My first dataset analysis

November 16, 2024

```
[1]: install.packages("psych") ## Download and install psych for graphs
     library(psych) ## Load psych into this current instance of R but its slowing it␣
     ↪down

     df <- read.csv("trials.csv")
      #Add a new column named "BMI"
     df$BMI <- df$Weight / (df$Height^2 / 10000)

     #Print the modified data frame
```

also installing the dependencies 'mnormt', 'GPArotation'

Updating HTML index of packages in '.Library'

Making 'packages.html' …
 done

Part One First, lets produce a summary table of baseline patient characteristics for each treatment arm separately (i.e. in separate columns).Then, summarise Age, Sex, BMI, Diabetic status, SBP and DBP at the start of the study using appropriate summary statistics for each variable. Indicate in the table which statistics you use for each variable to show whether patients allocated into two arms are comparable.

```
[2]: names(df)#shows the different columns of the table
```

1. 'Patid' 2. 'Age' 3. 'Sex' 4. 'Trt' 5. 'SBP1' 6. 'DBP1' 7. 'SBP2' 8. 'DBP2' 9. 'Diab' 10. 'Weight' 11. 'Height' 12. 'BMI'

Statistics used for each variable:

Age: mean, range

Sex: percentage ratio

BMI: mean, standard deviation

Diabetic status: ratio(positive and in percentage)

SBP: mean, standard deviation

DBP: mean, standard deviaton

```
[3]: table(df$Diab)
```

```
     0   1
   191 109
```

```
[4]: table(df$Sex)
```

```
     0   1
   153 147
```

```
[5]: meanAge <-(mean(df$Age))
     min_age <- min(df$Age)
     max_age <- max(df$Age)
     sdAge <- sd(df$Age)
     ageRange<- max_age - min_age

     meanSBP <- (mean(df$SBP1))
     sdSBP <- (sd(df$SBP1))
     min_SBP <- min(df$SBP1)
     max_SBP <- max(df$SBP1)
     SBPrange<- max_SBP - min_SBP

     meanDBP <- (mean(df$DBP1))
     sdDBP<-(sd(df$DBP1))
     min_DBP <- min(df$DBP1)
     max_DBP <- max(df$DBP1)
     DBPrange<- max_DBP - min_DBP

     meanBMI<-(mean(df$BMI))
     sdBMI <-(sd(df$BMI))
     min_BMI <- min(df$BMI)
     max_BMI <- max(df$BMI)
     BMI_range<-max_BMI - min_BMI

     femalePercentage <- (147/300)*100
     malePercentage <- (153/300)*100

     maleDiab <- (malePercentage/191)*100 #showing the % value of males positive␣
      ↪with diabetes
     femaleDiab <- (femalePercentage/109)*100 #showing the % value of females␣
      ↪positive with diabetes

     maleDiab <- round(maleDiab,2) #showing the % value of males positive with␣
      ↪diabetes
```

```
femaleDiab <- round(femaleDiab,2) #showing the % value of females positive with␣
↪diabetes
```

Creating a chracterstics table for the partcipants

```
[6]: Characteristics <- c("Age", "Sex","BMI","Diabetic Status(postive %)","Systolic␣
↪Blood Pressure before(mmHg)","Diastolic Blood Pressure before(mmHg)")
mean <- c(meanAge,"na",meanBMI,"na",meanSBP,meanDBP)
sd <- c(sdAge,"na",sdBMI,"na",sdSBP,sdDBP)
range<- c(ageRange,"na",BMI_range,"na",SBPrange,DBPrange)
Males<-c("na",153,"na",maleDiab,"na","na")
Females<-c("na",147,"na",femaleDiab,"na","na")


summaryTable <- data.frame(Characteristics, mean, sd,range,Males,Females)
summaryTable
```

A data.frame: 6 × 6

Characteristics <chr>	mean <chr>	sd <chr>	range <chr>
Age	59.6266666666667	12.6382901063235	72
Sex	na	na	na
BMI	28.3101664401655	7.89670613075354	45.61084
Diabetic Status(postive %)	na	na	na
Systolic Blood Pressure before(mmHg)	130.713333333333	15.3068226057261	103
Diastolic Blood Pressure before(mmHg)	86.03	16.1536031993197	109

Now we will produce a scatterplot to visually explore whether there is an association between DBP before and after treatment. The data is mostly clustered in one area, there isn't a very clear relationship between the DBP before and after the trial, but it's not curve shaped so its linear. The ab line is also slightly increased over the course of the data which further supports a linear relationship.

```
[7]: plot(df$DBP1, df$DBP2,
  pch = c(16, 1), # Different point characters
  col = c("red", "blue"), # Colors for each point set
  xlab = "Before",
  ylab = "After",
  main = "Scatterplot of Diastolic Blood Pressure"
)
#displaying the labels for the graph
legend(x="right", legend = c("before","after"),col = c("red", "blue"), pch =␣
↪c(16, 1))
```

