Shape

Description automatically generated with medium confidence

|  |  |
| --- | --- |
| Course code | ECON1193B |
| Course name | Business Statistics 1 |
| Lecturer’s Name | Hieu Thai Trung |
| Student’s Name | Nguyen Huynh Quang |
| Student’s ID | S3927039 |
| Assignment | Individual Case Study - Inferential Statistics |
| Dataset | Mortality rate neonatal – Dataset 10 |
| Words count | 2761 |

1. Introduction

* Neonatal mortality rate is defined as the number of newborn deaths within the first 28 days of life per one thousand live births (WHO n.d). Birth defects are regarded to be the leading cause of newborn death because they can occur in any pregnancy. In 2020, there were 2,4 million neonatal fatalities, which is an improvement compared to 1990, when there were 5 million newborn deaths, when the rate was only half as high (WHO n.d). UNICEF reports that the first 28 days of a newborn's life are associated with a high mortality rate of 17 deaths per 1000 live births, whereas the rate drops after the first month and before the age of 1 to 9 deaths per 1000 children (UNICEF n.d). The newborn death rate reflects not only baby health, but also the nation's overall health, cleanliness, health care, nutrition, and education (World Vision, 2021).
* According to the World Health Organization, the majority of infant deaths occur in low- and middle-income countries; Africa and South Asia regions account for 80 percent of global infant deaths. Sierra Leone and the Central African Republic, both of which had per capita gross national incomes of $510 in 2020, had the highest Newborn Mortality Rate in 2020, with 80.10 and 77.50 deaths per 1000 live births, respectively. Meanwhile, nations with a high per capita GNI, such as Iceland and San Marino (GNI per capita: $64,410 and $40,801, respectively), had the lowest infant mortality rate, with 1.54 and 1.56 deaths per 1,000 live births, respectively (UNICEF n.d). National wealth and other social and economic factors have a substantial impact on population health (Houweling, 2005 & Schell, 2007). Preston in 1975, Prichett and Summer in 1996 discovered that in low-income countries, people have limited access to medical resources and treatment, whereas in high-income countries, the healthcare system is superior, resulting in a significantly higher neonatal survival rate in developed nations than in low- and middle-income nations.
* In addition to the factors listed above that influence the newborn mortality rate, there are additional socioeconomic and medical issues. In 2016, the Ethiopian Demographic and Health Survey was done and it revealed three factors that influence infant mortality. First, higher education and greater economic resources families are more likely to have a lower infant mortality rate than those with no formal education or only a primary education. The study also found that male births had a higher death rate than female births due to genetic and protein variations between the sexes. A study conducted in South Asia revealed that the newborn male mortality rate was 33.2/1000 live births, while the newborn female mortality rate was 27.4/1000 live births (Aghai, Z.H., Goudar, S.S., Patel, A. et al).

1. Descriptive statistic and probability
2. Probability

* In the presented data sample of 28 nations, the number of countries is divided in different groups according to GNI (Table 1). The neonatal death rate is also categorised into two groups (Table 2). The conventional value for classifying a newborn death rate as low or high is 15,00.

|  |  |
| --- | --- |
| Low - Income Countries (LI) | GNI per capita: less than US$1000 |
| Middle - Income Countries (MI) | GNI per capita: between US$1000 and US$12,500 |
| High - Income Countries (HI) | GNI per capita: more than US$12,500 |

***Table 1:*** Level of national income based on GNI.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Low NMR (L) | High NMR (H) | Total (X) |
| Low-Income Countries (LI) | 0 | 5 | 5 |
| Middle-Income Countries (MI) | 9 | 4 | 13 |
| High-Income Countries (HI) | 5 | 0 | 5 |
| Total | 14 | 9 | 23 |

***Table 2:*** Cognitency table for each level of national income and neonatal mortality. rate (number of deaths per 100 live births) in 2017

* In theory, two events A and B are independent when event A occurs but have no effect on the probability of event B. In mathematic, the way to express independent events is: or (UC Berkeley). In this scenario, it is a comparison between the probability of high neonatal mortality rate P(H) and the probability of number of countries having high neonatal mortality rate and high GNI per capita P(H|HI).
* High-Income Countries:

