

## Q2

2024-04-17

### Question 2

Import this Google sheet in R/Python and for each of the parameters (P1 to P10) perform a t-test and ANOVA. Share the link of your results and the script.

#### installing packages required for the operations performed

googlesheets4 is used for read the data from google sheets

tidyverse is usedl to redhaping the data from wide to long

dplyr is used for data manipulations

tinytex is used for get the output in pdf format

knitr is used for output format and styles

rmarkdown is used for create and genarate rmarkdown files

```
library(dplyr)
library(ggplot2)
library(tidyverse)
library(googlesheets4)
library(rmarkdown)
library(knitr)
library(tinytex)
```

To read the data from google sheets and printing the raw data

```
df <-read_sheet("https://docs.google.com/spreadsheets/d/1ndg1XMmiTsMLITNCPTapVX6UId6H99mctIC--6DPM2Q/edit")
print.data.frame(df)
```

	Sr. No.	Parameter	Group A -Sample1	Group A -Sample2	Group A -Sample3
1	1	P1	34	14	22
2	2	P2	6	61	59
3	3	P3	11	13	35
4	4	P4	43	17	39
5	5	P5	20	42	54
6	6	P6	30	16	33
7	7	P7	34	60	28
8	8	P8	28	57	57
9	9	P9	27	68	36
10	10	P10	5	41	30

  

	Group B -Sample1	Group B -Sample2	Group B -Sample3	Group B-Sample4
1	53	57	38	29
2	74	58	40	54
3	74	72	57	31
4	67	50	47	31
5	60	73	58	39
6	40	57	68	61
7	45	70	26	49
8	65	35	58	47
9	45	64	42	73
10	51	74	43	55

  

	Group C-Sample1	Group C-Sample2	Group C-Sample3
1	67	79	91
2	86	80	65
3	87	62	77
4	65	69	60
5	87	94	56
6	85	100	82
7	59	54	52
8	91	85	61
9	81	66	81
10	53	63	82

To transform the data from wide to long to arrange the data to perform ttest

```
ldf <- as.data.frame(df %>% gather("Group A -Sample1", "Group A -Sample2", "Group A -Sample3",  
                                "Group B -Sample1", "Group B -Sample2", "Group B -Sample3",  
                                "Group B-Sample4", "Group C-Sample1", "Group C-Sample2",  
                                "Group C-Sample3", key=cat, value=conc))  
  
ldf$Parameter <- as.factor(ldf$Parameter)  
ldf$group <- as.factor(substr(ldf$cat, 7,7))  
ldf$sample <- as.numeric(gsub("\\D", "", ldf$cat))  
ldf$id <- paste(ldf$Parameter, ldf$group, sep = "_")
```

### To calculate the count, mean and sd for each parameter by group wise

```
ls <- ldf %>% group_by(Parameter, group) %>%  
  summarise(n = n(), avg = mean(conc), sd = sd(conc))  
  
print.data.frame(ls)
```

	Parameter	group	n	avg	sd
1	P1	A	3	23.33333	10.066446
2	P1	B	4	44.25000	13.047988
3	P1	C	3	79.00000	12.000000
4	P10	A	3	25.33333	18.448125
5	P10	B	4	55.75000	13.149778
6	P10	C	3	66.00000	14.730920
7	P2	A	3	42.00000	31.192948
8	P2	B	4	56.50000	13.988090
9	P2	C	3	77.00000	10.816654
10	P3	A	3	19.66667	13.316656
11	P3	B	4	58.50000	19.841035
12	P3	C	3	75.33333	12.583057
13	P4	A	3	33.00000	14.000000
14	P4	B	4	48.75000	14.750706
15	P4	C	3	64.66667	4.509250
16	P5	A	3	38.66667	17.243356
17	P5	B	4	57.50000	14.011900
18	P5	C	3	79.00000	20.223748
19	P6	A	3	26.33333	9.073772
20	P6	B	4	56.50000	11.902381
21	P6	C	3	89.00000	9.643651
22	P7	A	3	40.66667	17.009801
23	P7	B	4	47.50000	18.046237
24	P7	C	3	55.00000	3.605551
25	P8	A	3	47.33333	16.743158
26	P8	B	4	51.25000	13.124405
27	P8	C	3	79.00000	15.874508

