

- Suppose you estimate an OLS regression of Y on X . Now consider a new variable $X^* = 10X$. How does the estimated slope coefficient for the regression of Y on X^* compare to the slope in the original regression?
 - It remains unchanged.
 - It becomes 10 times larger.
 - It becomes 10 times smaller.
 - It reverses sign.

- Let V_t denote the nominal value of production in period t , and let P_t be a price index with base year 0. Define the deflated (real) series

$$V_t^{real} = \frac{V_t}{P_t}.$$

A separate “quantity index” is defined as

$$Q_{t,0} = \frac{\sum_j p_0^j q_t^j}{\sum_j p_0^j q_0^j},$$

where p_0^j are base-year prices.

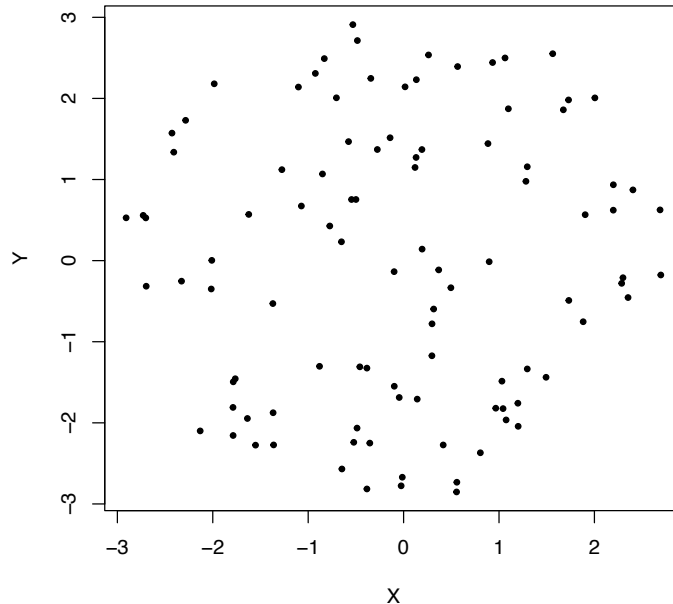
Which statement correctly distinguishes the economic meaning of V_t^{real} versus $Q_{t,0}$?

- V_t^{real} and $Q_{t,0}$ must coincide because both measure the change in nominal output after removing price movements.
 - V_t^{real} removes price effects using a general price level, whereas $Q_{t,0}$ measures changes in physical quantities using base-year relative prices as weights.
 - V_t^{real} isolates the pure quantity effect, while $Q_{t,0}$ removes the quantity effect and leaves only price dynamics.
 - The two measures differ only by a constant scaling factor, so they will have identical growth rates even if their levels differ.
- Suppose you are working for the State Statistical Office and conducting a large household interview survey. Respondents report their education level using the following categories:

none, primary, secondary, tertiary.

How should this variable be correctly classified?

- Qualitative — nominal
 - Quantitative — discrete (because education can be measured in complete years)
 - Qualitative — ordinal
 - Quantitative — continuous (because education evolves over time)
- Which of the following violates valid continuous class construction?
 - $[0, 50), [50, 100), [100, 150)$
 - $(0, 50], (50, 100], (100, 150]$
 - $[0, 50], (50, 100], (100, 200]$
 - $[0, 100), [100, 200), [180, 250)$
 - The graph below shows a simulated scatterplot of two variables, X and Y . Which statement about the correlation between X and Y is most accurate?



Based on the scatterplot shown,

- (a) The correlation will be close to +1, because the points are centered around the origin, so larger X values tend to occur with larger Y values on average.
- (b) The correlation will be close to -1 , because points with large positive X values often correspond to negative Y values and vice versa, creating a balancing negative association.
- (c) The correlation will be close to 0, because the pattern shows no consistent increase or decrease in Y as X changes.
- (d) The correlation cannot be computed, because the relationship is nonlinear and the cloud of points does not follow a functional trend.

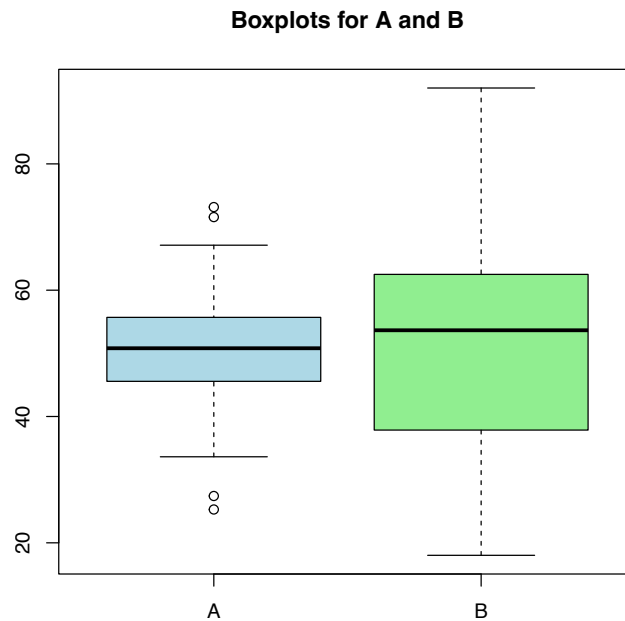
6. The following table shows monthly study hours of a sample of 60 students:

Hours	Frequency (f_j)	Relative Frequency
0–5	5	0.0833
5–15	12	0.2000
15–25	20	0.3333
25–35	13	0.2167
35–50	10	0.1667

which of the following statements is correct?

- (a) Mean hours ≈ 22.46 , Median hours ≈ 21.5 , Modal class = 15–25
- (b) Mean hours ≈ 24.54 , Median hours ≈ 22.2 , Modal class = 25–35
- (c) Mean hours ≈ 20.00 , Median hours ≈ 20.0 , Modal class = 0–5
- (d) Mean hours ≈ 26.00 , Median hours ≈ 25.0 , Modal class = 35–50

7. Two boxplots for variables A and B are shown below.



Which variable most likely has the *larger variance*?

