

1)a) ① constant variance
 ② Independence of variance
 ③ Normality } assumptions of one-way ANOVA

b) ① comparing means of two related groups
 ② example: before & after observations

$$\begin{aligned} c) \quad y &= 1500 - 800(BE) + 500(IE) \\ &= 1500 - 800\overset{1}{(BE)} + 500\overset{1}{(IE)} \\ i) \quad \text{Adv. education} &\qquad \qquad \qquad \text{binary (not necessary)} \\ ii) \quad \bar{y} &= 1500 - 800(1) + 500(0) \\ &= 700 \end{aligned}$$

d) i) ANOVA has one dependent variable
 MANOVA has more than one dependent variable

ii) ANOVA has factor or combination of factors as independent variables
 ANCOVA has a mixture of quantitative and qualitative ind. variables

2) Paired t-test - the cyclists are the same / one cyclist drinks choco and he also drinks the carbs on another day

- ① Take the diff between choco milk and carbs replacement
- ② Check the diff for normality using Shapiro-Wilk test:

H_0 : The diff between choco milk and carbs are normally distr.
 H_1 : not normally distr.

If p-value > 0.05, do not reject H_0 and conclude. \therefore proceed w/ paired t-test

③ $H_0: \mu_d = 0$

$H_1: \mu_d \neq 0$

If p-value > 0.05, do not reject H_0 and conclude:

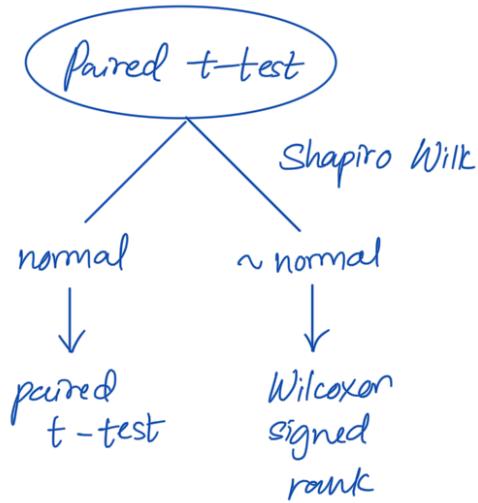
there is no sig. evid. that the μ time of exhaustion is greater after choclo milk than after carbo drink. (The otherwise for p-value < 0.05.)

④ If p-value for Shapiro Wilk-test < 0.05, reject H_0 and conclude.

\therefore proceed with Wilcoxon signed rank test.

⑤ H_0 : The two pop. have the same distr.

H_1 : The two pop. do not have the same distr.



3) a) $y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 \mathbb{1}(X_3) + \beta_4 X_4$

where

y = final exam scores

X_1 = no. of study hrs/week

X_2 = avg no. of sleep

X_3 = attend tutor (attend / not attend)

X_4 = no. of missed assignments

b) $y = 40 + 2X_1 + 3X_2 + 5X_3 - 4X_4$

interpretations:

Holding all other variables constant:

X_1 : for each additional hr of student studies per week (X_1),
the final exam score (y) is expected to increase by 2 pts.

X_2 : ...

X_3 : ...

X_4 : ...

c) $\hat{y} = 40 + 2(15) + 3(7) + 5(1) - 4(1) = 92$

d) LINE

4) i) The yield for year 1931 is generally higher than year 1932 for all sites except Morris

ii) The interaction between site:year is sig. as the p-value is $7.1416 \times 10^{-11} < 0.05$,
although the main effects i.e. site and year are non-sig. (p-value > 0.05)
we keep them in the model since their interactions are sig.

iii) Duluth - Grand Rapids

Duluth - Uni farm

Morris - Uni farm

Crookston - Uni farm

Crookston - Morris

iv) Residuals vs. fitted plot:

shows random scatter: equal variance of the resid.

Q-Q plot:

most of the plots are in line: resid. follows normal distr.