=

* Similarly, the same mathematical way is used to check countries which have high neonatal mortality rate (NMR) with Middle-Income and Low-Income.
* Middle-Income Countries:

= 0,307

* Low-Income Countries:
* According to the computed conclusion, there is a difference in the association between the probability of a high neonatal death rate and countries with different groups of countries.While High GNI Countries appear statistically independent from NMR, Middle-Income and Low-Income Countries have newborn mortality rates that are dependent on GNI.
* Deeply analyzed the data, we can find out the probability of high neonatal mortality rate in 3 level of GNI through this comparison:

* Based on this comparison, we may conclude that Low-Income and Middle-Income Countries have the likelihood of experiencing a high neonatal mortality rate, whereas High-Income Countries have no such possibility.

1. Descriptive statistics
2. Central Tendency

|  |  |  |  |
| --- | --- | --- | --- |
|  | Low-Income (LI) | Middle-Income (MI) | High-Income (HI) |
| Mean | 21.96 | 12.95 | 5.42 |
| Median | 21.70 | 11.50 | 4.10 |
| Mode | #N/A | 5.3 | #N/A |

***Table 3***: Table for central tendency on neonatal mortality rate categorized by level of countries GNI in 2017.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Low-Income (LI) | Middle-Income (MI) | High-Income (HI) |
| Q1 | 20.4 | 6.7 | 3.6 |
| Q3 | 25.3 | 20.9 | 4.7 |
| IQR | 4.9 | 14.2 | 1.1 |
| Upper Limit | 32.65 | 42.2 | 6.35 |
| Lower Limit | 13.05 | -14.6 | 1.95 |
| Outliers | No | No | Yes |

***Table 4***: Table for calculating quartiles, upper and lower limit and identifying outliers of neonatal mortality rate.

* The mean cannot be considered an accurate measure of central tendency because it is highly sensitive to outliers and counts all data points (J.Pharmacol, 2011). In contrast, the median is regarded as the ideal statistic for skewed distributions since it is less affected by extreme values. In contrast, mode cannot be used for comparison because neither low- nor high-GNI nations have specified modes (Table 4). Based on table 3, we may deduce that countries with low GNI have a greater neonatal death rate, whereas middle- and high-income countries maintain a lower rate. 50% in the Low GNI group have a median newborn mortality rate of 21,70 per 1000 live births. While 50% of the Middle and High GNI groups had only 11.50 and 4.10 deaths per 1,000 live births, respectively.

1. Variation

|  |  |  |  |
| --- | --- | --- | --- |
|  | Low-Income (LI) | Middle-Income (MI) | High-Income (HI) |
| Range | 11.60 | 18.70 | 9.50 |
| IQR | 4.9 | 14.2 | 1.1 |
| Sample Variance | 20.52 | 45.86 | 14.54 |
| Standard Deviation | 4.53 | 6.77 | 3.81 |
| Coefficient of Variation | 0.21 | 0.52 | 0.70 |

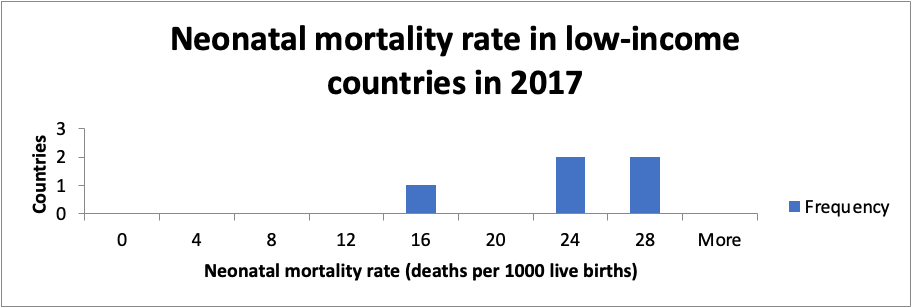
***Table 5***: Table of variation for neonatal mortality rate on each level of GNI per capita in 2017.

