and is related to the categorical predictors ('Body Class' and 'Expression'), ANOVA might not fully account for this relationship.
2. **ANCOVA (c):**
   * **Advantages:** ANCOVA extends ANOVA by including one or more continuous variables (covariates) alongside categorical predictors. It enables the assessment of group differences while controlling for the effects of continuous variables.
   * **Suitability:** ANCOVA is suitable when there's a need to control for the influence of continuous variables ('Body Length') on the outcome ('Yield (mg)') while examining group differences associated with categorical predictors ('Body Class' and 'Expression').
   * **Limitations:** ANCOVA assumes a linear relationship between the covariate(s) and the outcome variable. If this assumption is violated, the results might be misleading. Additionally, ANCOVA assumes that the covariate(s) do not interact with the categorical predictors, which should be assessed and considered in the analysis.

The better approach depends on whether 'Body Length' is considered an influential variable that impacts venom yield and if it is related to the categorical predictors. If 'Body Length' is deemed an essential variable influencing venom yield and is related to the categorical predictors, ANCOVA would be a better choice to control for its influence. If 'Body Length' is not considered influential or is independent of the categorical predictors, ANOVA might be more appropriate for exploring group differences in venom yield.

3(a)

# Define the hazard function

hazard <- function(t) {

return(1 - exp(-t))

}

# Survival function

survival\_function <- function(t) {

return(exp(-integrate(hazard, lower = 0, upper = t)$value))

}

# Failure probability density function

failure\_density <- function(t) {

return(hazard(t) \* survival\_function(t))

}

# Create a sequence of time points

time\_points <- seq(0, 10, by = 0.1)

# Calculate S(t) and f(t) for the time points

survival\_values <- sapply(time\_points, survival\_function)

failure\_density\_values <- sapply(time\_points, failure\_density)

# Plotting the survival function S(t)

plot(time\_points, survival\_values, type = 'l', xlab = 'Time (years)', ylab = 'Survival Probability', main = 'Survival Function S(t)')

# Plotting the failure probability density function f(t)

plot(time\_points, failure\_density\_values, type = 'l', xlab = 'Time (years)', ylab = 'Failure Density', main = 'Failure Probability Density Function f(t)')

A graph of a function

Description automatically generated

A graph of a function

Description automatically generated

* The "Survival Function S(t)" illustrates the declining trend of survival probabilities as time progresses, providing insights into the likelihood of survival at various time intervals.
* The "Failure Probability Density Function f(t)" highlights the hazard rate, showing how the risk of failure or event occurrence changes over time.

3(b)

> # Given data

> time\_intervals <- c(0, 1, 2, 3, 4, 5, 6, 7, 8, 9)

> ninint <- 100

> nlost <- c(1, 0, 3, 4, 11, 8, 8, 15, 17, 10)

> nevent <- c(0, 0, 2, 0, 0, 5, 3, 4, 2, 1)

>

> # Generating the life table

> my\_table <- lifetab(time\_intervals, ninint, nlost, nevent)

Warning messages:

1: In c(diff(-1 \* Sj), NA)/diff(tis) :

longer object length is not a multiple of shorter object length

2: In nevent/diff(tis) :

longer object length is not a multiple of shorter object length

3: In Sj \* qj/diff(tis) :

longer object length is not a multiple of shorter object length

4: In hmj \* diff(tis) :

longer object length is not a multiple of shorter object length

>

> print(my\_table)

nsubs nlost nrisk nevent surv pdf hazard se.surv se.pdf se.hazard

0-1 100 1 99.5 0 1.0000000 0.00000000 0.00000000 0.00000000 NaN NaN

1-2 99 0 99.0 0 1.0000000 0.00000000 0.00000000 0.00000000 NaN NaN

2-3 99 3 97.5 2 1.0000000 0.02051282 0.02072539 0.00000000 0.01435522 0.01465428

3-4 94 4 92.0 0 0.9794872 0.00000000 0.00000000 0.01435522 NaN NaN

4-5 90 11 84.5 0 0.9794872 0.00000000 0.00000000 0.01435522 NaN NaN

5-6 79 8 75.0 5 0.9794872 0.06529915 0.06896552 0.01435522 0.02822868 0.03082397

6-7 66 8 62.0 3 0.9141880 0.04423490 0.04958678 0.03123227 0.02495929 0.02862014

7-8 55 15 47.5 4 0.8699531 0.07325921 0.08791209 0.03878172 0.03520520 0.04391356