- (a) They must have the same variance, because their medians and central boxes appear approximately aligned.
- (b) Variable B, because its box is wider, its whiskers extend further, and it exhibits a larger set of high-end outliers.
- (c) Impossible to determine: boxplots reveal quartiles but cannot be used to reason about variance or spread in any way.
- (d) Variable A, because its lower whisker extends farther relative to its IQR, indicating greater spread of observations.

8. A relative-frequency histogram has two adjacent classes. Class A has width 10 and height 0.06; class B has width 40 and height 0.015. Which statement is correct?

- (a) Both bars have the same area.
- (b) Class B has a larger area than class A.
- (c) Class A has larger area than class B.
- (d) Areas cannot be compared without the total sample size.

9. You are estimating the median from grouped data using standard grouped-quantile interpolation. Consider the following possible changes to how data are grouped. Which operation will generally reduce the accuracy of the grouped median estimate?

- (a) Decreasing class widths to create more narrow intervals, even though this may increase sampling noise within each class.
 - (b) Increasing class widths so that individual observations become spread over fewer, wider classes, even if midpoints still look “reasonable.”
 - (c) Switching to the original ungrouped data, eliminating class structure entirely.
 - (d) Keeping the same midpoints and class frequencies while modifying only the boundaries of the classes so that widths no longer match the implied midpoint spacing.
10. A sample correlation coefficient between X and Y is estimated to be very close to zero. Which statement most accurately describes what this implies about the relationship between the variables?
- (a) It rules out any linear association, but nonlinear dependence or other forms of systematic association may still be present.
 - (b) It implies that X and Y are statistically independent, meaning knowledge of one provides no information about the other.
 - (c) It indicates a data-quality problem because valid datasets rarely produce correlations near zero unless variables are constant.
 - (d) It shows that both variables must have similar variances; otherwise the correlation could not be close to zero.
11. A statistical agency publishes a fixed-base index number $i_{t,0}$ comparing the level of a variable X_t in period t to its level X_0 in the base period. Suppose an analyst reports the following information:

$$i_{t,0} = 0.87, \quad i_{t,1} = 1.06, \quad i_{1,0} = 0.92.$$

Which interpretation is *correct* for the fixed-base index $i_{t,0} = 0.87$?

- (a) X_t is 13% lower than X_0 because $i_{t,0} < 1$ means the numerator of X_t/X_0 is smaller than the denominator.
 - (b) Since $i_{t,0} < 1$, this implies X_t must be negative or the index would not fall below 1.
 - (c) Because $i_{t,0} = 0.87$ and $i_{1,0} = 0.92$, period t must lie between 0 and 1 in calendar time.
 - (d) An index below 1 means X_0 is 13% lower than X_t , so period t exceeds the base-value level.
12. Which property does a time-series chain index (constructed by linking consecutive short-period indices) satisfy?
- (a) Each link is averaged (arithmetically) with all previous links to obtain the multi-period change.
 - (b) The multi-period index over several periods equals the product of the successive one-period (link) factors.
 - (c) Only the first and last one-period links determine the multi-period chained index; intermediate links can be ignored.
 - (d) A chain index can never fall below 1, because each link is a positive ratio and indices are normalized.
13. You are given the following fixed-base price indices (base year = 2010):

$$i_{2018,2010} = 1.40, \quad i_{2020,2010} = 1.75.$$

You are asked to express the 2018 price level relative to the 2020 price level, i.e. compute

$$i_{2018,2020}.$$

Which of the following is correct?

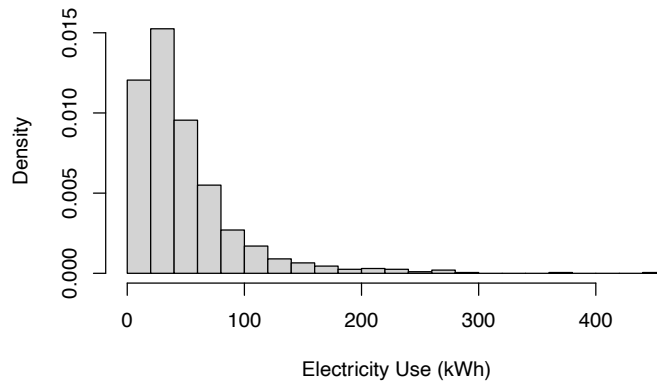
- (a) $i_{2018,2020} = \frac{1.40}{1.75} \approx 0.800$
 - (b) $i_{2018,2020} = 1.40 \times 1.75 = 2.450$
 - (c) $i_{2018,2020} = \frac{1.75}{1.40} \approx 1.250$
 - (d) $i_{2018,2020} = 1.40 - 1.75 = -0.35$
14. Real GDP grows by 14% over a given period, while population grows by 9%. Using the standard approximation for per-capita growth rates, what is the approximate growth rate of real GDP per capita?
- (a) About 5%, because per-capita growth is approximately GDP growth minus population growth.
 - (b) It cannot be determined without initial levels of GDP and population.
 - (c) Real GDP per capita must have decreased because population grew.
 - (d) Exactly 14%, because population growth does not affect per-capita GDP.
15. Suppose we run an OLS regression of Y on X using the model

$$Y = \beta_0 + \beta_1 X + u.$$

Now imagine that every value of X in the dataset is multiplied by 0.1. How does this transformation affect the estimated OLS slope $\hat{\beta}_1$?

- (a) It multiplies by 0.1.
 - (b) It multiplies by 10.
 - (c) It remains unchanged.
 - (d) It becomes undefined.
16. The histogram below shows household electricity use in a large sample. Based on the shape of the histogram, which description of the quartiles is most plausible?

Histogram of Household Electricity Use



- (a) Q1 close to the mode, Q3 far from the median.
- (b) Q1 far above the mode, Q3 close to the mode.
- (c) Q1 and Q3 equally distant from the median.
- (d) $Q1 > Q2 > Q3$.

17. You work for the National Statistical Agency and must prepare a rapid briefing on how income inequality will change after a policy reform. Because you do not yet have time to recompute the full Gini index from microdata, you rely on known properties of the Gini index to explain to your colleagues how inequality will change under certain transformations.

Suppose the reform consists of a uniform lump-sum transfer: every individual in the population receives the same fixed payment. Before doing any computations, how should you expect the Gini index to change?