* Interquartile Range (IQR) is regarded as an appropriate measure of variation in statistical theory because it represents the range of value between Q1 and Q3 and is not affected by outliers (Raymondo, 1999). Middle-Income Countries had the highest IQR with 14,2 deaths per 1000 live births, followed by Low-Income Countries and High-Income Countries with 4,9 and 1,1 deaths per 1000 live births, respectively (see Table 5). The data for Middle-Income countries are more widely dispersed and scattered, whereas those for High-Income countries are more likely to be dispersed around the median.

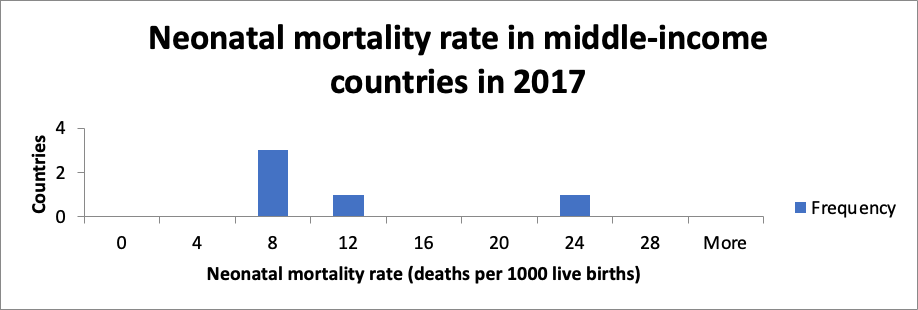
1. Histogram

|  |  |  |  |
| --- | --- | --- | --- |
| Neonatal mortality rate | Low-Income (LI) | Middle-Income (MI) | High-Income (HI) |
| From 0 to 4 | 0 | 0 | 0 |
| From 4 to 8 | 0 | 0 | 2 |
| From 8 to 12 | 0 | 3 | 2 |
| From 12 to 16 | 0 | 1 | 0 |
| From 16 to 20 | 1 | 0 | 1 |
| From 20 to 24 | 0 | 0 | 0 |
| From 20 to 24 | 2 | 1 | 0 |
| From 24 to 28 | 2 | 0 | 0 |
| More | 0 | 0 | 0 |

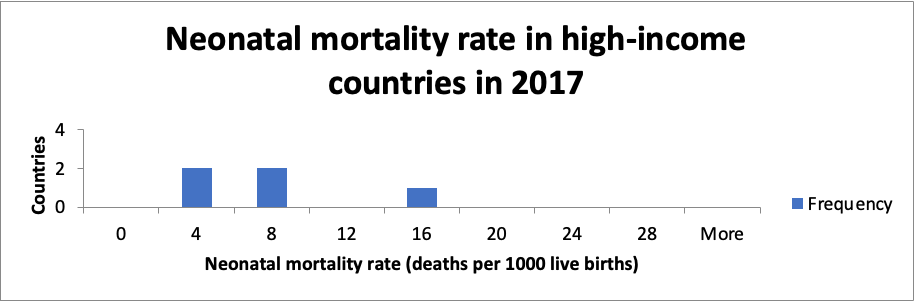
***Table 6***: Table of frequency distribution of neonatal mortality rate in three different groups of countries in 2017.



***Figure 1***: Histogram of neonatal mortality rate in low-income countries in 2017.



***Figure 2***: Histogram of neonatal mortality rate in middle-income countries in 2017.



***Figure 3***: Histogram of neonatal mortality rate in high-income countries in 2017.

* The data for middle-income and high-income countries are skewed to the right, whilst data for low-income countries are slanted to the left. Low- and high-income nations tend to be spread towards the two ends of the analyzed range, whereas middle-income nations have a more uniform distribution. One of the aspects that must be considered is that data from low-income nations are distributed at a higher rate of newborn death than those from high and middle-income countries.

