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 **13** questions for **24** marks
*Beantwoord al **13** vrae vir **24** punte*

Total time: 90 minutes
Totale tyd: 90 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.

Census data

The following questions deals with the given sample as taken from the 2011 Census *Public Use Micro Sample* (PUMS). You are given the metadata (`metadata.pdf`), the record layout (`recordLayout.xls`), the municipal locations (`municipalities.csv`), and the compressed (GZIP) files of persons (`persons.txt.gz`) and households (`households.txt.gz`).

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

2 2. Using only the `persons-june2020.txt.gz` data, what is the median age of the sample? _____

2 3. How many individuals in the sample were born in Gauteng? _____

2 4. Using only the `persons-june2020.txt.gz` data, how many households have a female head? _____

2 5. What proportion of households own their dwelling and it is paid off? Give your answer as the 90% confidence limits and round each to four decimal places. For example, if you calculated the lower limit to be 12.34%, give your answer as `0.1234`. Lower limit: _____; upper limit: _____

Student arrivals

The given file, `transactions.csv`, 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 time of day first in the compact format `HHMMSS` and then the date as day/month/year. There is a variable indicating whether this is an `entry` into the campus or an `exit` leaving the campus. The `success` variable indicates a `1` if the transaction was completed with only one tap of the student card, and `0` 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 6. What distribution best describes the transaction duration?

A. Normal distribution with $\mu = 6.51$s and $\sigma = 0.80$.

B. Uniform distribution with min = 3.75s; median = 6.51s and max = 9.56s.

C. Exponential distribution with $\lambda = 1/6.51 = 0.154$.

D. Poisson distribution with $\lambda = 6.51$.

E. No distinguishable distribution (completely random).

2 7. You read a discussion on the third year WhatsApp group where a student claim the answer to question 6 was a uniform distribution. Do you believe the claim? ◯ Yes ◯ No

2 8. Motivate your answer using a χ^2 test with 8 breaks. That is, use the `breaks=8` argument for your histogram.

2 9. Explain the difference between the number of entries and exits on Tuesday, 3 March 2020.

2 10. The University conducted a manual study using students, who did not complete BAN 313, to analyse the same transactional data you received for 28 February 2020. They calculated that the day's average time between arrival for *entries* into the campus is $\frac{1}{\lambda} = 3.177$ min. Do you agree with their calculation?
◯ Yes ◯ No

2 11. Does the students' calculation in question 10 capture the hourly variation? ◯ Yes ◯ No

2 12. Motivate your answer in question 11 quantitatively.

2 13. 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

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

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

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

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

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

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

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

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

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

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

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

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

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