- (a) The Gini index will decrease, because adding the same amount to everyone reduces relative differences in income.
- (b) The Gini index will increase, because richer individuals gain a larger share of total income after the transfer.
- (c) The Gini index will remain unchanged, because adding a constant to each income does not affect the Gini index.
- (d) It is impossible to know how the Gini index changes without full recalculation from microdata.

18. You are given the grouped frequency table below from an income survey.

Class	Frequency
0–10	6
10–25	15
25–40	18
40–60	12
60andmore	9

Which of the following is closest to the grouped mean?

- (a) 28.7
- (b) 35.1
- (c) 31.4
- (d) 26.9

19. The table below summarizes weekly study hours for a sample of 50 students. Relative frequencies are shown and class widths are given.

Hours	Relative Frequency	Class width (a_j)
0–5	0.08	5
5–15	0.20	10
15–20	0.16	5
20–35	0.30	15
35–50	0.26	15

Compute the approximate modal value (grouped-mode) of study hours using the standard interpolation for the modal class.

- (a) 19.5
- (b) 26.5
- (c) 16.5
- (d) 17.5

20. A national statistics office reports a Laspeyres price index $L_{t,0}$ with base year 0. They decide to *rebase* the index so that year 1 becomes the new base year. Let $L_{1,0}$ denote the previously published index for year 1 using base year 0. After rebasing, all new index values must satisfy $L_{1,1} = 1$.

Which statement correctly describes how the rebasing affects the entire index series $\{L_{t,0}\}$?

- (a) Each index value increases, because a more recent base year adjusts for substitution and therefore raises the scale of the series.
- (b) Each index value decreases, because normalizing at year 1 removes inflation accumulated between years 0 and 1.
- (c) Each index value is obtained by dividing $L_{t,0}$ by the old value $L_{1,0}$, so that all series values are rescaled relative to year 1.
- (d) All index values remain unchanged except for the new base year, because rebasing only renames the base and does not alter the value of other years.

21. When economists summarize multi-year growth using a single “average annual growth rate,” they replace the sequence of year-to-year changes with one constant rate applied over all k years.

Which statement best explains why this constant compound rate is the appropriate summary measure?

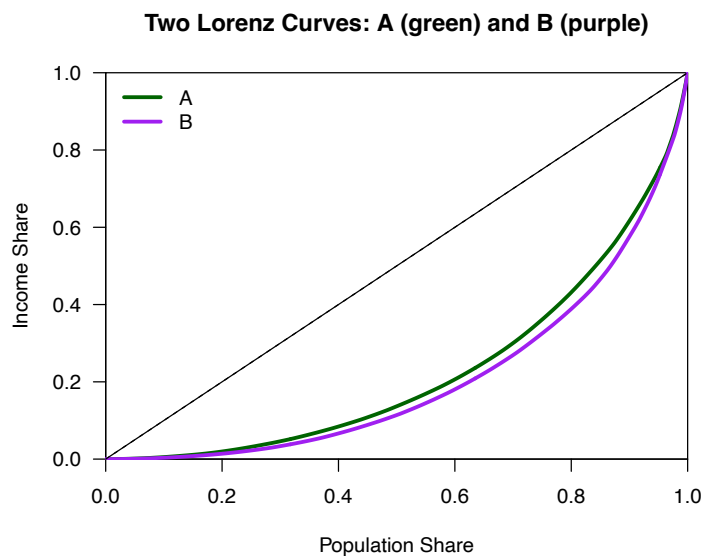
- (a) It equals the simple arithmetic mean of the k annual percentage changes, ensuring that unusually high or low single-year rates do not distort the long-run picture.
- (b) It is the unique annual rate which, when applied repeatedly for k years, reproduces exactly the observed overall proportional change from X_t to X_{t+k} .
- (c) It ensures that each year’s contribution to long-run growth is evaluated relative to the initial level X_t , thereby keeping all yearly percentage changes comparable.
- (d) It assumes the total change $X_{t+k} - X_t$ is spread evenly across years, treating growth as linear so that each year contributes an equal absolute increase.

22. Suppose you want a frequency polygon to provide a good numerical approximation to the underlying probability density function of a continuous variable. You have classes of unequal width and a large sample.

Which condition is necessary for the frequency polygon to approximate the density function correctly?

- (a) Compute heights using absolute frequencies and connect midpoints; the polygon will approximate the density as long as sample size is large.
- (b) Scale class heights by dividing frequencies by both class width and total sample size, ensuring that the total area under the polygon is approximately 1.
- (c) Use equal-width classes so the polygon is valid even if heights are not scaled as densities.
- (d) Use cumulative frequencies rather than class frequencies; cumulative polygons always track the underlying density.

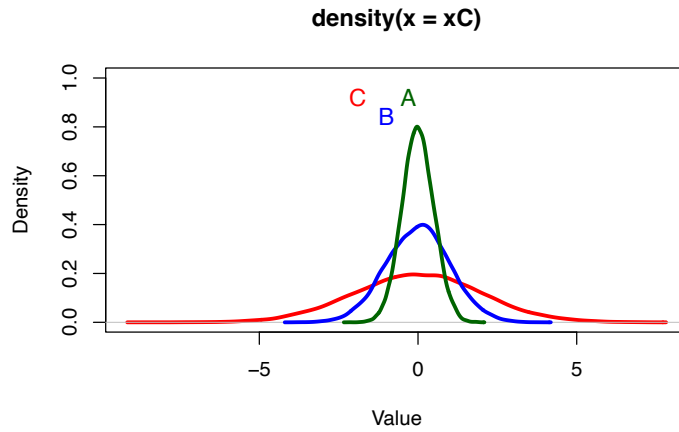
23. Below are two Lorenz curves, labelled A (green) and B (purple), based on two different income distributions.



Based on the Lorenz curves shown, which ranking of Gini indices is most plausible?

- (a) $Gini(A) > Gini(B)$: Curve A lies farther from the diagonal over a larger share of the population.
- (b) $Gini(A) < Gini(B)$: Curve B dips farther below the diagonal near the very top of the income distribution, which dominates overall inequality.
- (c) $Gini(A) = Gini(B)$: The curves cross, so the total areas between them and the diagonal cancel out exactly.
- (d) It is theoretically impossible to compare Gini indices from Lorenz curves that intersect even once.