1. Confidence Interval
2. Calculation

* A confidence interval computes an estimated range of values for a population parameter whose value is unknown; this range is derived from sample data (Easton and John H.Macoll et al). In this circumstance, the confidence interval is calculated using the t-value because the population sample deviation is unknown. Assumes the dataset has a level of confidence of 95% and a significance level of 5%.

|  |  |
| --- | --- |
| Significance Level | 5% |
| Confidence Level | 95% |
| Population SD | unknown |
| Sample SD (S) | 7.91 |
| Sample Mean () | 13.27 |
| Sample Size (n) | 23 |
| t-value (two-tailed) | 2.074 |
| Lower bound | 9.85 |
| Upper bound | 16.69 |
| Degree of freedom | 22 |

***Table 7***: Statistics represents confidence intervals for neonatal mortality rate in 23 countries in 2017.

* Degree of freedom: d.f = n - 1 = 23 - 1 = 22
* Level of significance:
* Confidence intervals :

1. Conclusion

* Based on the calculation above, we can conclude that 95% certain that the mean for neonatal mortality rate in 23 countries in 2017 will range from 9.85 to 16.69 deaths per 1000 live births.

1. Discussion of assumption

* Because the sample size of the dataset is 23, which is less than 30. For this reason, the Central Limit Theorem cannot apply in this scenario (Shuan Turney et al., 2022). However, we can assume that the population follows a normal distribution.

1. Hypothesis testing

* The neonatal mortality rate of the world was estimated to be at 18.6 deaths per 1000 live births.
* Step 1: Central Limit Theorem (CLT)

The sample size of the dataset is 23, which is less than 30. For this reason, CLT cannot be applied.

* Step 2: Determine null and alternative hypothesis

Null hypothesis:

Alternative hypothesis

* Step 3: Type of testing

Confidence level:

Sample size: n = 23

Since there is an unequal sign () of , we will choose a two-tailed test

* Step 4: Determine whether use Z-table or T-table

Because the population standard deviation is unknown and the population is normally distributed, the t-table is chosen for this test.

* Step 5: Determine critical value (CV)

Degree of freedom: d.f = n – 1 = 22

Level of significance:

Critical value:

* Step 6: Calculate test statistic t:
* Step 7: Statistical decision

Since the test result fall in the rejection region (, we can reject the null hypothesis .

* Step 8: Make a managerial decision

As we can reject the null hypothesis, we are certain that there is evidence that proves the world average neonatal mortality rate will change other than 18.6 deaths per live births.

* Step 9: Determine type of errors

Due to the fact that we have committed Type I errors by rejecting Ho, we can report that the global mean neonatal mortality rate in 2017 was different than in 2016. Type I errors occur when we sense a difference when there’s not statistically difference in truth (Berkeley n.d). To reduce Type I errors, we can obtain a larger sample size and increase our level of confidence (NEDARC n.d). Another technique to prevent Type I mistake is to reduce the significance level, which means significance testing is abandoned (J.Rothman et al).

1. Halved the sample size

* Since the sample size is expected to be cut in half, the confidence level and margin of error will decrease (Chris Diezel et al, 2018). As the sample size decreases to half, it is possible that the t-value will increase (Jim Frost et al). Consequently, the null hypothesis will fall into the rejection zone, indicating that there is likely evidence to reject the null hypothesis. Due to the halves of the sample size, the statistical decision is unaffected, the t-value increases, and the outcome will be less precise.

1. Conclusion

* Neonatal mortality rate not only reflects overall newborn health, but also reflect the health of a community, socioeconomic status and national healthcare system. The calculation of the dataset which includes 23 countries will give us an overview of neonatal mortality rate in 3 different countries.
* Initially, the probability testing for three groups of countries' GNI yields the anticipated outcome. In high GNI countries, the neonatal mortality rate is independent of Gross National Income, however there is a correlation between neonatal mortality rate and GNI per capita in low- and middle-income countries. The test also demonstrates that high-income nations have no chance of experiencing a high infant mortality rate, with a probability of zero. In the meanwhile, the majority of low-income nations are likely to have a high neonatal mortality rate, with probabilities of 0.56 and 0.307, respectively. This is best described by each group's respective national healthcare system and health spending budget.
* Secondly, the median is the best measure of central tendency because it seeks the central value of the analyzed range and is less impacted by extreme values, but the IQR is the best indicator of variation since it reports the range of values from Q1 to Q3. 50% of high-income nations had an NMR of 4.10/1000 live births, while 50% of low- and middle-income countries have 11.50 and 27.50 deaths/1000 live births, respectively. Middle-Income countries have the greatest IQR with 14.2 per 1000 live births, while high-income countries have the lowest IQR with 1.1 per 1000 live births.
* Finally, we are able to demonstrate that the global average neonatal death rate in 2017 was distinct from that of 2016. By calculating the confidence interval and conducting hypothesis testing, we can conclude that the global average neonatal mortality rate for 2017 ranges from 9.85 to 16.69 deaths per 1000 live births. In the meanwhile, halving the sample size may reduce the accuracy of hypothesis testing.

