University of Pretoria
Department of Industrial and Systems Engineering

Industrial Analysis
Bedryfsanalise

BAN 313

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

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

Answer all questions on *clickUP*.

Beantwoord al die vrae op clickUP

Complete all **11** questions for **20** marks
*Beantwoord al **11** vrae vir **20** 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 82 338 0565 and via the *clickUP Collaborate* session.

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

Census data

The following questions deal 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-july2020.txt.gz`) and households (`households-july2020.txt.gz`).

[2] 2. You are tasked with doing some research on the state of the country's people with disability. If you want to read the variable that gives evidence on persons having trouble washing, dressing or feeding themselves, you will start reading each record at position _____ and end with the position _____.

[2] 3. How many individuals in the sample were born in July? _____.

[2] 4. How many persons in the sample, born in 1975, specified themselves as being female? _____

[2] 5. You are tasked with organising a *Council of Municipalities* meeting. To be fair, you decide to calculate the most central location, weighing each municipality equally. What is the median coordinate of this central location? Give your answer in decimal degrees and round to four decimals, for example 25.1234. Longitude (x): _____; Latitude (y): _____.

[2] 6. What is the closest town to the calculated coordinate?

[2] 7. Government is considering interventions to alleviate the severe drought, especially for vulnerable households. How many households in the sample, living in the Nama Khoi municipality (in the Northern Cape), have access to piped water within 500m of their dwelling? _____

2 8. What distribution best describes the age (in completed years) of the individuals in the entire sample?

A. Normal distribution with $\mu = 27.96$ and $\sigma = 19.95$.
B. Uniform distribution with min $= 0$ and max $= 118$.
C. Triangular distribution with min $= 0$; median 25 and max $= 118$.
D. Exponential distribution with $\lambda = 1/27.96 = 0.0358$.
E. Poisson distribution with $\lambda = 27.96$.
F. None of the above.

2 9. In the debate of what age groups to focus on for the drought relief, you hear a prominent politician who is associated with *Statistics South Africa* claim that the age distribution of South Africans follows a Poisson distribution. Do you agree with the politician? Answer `True` if you agree and `False` otherwise.

_____

2 10. Motivate your answer in question 9 using a χ^2 test with 10 breaks. That is, use the `breaks=10` argument for your histogram.

2 11. 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

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

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

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

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

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

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

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

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

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

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

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

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

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