24. The three density curves A, B, and C shown below all have the SAME mean (0) but different standard deviations.



Which matching of density curve to standard deviation is correct?

- (a) $A = 0.5, B = 1.0, C = 2.0$
- (b) $A = 2.0, B = 1.0, C = 0.5$
- (c) $A = 1.0, B = 0.5, C = 2.0$
- (d) $A = 1.0, B = 2.0, C = 0.5$

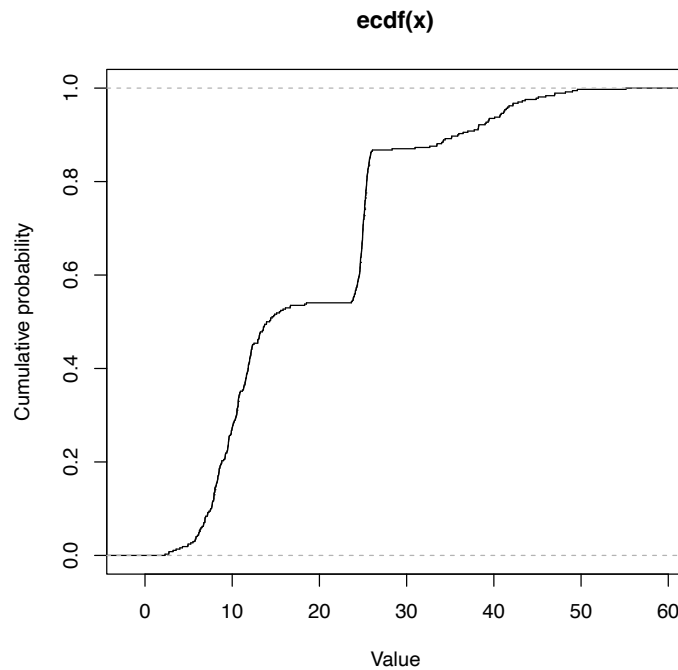
25. You are studying the relationship between study hours (X) and exam score (Y). From a large sample you have the following summary statistics:

$$\begin{aligned} \text{Var}(X) = 16, \quad \text{Var}(Y) = 100, \quad \text{Cov}(X, Y) = 28, \\ \bar{X} = 12, \quad \bar{Y} = 65. \end{aligned}$$

which of the following is the correct estimated regression line?

- (a) $\hat{Y} = 44 + 1.75X$
- (b) $\hat{Y} = 65 + 0.25X$
- (c) $\hat{Y} = 12 + 1.00X$
- (d) $\hat{Y} = 44 - 1.75X$

26. The plot below displays the empirical cumulative distribution function (CDF) for a sample of household responses. There is a pronounced sharp vertical increase (jump) in the CDF near $x \approx 25$.



Which of the following interpretations of the sharp CDF jump is most accurate?

- (a) The CDF increases rapidly there because a disproportionately large number of observations fall in a very narrow interval around that value.

- (b) The steep jump indicates that the density must be decreasing sharply in that region, causing probability mass to accumulate more quickly on the right side.
- (c) A jump of this form implies the presence of a plateau (flat segment) in the underlying density, meaning the density is nearly constant over that region.
- (d) Such a jump forces the median to be inside that interval, because a large concentration of values necessarily pins the 0.5 quantile to that location.

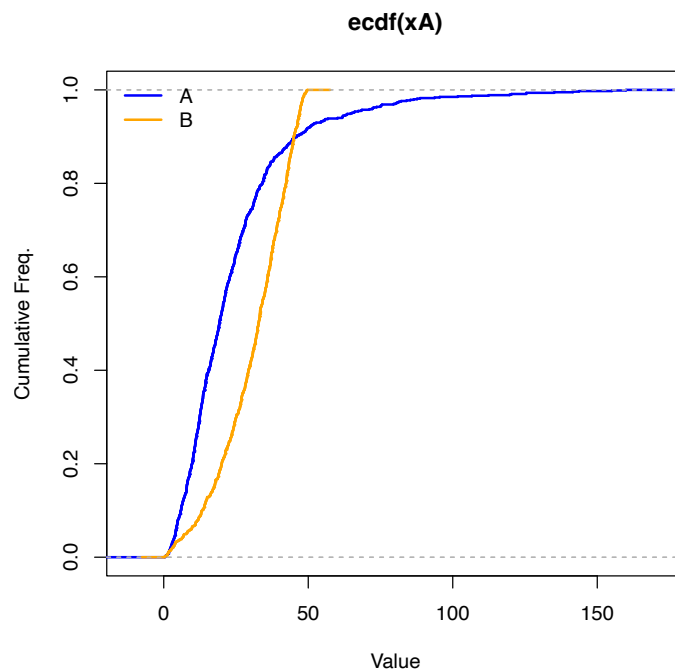
27. A statistical agency considers using a fixed-base quantity index to measure real output growth:

$$Q_{t,0} = \frac{\sum_j p_0^j q_t^j}{\sum_j p_0^j q_0^j}.$$

The index uses base-year prices as weights for all later periods. Over long horizons, relative prices change substantially and consumption/production patterns shift.

Which statement best describes a key theoretical limitation of using this fixed-base quantity index over long periods?

- (a) The index becomes dimensionally inconsistent because the ratio $\frac{q_t^j}{q_0^j}$ cannot be meaningfully compared when technology changes.
 - (b) The index tends to understate real output growth because unchanged base-year prices place too much weight on goods whose prices have risen, making real output appear smaller.
 - (c) The index may yield misleading growth estimates because base-year price weights become increasingly unrepresentative as the economy evolves, distorting measured real quantities.
 - (d) The index becomes invalid once any good disappears from the market, because all fixed-base indices require the full basket to remain available indefinitely.
28. The plot below shows two empirical cumulative distribution functions (CDFs), labelled F_A (blue) and F_B (orange). Both correspond to continuous variables measured on the same scale.



Based on the shapes of the two CDFs, which statement is most accurate regarding skewness?

