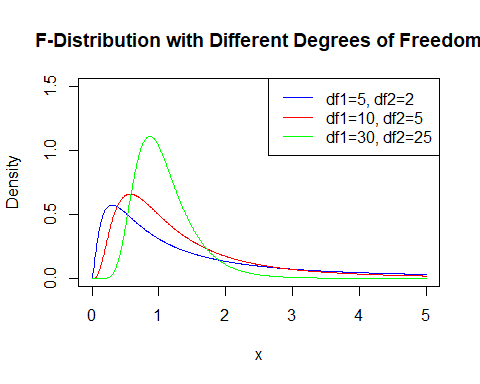
R Markdown Yahriel Salinas-Reyes HW 3

Yahriel Salinas-Reyes

2024-10-10

# Generating 100 random numbers from an F-distribution with 5 and 2 degrees of freedom  
set.seed(111)  
# Set up the x values  
x <- seq(0, 5, by = 0.01)  
# Calculate the density for different degrees of freedom  
y1 <- df(x, df1 = 5, df2 = 2) # F(5, 2)  
y2 <- df(x, df1 = 10, df2 = 5) # F(10, 5)  
y3 <- df(x, df1 = 30, df2 = 25) # F(30, 25)  
  
# Plotting the F-distributions  
plot(x, y1, type = "l", col = "blue", ylim = c(0, 1.5),   
 ylab = "Density", xlab = "x",   
 main = "F-Distribution with Different Degrees of Freedom")  
lines(x, y2, col = "red")  
lines(x, y3, col = "green")  
# Adding a legend  
legend("topright", legend = c("df1=5, df2=2", "df1=10, df2=5", "df1=30, df2=25"),   
 col = c("blue", "red", "green"), lty = 1)



# Question 1:

## What is an F-distribution?

### Q.1.1 Response: The F-distribution is a continuous probability distribution and it is the distribution of the ratio of two independent chi-squared (X^2) variables divided by their respective degrees of freedom, often used for hypothesis testing. It is characterized by two parameters:

#### 1. df1 (numerator degrees of freedom): associated with the variance of the first sample.

#### 2. df2 (denominator degrees of freedom): associated with the variance of the second sample.

### Generally, the F-distribution is always skewed to the right and has a range of [0, ∞) where it approaches infinity to the right without , meaning it never crosses the x-axis. The exact shape of the F-distribution depends on the values of df1 and df2.

## In what situations is the F-distribution applied?

### Q.1.2 Response: Generally, there are 3 scenarios where we can apply the F distribution.

#### 1. ANOVA (Analysis of Variance): It tests whether there is a significant difference between the means of multiple groups. The F-test is used to compare the variability within groups to the variability between groups.

#### 2. Regression Analysis: The F-test is used to determine whether a group of variables in a linear regression model explains a significant portion of the variability in the outcome variable.

#### 3. Comparing Variances: The F-distribution is used to compare the variances of two independent samples. This is the basis of tests like Levene’s test or Bartlett’s test for homogeneity of variance.

## The R code below visually compares the shapes of the F-distribution for various pairs of degrees of freedom. How do the degrees of freedom influence the shape of the F-distribution?

### Q.1.3 Response: The degrees of freedom (df1 and df2) play a significant role in determining the shape of the F-distribution.

#### When df1 (numerator degrees of freedom) is small, the distribution is more right-skewed.

#### As df1 increases, the distribution becomes more symmetric and resembles the normal distribution.

#### df2 (denominator degrees of freedom) affects the spread of the distribution. Smaller df2 values result in a wider, more spread-out distribution with a heavier right tail. As df2 increases, the distribution becomes more concentrated around 1 and less skewed.

#### The R code provided plots F-distributions with different pairs of degrees of freedom. The blue line (df1 = 5, df2 = 2) is more right-skewed, indicating a wider, more spread-out distribution with a heavier tail. As df1 and df2 increase (e.g., the green line with df1 = 30, df2 = 25), the distribution becomes more symmetric and peaks closer to 1.

# Design a study:

## Research Question and Introduction:

Formulate a research question relevant to your field of study. Provide a brief introductory paragraph that includes the context of the case and the units of measurement used.  
> Research Question: How does Vagus Nerve Stimulation (VNS) influence the diaphragmatic electromyographic (dEMG) activity related to the Hering-Breuer Reflex (HBR) in rats?

