C. DETAILED OUTLINE

I. Introduction -

- 2-3 first sentences: Define crude birth rate

Example:

The crude birth rate measures the total number of live births per 1,000 persons in a given population over the course of a year. It provides a key metric for gauging fertility and population growth trends within a country or region. The crude birth rate can be compared to more refined fertility measures like age-specific birth rates and total fertility rates.

- 4-5 next sentences: Explain importance of monitoring birth rates
    + Highlights demographic trends, population growth
    + Informs public health planning, resource allocation
    + Allows analysis of determinants of fertility
    + Impacts economic growth, ages structure, dependency ratios

Example:

Carefully monitoring crude birth rates is crucial for several reasons. The rates highlight important demographic shifts in population sizes and growth trajectories. Trends and forecasts for birth rates also allow governments to effectively plan healthcare capacity, education investments, and other social services based on projections of dependent population cohorts. Additionally, statistical analysis of changing crude birth rates facilitates study of the various cultural, economic, and policy determinants driving fertility. Variations in birth rates profoundly shape national economic growth, age distributions, and dependency ratios.

- 4-5 next sentences: Review relationship between birth rate and GNI per capita
    + Fertility tends to fall as income rises (demographic transition model)
    + But many factors interact - complex relationship
    + Review academic studies on income-fertility link
    + Note exceptions, changing trends

Example:

The demographic transition model hypothesizes that national fertility rates historically decline as countries transition from low to high levels of income per capita. However, empirical studies find the relationship is complex, influenced by many other mediating factors. Academic research reveals examples of higher income countries where fertility has stagnated rather than continuing to fall. The link between income and fertility has become more heterogeneous and non-linear in recent decades.

- 4-5 next sentences: Describe other factors influencing birth rates
    + Female education levels
    + Infant and child mortality rates
    + Urbanization
    + Religion, culture, family planning access
    + Government policies

Example:

Aside from income, key influences on crude birth rates include female education levels and labor force participation, infant and child mortality rates, extent of urbanization, access to contraception, prevailing religious and cultural norms, and family planning policies. Government interventions through regulatory restrictions, subsidies, or awareness campaigns also shape fertility intentions and behaviors.

II. Probability & Descriptive Analysis

1. Probability

Here's a step-by-step guide on how to check independence and determine which country categories are more likely to have a high prevalence of overweight in children using probability:

1.1. Define Events

- Let A be the event of a high birth rate.
- Let B be the event of a specific GNI per capita category (e.g., Low, Middle, High).

1.2. Calculate Marginal Probabilities

- Calculate the marginal probability of A (the probability of high birth rate).

$$P(A) = \frac{Number\ of\ occurrences\ of\ high\ birth\ rate}{Total\ number\ of\ observations}$$

- Calculate the marginal probability of B (the probability of being in a specific GNI per capita category).

1.3. Calculate Joint Probability

- Calculate the joint probability of both events A and B occurring.

$$P(A \cap B) = \frac{Number\ of\ occurrences\ of\ high\ birth\ rate\ in\ the\ specific\ GNI\ per\ capita\ category}{Total\ number\ of\ observations}$$

1.4. Check Independence

- If $P(A \cap B) = P(A) \cdot P(B)$, the events are independent.
- If $P(A \cap B) \neq P(A) \cdot P(B)$, the events may be dependent.

1.5. Repeat for Different GNI Categories

- Repeat the process for each GNI per capita category (Low, Middle, High) to assess independence for each group.

1.6. Compare Conditional Probabilities

- Compare $P(A|B)$ across different GNI per capita categories.
- If $P(A|B)$ is higher in one category compared to others, that category is **more likely to have a high prevalence of overweight in children.**

1.7. Interpret Results (6-7 sentences)

- For independence:
  + If equality holds, you can conclude that high birth rate and GNI per capita are statistically independent.
  + If the equality does not hold, it suggests a potential association or dependence between the two variables.
- For determination of country has the highest likelihood:
  + If $P(A|B)$ is highest in Low GNI per capita countries, you may conclude that children in Low GNI per capita countries are more likely to have a high prevalence of overweight.

Example:

Tests of statistical independence reveal that crude birth rate and income level are dependent events, as the probability of high birth rates within each income category was not equal. Further examination of the probabilities shows a clear link between national income and birth rates. Low-income countries had the highest probability of high birth rates at 76.9%, while middle-income and high-income countries had substantially lower probabilities of 42.9% and 0% respectively. This analysis demonstrates that lower income countries tend to experience markedly higher crude birth rates compared to their higher income counterparts. The economic development level of a nation appears to be strongly associated with prevailing fertility rates based on the calculated conditional probabilities across income groups.