- (a) Distribution A is more positively skewed, and Distribution B is more negatively skewed.
 - (b) Distribution A is more negatively skewed, and Distribution B is more positively skewed.
 - (c) Both distributions must be symmetric, since their CDFs eventually reach 1.
 - (d) Skewness cannot be deduced from CDFs; only histograms can reveal asymmetry.
29. A country's GDP (in billions) is reported as follows:

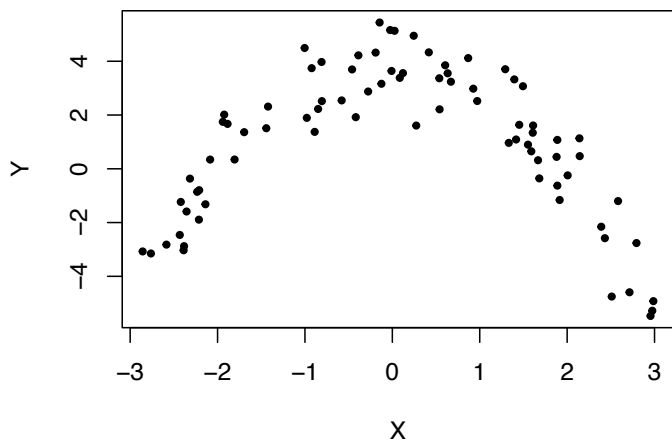
$$X_{2018} = 820, \quad X_{2019} = 874, \quad X_{2020} = 832.$$

Which of the following is the correct **total growth rate** over the two-year period?

- (a) -3.9%
- (b) +6.6%
- (c) +2.1%

(d) +1.4%

30. You are given the following scatterplot of variables X and Y . Based on this scatterplot, which statement best describes the slope β_1 of the original linear regression?



- (a) The slope is moderately positive because higher X values tend to produce higher Y values on the right side of the plot.
(b) The slope is moderately negative because lower X values tend to produce lower Y values on the left side of the plot.
(c) The slope is close to zero because the symmetric curved pattern causes positive and negative contributions to cancel in a linear fit.
(d) The slope is strongly nonzero because the relationship is clearly nonlinear.
31. Consider a dataset with mean μ , variance σ^2 , and mode m . A new variable Y is obtained from the original observations X through the transformation

$$Y = \frac{X}{2} + \frac{2}{3}.$$

Which of the following correctly describes the mean, variance, and mode of Y in terms of μ , σ^2 , and m ?

- (a) Mean = $\frac{1}{2}\mu + \frac{2}{3}$, variance = $\frac{1}{4}\sigma^2$, mode = $\frac{1}{2}m + \frac{2}{3}$.
(b) Mean = $\frac{1}{2}\mu + \frac{2}{3}$, variance = $\frac{1}{2}\sigma^2$, mode = $\frac{1}{2}m$.
(c) Mean = $\frac{1}{2}(\mu + \frac{2}{3})$, variance = $\frac{1}{4}\sigma^2$, mode unchanged.
(d) Mean = $\frac{1}{2}\mu$, variance unchanged, mode = $m + \frac{2}{3}$.
32. Suppose you have the following three annual chain index factors for successive years:

$$1.02, \quad 0.99, \quad 1.05.$$

What is the cumulative chain index over the entire period?

- (a) $1.02 + 0.99 + 1.05$
(b) $1.02 \times 0.99 \times 1.05$
(c) $\frac{1.02 + 0.99 + 1.05}{3}$
(d) $\sqrt[3]{1.02 \times 0.99 \times 1.05}$
33. A statistician reports three fixed-base index numbers for a variable X :

$$i_{t,0}, \quad i_{k,0}, \quad i_{t,k}.$$

You are told only that

$$i_{t,0} = 1.32, \quad i_{k,0} = 0.88.$$

No information about the ordering of t and k is provided.

Which of the following *could* be the correct value of the cross-index $i_{t,k}$?

- (a) $i_{t,k} = \frac{0.88}{1.32}$
(b) $i_{t,k} = \sqrt{\frac{1.32}{0.88}}$
(c) $i_{t,k} = 1.32 - 0.88$

(d) $i_{t,k} = \frac{1.32}{0.88}$

34. Which of the following statements about covariance is correct?

- (a) Covariance captures only nonlinear association; linear relationships do not affect it.
- (b) Covariance remains unchanged when either variable is multiplied by any nonzero constant.
- (c) If X and Y are independent, then their covariance is zero, although the converse need not hold.
- (d) Covariance is bounded between -1 and 1 because it standardizes deviations by their variances.

35. A dataset has mean 100 and standard deviation 20. We form

$$Z = \frac{X - 50}{5}.$$

What is $\text{Var}(Z)$?

- (a) 16
- (b) 20
- (c) 400
- (d) 25

36. A grouped dataset has the following classes and frequencies:

Class	Frequency
0–5 :	8
5–15 :	14
15–30 :	22
30–50 :	11
50 and more :	5

Assume the last class has the same width as the previous class. Which class contains the grouped median?

- (a) 5–15
- (b) 15–30
- (c) 30–50
- (d) 50+

37. Consider the Lorenz curve of an income distribution. Suppose the curve lies very close to the 45° line across its entire range, with only a slight deviation near the end.

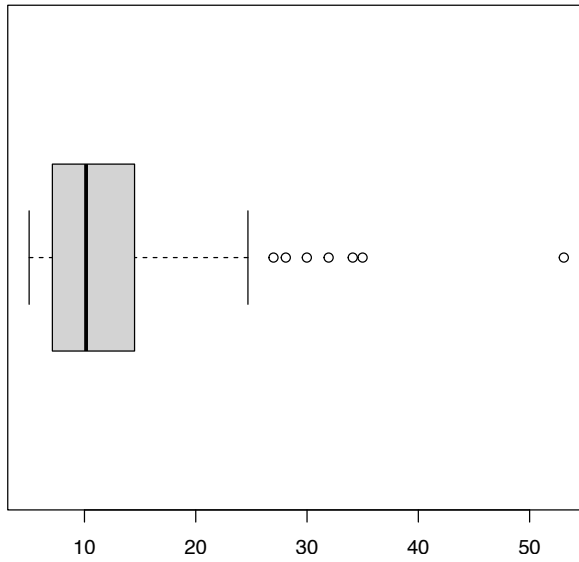
Which statement provides the most accurate interpretation?