Scatterplot of Diastolic Blood Pressure
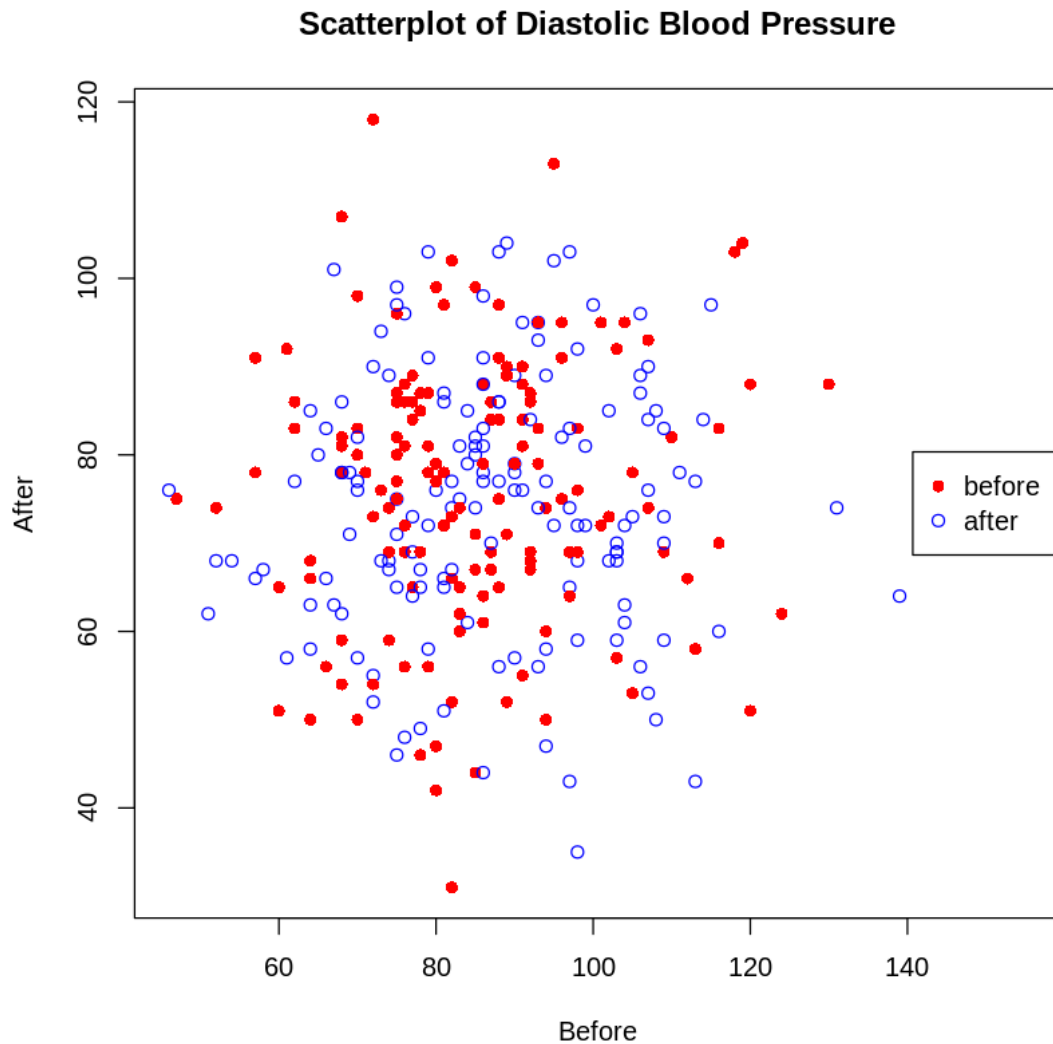
Lets calculate a suitable statistic to explore whether there is a linear association between DBP before and after treatment. Is the statistics providing you good evidence to apply linear regression to above mentioned association? Lets see... The coefficient calculated is 0.034 meaning the direction of the correlation is positive, but a weak positive correlation. However, when calculating the t statistic to get the p-value(0.54), it was larger than $p < 0.05$, proving no significant or concrete association between DBP before the trial and DBP after the trial.

```
[8]: DBP_summaryTable <- data.frame(
         Treatment = df$Trt,
         Diastolic_before = df$DBP1,
         Diastolic_after = df$DBP2)
     print(DBP_summaryTable)
```