2. Descriptive Statistics

2.1 Measures of Central Tendency

(i) Calculation

- To calculate three measures of central tendency, you can utilize formula in Excel (Google Sheets)
    + Mean: = AVERAGE(range)
    + Median: = MEDIAN(range)
    + Mode: Excel doesn't have a direct mode function, so you might need to use a combination of functions. For example, you can use the following array formula for mode: =MODE.MULT(range)

- Construct table with income categories and measures of central tendency

Example:

Central Tendency	LI	MI	HI
Mean	x	y	z
Median	x	y	z
Mode	None	None	z
Q1	x	y	z
Q3	x	y	z
IQR	x	y	z
Lower limit	x	y	z
Upper limit	x	y	z
Outliers	None	None	None

(i) Interpretation

- 2-3 next sentences: Determine the best measure, you have to detect outliers
    + Calculate the IQR, which is the range between the first quartile (Q1 =QUARTILE.INC(dataset, 1)) and the third quartile (Q3 =QUARTILE.INC(dataset, 3)). Points outside the range Q1 −(1.5×IQR) to (Q3+1.5×IQR) are considered potential outliers.

- How to choose the best measure:
    + Mean: Use the mean when your data is approximately symmetric and does not have extreme outliers.
    + Median: Choose the median when your data is skewed or contains outliers. The median is less sensitive to extreme values.
    + Mode: Use the mode when you want to identify the most frequently occurring value. However, the mode might not be suitable for continuous data or data with no clear mode.

Example:

An examination of the data reveals no extreme outliers across the three income categories, indicating the mean is an appropriate measure of central tendency for analyzing crude birth rates.

- 6-7 next sentences: Structure for analysing the crude birth rates of both countries using only the BEST measure.
    + Compare the values of different categories - which country has a higher best measure of central tendency for crude birth rates?
    + Comment on how close or far apart the two values are. Are the average crude birth rates drastically different or relatively similar?
    + Relate the best measures back to the context - do they confirm expectations in each country category?
    + Avoid just describing the values - interpret and analyze what the best measure reveals about differences and trends in crude birth rate between the two countries.

Example:

The calculated mean birth rates differ markedly between income levels. Specifically, the middle income mean rate is approximately 1.5 times higher than the high income mean rate. However, it is only 1.7 times the low income mean rate. This demonstrates that economically disadvantaged nations have substantially higher fertility compared to their prosperous counterparts. The median values further show that 50% of low income countries have a birth rate above 27.09 per 1,000, while 50% of high income countries have a rate above just 10.6 per 1,000. The pronounced variation in means and medians aligns with prior research indicating economic factors likely influence fertility decisions. Overall, the descriptive analysis clearly demonstrates poorer nations have considerably higher crude birth rates compared to richer countries.

2.2 Measures of Variation

(i) Calculation

- Calculate range, interquartile range, variance, coefficient of variation and standard deviation using Excel formulas (Google Sheets)
    + Range: =MAX(range)-MIN(range)
    + Interquartile Range = Calculated in Part 2
    + Variance: =VAR(range)
    + Standard deviation: =STDEV(range)
    + Coefficient of Variation = Standard Deviation/Mean

- Construct table with income categories and measures of variation

Example:

Variation	**LI**	**MI**	**HI**
Range	28,874	13,767	6,516
IQR	13,287	8,117	1,945
Sample Variance (SV)	77,548	23,001	3,581
Standard Deviation (SD)	8,806	4,796	1,892
Coefficient of Variation (CV) (%)	32,507	29,790	17,816

(i) Interpretation

- 1 next sentence: How to choose the best measure:
    + Range: Affected by outliers
    + Variance: Outlier impact squared
    + Standard deviation: Best for typical distributions
    + IQR: Uses middle 50%, so filters outliers
    + CV: Standardizes by mean, allows comparison

Example:

SD is the ideal indicator for observing the dispersion of data around the Mean because the units are identical and there are no outliers.

- 3-4 next sentences: Structure for analysing the crude birth rate of both countries using only the BEST measure.
    + Compare standard deviations - which country group has the highest crude birth rate?
    + Interpret differences in variation - are rates relatively stable or fluctuating wildly?