- (a) The distribution exhibits high inequality because most income is earned by a small fraction of the population.
- (b) The distribution shows perfect equality, since the Lorenz curve coincides exactly with the diagonal.
- (c) The distribution exhibits low inequality, meaning income shares are fairly evenly distributed across the population.
- (d) The Lorenz curve alone is insufficient to judge inequality because both high and low inequality curves may appear close to the diagonal.

38. Measured GDP rises from 500 (year t) to 540 (year $t + 1$). You must report a single-period growth rate for this change. Which of the following is the annual growth rate for GDP between t and $t + 1$?

- (a) 40%
- (b) 8%
- (c) Cannot be determined (need price deflators / nominal vs real)
- (d) 7.4%

39. Consider the following box and whisker plot



Based on the shape of this boxplot, which numeric relation among mean, median, and mode is most likely?

- (a) $\text{mean} < \text{median} < \text{mode}$
- (b) $\text{mean} > \text{median} > \text{mode}$
- (c) $\text{mean} = \text{median} = \text{mode}$
- (d) $\text{median} < \text{mean} < \text{mode}$

1. Consider a continuous variable for which you want to estimate the mean, median, or mode using a grouped frequency table.

Which of the following statements is correct regarding the effect of changing the class width (size of the intervals)?

- (a) Changing the class width has no effect on the estimated statistics.
 - (b) Decreasing the class width generally improves the approximation to the true mean, median, or mode, because smaller intervals capture more detail about the distribution.
 - (c) The estimated statistics are exact for any class width if the variable is continuous.
 - (d) Increasing the class width always improves the accuracy of the estimated statistics.
2. Suppose $X_{2018} = 100$, $X_{2021} = 133$. What is the average annual rate of growth for the period?
- (a) $(133/100)^{1/3} - 1$
 - (b) $(133 - 100)/3$
 - (c) $(133/100) - 1$
 - (d) $\sqrt{133/100} - 1$

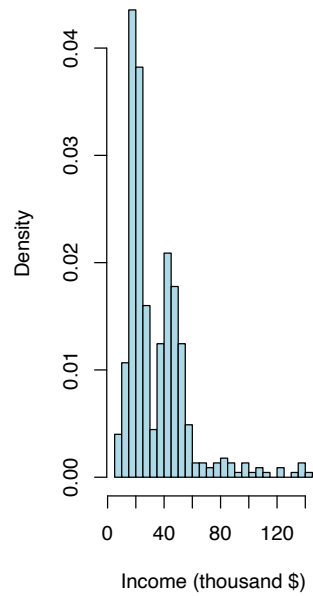
3. A statistician reports the following fixed-base indices for a country:

$$i_{2018,2014} = 1.08, \quad i_{2019,2014} = 1.12.$$

What is the growth from 2018 to 2019?

- (a) 3.7%
 - (b) 96.4%
 - (c) -4%
 - (d) 4%
4. Suppose you have a histogram representing relative frequencies. The first class has a width of 100 and a relative frequency of 0.20, while the second class has a width of 400 and a relative frequency of 0.20. Which statement about the histogram is correct?
- (a) Both bars must have the same area but different heights
 - (b) The bar of the wider class must be taller
 - (c) The wider class must have a larger area but lower height
 - (d) Both bars must have the same height
5. Fixed-base Laspeyres indices use expenditure weights from a chosen base year, applying them unchanged to all subsequent periods. Which of the following captures the main theoretical disadvantage of this approach when measuring price change over long horizons?
- (a) The index cannot be interpreted consistently across years because the ratio $\frac{p_t^j}{p_0^j}$ loses its meaning when base prices become outdated.
 - (b) The index systematically understates inflation because consumers substitute toward goods whose prices have risen the most.
 - (c) The fixed expenditure shares increasingly fail to represent actual consumption patterns as relative prices and quantities evolve, leading to bias in measured price change.
 - (d) The method requires frequent collection of up-to-date expenditure data to maintain accuracy in weighting the consumption bundle.
6. A country's unemployment rate rises from 6% to 8.5% between 2020 and 2021. Which of the following statements correctly distinguishes the percentage change from the percentage point change in unemployment?
- (a) The unemployment rate increased by 2.5 percentage points, corresponding to a percentage change of $\frac{2.5}{6} \approx 41.7\%$.
 - (b) The unemployment rate increased by 41.7%, which is also the percentage-point increase since percentage change and percentage points coincide when measured in rates.
 - (c) The unemployment rate increased by 2.5 percentage points, and therefore the percentage change is exactly 2.5%.
 - (d) The unemployment rate increased by 2.5%, corresponding to a percentage change of $\frac{8.5-6}{8.5} \approx 29.4\%$.
7. A histogram of incomes shows a single peak around \$20k, a second much smaller peak around \$45k, and a long tail extending to very high incomes. Which statement is the best interpretation?

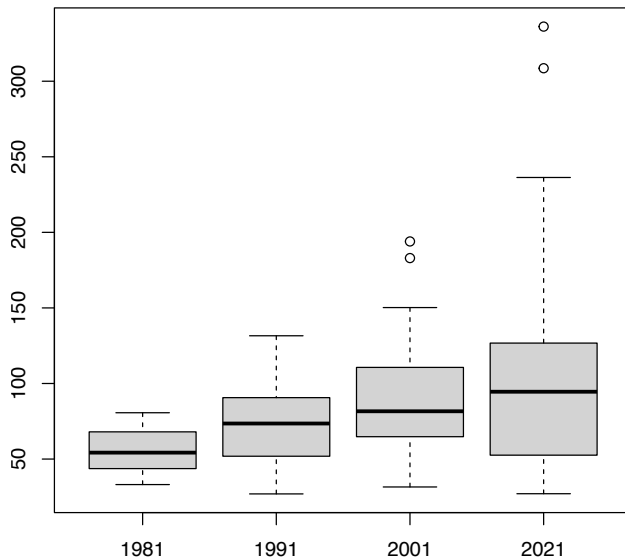
Histogram of rent



- (a) The distribution is unimodal but heavy-tailed; median > mean.
- (b) The distribution is multimodal and positively skewed (mean > median).
- (c) The distribution is unimodal and asymmetric.
- (d) The distribution must have mean = median = mode.