       Treatment Diastolic_before Diastolic_after

1	1	83	65
2	1	90	79
3	1	119	104
4	1	103	59
5	0	72	118
6	0	79	103
7	1	68	81
8	1	113	43
9	0	62	86
10	0	86	91
11	0	60	51
12	0	104	61
13	1	91	81
14	0	93	95
15	0	92	87
16	0	84	79
17	1	88	84
18	1	69	71
19	0	87	69
20	1	97	103
21	0	79	78
22	1	86	81
23	1	75	82
24	0	73	94
25	0	93	95
26	0	85	81
27	1	91	55
28	0	78	67
29	1	71	78
30	0	88	86
31	1	107	74
32	1	98	35
33	0	81	78
34	0	72	90
35	0	91	88
36	0	116	60
37	0	94	60
38	1	98	68
39	1	77	65
40	0	82	77
41	0	85	99
42	0	72	55
43	1	94	50
44	1	90	76
45	1	77	86
46	0	102	68
47	0	110	82
48	1	67	63

49	1	68	107
50	0	61	57
51	0	92	67
52	0	85	82
53	0	80	42
54	1	94	58
55	0	88	91
56	1	103	70
57	1	75	82
58	0	106	96
59	1	74	74
60	0	108	85
61	1	76	56
62	1	66	66
63	0	86	64
64	1	75	65
65	1	120	51
66	0	77	64
67	1	87	67
68	0	75	99
69	1	72	73
70	0	46	76
71	0	88	91
72	1	64	63
73	1	87	84
74	1	98	59
75	0	80	99
76	0	78	49
77	1	75	86
78	1	93	93
79	0	79	81
80	0	62	77
81	0	73	76
82	0	58	67
83	1	74	59
84	0	87	70
85	0	70	50
86	1	100	97
87	1	86	61
88	0	75	75
89	1	80	79
90	0	76	96
91	0	94	74
92	0	79	72
93	1	64	50
94	1	90	78
95	0	116	83
96	0	107	84

97	0	105	53
98	0	94	89
99	0	52	74
100	1	68	78
101	1	57	91
102	0	64	85
103	1	64	68
104	0	82	74
105	0	79	56
106	0	88	56
107	1	92	69
108	0	107	53
109	1	98	76
110	1	74	67
111	0	96	75
112	1	74	68
113	0	75	77
114	1	97	65
115	1	91	84
116	1	52	68
117	1	76	86
118	1	92	84
119	1	95	113
120	1	103	69
121	0	98	69
122	0	111	78
123	0	75	87
124	0	75	71
125	0	77	84
126	1	115	97
127	0	78	87
128	1	109	59
129	1	76	69
130	0	89	104
131	1	82	52
132	0	78	65
133	1	89	90
134	0	85	80
135	0	75	80
136	1	81	66
137	1	105	78
138	1	102	85
139	1	85	67
140	1	83	81
141	1	76	81
142	0	86	78
143	1	75	96
144	0	93	56

145	0	78	85
146	0	94	77
147	1	75	75
148	0	80	76
149	0	85	44
150	1	85	74
151	0	68	82
152	0	95	102
153	0	116	70
154	0	66	83
155	0	92	68
156	1	90	89
157	1	82	66
158	1	70	57
159	0	68	78
160	1	88	103
161	0	85	71
162	1	68	62
163	1	155	72
164	0	103	68
165	0	93	83
166	0	99	72
167	0	107	93
168	1	106	56
169	1	80	77
170	1	108	50
171	1	83	60
172	1	83	75
173	1	101	72
174	1	68	86
175	0	96	95
176	0	131	74
177	0	118	103
178	1	97	43
179	0	83	74
180	1	105	73
181	0	61	92
182	0	67	101
183	0	78	46
184	1	84	61
185	1	82	102
186	0	75	97
187	0	86	79
188	0	70	77
189	1	113	58
190	1	81	51
191	1	68	59
192	1	109	83

193	1	72	73
194	0	81	87
195	0	64	66
196	1	114	84
197	1	130	88
198	1	68	78
199	0	81	97
200	0	72	52
201	0	74	69
202	0	93	74
203	0	98	83
204	1	104	63
205	0	88	65
206	1	109	73
207	0	91	90
208	1	79	91
209	1	104	95
210	1	96	82
211	1	97	69
212	0	107	90
213	1	66	56
214	1	65	80
215	1	88	75
216	0	90	57
217	1	77	89
218	1	70	76
219	0	89	71
220	1	88	77
221	0	103	92
222	1	86	83
223	0	86	88
224	1	81	65
225	1	81	72
226	0	103	69
227	1	103	57
228	1	76	48
229	1	60	65
230	1	54	68
231	0	77	84
232	0	64	58
233	0	89	52
234	0	77	69
235	0	88	97
236	1	106	89
237	0	70	98
238	0	98	92
239	1	90	79
240	0	99	81

