Additional practical questions

Contents

1 Identifying an experimental design, comparing two groups 1

1.1 Variables and experimental unit . . . 1

1.2 Think whether the following senteces are TRUE or FALSE . . . 1

1.3 Optimal representation of the data . . . 2

1.4 Choosing between tests . . . 2

1.5 Interpretation of tests . . . 3

1.6 We ignored the Group in the tests: should we or shouldn't we? . . . 3

1.6.1 Steps . . . 3

1 Identifying an experimental design, comparing two groups

1.1 Variables and experimental unit

Look at this dataset and try to identify what was the design and what the experimenter was trying to do.

Measure	Subject.ID	Group	Protein.before	Protein.after
1	1	Glucose	619.08	629.09
2	1	Glucose	694.37	634.47
1	2	Glucose	727.49	692.43
2	2	Glucose	700.44	617.92
1	3	Glucose	673.21	649.19
2	3	Glucose	656.48	565.61
1	4	Glucose	582.30	713.03
2	4	Glucose	586.14	807.02
1	5	Glucose	683.21	629.19
2	5	Glucose	646.48	595.61
1	6	TGN+Glucose	689.18	643.31
2	6	TGN+Glucose	679.01	655.30
1	7	TGN+Glucose	687.99	591.84
2	7	TGN+Glucose	992.60	635.98
1	8	TGN+Glucose	619.99	637.90
2	8	TGN+Glucose	603.10	487.24
1	9	TGN+Glucose	716.51	509.87
2	9	TGN+Glucose	716.58	742.43
1	10	TGN+Glucose	629.99	617.90
2	10	TGN+Glucose	643.10	587.24

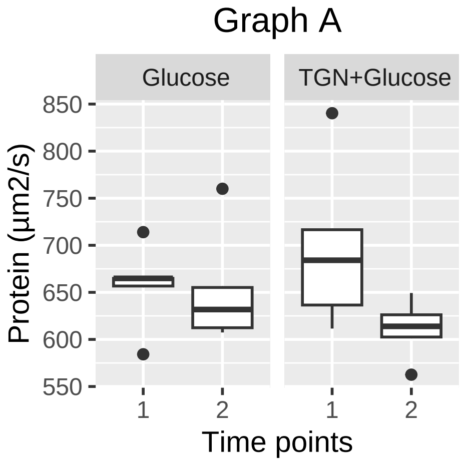
1.2 Think whether the following senteces are TRUE or FALSE

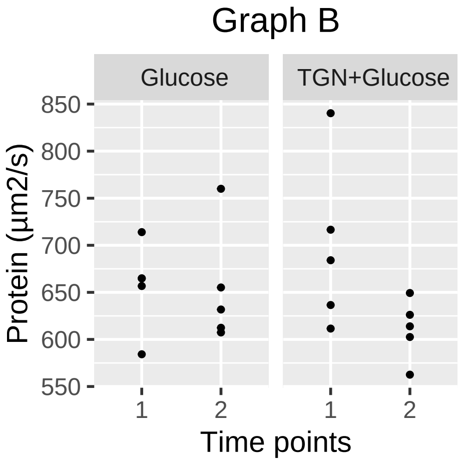
(a) We have 10 independent measurements, each one with a replicate (Protein.before, Protein.after).
(b) We have 40 independent measurements.
(c) We have 1 numerical variable.

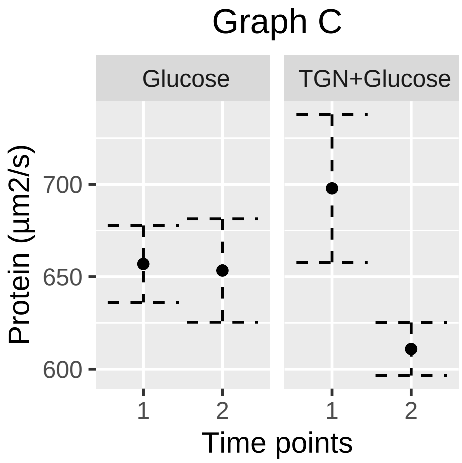
Now, try to describe the design. Be explicit about technical and biological replicates.