8. The boxplot below represents GDP values for 4 different years. Which statement best describes the change in dispersion over time?

GDP at Current Prices (Simulated)



- (a) Dispersion increases over time, as both the inter quartile range and the spread of extremes widen systematically across the four years.
- (b) Dispersion decreases over time, because the increase in median GDP outpaces the expansion of the quartiles, making the spread relatively smaller.
- (c) Dispersion remains essentially unchanged, since the medians differ but the central 50 percent stays roughly constant across years.
- (d) It is impossible to tell, because the central boxes appear similar while the whiskers vary inconsistently.

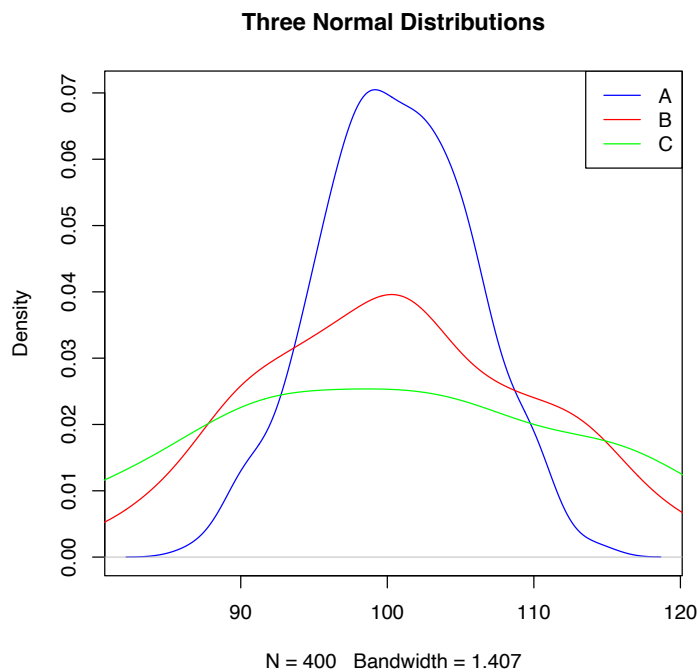
9. An economist collects the following data from a household survey:

Variable	Type
Household Income	Quantitative - Continuous
Number of Children	Quantitative - Discrete
Gender of Head of Household	Qualitative - Nominal
Education Level	Qualitative - Ordinal
Age of Head of Household	Quantitative - Continuous

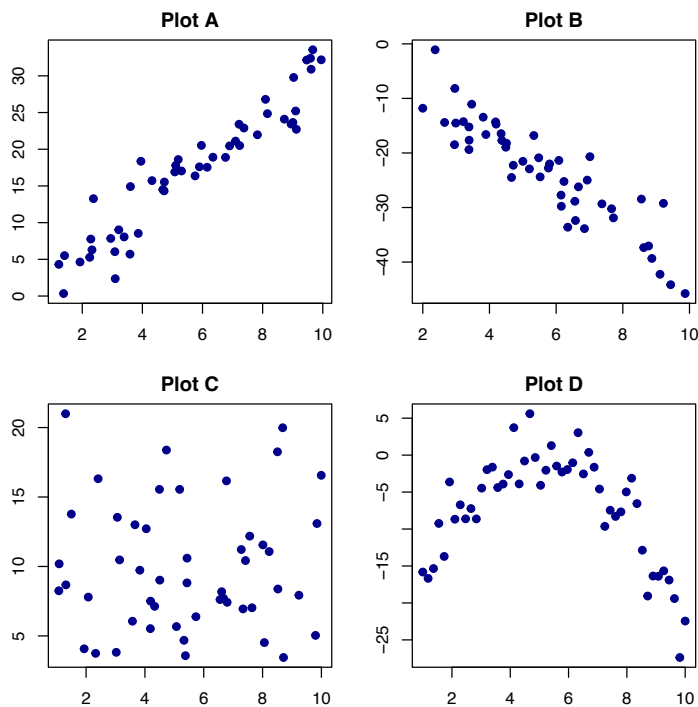
Which of the following statements about these variable types is correct?

- (a) All variable types are correctly assigned except for Household Income
- (b) Number of Children should be Qualitative - Ordinal, not Quantitative
- (c) Gender of Head of Household should be Quantitative - Nominal, not Qualitative
- (d) All variable types are correctly assigned

10. The density curves below show three distributions with different dispersions but similar means. Which distribution has the largest standard deviation?



- (a) Impossible to tell
 - (b) C
 - (c) B
 - (d) A
11. Suppose you add a constant c to every observation and then multiply all observations by a positive constant k . How do the mean and variance change?
- (a) Both mean and variance increase.
 - (b) Mean increases; variance increases.
 - (c) Mean increases; variance stays the same.
 - (d) Mean stays the same; variance increases.
12. Given grouped data with class midpoints MP_j and frequencies f_j you compute the grouped mean \bar{x}_g . Which of the following statements is true regarding the error between \bar{x}_g and the true sample mean \bar{x} computed from raw data (i.e., the one with the entire dataset and not grouped data)?
- (a) The grouped mean \bar{x}_g is always equal to the true mean.
 - (b) The grouped mean error does not depend on the distribution shape.
 - (c) The grouped mean is unbiased regardless of starting and ending point of classes.
 - (d) The error $|\bar{x}_g - \bar{x}|$ increases as class widths increase, all else equal.
13. Which plot shows the strongest positive linear association?



- (a) Plot C, because the cloud is highly scattered and centred, indicating a strong linear trend.
 (b) Plot A, because the points follow a tight upward-sloping linear pattern, indicating a strong positive correlation.
 (c) Plot B, because the downward trend is strong and indicates a large positive association.
 (d) Plot A, because it shows the smallest variance in the X-direction, making the relationship perfectly linear.

14. The following table shows the number of pages read per week by 60 students:

Pages	Relative Frequency
0–10	0.10
10–25	0.25
25–40	0.30
40–60	0.20
60–80	0.15

Compute the approximate mode of pages read.