1. Extension

* As Vietnam is included in the dataset, we would like to investigate the current newborn mortality rate in various cities. In this situation, stratified sampling would be a suitable random sample method since it divides the population into two or more strata depending on their shared traits (BRM n.d). This type of random sampling may produce very accurate findings as well as individual estimates for each stratum.
* In order to collect data, a survey is done online and offline to obtain sample data. To conduct an online survey, an online form is produced and emailed to various homes in several Vietnamese cities in order to collect information from throughout the country. In rural places, in addition to the online survey, the offline survey is undertaken since Internet connection is restricted. By conducting online surveys, researchers may obtain faster and greater response rates at a lower cost (Vikas Menon et al). In rural places with limited Internet connectivity, offline surveys must be conducted in addition to online ones. This will strengthen the survey's fairness between urban and rural respondents. In order to compute the newborn death rate in different cities in Vietnam, data would be sorted and inaccurate data would be eliminated following data collection. During the observation, there will be several mistakes. The most common inaccuracy that is likely to occur is that many participants will have bias with the survey, resulting in inaccurate results (Chittaranjan Andrade et al). To address this issue, participants will get only the information required to complete the job.
* Once the research is complete, it’s essential to compare the results with the previous research which were conducted by government agencies in order to enhance the precision as well as detect errors.

1. Reference

“Central Limit Theorem” *Corporate Finance Institute*, viewed 10 December 2022, <https://corporatefinanceinstitute.com/resources/data-science/central-limit-theorem/>.

Watts, V “8.7 Hypothesis Tests for a Population Mean with Unknown Population Standard Deviation – Introduction to Statistics,” *8.7 Hypothesis Tests for a Population Mean with Unknown Population Standard Deviation – Introduction to Statistics*, viewed 10 December 2022, <https://ecampusontario.pressbooks.pub/introstats/chapter/8-7-hypothesis-tests-for-a-population-mean-with-unknown-population-standard-deviation/>.

“Significance, Errors, Power, and Sample Size: The Blocking and Tackling of Statistics - PubMed” 2018, *PubMed*, viewed 10 December 2022, <https://pubmed.ncbi.nlm.nih.gov/29346210/>.

UC Berkeley “Multiple Hypothesis Testing and False Discovery Rate,.”

Banerjee, A, Chitnis, UB, Jadhav, SL, Bhawalkar, JS & Chaudhury, S “Hypothesis testing, type I and type II errors,” *PubMed Central (PMC)*, viewed 10 December 2022, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2996198/>.

Rothman, KJ 2010, “Curbing type I and type II errors,” *PubMed Central (PMC)*, viewed 10 December 2022, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2850991/>.

The National EMSC Data Analysis Resource Center, University of Utah, N- 2022, “NEDARC - Confidence Intervals,” *NEDARC - Confidence Intervals*, viewed 10 December 2022, <https://www.nedarc.org/statisticalhelp/advancedstatisticaltopics/confidenceintervals.html>.

“The Effects of a Small Sample Size Limitation” 2018, *Sciencing*, viewed 10 December 2022, <https://sciencing.com/effects-small-sample-size-limitation-8545371.html>.