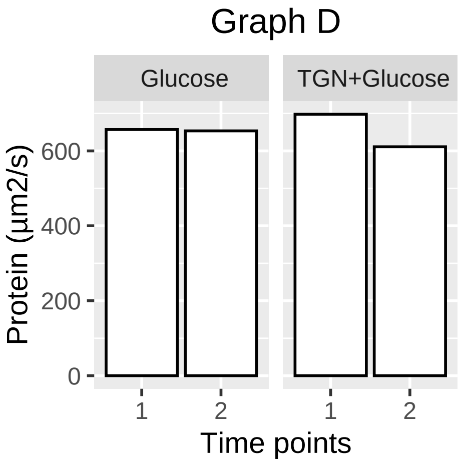
1.3 Optimal representation of the data

Argue which of the following graphs are a better representations of the data (the answer could be “I’d use a different graph”; in this case, explain what and how you’d represent)









In A, C, D: we have no idea if we are representing all 40 points, or the averages or what.

1.4 Choosing between tests

Taking into account that you want to compare the values of protein expression before an experimental manipulation (“x”) and after such manipulation (“y”) to n = 10 subjects, choose which of the following tests you would apply. In the following, data “df.m” contains not the data above but the data after averaging the two measurements. In other words:

```
test1 <- t.test(x = df.m$Protein.before, y = df.m$Protein.after,
                paired = TRUE, var.equal = FALSE)
test1
```

```
##
##  Paired t-test
##
## data:  df.m$Protein.before and df.m$Protein.after
## t = 1.474, df = 9, p-value = 0.1746
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  -24.18789 114.65589
## sample estimates:
## mean difference
##          45.234
```

```
test2 <- t.test(x = df.m$Protein.before, y = df.m$Protein.after,
                paired = FALSE, var.equal = FALSE)
test2
```

```
##
##  Welch Two Sample t-test
##
## data:  df.m$Protein.before and df.m$Protein.after
## t = 1.632, df = 16.53, p-value = 0.1216
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -13.37204 103.84004
## sample estimates:
## mean of x mean of y
##  677.3625  632.1285
```

a) test1 because is paired
b) test2 because in not paired

1.5 Interpretation of tests

From the previous outputs, we can deduce that:(Answer with TRUE or FALSE) a) Difference between mean values can be considered significant because the 95% CI include 0 b) Because the p-value in test2 is higher than in test1, this indicates that the probability of obtaining a significant difference between mean values is higher in test2 c) There is no sufficient statistical evidence to support that the means are different

