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 **26** marks
*Beantwoord al **13** vrae vir **26** 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`).

2 1. For how many households do you have complete records? That is, where there are records for the household and its individual members. ______________

2 2. If we define a *large* household as one with 7 or more members, what proportion of households in Mafikeng would be considered *large*? Use only the `households.txt.gz` and `municipalities.csv` files to answer the question and round your answer to 4 decimal places. For example, if your proportion is 12.34%, give your answer as `0.1234` and make sure to use a decimal point (not a comma). ______________

2 3. Since we only gave you a sample of the census data, give the 90% confidence interval for your estimate in question 2. Similar to the previous question, round your answer for each interval limit to four digits.
Lower limit: ______________; upper limit: ______________.

2 4. A recent healthcare survey suggests that the proportion of large families in Mafikeng is 12.5%. Do you agree with the survey? Answer `True` if you agree and `False` otherwise. ______________

2 5. Motivate your answer in question 4 statistically.

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 day/month/year and time of day in the format hour:minutes:seconds. 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.

2 6. True or false? The proportion of transactions that are successful first time round is dependent on the transaction type? ______________

2 7. Motivate your answer in question 6 statistically.

2 8. Filter the transactions to those occurring on weekdays in the time interval [10:00:00; 11:00:00). How many records do you have? ______________

2 9. Using the weekday records in the time interval [10:00:00; 11:00:00), how would you describe the time between arrival distribution for persons entering the campus. Mark all that are correct.

A. Unimodal.

B. Bimodal.

C. Multimodal.

D. Left-skewed.

E. Symmetric.

F. Right-skewed.

2 10. What would be your best estimate for the distribution in question 9?

A. Normal distribution with $\mu = 94.6$s and $\sigma = 95.6$.

B. Uniform distribution with min = 1s; median = 64s and max = 590s.

C. Exponential distribution with $\lambda = 1/94.6 = 0.0106$.

D. Poisson distribution with $\lambda = 94.6$.

E. No distinguishable distribution (completely random).

2 11. Is the mean time between arrivals for students entering the campus on a weekday in the interval [10:00:00; 11:00:00) the same as the mean time between arrivals for students leaving? Answer `True` for yes and `False` for no. ______________

2 12. Motivate your answer in question 11 statistically.

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

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

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

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

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

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

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

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

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

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

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

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

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

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