(24.11.2021)

Ex Sh. 4

Q.1 (i) Neg. correl between wait and next wait
" " " " " duration
Pos. " " next wait " "

But we can say more:
- distributions are bimodal, distinct point clouds
- short waiting times imply long durations, long " " can come with short or long durations of next eruption
- a long duration implies a long next waiting time, short duration → short next waiting time

(ii) $\hat{y}$ Nextwaiting [min] = 34.99 + 10.77 Duration [min]

(iv) Interpretation:
Slope: For each increase of 1 min duration, the predicted next waiting increases by 10.77 min
unit min/min (= no unit)
Intercept: For a duration of 0 min, the next waiting time would be predicted to be 34.99 min

(vi) see code

(vii) Eruption lasted 4 min
Predicted next waiting

$$\underset{\text{min}}{\hat{y}} = \underset{\text{min}}{34.99} + \underset{\substack{\text{min/min} \\ \text{min per min}}}{10.77} \cdot \underset{\text{min}}{4} = 78.05$$

(viii) Switch both to s:

- $\underset{s}{\hat{y}} = \underset{s}{34.99 \cdot 60} + \underset{s/s}{10.77}\ \underset{s}{x}$

$$34.99 \text{ min} = 34.99 \cdot 60 \text{ s}$$

- $\underset{\text{min}}{\hat{y}} = \underset{\text{min}}{34.99} + \underset{\text{min}/s}{10.77/60}\ \underset{s}{x}$

$$\text{slope} = \frac{S_{xy}}{S_x^2} \mathrel{\hat{=}} \frac{(\text{unit min})(\text{unit s})}{(\text{unit s})^2} = \frac{\text{min}}{s}$$

Calculation task:

(v) What does the Spearman Rank correlation measure?

$$\left.\begin{array}{l} X \longrightarrow R_X \\ Y \longrightarrow R_Y \end{array}\right\} \text{ ranks}$$

Spearman = Bravais-Pearson correlation of ranks

Bravais-Pearson: measures direction and strength of linear relation

Spearman: +1 R_X and R_Y are on straight line with positive slope

e.g.

R_X	1	2	3	4	5
R_Y	1	2	3	4	5
x_i	23	25	29	37	49
y_i	2	5	7	12	14

Original data have a positive strictly monotonic relation $x_j > x_i \Leftrightarrow y_j > y_i$

Statistics: Sample functions

Distributions you know: normal $N(\mu, \sigma^2)$

binomial, geometric, hypergeometric, neg. binomial, Poisson } discrete

Model: $X \sim N(\mu, \sigma^2)$ X follows a normal distribution with expectation μ and variance σ^2

Property of Normal distribution:

$$Y = a + bX \sim N(a + b\mu, b^2\sigma^2)$$

Special case: Standardize to $N(0,1)$

$$Z = \frac{X-\mu}{\sigma}$$

Simple Random Sample:

- randomly draw data from a model e.g. assuming that a dimension in a $\underbrace{\text{production process}}_{\text{infinite population}}$ behaves like that model
  - Sample size n
  - independent sample
    - assuming the process follows constantly the same model (no wear of tools, no readjustment of machine settings, ...), units produced can be considered as independent

      filling bottles of beer

      X = amount in a bottle

      X_i are independent, when leaving production
    - filling a box with 11 bottles that

e.g. $\mu = 500\,ml$
$\sigma^2 = 400\,ml^2$

came off the line one after the other; $X_1, \dots, X_{\underbrace{11}_{n}} \sim N(\mu, \sigma^2)$ iid

independent identically distributed

Total amount of beer in the box

$$T = \sum_{i=1}^{11} X_i \sim N(11\mu, 11\sigma^2)$$

because $X_1, \dots, X_{11}$ iid.

- Someone could interfere and destroy independence, e.g. by picking bottles such that filling amounts look homogeneous
  - ⟹ bottles within a box more similar than you would expect under independence
  - ⟹ Variance of T would increase, because within box variance goes down but between box differences increase