28	P9	A	3	43.66667	21.548395
29	P9	B	4	56.00000	14.944341
30	P9	C	3	76.00000	8.660254

### data subsetting for each group individually

```
ldf_a <- ldf[ldf$group=="A",]  
ldf_b <- ldf[ldf$group=="B",]  
ldf_c <- ldf[ldf$group=="C",]
```

Applying ttest for each parameter different combination A vs B, A vs C and B vs C

```
for(i in levels(ldf$Parameter)){
  str<- as.character(i)

  cat("Two sample T-test for A vs B for: ", i, "\n")

  print(t.test(ldf_a[ldf_a$Parameter==str, "conc"], ldf_b[ldf_b$Parameter==str,"conc"],
    alternative = "two.sided", var.equal = FALSE))
  cat(rep("_", 70))
  writeLines("\n")

  cat("Two sample T-test for A vs C for: ", i, "\n")

  print(t.test(ldf_a[ldf_a$Parameter==str, "conc"], ldf_c[ldf_c$Parameter==str,"conc"],
    alternative = "two.sided", var.equal = FALSE))
  cat(rep("_", 70))
  writeLines("\n")

  cat("Two sample T-test for B vs C for: ", i, "\n")
  print(t.test(ldf_b[ldf_b$Parameter==str, "conc"], ldf_c[ldf_c$Parameter==str,"conc"],
    alternative = "two.sided", var.equal = FALSE))
  cat(rep("_", 70), "\n" )
  writeLines("\n")
}
```

Two sample T-test for A vs B for: P1

Welch Two Sample t-test

```
data: ldf_a[ldf_a$Parameter == str, "conc"] and ldf_b[ldf_b$Parameter == str, "conc"]
t = -2.394, df = 4.9627, p-value = 0.06246
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -43.427441  1.594108
sample estimates:
mean of x mean of y
 23.33333  44.25000
```

-----

Two sample T-test for A vs C for: P1

Welch Two Sample t-test

```
data: ldf_a[ldf_a$Parameter == str, "conc"] and ldf_c[ldf_c$Parameter == str, "conc"]
t = -6.1557, df = 3.8826, p-value = 0.003895
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -81.07670 -30.25663
sample estimates:
```

mean of x mean of y  
23.33333 79.00000

-----  
Two sample T-test for B vs C for: P1

Welch Two Sample t-test

data: ldf\_b[ldf\_b\$Parameter == str, "conc"] and ldf\_c[ldf\_c\$Parameter == str, "conc"]  
t = -3.6516, df = 4.671, p-value = 0.01659  
alternative hypothesis: true difference in means is not equal to 0  
95 percent confidence interval:  
-59.740249 -9.759751  
sample estimates:  
mean of x mean of y  
44.25 79.00

-----  
Two sample T-test for A vs B for: P10

Welch Two Sample t-test

data: ldf\_a[ldf\_a\$Parameter == str, "conc"] and ldf\_b[ldf\_b\$Parameter == str, "conc"]  
t = -2.43, df = 3.478, p-value = 0.08156  
alternative hypothesis: true difference in means is not equal to 0  
95 percent confidence interval:  
-67.326319 6.492985  
sample estimates:  
mean of x mean of y  
25.33333 55.75000

-----  
Two sample T-test for A vs C for: P10

Welch Two Sample t-test

data: ldf\_a[ldf\_a\$Parameter == str, "conc"] and ldf\_c[ldf\_c\$Parameter == str, "conc"]  
t = -2.9836, df = 3.8133, p-value = 0.04313  
alternative hypothesis: true difference in means is not equal to 0  
95 percent confidence interval:  
-79.25188 -2.08145  
sample estimates:  
mean of x mean of y  
25.33333 66.00000