1.6 We ignored the Group in the tests: should we or shouldn't we?

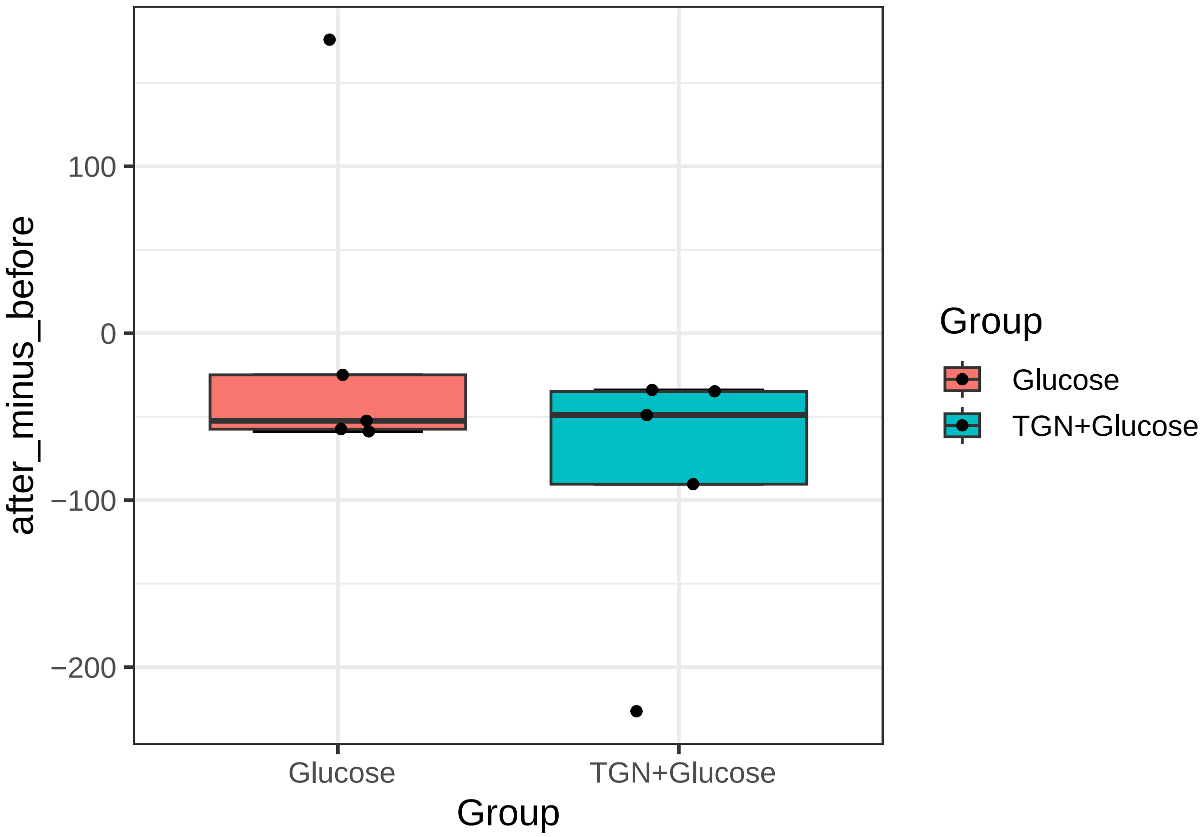
In the tests above, we ignored the group (Glucose vs.TGN+Glucose) (but note that, in the figures, we kept the two groups as separate). It is conceivable that the behavior of the two groups could differ. Incorporating additional variables is something we discuss in the next lesson.

But we can actually ask (and do a formal test of) whether the within-individual differences have the same mean in the two groups. This is very easy: carry out a two-sample t-test comparing the within individual differences between the two groups. What exactly would this test be testing? Can you do it? How?

1.6.1 Steps

- Recode Subject.ID as factor.
- Aggregate Protein.after and Protein.before by Subject.ID and Group (if we do it by Group too, even if not necessary, we avoid Group being recoded. Now, why did I say that aggregating by Group is not really necessary?)
- VERIFY a few cases (e.g., subjects 1, 10, 2, 5, 6, 8; this verification strategy is a sensible one: check cases at end of file, cases right on the border of the limits of the groups, and a few others; here, we check way too many, as checking 6 out of 10 is … well, 60%; but if you have 1000 subjects, checking 6 is well worth it).
- Compute the difference Protein.after minus Protein.before.
- Plot that difference, possibly by groups: i.e., plot two boxplots.
- Carry out a t-test that compares the within-individual differences between the two groups.

This is all doable from the R Commander interface. We won't show you the steps (try it yourself), just the output of a figure and the results of the test.



```
##
##  Welch Two Sample t-test
##
## data:  after_minus_before by Group
## t = 1.4357, df = 7.6438, p-value = 0.1907
## alternative hypothesis: true difference in means between group  Glucose and group  TGN+Glucose is no
## 95 percent confidence interval:
##  -51.61716 218.29716
## sample estimates:
##     mean in group  Glucose mean in group  TGN+Glucose
##                     -3.564                    -86.904
```

Now, step back and look at the boxplots with the overimposed individual observations (the within-individual differences): How many points are there? How much variation is in there? Do you think we can be very confident in the results?