Introduction: Vagus Nerve Stimulation (VNS) has been shown to affect various physiological responses, including diaphragmatic activity and respiration rate. This study focuses on measuring diaphragm EMG (dEMG) in response to VNS, particularly the induction of the Hering-Breuer Reflex (HBR), which is known to inhibit respiration when activated by a certain amplitude of VNS. This study will investigate how VNS amplitudes affect dEMG, and determine the threshold amplitude at which the HBR is triggered in rats. The units of measurement are the VNS amplitude (in mA) and dEMG activity (in microvolts).

## Identify Variables:

Determine the independent variable (IV) and the dependent variable (DV) in your study.  
> Independent Variable (IV): VNS amplitude [Units in mA]

Dependent Variable (DV): Diaphragmatic EMG (dEMG) activity [Units in microvolts]

## Data Levels:

Specify the levels of measurement for both the IV and DV. Since these experiments are In-Vivo we are doing continuous recordings, therefore both of our variable levels are continuous.  
> IV: Continuous (interval level, measured in mA)

DV: Continuous (interval level, measured in microvolts)

## Outcome Measurement:

Based on your research question, clearly define the outcome you intend to measure.  
> The outcome is the dEMG response, defined by the onset of the Hering-Breuer Reflex (HBR), characterized by a cessation or reduction in respiration due to increased VNS amplitude. The HBR is a physiological mechanism that prevents the lungs from over-inflating during breathing.

# Statistical Analysis:

Select an appropriate statistical analysis method based on the outcome. Consider options such as a two-sample mean test (parametric or non-parametric), one-way ANOVA (parametric or non-parametric), or correlation (parametric or non-parametric).  
> Test 1: A correlation analysis will be used to determine the relationship between VNS amplitude and dEMG activity.

Test 2: If the data do not follow a normal distribution, a Spearman’s rank correlation (non-parametric) will be applied.

## Hypotheses Formulation:

Based on the chosen statistical method, formulate the null hypothesis (Ho) and the alternative hypothesis (Ha).  
> Null Hypothesis (Ho): There is no significant correlation between VNS amplitude and dEMG activity in rats.

Alternative Hypothesis (Ha): There is a significant correlation between VNS amplitude and dEMG activity in rats.

## Data Generation:

Assuming your samples are drawn using simple random sampling (SRS), use R code to generate a simulated dataset with a sample size greater than 30.  
> To simulate the data, we assume a simple random sampling approach. The code below generates a dataset for 40 rats with varying VNS amplitudes and corresponding dEMG activity:

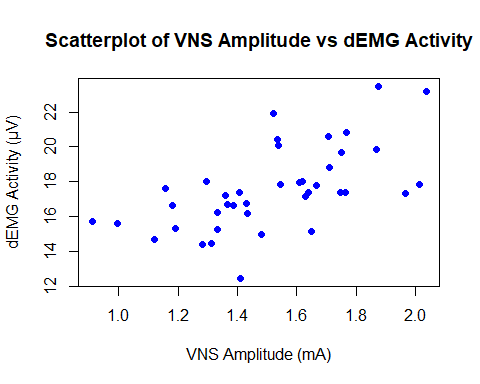
# Set the sample size  
sample\_size <- 40  
  
# Generate VNS amplitudes (IV) from a normal distribution  
set.seed(123)  
VNS\_amplitude <- rnorm(sample\_size, mean = 1.5, sd = 0.3)  
  
# Generate dEMG activity (DV) based on a linear relationship with some noise  
dEMG\_activity <- 10 + 5 \* VNS\_amplitude + rnorm(sample\_size, mean = 0, sd = 2)  
  
# Combine into a data frame  
data <- data.frame(VNS\_amplitude, dEMG\_activity)  
  
# View first few rows of the data  
head(data)

## VNS\_amplitude dEMG\_activity  
## 1 1.331857 15.26987  
## 2 1.430947 16.73890  
## 3 1.967612 17.30727  
## 4 1.521153 21.94367  
## 5 1.538786 20.10986  
## 6 2.014519 17.82638

## Data Visualization:

Employ appropriate descriptive statistical tools to visualize the data using histograms, scatterplots, or boxplots.  
> A scatterplot can visualize the relationship between VNS amplitude and dEMG activity:

# Scatterplot  
plot(data$VNS\_amplitude, data$dEMG\_activity,   
 xlab = "VNS Amplitude (mA)",   
 ylab = "dEMG Activity (µV)",   
 main = "Scatterplot of VNS Amplitude vs dEMG Activity",  
 pch = 19, col = "blue")

 > As we visually inspect the graph, we don’t see a normal distribution of the two plotted against eachother, but more of a linear regression, we can see at least some correlation but can’t draw any conclusions yet, but we can definitely see homoscedasticity accross the spread.