Example:

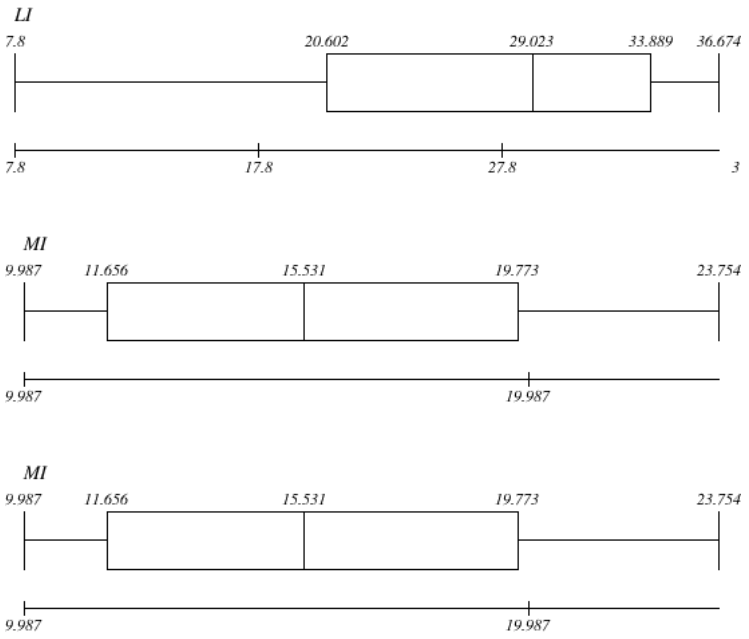
- Therefore, this statistic may be used to investigate value dispersion. While MI's SD (4,796%) is 2.5 times that of HI's (1,892%), it only makes up about half of LI's (8,806%). Thus, LI's birth rate is moderately higher than MI's and HI's, implying that LI and MI had greater variation around the Mean than the HI.

2.3 Box-and-Whisker plot

(i) Creating the Box Plots

- Use Excel to generate box plots for each income category
- X-axis is income category
- Y-axis is crude birth rate
- Ensure box plots are scaled identically

Example:



(ii) 4-5 next sentences: Analyzing the Box Plots

- Compare medians across income categories
- Examine interquartile range spreads
- Note any outliers
- Visually assess skew based on median location and tail lengths
- Relate to previous central tendency analysis
- Interpret differences in context of income-fertility relationship
- Align with other graphical analysis
- Summarize how box plots further inform income-birth rate relationship

Example:

- The low income category has the highest median birth rate.
- It also has the widest spread based on its interquartile range.
- Positive skew seen with longer right tail for low income rates.
- The pronounced skew fits with frequent high birth rates amid poverty.
- Box plots are consistent with central tendency analysis and positively skewed distribution.
- In summary, box plots further demonstrate low income countries have substantially higher and more variable crude birth rates.

III. Confidence Intervals

A. Computation

Step 1: Choose Confidence Level:

- Decide on the confidence level you want for your interval. Common choices are 95% or 99%. A 95% confidence level is often used, which means you are 95% confident that the true average falls within the calculated interval.

Step 2: Determine the Z-Score or T-Score:

- Based on your chosen confidence level, determine the critical value (Z-score or T-score) from the standard normal distribution or the t-distribution. This is based on the number of data points and the desired confidence level.

Note:

- Z-score: a statistical measurement that describes a value's relationship to the mean of a group of values, measured in terms of standard deviations from the mean. In the context of confidence intervals, it's used when you are working with large sample sizes (usually over 30) and when the population standard deviation is known.
- T-score is similar to a Z-score but is used when the sample size is small (usually less than 30) or when the population standard deviation is unknown. The T-score is based on the t-distribution, which adjusts for the uncertainty introduced by these factors

3. Calculate the Margin of Error:

- The margin of error (ME) is given by the formula:

ME = Crititical Value x $\frac{s}{\sqrt{n}}$

where s is the standard deviation and n is the number of countries.

4. Calculate the Confidence Interval:

- Use the margin of error to calculate the confidence interval:

$$\text{Confidence Interval} = (\bar{x} - \text{ME}, \bar{x} + \text{ME})$$

This interval provides a range within which the true average crude birth rate is likely to fall.

Note:

- $\bar{x}$: This is the sample mean. It's the average value of your data set. If you're measuring something like the average height in a sample group, x would be that average height.

5. Interpret the Results:

- 2-3 next sentences: Interpret the confidence interval. For example, if you calculated a 95% confidence interval, you can say that you are 95% confident that the true average crude birth rate for the world falls within the calculated interval.

Example:

Sample Size (n)	28
Significance Level (α)	5%
Confidence Level (1 - α)	95%
Population Standard Deviation (σ)	Unknown
Sample Standard Deviation (S)	9,725
Sample Mean ($\overline{X}$)	19,637
Degree of freedom (d.f = n - 1)	27
T-value	$\pm$ 2,052

Note:

To determine if data appears approximately normal, you can use both visual and statistical methods:

Visual Inspection:

- Histogram: Plot a histogram of your data. If the data is approximately normal, the histogram should show a bell-shaped curve.
- Q-Q Plot: In a Quantile-Quantile Plot, if your data is approximately normal, the points should fall roughly along a straight line.