-----  
Two sample T-test for B vs C for: P10

Welch Two Sample t-test

```

data: ldf_b[ldf_b$Parameter == str, "conc"] and ldf_c[ldf_c$Parameter == str, "conc"]
t = -0.95349, df = 4.1231, p-value = 0.3928
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  -39.74855  19.24855
sample estimates:
mean of x mean of y
   55.75    66.00

```

---

Two sample T-test for A vs B for: P2

Welch Two Sample t-test

```

data: ldf_a[ldf_a$Parameter == str, "conc"] and ldf_b[ldf_b$Parameter == str, "conc"]
t = -0.75053, df = 2.6092, p-value = 0.5148
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  -81.53627  52.53627
sample estimates:
mean of x mean of y
   42.0     56.5

```

---

Two sample T-test for A vs C for: P2

Welch Two Sample t-test

```

data: ldf_a[ldf_a$Parameter == str, "conc"] and ldf_c[ldf_c$Parameter == str, "conc"]
t = -1.8362, df = 2.4741, p-value = 0.1829
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  -103.65002  33.65002
sample estimates:
mean of x mean of y
   42        77

```

---

Two sample T-test for B vs C for: P2

Welch Two Sample t-test

```

data: ldf_b[ldf_b$Parameter == str, "conc"] and ldf_c[ldf_c$Parameter == str, "conc"]
t = -2.1863, df = 4.9607, p-value = 0.0809
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  -44.660314  3.660314
sample estimates:
mean of x mean of y

```

56.5        77.0

-----

Two sample T-test for A vs B for: P3

Welch Two Sample t-test

```
data: ldf_a[ldf_a$Parameter == str, "conc"] and ldf_b[ldf_b$Parameter == str, "conc"]
t = -3.094, df = 4.9873, p-value = 0.02713
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -71.121523 -6.545144
sample estimates:
mean of x mean of y
 19.66667  58.50000
```

-----

Two sample T-test for A vs C for: P3

Welch Two Sample t-test

```
data: ldf_a[ldf_a$Parameter == str, "conc"] and ldf_c[ldf_c$Parameter == str, "conc"]
t = -5.2626, df = 3.9872, p-value = 0.006299
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -85.07237 -26.26097
sample estimates:
mean of x mean of y
 19.66667  75.33333
```

-----

Two sample T-test for B vs C for: P3

Welch Two Sample t-test

```
data: ldf_b[ldf_b$Parameter == str, "conc"] and ldf_c[ldf_c$Parameter == str, "conc"]
t = -1.369, df = 4.9465, p-value = 0.2299
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -48.54452  14.87785
sample estimates:
mean of x mean of y
 58.50000  75.33333
```

-----

Two sample T-test for A vs B for: P4

Welch Two Sample t-test



```

data: ldf_a[ldf_a$Parameter == str, "conc"] and ldf_b[ldf_b$Parameter == str, "conc"]
t = -1.4394, df = 4.5938, p-value = 0.2145
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -44.64231 13.14231
sample estimates:
mean of x mean of y
 33.00    48.75

```

-----

Two sample T-test for A vs C for: P4

Welch Two Sample t-test

```

data: ldf_a[ldf_a$Parameter == str, "conc"] and ldf_c[ldf_c$Parameter == str, "conc"]
t = -3.7291, df = 2.4105, p-value = 0.04837
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -62.8433171 -0.4900162
sample estimates:
mean of x mean of y
33.00000 64.66667

```

-----

Two sample T-test for B vs C for: P4

Welch Two Sample t-test

```

data: ldf_b[ldf_b$Parameter == str, "conc"] and ldf_c[ldf_c$Parameter == str, "conc"]
t = -2.035, df = 3.7078, p-value = 0.1171
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -38.324000 6.490667
sample estimates:
mean of x mean of y
48.75000 64.66667

```

-----

Two sample T-test for A vs B for: P5

Welch Two Sample t-test

```

data: ldf_a[ldf_a$Parameter == str, "conc"] and ldf_b[ldf_b$Parameter == str, "conc"]
t = -1.5471, df = 3.8431, p-value = 0.1996
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -53.18368 15.51701
sample estimates:
mean of x mean of y