241	1	97	64
242	1	107	76
243	0	82	73
244	0	74	89
245	1	70	83
246	0	69	78
247	0	87	86
248	0	77	73
249	0	109	69
250	0	86	77
251	0	83	62
252	1	86	44
253	1	124	62
254	1	73	68
255	1	82	31
256	1	139	64
257	1	47	75
258	1	95	72
259	0	72	54
260	1	84	85
261	0	76	88
262	0	104	72
263	1	101	95
264	1	79	58
265	1	76	72
266	0	81	86
267	0	89	89
268	0	86	88
269	1	79	78
270	1	75	46
271	0	57	78
272	1	91	76
273	1	80	47
274	0	94	47
275	0	62	83
276	1	113	77
277	0	76	72
278	1	97	83
279	1	92	68
280	1	91	95
281	0	92	86
282	1	82	67
283	1	112	66
284	1	86	98
285	0	68	54
286	1	109	70
287	0	102	73
288	1	70	82

289	0	79	87
290	1	88	86
291	1	120	88
292	1	51	62
293	0	96	91
294	1	97	74
295	0	78	69
296	1	98	72
297	1	70	80
298	0	57	66
299	1	93	79
300	0	106	87

[9]:
```
summary(DBP_summaryTable)
```

```
   Treatment       Diastolic_before Diastolic_after
 Min.   :0.0000   Min.   : 46.00    Min.   : 31.00
 1st Qu.:0.0000   1st Qu.: 75.00    1st Qu.: 66.00
 Median :1.0000   Median : 85.00    Median : 76.00
 Mean   :0.5067   Mean   : 86.03    Mean   : 75.11
 3rd Qu.:1.0000   3rd Qu.: 96.00    3rd Qu.: 85.25
 Max.   :1.0000   Max.   :155.00    Max.   :118.00
```

[10]:
```
# Generating Sample Data
set.seed(123) # Ensure reproducibility. The same "seed" will give the same␣
 ↪output every time
beforeGroup <- rnorm(300, mean = 86, sd = 16) # Control group
afterGroup <- rnorm(300, mean = 75, sd = 15) # Treatment group

# Viewing a summary of the data
summary(beforeGroup)
summary(afterGroup)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  49.05   76.79   85.30   86.55   96.12  137.86
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  32.85   65.95   75.68   75.14   85.33  113.57
```

[11]:
```
# Define the data
DBP_before <- DBP_summaryTable$Diastolic_before
DBP_after <- DBP_summaryTable$Diastolic_after

# Calculations
sum_x <- sum(DBP_before)
sum_y <- sum(DBP_after)
sum_xy <- sum(DBP_before * DBP_after)
sum_x_squared <- sum(DBP_before^2)
```

```
sum_y_squared <- sum(DBP_after^2)
n <- length(DBP_before) # Number of participants (observations)

# Apply the Pearson correlation coefficient formula
numerator <- n * sum_xy - sum_x * sum_y
denominator <- sqrt((n * sum_x_squared - sum_x^2) * (n * sum_y_squared -␣
  ↪sum_y^2))

r <- numerator / denominator

# Print the calculated correlation coefficient
print(r)
if (r < 0){
    print("direction of correlation is negative")
}else if (r==0){
    print("no correlation")
}else{
    print("direction of correlation is positive")
}
#correlation is weak positive
```

```
[1] 0.03494596
[1] "direction of correlation is positive"
```

```
[12]: #calculating t statistic of correlation coefficient to test significance
      #variables
      r<- 0.03494596
      n<- length(DBP_before)
      t<- (r* sqrt(n-2))/sqrt(1-(r^2))
      print(t)
```

```
[1] 0.6036295
```

```
[13]: #looking for a difference so its a two-tailed p-value, subtract the cumulative␣
      ↪probability from 1 and multiply by 2
      dof= n - 2
      p_value <- 2 * (1 - pt(abs(t), dof))
      p_value

      if (p_value < 0.05) {
        print("there is an evident association between DBP before and after␣
      ↪treatment")
      }else{
          print("there's no concrete association between DBP before and after")
      }
```

0.546549188934066

[1] "there's no concrete association between DBP before and after"

```
[14]: #for extra detail
      cor.test(DBP_before, DBP_after)
```

        Pearson's product-moment correlation

data: DBP_before and DBP_after
t = 0.60363, df = 298, p-value = 0.5465
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.07860595 0.14760269
sample estimates:
        cor
0.03494596