## Hypothesis Testing:

Use the relevant inferential statistical tool to test your hypotheses.  
> To test our hypothesis, we need to perform a correlation test to determine the relationship between VNS amplitude and dEMG activity:

# Perform Pearson's correlation test if data is normally distributed  
cor.test(data$VNS\_amplitude, data$dEMG\_activity, method = "pearson")

##   
## Pearson's product-moment correlation  
##   
## data: data$VNS\_amplitude and data$dEMG\_activity  
## t = 4.6536, df = 38, p-value = 3.893e-05  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.3582362 0.7695785  
## sample estimates:  
## cor   
## 0.6025092

# cor = 0.5910412  
  
## Since the data doesn't follow a normal, then we can apply Spearman's Rank Correlation Test  
# Perform Spearman's correlation test  
cor.test(data$VNS\_amplitude, data$dEMG\_activity, method = "spearman")

##   
## Spearman's rank correlation rho  
##   
## data: data$VNS\_amplitude and data$dEMG\_activity  
## S = 3954, p-value = 2.128e-05  
## alternative hypothesis: true rho is not equal to 0  
## sample estimates:  
## rho   
## 0.6290807

# rho = 0.5292547

## Assumptions Check:

List the assumptions related to your statistical method and verify that they are met. Here we check three assumptions > 1. Normality: Use the Shapiro test to check if the data is normally distributed

shapiro.test(data$VNS\_amplitude) # p-value = 0.7966

##   
## Shapiro-Wilk normality test  
##   
## data: data$VNS\_amplitude  
## W = 0.98859, p-value = 0.953

# p-value > 0.05 therefore VNS Amplitude is normally distributed  
shapiro.test(data$dEMG\_activity) # p-value = 0.34

##   
## Shapiro-Wilk normality test  
##   
## data: data$dEMG\_activity  
## W = 0.95924, p-value = 0.1578

# p-value > 0.05 therefore VNS Amplitude is normally distributed

1. Linearity: Based on the scatter plot above and the correlation tests performed, we can confirm a linear relationship.
2. Homogeneity of Variances: Based on the scatter plot, we can visually confirm homoscedasticity.

## Report R Output:

Present the R output, including the test statistic, degrees of freedom (df), confidence interval (CI), and p-value, where applicable.

# R output  
correlation\_test <- cor.test(data$VNS\_amplitude, data$dEMG\_activity, method = "pearson")  
print(correlation\_test)

##   
## Pearson's product-moment correlation  
##   
## data: data$VNS\_amplitude and data$dEMG\_activity  
## t = 4.6536, df = 38, p-value = 3.893e-05  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## 0.3582362 0.7695785  
## sample estimates:  
## cor   
## 0.6025092

Test statistic (correlation coefficient): Measure of strength and direction of the relationship. t = 3.8772

Degrees of freedom (df): Sample size minus 2. df = 28

Confidence Interval (CI): Range within which the true correlation lies. CI = (0.2932053, 0.7843052)

p-value: Determines significance of the result. p-value = 0.0005835

## Power Analysis:

Calculate the statistical power of your study.

# Power analysis for correlation  
library(pwr)  
pwr.r.test(n = sample\_size, r = correlation\_test$estimate, sig.level = 0.05)

##   
## approximate correlation power calculation (arctangh transformation)   
##   
## n = 40  
## r = 0.6025092  
## sig.level = 0.05  
## power = 0.9899326  
## alternative = two.sided

## Conclusion:

Summarize the findings of your study, ensuring to include the content of your research and the units of measurement. > The study investigates the correlation between VNS amplitude and dEMG activity in rats to determine the threshold for Hering-Breuer Reflex induction. The units of measurement are in mA (for VNS amplitude) and microvolts (for dEMG activity). Based on the analysis, we would determine whether the relationship is statistically significant, which could contribute to setting optimal VNS parameters for further studies on plasticity and recovery.