World Health Organisation 2022, “Newborn Mortality,” *Newborn Mortality*, viewed 10 December 2022, <https://www.who.int/news-room/fact-sheets/detail/levels-and-trends-in-child-mortality-report-2021#:~:text=There%20are%20approximately%206700%20newborn,up%20from%2040%25%20in%201990>.

“Infant mortality: tragic, unjust … and usually preventable” 2021, *Infant Mortality Rate: Facts & How to Help | What World Vision Does*, viewed 10 December 2022, <http://www.worldvision.ca/stories/health/infant-mortality-rate>.

Frost, J 2017, “How t-Tests Work: t-Values, t-Distributions, and Probabilities,” *Statistics By Jim*, viewed 10 December 2022, <http://statisticsbyjim.com/hypothesis-testing/t-tests-t-values-t-distributions-probabilities/>.

“Neonatal mortality - UNICEF DATA” *UNICEF DATA*, viewed 10 December 2022, <https://data.unicef.org/topic/child-survival/neonatal-mortality/>.

“Infant Mortality Rate by Country 2022” *Infant Mortality Rate by Country 2022*, viewed 10 December 2022, <https://worldpopulationreview.com/country-rankings/infant-mortality-rate-by-country>.

Wolde, HF, Gonete, KA, Akalu, TY, Baraki, AG & Lakew, AM 2019, “Factors affecting neonatal mortality in the general population: evidence from the 2016 Ethiopian Demographic and Health Survey (EDHS)—multilevel analysis - BMC Research Notes,” *BioMed Central*, viewed 10 December 2022, <https://bmcresnotes.biomedcentral.com/articles/10.1186/s13104-019-4668-3>.

Aghai, ZH, Goudar, SS, Patel, A, Saleem, S, Dhaded, SM, Kavi, A, Lalakia, P, Naqvi, F, Hibberd, PL, McClure, EM, Nolen, TL, Iyer, P, Goldenberg, RL & Derman, RJ 2020, “Gender variations in neonatal and early infant mortality in India and Pakistan: a secondary analysis from the Global Network Maternal Newborn Health Registry - Reproductive Health,” *BioMed Central*, viewed 10 December 2022, <https://reproductive-health-journal.biomedcentral.com/articles/10.1186/s12978-020-01028-0>.

“Stratified Sampling - Research Methodology” *Research-Methodology*, viewed 10 December 2022, <https://research-methodology.net/sampling-in-primary-data-collection/stratified-sampling/>.

Sharma, G “Pros and cons of different sampling techniques,.”

Magwai, T, Oginga, FO, Chiliza, B, Mpofana, T & Xulu, KR 2022, “Genome-wide DNA methylation in an animal model and human studies of schizophrenia: a protocol for a meta-analysis,” *BMJ Open Science*, vol. 6, no. 1, p. e100264.

Andrade, C 2020, “The Limitations of Online Surveys,” *Indian Journal of Psychological Medicine*, vol. 42, no. 6, pp. 575–576.

“GNI per capita, Atlas method (current US$) - Sierra Leone | Data” *GNI per capita, Atlas method (current US$) - Sierra Leone | Data*, viewed 10 December 2022, <https://data.worldbank.org/indicator/NY.GNP.PCAP.CD?locations=SL>.

“GNI per capita, PPP (current international $) - Central African Republic | Data” *GNI per capita, PPP (current international $) - Central African Republic | Data*, viewed 10 December 2022, <https://data.worldbank.org/indicator/NY.GNP.PCAP.PP.CD?locations=CF>.

“GNI per capita, Atlas method (current US$) - Iceland | Data” *GNI per capita, Atlas method (current US$) - Iceland | Data*, viewed 10 December 2022, <https://data.worldbank.org/indicator/NY.GNP.PCAP.CD?locations=IS>.

“San Marino GNI per capita, 1970-2021 - knoema.com” *Knoema*, viewed 10 December 2022, <https://knoema.com//atlas/San-Marino/topics/Economy/National-Accounts-Gross-National-Income/GNI-per-capita>.

Aldous, D “STAT C205B/MATH C218B: Probability Theory - Lecture 3,.”

UC Berkeley “STATC141 - Multiple Hypothesis Testing and False Discovery Rate,.”