First Assignment

by PsychoStatisticia

1. Car manufacturers make efforts to reduce CO2-emissions (measured in g/km). Via a new production process, 2 new filter systems have been independently designed: systems A and B. To test them, 2 series of n = 10 comparable cars were evaluated: the first series was equipped with filter system A and the second series with system B.

Table 1: Comparison of Systems A and B Before and After Intervention

System A		System B	
before ($S^2 = 36.22$)	**after** ($S^2 = 78.30$)	**before** ($S^2 = 13.65$)	**after** ($S^2 = 58.78$)
164.08	150.78	172.95	176.17
170.40	159.36	165.48	168.61
167.40	164.17	171.04	165.17
164.39	172.77	174.52	166.75
162.86	150.67	163.58	155.28
160.38	166.31	165.64	161.56
170.72	149.56	170.50	164.79
173.17	151.04	171.91	159.40
180.55	154.90	168.92	181.40
164.03	170.34	166.02	164.56

Eight students analyse the data in R in eight different ways. In the analysis results below, *'before'* is a variable containing the 20 measurements before installation of the filter system, *'after'* is a variable containing the 20 measurements after installation of the filter system, *'test'* is a variable which is 0 for system A and 1 otherwise, and *diff = after - before*.

```
> t.test(before~test,paired=F) # Analysis 1
t = -0.5633, df = 14.94, p-value = 0.5816
alternative hypothesis: true difference in means is not equal to 0
sample estimates:
mean in group 0 mean in group 1
        167.798         169.056


> t.test(before~test,paired=T) # Analysis 2
t = -0.6188, df = 9, p-value = 0.5514
alternative hypothesis: true difference in means is not equal to 0
sample estimates:
mean of the differences -1.258
```

```
> t.test(after~test,paired=F) # Analysis 3
t = -1.993, df = 17.642, p-value = 0.06196
alternative hypothesis: true difference in means is not equal to 0
sample estimates:
mean in group 0 mean in group 1
        158.990         166.369


> t.test(after~test,paired=T) # Analysis 4
t = -1.9384, df = 9, p-value = 0.08453
alternative hypothesis: true difference in means is not equal to 0
sample estimates:
mean of the differences -7.379


> t.test(before[test==0],after[test==0],paired=F) # Analysis 5
t = 2.6027, df = 15.859, p-value = 0.01934
alternative hypothesis: true difference in means is not equal to 0
sample estimates:
mean of x mean of y
   167.798    158.990


> t.test(before[test==0],after[test==0],paired=T) # Analysis 6
t = 2.2154, df = 9, p-value = 0.05397
alternative hypothesis: true difference in means is not equal to 0
sample estimates:
mean of the differences 8.808


> t.test(diff~test,paired=F) # Analysis 7
t = -1.3329, df = 14.41, p-value = 0.2032
alternative hypothesis: true difference in means is not equal to 0
sample estimates:
mean in group 0 mean in group 1
        -8.808          -2.687


> t.test(diff~test,paired=T) # Analysis 8
t = -1.2036, df = 9, p-value = 0.2594
alternative hypothesis: true difference in means is not equal to 0
sample estimates:
mean of the differences -6.121
```

a) What does each of the above analyses investigate? Attention! It could be possible that some of the above analyses are wrong. You need to explain if any analysis is wrong.

b) Which of these analyses would you prefer to report for evaluating whether two filter systems have different CO2-emission? Interpret the output.

c) What assumptions must be met for this analysis?

d) Explain how the numbers t, df and p-value were obtained in the output. Double check the calculations where possible.

2. Given the R simulation code below:

```
set.seed(123123)
n1=30
n2=30
mu1=0
mu2=0
sigma1=2
sigma2=2
Nsim=10000
p.result=rep(NA, Nsim)

for (i in 1:10000){
  group.1 <- rnorm(n1, mean=mu1, sd=sigma1)
  group.2 <- rnorm(n2, mean=mu2, sd=sigma2)

  t.result <- t.test(group.1, group.2, var.equal=TRUE, paired=FALSE)
  p.result[i] <- t.result$p.value
}

p.result< .05 # what does this result mean?
sum(p.result<.05) # what does this result mean?
sum(p.result<.05)/Nsim # what does this result mean?
```

a) Understand and run the code given above by visiting R documentation or simply search for anything you are confused with. Please answer the 3 comments of the final 3 lines of the code.

b) Change the sample sizes of 2 groups to see if the result will change.

c) What if we change *sigma1=10*? Are we supposed to get the same result? Test it by yourself! Remember to interpret the result and explain why. If the result is "abnormal", try to find a reasonable way to put it back to "normal".

d) (Challenging) Try to modify the code above to calculate the power of a t.test. Also, by the code you draft up, could you show the general relationship between the power and those factors influencing the power we mentioned (i.e, the effect size, the alpha level and the sample sizes) in the lecture?

e) (Optional) We have mentioned in one of the lectures that it is preferred that a researcher set a priority to Welch t-test (the t-test to usewhen the homogeneity of variances is violated) instead of pooled-variance t-test if the homogeneity of variances is unknown. By the above R simulation codes you have read and written by yourself, try to explain what's the loss to keep using Welch t-test if the homogeneity of variances is known.