Now lets write down the model equation for the simple linear regression with DBP at the start of the study as the only covariate. $y_i = a + Bx + e_i$ where: • y is the DBP after trial • a is the intercept • Bx is the slope of DBP before trial • e is the standard error

```
[15]: #calculating beta_0 the interpect an beta_1 the slope

      # Calculate the necessary components
      mean_x <- mean(DBP_before)
      mean_y <- mean(DBP_after)

      # Calculate the slope (beta_1)
      beta_1 <- (n * sum_xy - sum_x * sum_y) / (n * sum_x_squared - sum_x^2)

      # Calculate the intercept (beta_0)
      beta_0 <- mean_y - beta_1 * mean_x

      # Print the coefficients
      print(paste("Slope: ", round(beta_1, 2)))
      print(paste("Intercept: ", round(beta_0, 2)))
```

[1] "Slope: 0.03"
[1] "Intercept: 72.37"

```
[16]: plot(DBP_before, DBP_after, main = paste("Intercept: ", round(beta_0, 2), " ␣
      ↪Slope", round(beta_1, 2)))
      abline(a = 72.37, b = 0.03, col = "red", lwd = 4) # Here, a=Intercept, b=Slope.
```

Intercept: 72.37 Slope 0.03
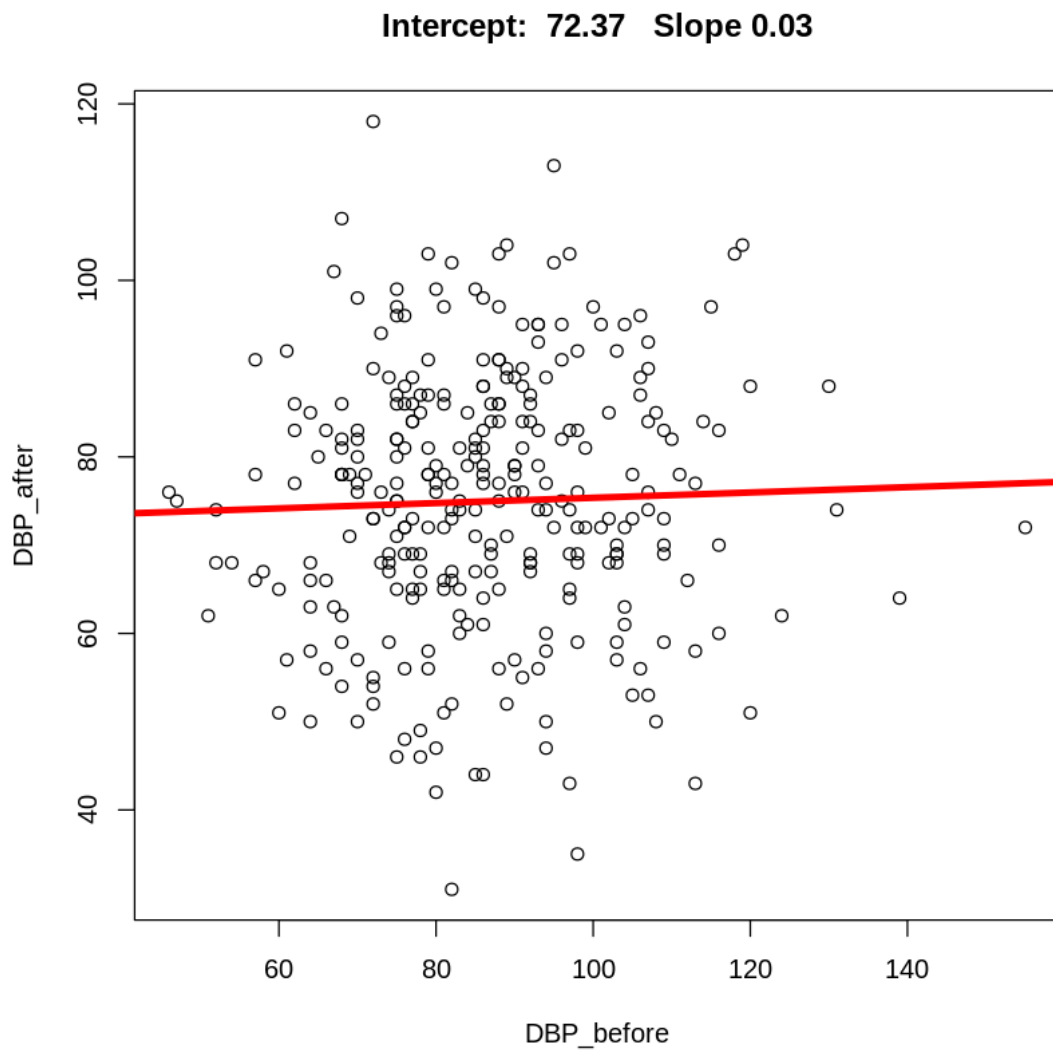
[17]:
```
x <- DBP_before
y <-DBP_after

# Fit the model
model <- lm(y ~ x)
```