```

38.66667 57.50000

-----  
Two sample T-test for A vs C for: P5

Welch Two Sample t-test

data: ldf\_a[ldf\_a\$Parameter == str, "conc"] and ldf\_c[ldf\_c\$Parameter == str, "conc"]  
t = -2.6286, df = 3.9025, p-value = 0.05976  
alternative hypothesis: true difference in means is not equal to 0  
95 percent confidence interval:  
-83.358892 2.692226  
sample estimates:  
mean of x mean of y  
38.66667 79.00000

-----  
Two sample T-test for B vs C for: P5

Welch Two Sample t-test

data: ldf\_b[ldf\_b\$Parameter == str, "conc"] and ldf\_c[ldf\_c\$Parameter == str, "conc"]  
t = -1.5789, df = 3.4051, p-value = 0.2017  
alternative hypothesis: true difference in means is not equal to 0  
95 percent confidence interval:  
-62.05924 19.05924  
sample estimates:  
mean of x mean of y  
57.5 79.0

-----  
Two sample T-test for A vs B for: P6

Welch Two Sample t-test

data: ldf\_a[ldf\_a\$Parameter == str, "conc"] and ldf\_b[ldf\_b\$Parameter == str, "conc"]  
t = -3.8048, df = 4.9723, p-value = 0.0127  
alternative hypothesis: true difference in means is not equal to 0  
95 percent confidence interval:  
-50.581766 -9.751567  
sample estimates:  
mean of x mean of y  
26.33333 56.50000

-----  
Two sample T-test for A vs C for: P6

Welch Two Sample t-test

```
data: ldf_a[ldf_a$Parameter == str, "conc"] and ldf_c[ldf_c$Parameter == str, "conc"]
t = -8.1972, df = 3.9853, p-value = 0.001226
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -83.92332 -41.41001
sample estimates:
mean of x mean of y
 26.33333  89.00000
```

-----

Two sample T-test for B vs C for: P6

Welch Two Sample t-test

```
data: ldf_b[ldf_b$Parameter == str, "conc"] and ldf_c[ldf_c$Parameter == str, "conc"]
t = -3.9879, df = 4.9089, p-value = 0.01084
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -53.56684 -11.43316
sample estimates:
mean of x mean of y
   56.5      89.0
```

-----

Two sample T-test for A vs B for: P7

Welch Two Sample t-test

```
data: ldf_a[ldf_a$Parameter == str, "conc"] and ldf_b[ldf_b$Parameter == str, "conc"]
t = -0.51238, df = 4.6112, p-value = 0.632
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -42.00363  28.33697
sample estimates:
mean of x mean of y
 40.66667  47.50000
```

-----

Two sample T-test for A vs C for: P7

Welch Two Sample t-test

```
data: ldf_a[ldf_a$Parameter == str, "conc"] and ldf_c[ldf_c$Parameter == str, "conc"]
t = -1.4278, df = 2.1794, p-value = 0.2798
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -54.29292  25.62625
sample estimates:
mean of x mean of y
 40.66667  55.00000
```

-----  
Two sample T-test for B vs C for: P7

Welch Two Sample t-test

```
data: ldf_b[ldf_b$Parameter == str, "conc"] and ldf_c[ldf_c$Parameter == str, "conc"]
t = -0.80992, df = 3.3138, p-value = 0.4721
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -35.45251  20.45251
sample estimates:
mean of x mean of y
   47.5     55.0
```

-----  
Two sample T-test for A vs B for: P8

Welch Two Sample t-test

```
data: ldf_a[ldf_a$Parameter == str, "conc"] and ldf_b[ldf_b$Parameter == str, "conc"]
t = -0.33523, df = 3.7387, p-value = 0.7554
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -37.26940  29.43606
sample estimates:
mean of x mean of y
 47.33333  51.25000
```

