University of Pretoria
Department of Industrial and Systems Engineering

Industrial Analysis Bedryfsanalise BAN 313

Internal examiner: *Interne eksaminator:* Prof. Johan W. Joubert

External examiner: *Eksterne eksaminator:* Mr. Werner W. van Niekerk

Answer all questions on *clickUP*.

Beantwoord al die vrae op clickUP

Complete all **12** questions for **22** marks
*Beantwoord al **12** vrae vir **22** punte*

Total time: 120 minutes
Totale tyd: 120 minute

This is strictly an individual assessment. You are welcome to access any documented material, but no communication with (any) other individuals via any mode or means. A summarised formula sheet is made available at the end of this question paper.

Please take note of the last question, which requires that you upload the R/RMarkdown file that you used to complete your calculations. This must be a ***single file***, so ensure that you plan and set up your R session accordingly.

The internal examiner is available during the course of the test on +27 12 420 2843 and via the *clickUP Collaborate* session.

[0] 1. Carefully read the University's integrity statement and answer truthfully.

Student arrivals

The given file, `transactions.txt`, contains a transactional data dump of students entering and leaving at the University Road drop-off gate. That is, the entrance closest to the Engineering buildings. There are three turnstiles. The first is North, denoted with '`N`', and is closest to the Mineral Sciences Building. The second is in the centre, denoted with '`C`', and closest to the Mining Engineering Study Centre. The third is South, denoted by '`S`', and is closest to the Engineering II Building.

The time stamp indicates the date and time as `dd/mm/yyyy HH:MM:SS`. There is a variable indicating whether this is an `entry` into the campus or an `exit` leaving the campus. The `Success` variable indicates a `true` if the transaction was completed with only one tap of the student card, and `false` if two or more taps were required.

The final variable indicates the system's estimate for the entry or exit transaction duration (in seconds). That is the time from when a student's card is first tapped until the turnstile is activated and locked again behind the student.

[2] 2. What distribution best describes the transaction duration for the entire data set?

A. Normal distribution with $\mu = 6.05$s and $\sigma = 1.45$s.
B. Uniform distribution with min = 3.52s and max = 8.55s.
C. Triangular distribution with min = 3.52s; median = 6.06s and max = 8.55s.
D. Exponential distribution with $\lambda = 1/6.05 = 0.165$.
E. Poisson distribution with $\lambda = 6.05$.
F. No distinguishable distribution (completely random).

[2] 3. Calculate the 99% confidence limits for the mean of the transaction duration of the entire data set. Round your answers to three decimal places. For example, `1.234`. Lower limit: ______________; upper limit: ______________.

2 4. You read a discussion on the third year WhatsApp group where a student claims the answer to question 2 was a uniform distribution. Do you believe the claim? ◯ Yes ◯ No

2 5. Motivate your answer in question 4 using a χ^2 test with 6 breaks. That is, use the `breaks=6` argument for your histogram.

2 6. What is the total number of entries through the northern turnstile? _______________

2 7. Some students attribute the lower usage of the northern turnstile for entries to an error-prone card reader. Do you agree with them that the proportion of first-time successes is a function of the gate?
◯ Yes ◯ No

2 8. Motivate your answer in question 7 statistically.

2 9. Consider students leaving the campus on Wednesday, 4 March 2020 through the center gate, `C`. What proportion of the day's transactions does it represent? Round your answer to four decimal places. For example, if you believe it is 12.34%, give your answer as `0.1234`. _______________

2 10. What is the arrival rate, λ, in *entries per hour*, on weekdays between 08:00 and 09:00? Round your answer to one decimal place. For example, if you believe it is 12.34, give your answer as `12.3`. _______________

2 11. What distribution should the arrival rate of question 10 typically follow?

A. Normal distribution.

B. Uniform distribution.

C. Exponential distribution.

D. Poisson distribution.

E. No distinguishable distribution (completely random).

F. We don't have enough (or the correct) data to determine that.

2 12. Submit your supporting code (R or RMarkdown document) as a **single file**, using your student number as the filename. For example, `01234567.R` or `01234567.Rmd`.

⋆⋆⋆ **end of paper** ⋆⋆⋆
einde van vraestel

Formulas

$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ $\quad s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$ $\quad var = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$

$\Pr(A^c) = 1 - \Pr(A)$ $\quad \Pr(\text{A and B}) = \Pr(\text{A}) \times \Pr(\text{B})$

$\Pr(\text{A or B}) = \Pr(\text{A}) + \Pr(\text{B}) \text{ - } \Pr(\text{A and B})$ $\quad \Pr(A|B) = \frac{\Pr(\text{A and B})}{\Pr(\text{B})}$

$z = \frac{x - \mu}{\sigma}$ $\quad x = \mu + z\sigma$

$Q_1 - 1.5 \times IQR, \quad Q_3 + 1.5 \times IQR$

$\hat{p} \pm z_{score} \times SE_{\hat{p}}$ $\quad SE_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ $\quad z = \frac{\hat{p} - p_0}{SE_0}$ $\quad SE_0 = \sqrt{\frac{p_0(1-p_0)}{n}}$

$\bar{x} \pm t_{score} \times SE_{\bar{x}}$ $\quad t = \frac{\bar{x} - \mu_0}{SE_{\bar{x}}}$ $\quad SE_{\bar{x}} = \frac{s}{\sqrt{n}}$ $\quad df = n - 1$

$(\hat{p}_1 - \hat{p}_2) \pm z_{score} \times SE$ $\quad SE = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$

$z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{SE_0}$ $\quad SE_0 = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$ $\quad \hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$

$(\bar{x}_1 - \bar{x}_2) \pm t_{score} \times SE$ $\quad t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{SE}$ $\quad SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ $\quad df = \min(n_1 - 1, n_2 - 1)$

$\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$ $\quad \text{expected} = \frac{\text{row} \times \text{column}}{\text{total}}$ $\quad df = (r-1) \times (c-1)$

$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ $\quad \hat{\beta}_1 = r\left(\frac{s_y}{s_x}\right)$ $\quad \text{residual} = y - \hat{y}$ $\quad s = \sqrt{\frac{\sum(y - \hat{y})^2}{n-2}}$

$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$ $\quad t = \frac{\hat{\beta}_1 - 0}{SE_{\hat{\beta}_1}}$ $\quad \hat{\beta}_1 \pm t_{score} \times SE_{\hat{\beta}_1}$ $\quad df = n - 2$