[18]:
```
summary(model)
```

Call:
lm(formula = y ~ x)

Residuals:

```
     Min      1Q  Median      3Q     Max
 -43.985  -8.961   0.744  10.616  43.334


Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 72.37215    4.62026  15.664   <2e-16 ***
x            0.03186    0.05279   0.604    0.547
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.74 on 298 degrees of freedom
Multiple R-squared:  0.001221,     Adjusted R-squared:  -0.00213
F-statistic: 0.3644 on 1 and 298 DF,  p-value: 0.5465
```

[19]:
```
Coefficients <- c("Intercept", "slope(DBP slope before treatment)")
Estimates <- c(beta_0, beta_1)
StandardError <- c(4.62026 , 0.05279)
Cl<- c(-0.07860595,0.14760269)
p_value<-c("2e-16",0.547)
Estimates<-round(Estimates,3)
StandardError<-round(StandardError,3)
Cl<- round(Cl,3)
modelTable <- data.frame(Coefficients, Estimates, StandardError,Cl,p_value)
modelTable
```

	Coefficients	Estimates	StandardError	Cl	p_value
	<chr>	<dbl>	<dbl>	<dbl>	<chr>
A data.frame: 2 × 5	Intercept	72.372	4.620	-0.079	2e-16
	slope(DBP slope before treatment)	0.032	0.053	0.148	0.547

So what have we noticed?

When DBP before the trial is 0, the base DBP level after the trial is 72.372mmHg.

The slope coefficient is a magnitude 0.032, being the average change in the DBP after the trial.

There is no statistical significance, because the value calculated was bigger than $p<0.05$, so there is more than a 5% chance of the observed association occurring by chance, making the results insignificant.

What assumptions does our model make and how can we further investigate them? Time to produce an appropriate residual plots and interpret;) Results? There is a linear relationship shown between DBP before and after the trial, and evident from the random cluster that each observed parameter is independent. This satisfies the assumption of linearity and independence from the model. But, the consistency varies due to the amount of outliers, so the assumption that the scattering is consistent is unsatisfactory.

[20]:
```
# Fit the linear model for hp ~ mpg in mtcars dataset
residual_model <- lm(DBP_before ~ DBP_after, data = DBP_summaryTable)
```

```r
# Create a new dataframe that includes the original data, the fitted values,␣
 ↪and the residuals
DBP_summaryTable$model_fitted <- fitted(residual_model)
 DBP_summaryTable$model_residuals <- residuals(residual_model)

# Open a new graphics window

# Plot 1: Scatter plot with all residuals
plot( DBP_summaryTable$Diastolic_before,  DBP_summaryTable$Diastolic_after,␣
 ↪xlab = "before (mmHg)", ylab = "after (mmHg)",
     main = "Scatter Plot with Residuals", pch = 16, col = "blue")
abline(model, col = "red") # Add the regression line
segments( DBP_summaryTable$Diastolic_before,  DBP_summaryTable$Diastolic_after,␣
 ↪ DBP_summaryTable$Diastolic_before,  DBP_summaryTable$model_fitted, col =␣
 ↪"blue") # Draw residuals
```