Statistical Tests:

- Shapiro-Wilk Test: This is a common test for normality. If the p-value is above a certain threshold (usually 0.05), it suggests that the data does not significantly deviate from a normal distribution.

B. Assumption

- 3-4 next sentences: State assumptions, here are several:
- The key assumption is that the sample data follows an approximate normal distribution.

    + This normality assumption enables the use of a z-score/t-score statistical test to calculate the margin of error and confidence interval range.
    + Assess the distribution of the sample data visually with histograms and statistically with normality tests.
    + If data appears approximately normal, state this as justification for assuming normality when calculating the confidence interval.
    + If data is significantly non-normal, transformations may be required or non-parametric methods used.

- Other minor assumptions:

+ Data is a random sample representative of the population.
+ Sample observations are independent.
+ The parameter being estimated (e.g. population mean) is a fixed constant.
+ Standard deviation can be accurately estimated from the sample.
+ Briefly note these assumptions apply but normality of the data distribution is the most critical.
+ Explain that confidence intervals based on normal distributions will be inaccurate if this assumption is violated.
+ Assessing and validating assumptions is key to ensuring statistically valid confidence intervals.

C. Discussion

- 6-8 sentences: Discuss the results of computation.

Example:

- If the birth rate's world standard deviation (SD) is determined, the critical z-value table is employed. Both population and sample SD measure variability, although their confidence levels differ. The t-value is calculated in the same way as the z-value, but for a sample. Population SD is constant over the population dataset (Taylor 2019), but sample SD exhibits more fluctuation. Sample SD is more unreliable than population SD.

IV. Hypothesis Testing

A. Hypothesis Testing

Step 1: State the Hypotheses:

- Null Hypothesis (H0): The world average crude birth rate in 2020 is equal to 19 (the value from the previous study).

$$H0 : \mu = 19$$

- Alternative Hypothesis (H1): The world average crude birth rate in 2020 is different from 19.

$$H1: \mu = 19$$

Step 2: Set the Significance Level:

Choose a significance level (α), commonly 0.05. This represents the probability of committing a Type I error (rejecting a true null hypothesis).

Step 3: Check if the Hypothesized Value is within the Confidence Interval:

- If the hypothesized value (19 in this case) falls within the confidence interval, you would fail to reject the null hypothesis.
- If the hypothesized value is outside the confidence interval, it suggests evidence against the null hypothesis.

Step 4: Perform the Hypothesis Test:

- If the hypothesized value is outside the confidence interval, calculate a test statistic (e.g., z-score or t-score) to determine how many standard errors the hypothesized value is from the mean.
- If the test statistic falls beyond the critical values, you reject the null hypothesis.

Step 5: Make a Decision:

- If the test statistic falls beyond the critical values, reject the null hypothesis and conclude that there is evidence that the world average crude birth rate in 2020 is different from 19.
- If the test statistic falls within the critical values, fail to reject the null hypothesis, suggesting that there is not enough evidence to conclude a significant difference.

Possible Errors:

- Type I Error (False Positive): Rejecting the null hypothesis when it is true (i.e., concluding that the world average crude birth rate in 2020 is different from 19 when it's actually not).

- Type II Error (False Negative): Failing to reject the null hypothesis when it is false (i.e., failing to detect a difference in the world average crude birth rate in 2020 when there is a difference).

Consequences:

- Committing a Type I error may lead to incorrect policy decisions or interventions based on the assumption that the crude birth rate has changed when it has not.
- Committing a Type II error may result in a failure to implement necessary interventions or policies when the birth rate has actually changed.

Minimizing Errors:

- Increase Sample Size: A larger sample size can reduce the margin of error and increase the power of the test, making it easier to detect a true difference.
- Adjust Significance Level: Depending on the consequences of Type I and Type II errors, you may choose a more conservative or liberal significance level.
- Replicate Studies: Conducting additional studies or replicating the study to verify results can increase confidence in the findings.