8-9 36 17 27.5 2 0.7966939 0.05794138 0.07547170 0.04990109 0.03961934 0.05332854

9-1 17 10 12.0 1 0.7387525 NA NA 0.06080799 NA NA

>

> # Extracting S, f, and h

> S <- my\_table[, 5]

> f <- my\_table[, 6]

> h <- my\_table[, 7]

>

> # Adjusted time intervals

> t <- 0.5 + c(0:9)

>

> # Setting up a single plot area to combine all plots

> par(mfrow = c(3, 1)) # 3 rows, 1 column

>

> # Plotting all functions in one figure

> plot(t, S, type = 'l', col = 'blue', xlab = 'Time (years)', ylab = 'Survival Probability', main = 'Survival Function S(t)')

>

> plot(t, f, type = 'l', col = 'green', xlab = 'Time (years)', ylab = 'Failure Probability Density', main = 'Failure Density Function f(t)')

>

> plot(t, h, type = 'l', col = 'red', xlab = 'Time (years)', ylab = 'Hazard Function', main = 'Hazard Function h(t)')

>

A screenshot of a computer

Description automatically generated

Life table:

A number of numbers on a white background

Description automatically generated

A screenshot of a graph

Description automatically generated

***Interpretation*** :

The Survival function suggest the units have reasonable good life span with survival probability that decreases gradually over time.

The failure density appears to have peaks and troughs, indicating that there are specific times when the units are more likely to fail and other times when failures are less frequent

The hazard function curve is not constant, indicating that the risk of failure is not uniform

Showing some variability due to aging process of components used in AC .

***Conclusion on whether model in part(a) is accurate or not:***

The model provided in (a) is a basic theoretical representation of survival analysis based on a hazard function, survival function, and failure probability density function. However, without specific context or real data to compare these theoretical functions against observed survival data, it's difficult to evaluate the accuracy or applicability of this model in a real-world scenario.

Here are some considerations to assess the model:

1. **Theoretical Soundness:** The model is built upon fundamental concepts of survival analysis, including the hazard function, survival function, and failure density function. The functions are derived mathematically and follow the standard definitions used in survival analysis.
2. **Hazard Function:** The hazard function 1−exp(−*t*) used in this model is a simple form that indicates an increasing hazard over time. Depending on the context and characteristics of the data, this may or may not accurately represent the true hazard pattern. Real hazard functions can be more complex and may vary for different populations or conditions.
3. **Survival Function and Failure Density:** The survival function and failure density are calculated based on the hazard function and integral computations. These functions describe the cumulative survival probability and failure density over time. Their accuracy depends on the accuracy of the hazard function and the integration method used.
4. **Validation against Data:** To assess the accuracy of the model, it needs to be validated against empirical data from the specific population or context it aims to represent. Real-life survival data should be used to fit the model and check how well it predicts observed survival patterns.
5. **Assumptions:** The model assumes a specific form for the hazard function, which might not hold true for all scenarios. Survival analysis often involves making assumptions about the underlying hazard function, and the appropriateness of these assumptions should be evaluated based on the data.

In summary, while the provided code demonstrates fundamental concepts of survival analysis by defining and plotting theoretical survival and failure density functions based on a hazard function, its accuracy and applicability in real scenarios can only be assessed by comparing these functions against observed survival data from the relevant population or study context. Without such validation against empirical data, it's challenging to ascertain the model's accuracy or usefulness for real-world predictions or analysis.

Top of Form

Bottom of Form