Scatter Plot with Residuals
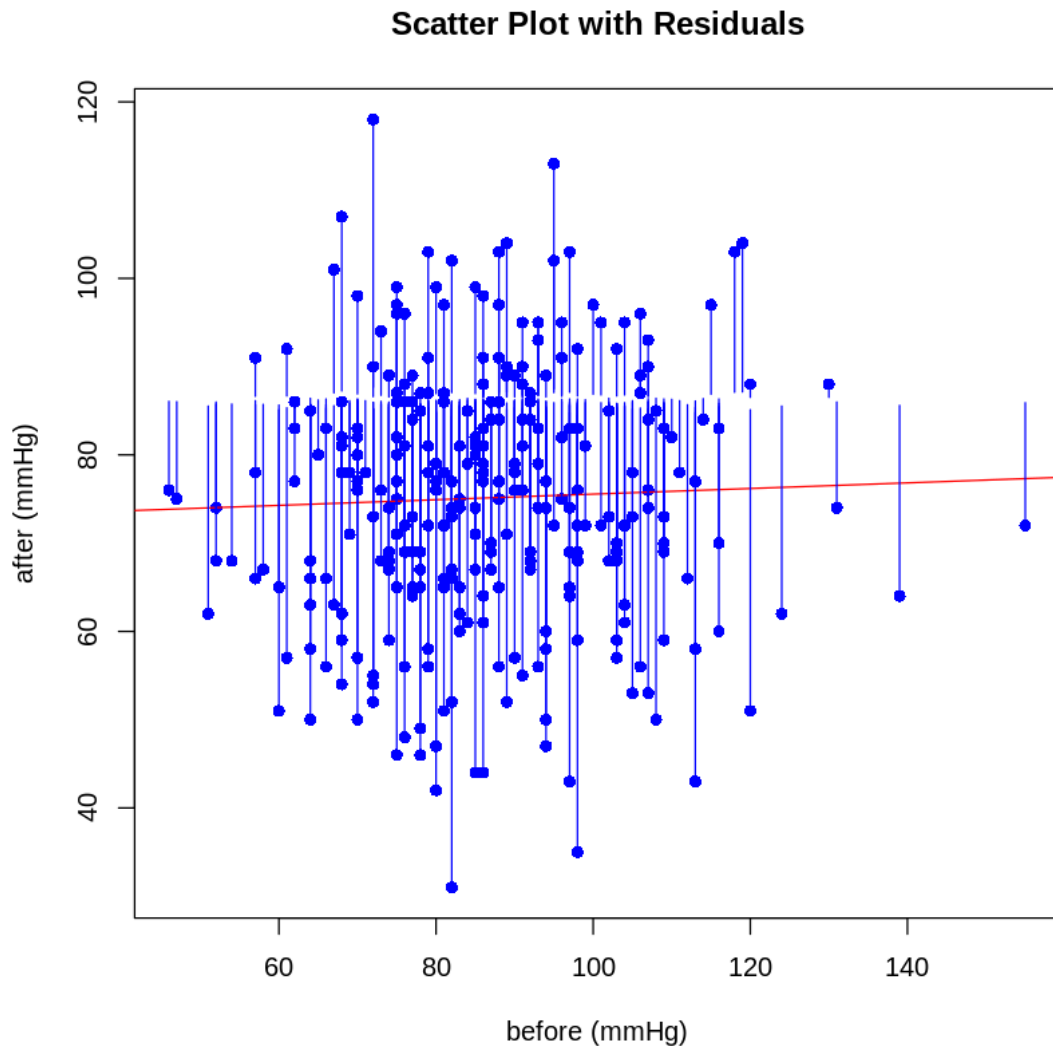
Tp make things more interesting, lets imagine the clinical team want to investigate the relationship between categorical DBP (defined as 0 = "DBP2 < 90", 1 = "DBP2   90") at the end of the study (DBP2) with treatment (Trt).

Lets generate a categorical DBP variable at the end of study where 0 = "DBP2 < 90" signifies DBP2 values less than 90 mmHg and 1 = "DBP2   90" signifies DBP values greater than or equal to 90 mmHg.

```
[21]: DBP_summaryTable$DBP2_category <- ifelse(DBP_summaryTable$Diastolic_after <↵
      ↪90,0,1)# create a variable that uses an if loop assess every observation in↵
      ↪the dataset
```

Now we can produce a 2 x 2 table to summarise treatment and categorical DBP variables. We will perform a statistical test (Chi squared test) to evaluate the hypothesis that there is no association

between treatment groups and DBP2 categories.

Results? Since the p-values are less than 0.05(95% CI), we can reject the null hypothesis of no association between treated participants and high DBP. The chi squared value isn't 0, so that suggests there's a statistically significant association.

```
[22]: #create table
      Treatment<- DBP_summaryTable$Treatment#where 0 = placebo and 1= medication
      DBP2 <- DBP_summaryTable$DBP2_category#where 0 = low and 1 = high
      contingency_table <- table(Treatment,DBP2)
      print(contingency_table)
```

```
         DBP2
Treatment   0   1
        0 118  30
        1 135  17
```

```
[23]: #chi squared manual way
      # Manually calculate row totals and column totals
      row_totals <- c(sum(contingency_table[1,]), sum(contingency_table[2,]))
      col_totals <- c(sum(contingency_table[,1]), sum(contingency_table[,2]))

      # Calculate grand total
      grand_total <- sum(contingency_table)

      # Manually compute expected frequencies for each cell
      expected_11 <- (row_totals[1] * col_totals[1]) / grand_total
      expected_12 <- (row_totals[1] * col_totals[2]) / grand_total
      expected_21 <- (row_totals[2] * col_totals[1]) / grand_total
      expected_22 <- (row_totals[2] * col_totals[2]) / grand_total

      expected <- matrix(c(expected_11, expected_21, expected_12, expected_22), nrow␣
       ↪= 2)
```

```
[24]: chi_statistic <- ((contingency_table - expected_11)^2 /␣
       ↪expected_11)+((contingency_table[1,2]-expected_12)^2 /expected_12)+
                      ((contingency_table[2,1]-expected_21)^2 /
       ↪expected_21)+((contingency_table[2,2]-expected_22)^2 / expected_22)

      df <- (nrow(contingency_table)-1)*(ncol(contingency_table)-1)
      #p value
      pVal<- pchisq(chi_statistic,df,lower.tail = FALSE)
```

```
[25]: cat("Chi-squared statistic:", chi_statistic, "\n")
      cat("Degrees of freedom:", df, "\n")
      cat("P-value:", pVal, "\n")
```

```
Chi-squared statistic: 4.685537 5.144996 76.33771 97.4424
Degrees of freedom: 1
```

```
P-value: 0.03041756 0.02331346 2.390746e-18 5.544593e-23
```