B. Possible impact

- Increasing sample size by 50% makes interval likely narrower
- This could change decision if 19 is excluded with larger sample, but not with original smaller sample
- Larger sample gives more accurate population estimate, reducing margin of error
- However, the specific value of 19 becomes no more or less likely to be the true mean
- Larger sample mainly reduces chance of Type II errors failing to reject an incorrect null
- But has minimal impact on Type I errors incorrectly rejecting a true null

V. Overall Conclusion

- 1-2 next sentences: Briefly restate the research topic in simple terms
- 10- 15 next sentences: Summarize 3 key statistical findings in plain language without jargon
    - For each finding:

- Restate the analysis done in simple terms (e.g. "compared countries' birth rates")
- State the main finding in layman's terms (e.g. "poorer countries had higher rates")
- Explain the takeaway in 1 sentence (e.g. "low income is linked to high fertility")
- Use relatable analogies or examples to illustrate technical concepts
- Avoid statistical terminology and replace with general descriptors
- Focus on broader meaning versus methodological details
- Highlight real world relevance and implications
- Close by broadening beyond just the specific findings to high-level takeaways

Example:

- This research examined factors influencing birth rates across countries.
    - Key findings:
- Compared countries' birth rates based on income levels:
- Poorer nations had higher average birth rates
- Low income is associated with greater fertility
- Analyzed how spread out countries' birth rates were:
- Poorer countries had wider differences in birth rates
- Poverty is linked to unstable fertility rates
- Tested past predictions on global birth rate:
- Current rate still aligned with past forecasts
- Global fertility rate appears relatively constant over time
- Overall, the analyses demonstrated strong links between income and birth rates. Poorer nations face greater challenges around high and unstable fertility. As countries develop, birth rates tend to fall and level off. These insights help inform policies aimed at improving economic well-being and public health outcomes.

VI. Extension

1. Sample type (6-7 next sentences):

- Use a probability sampling method like simple random sampling to ensure representative sample
- Simple random sampling gives each household equal chance of selection
- Avoid non-probability methods like convenience sampling which can bias results

2. Conducting the survey (8-10 next sentences):

- Obtain list of all households in Vietnam for sampling frame
- Use random number generator to randomly select sample from list
- Recruit and train interviewers on standardized data collection
- Visit selected households in-person and get consent for survey
- Administer structured questionnaire with birth rate and socioeconomic questions
- Offer small incentive for participation

3. Potential errors (8-10 next sentences):

- Sampling bias if sampling frame incomplete
- Non-response bias if low response rates
- Response bias if questions misinterpreted
- Interviewer bias if not properly trained

4. Reducing errors (8-10 next sentences):

- Verify completeness of household list
- Follow-up with non-respondents
- Pilot and refine questionnaire
- Thoroughly train interviewers

D. TIPS & TRICKS

1. **Writing Tips:**

- Stick closely to the provided structure and guidelines for each section
- Use formal academic writing style - avoid colloquialisms
- Define statistical terms and explain methodologies clearly
- Interpret and analyze results, don't just describe them
- Use topic sentences and smooth transitions between paragraphs
- Cite any data sources used
- Proofread carefully for spelling, grammar, punctuation errors

2. **Data Visualization Tips**

- Properly label charts with titles, axes, units etc.
- Choose appropriate chart types based on the data

- Format visualizations consistently and clearly
- Integrate graphs with text seamlessly
- Point out key insights from charts in your analysis

3. **Analysis Tips**

- Apply concepts and techniques covered in class
- Perform calculations carefully and show full work
- Interpret statistics to derive meaningful insights
- Synthesize results to draw overall conclusions
- Note any limitations, assumptions or caveats regarding analysis

E. FOOD FOR HUNGRY THOUGHTS

Based on your assigned countries and dataset, there are various journal articles and reports to be explored. Below are some reliable sources for you to find relevant articles and reports on your chosen topics.

- "The Relationship between National Income and Fertility" (Journal of Political Economy) https://www.journals.uchicago.edu/doi/abs/10.1086/260466
- "Socioeconomic Determinants of Fertility Decline" (World Development) https://www.sciencedirect.com/science/article/abs/pii/0305750X9290098N
- "Women's Education and Fertility: Results from 26 Demographic and Health Surveys" (Studies in Family Planning) https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1728-4465.2010.00235.x
- "Urbanization and Fertility Rates" (Health Economics) https://onlinelibrary.wiley.com/doi/10.1002/hec.3859
- "Fall in global fertility rates" (BBC) https://www.bbc.com/news/health-53409521
- "How Demographics Drive the Global Economy" (Wall Street Journal) https://www.wsj.com/articles/how-demographics-rule-the-global-economy-11569429774
- "Education, Income, and Fertility: Singapore's policy success" (East Asia Forum) https://www.eastasiaforum.org/2022/06/05/education-income-and-fertility-singapores-policy-success/

- "Africa bucking the trend on low fertility rates" (The Guardian) https://www.theguardian.com/global-development/2021/feb/18/africa-bucking-the-trend-on-low-fertility-rates-population-growth-sdgs