-----  
Two sample T-test for A vs C for: P8

Welch Two Sample t-test

```
data: ldf_a[ldf_a$Parameter == str, "conc"] and ldf_c[ldf_c$Parameter == str, "conc"]
t = -2.3772, df = 3.9887, p-value = 0.0764
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -68.692570   5.359237
sample estimates:
mean of x mean of y
 47.33333  79.00000
```

-----  
Two sample T-test for B vs C for: P8

Welch Two Sample t-test

```
data: ldf_b[ldf_b$Parameter == str, "conc"] and ldf_c[ldf_c$Parameter == str, "conc"]
```

```
t = -2.4618, df = 3.894, p-value = 0.07127
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  -59.385622   3.885622
sample estimates:
mean of x mean of y
   51.25    79.00
```

---

Two sample T-test for A vs B for: P9

Welch Two Sample t-test

```
data: ldf_a[ldf_a$Parameter == str, "conc"] and ldf_b[ldf_b$Parameter == str, "conc"]
t = -0.84985, df = 3.4076, p-value = 0.451
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  -55.54528  30.87862
sample estimates:
mean of x mean of y
  43.66667  56.00000
```

---

Two sample T-test for A vs C for: P9

Welch Two Sample t-test

```
data: ldf_a[ldf_a$Parameter == str, "conc"] and ldf_c[ldf_c$Parameter == str, "conc"]
t = -2.4115, df = 2.6297, p-value = 0.107
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  -78.61352  13.94685
sample estimates:
mean of x mean of y
  43.66667  76.00000
```

---

Two sample T-test for B vs C for: P9

Welch Two Sample t-test

```
data: ldf_b[ldf_b$Parameter == str, "conc"] and ldf_c[ldf_c$Parameter == str, "conc"]
t = -2.2245, df = 4.8342, p-value = 0.07853
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  -43.351856   3.351856
sample estimates:
mean of x mean of y
   56         76
```

-----

## ANOVA for each Parameter

##In the ANOVA tables the significance of the ANOVA model was specified as following

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

###‘\*\*\*’ means statistically significant at 0% level of significance

###‘\*\*’ means statistically significant at 0.001% level of significance

###‘\*’ means statistically significant at 1% level of significance

###‘.’ means statistically significant at 5% level of significance

###‘ ’ means statistically significant at 10% level of significance

###‘ ’ means statistically significant at 100% level of significance(there was no statistical significance)

```
for(i in levels(ldf$Parameter)){
  par <- i
  cat("ANOVA for: ", i, "\n")
  m <- aov(conc~group, ldf[ldf$Parameter==par, ])
  print(summary(m))
  cat(rep("_", 70), "\n")
  writeLines("\n")
}
```

ANOVA for: P1

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
group	2	4763	2381.5	16.65	0.00219 **
Residuals	7	1001	143.1		

---

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

-----

ANOVA for: P10

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
group	2	2725	1362.3	5.838	0.0322 *
Residuals	7	1633	233.3		

---

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

-----

ANOVA for: P2

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
group	2	1859	929.6	2.352	0.165
Residuals	7	2767	395.3		

ANOVA for: P3

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
group	2	4939	2469.3	9.331	0.0106 *
Residuals	7	1852	264.6		

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

ANOVA for: P4

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
group	2	1504	752.1	4.85	0.0477 *
Residuals	7	1085	155.1		

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

ANOVA for: P5

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
group	2	2444	1222	4.274	0.0612 .
Residuals	7	2002	286		

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

ANOVA for: P6

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
group	2	5894	2947.0	26.59	0.000536 ***
Residuals	7	776	110.8		

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

ANOVA for: P7

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
group	2	308.4	154.2	0.683	0.536
Residuals	7	1581.7	225.9		

ANOVA for: P8

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
group	2	1845	922.5	4.083	0.0668 .
Residuals	7	1581	225.9		



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

ANOVA for: P9

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
group	2	1603	801.7	3.209	0.103
Residuals	7	1749	249.8		

xelatex("q2.Rmd")