[26]:
```
#simplified
chisq.test(contingency_table, correct = FALSE)
```

```
        Pearson's Chi-squared test

data:  contingency_table
X-squared = 4.6855, df = 1, p-value = 0.03042
```

Let's Calculate the odds ratio of high DBP2 (categorical DBP value equals to 1) in the active treatment group compared to placebo with 95% CI and interpret the results. What is Odds Ratio? Results? The odds ratio(0.495), is smaller than 1, indicating the odds are low of a high DBP after treatment. In other words, patients who received the treatment, are 0.495 times more likely to have a higher DBP afterwards than those who received the placebo(the control group) which is very low.

[27]:
```
contingency_table
```

```
          DBP2
Treatment   0   1
        0 118  30
        1 135  17
```

[28]:
```
#based on the OR formula in class
#treatment group odds for high DBP2
trt_p <- 17/152
trt_1p <- 1 - trt_p
trtOdds<- trt_p/trt_1p
#control(placebo) group odds for high DBP2
control_p <- 30/148
control_1p <- 1 - control_p
controlOdds <- control_p/control_1p
#odds ratio
oddsRatio <- trtOdds/controlOdds
oddsRatio<- round(oddsRatio,3)
```

[29]:
```
if (oddsRatio > 1){
    print("higher odds of a high DBP after trial")
}else if (oddsRatio < 1){
    print("lower odds of a high DBP after trial")
}else{
    print("exposure doesn't affect odds of having high Diastolic Blood Pressure␣
    ↪at the end of the trial")
}
```

```
[1] "lower odds of a high DBP after trial"
```

```
[30]: #logit(y)= a + Bx
      # Logistic Regression Model
      DBP2_model <- glm(DBP2_category ~ Treatment, family = binomial, data =␣
        ↪DBP_summaryTable)



      # Display the summary of the logistic regression model
      summary(DBP2_model)
```

```
Call:
glm(formula = DBP2_category ~ Treatment, family = binomial, data =␣
  ↪DBP_summaryTable)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.6731  -0.6731  -0.4870  -0.4870   2.0932

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.3695     0.2045  -6.698 2.12e-11 ***
Treatment    -0.7026     0.3287  -2.138   0.0326 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 260.46  on 299  degrees of freedom
Residual deviance: 255.73  on 298  degrees of freedom
AIC: 259.73

Number of Fisher Scoring iterations: 4
```

```
[31]: # Extracting coefficients and their standard errors
      coefs <- coef(summary(DBP2_model))
      estimates <- coefs[, "Estimate"]
      stdE <- coefs[, "Std. Error"]

      # Calculate 95% Confidence Intervals
      ci_lower <- estimates - 1.96 * stdE
      ci_upper <- estimates + 1.96 * stdE

      # CI table
      confidence_intervals <- data.frame(
        Estimate = estimates,
        Lower_95_CI = ci_lower,
```

```
   Upper_95_CI = ci_upper
)


# Print the results
print(confidence_intervals)
```

```
              Estimate Lower_95_CI Upper_95_CI
(Intercept) -1.3694872   -1.770248 -0.96872650
Treatment   -0.7025742   -1.346803 -0.05834544
```

Lets write down the model equation for the logistic regression using treatment (Trt) to predict the binary DBP2 level.

$\text{logit}(y) = a + Bx$

Lets attempt to fit it a logistic regression for our model defined previously.

[32]:
```
#logit(y)= a + Bx
# Logistic Regression Model
DBP2_model <- glm(DBP2_category ~ Treatment, family = binomial, data =␣
  ↪DBP_summaryTable)
```

Things we've noticed

The negative coefficient for the treatment group suggests there's a negative association between treatment and high DBP. This means, when applying the treatment, there's a potential decrease in high DBP.

Both Confidence interval values are less than $p<0.05$, meaning we can be confident at a 95% level, that these effects are not just by chance, potentially the treatment has a positive impact on hypertension for those with high Diastolic Blood pressure.