- (a) 27
 (b) 33
 (c) 29
 (d) 31
15. A report provides the following index values for a country:

$$i_{2019,2015} = 1.20, \quad i_{2017,2015} = 1.10, \quad i_{2019,2017} = 1.05.$$

Based on the properties of index numbers, are these three values internally consistent?

- (a) No, because the implied growth from 2015 to 2019 is too large relative to the intermediate growth rates.
 (b) Yes, because the difference between 1.20 and the implied value is small enough to be due to rounding.
 (c) Yes, because only fixed-base indices must satisfy multiplicative consistency.
 (d) No, because the index for 2019 relative to 2017 should be smaller than the index for 2019 relative to 2015.
16. A correlation coefficient close to zero indicates:
- (a) That the data or model is likely misspecified, because meaningful economic relationships rarely produce correlations near zero.
 (b) That the variables are unrelated in every sense, since zero correlation rules out any form of dependence.
 (c) A very weak linear association, although other types of dependence may still exist.
 (d) A hidden nonlinear relationship must be present, because correlation detects only linear effects.
17. In the regression model $Y_i = b_0 + b_1 X_i + e_i$, what is the intuitive role of the residual e_i in ordinary least squares?

- (a) It is the difference between the observed value and the value predicted by the line, and OLS chooses the line that makes these differences as small as possible.
 (b) It measures how much X_i causes Y_i to change.
 (c) It is the part of Y_i that the model assumes is exactly zero.
 (d) It is the predicted value \hat{Y}_i produced by the regression line.

18. A dataset shows the distribution of monthly incomes in a city. The distribution is positively skewed. Which of the following best explains why the mean is greater than the median?
- (a) The median is always larger than the mean in a positively skewed distribution.
 - (b) Observations are evenly spread, so mean and median coincide.
 - (c) Most observations are clustered on the right, with a few very low values pulling the mean to the left.
 - (d) Most observations are clustered on the left, with a few very high values pulling the mean to the right.

19. The following table shows monthly study hours of a sample of 60 students:

Hours	Frequency (f_j)	Relative Frequency
0-5	4	0.067
5-15	10	0.167
15-25	18	0.300
25-35	15	0.250
35-50	13	0.217

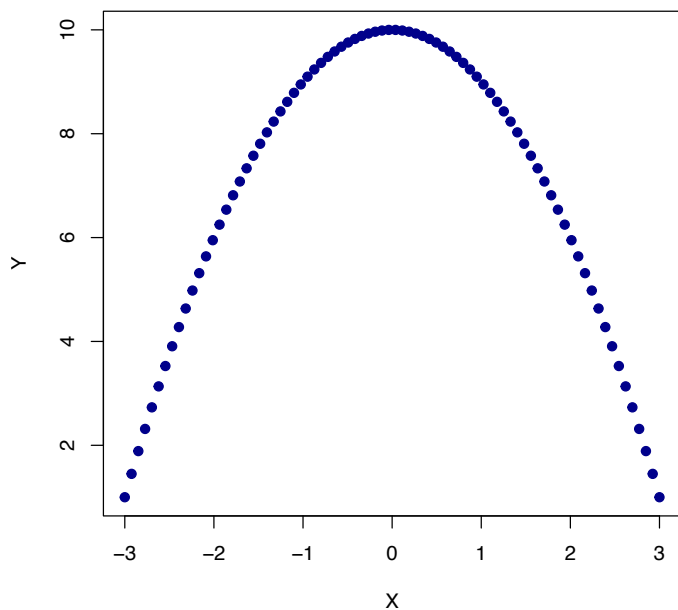
Based on this data, which of the following statements is correct?

- (a) Mean hours = 22, Median hours = 24, Modal class = 15-25
 - (b) Mean hours = 25, Median hours = 26, Modal class = 35-50
 - (c) Mean hours = 18, Median hours = 20, Modal class = 0-5
 - (d) Mean hours = 20, Median hours = 22, Modal class = 25-35
20. An index number compares a value in period t to a base value 0. Which statement is correct?
- (a) If $i_{t,0} = 1.12$, then X_t is 12% higher than X_0 .
 - (b) If $i_{t,0} = 1.12$, then X_0 is 12% higher than X_t .
 - (c) If $i_{t,0} = 112$, then X_t is 112 times larger than X_0 .
 - (d) If $i_{t,0} < 1$, the level of X_t must be negative.
21. Which of the following class definitions violates the rules of constructing valid classes for continuous variables?
- (a) $[0, 100)$, $[100, 200)$, $[200, 300)$
 - (b) $[0, 100)$, $[100, 200)$, $[180, 250)$
 - (c) $(0, 100]$, $(100, 200]$, $(200, 300]$
 - (d) $[0, 50]$, $(50, 100]$, $(100, 200]$
22. Let X_t^{nom} denote a variable measured in current prices and X_t^{real} the same variable expressed in constant prices using an arbitrary base year 0. Which of the following statements correctly identifies the economic meaning of the ratio

$$\frac{X_t^{nom}}{X_t^{real}} ?$$

- (a) It measures changes in physical quantities by stripping out all price effects.
 - (b) It captures both price and quantity changes simultaneously, reflecting the total value index.
 - (c) It represents the implicit price level required to convert real magnitudes into nominal ones, i.e. a price deflator.
 - (d) It is numerically equivalent to a Laspeyres price index regardless of the deflation method used.
23. Which statement correctly describes a property of the Gini index?
- (a) It equals zero only when exactly half the population receives all the income.
 - (b) It cannot be computed from grouped data.
 - (c) It always increases when all incomes increase by the same absolute amount.
 - (d) It is scale invariant: multiplying all incomes by a constant does not change the Gini index.

24. Below is a scatterplot of a variable Y against X .



Suppose you nonetheless the following regression model

$$Y_i = \beta_0 + \beta_1 X_i + u_i.$$

Based on the scatterplot above, which statement about the estimated slope $\hat{\beta}_1$ is most accurate?

- (a) $\hat{\beta}_1$ is undefined, because nonlinear data make the sample covariance S_{XY} meaningless.
- (b) $\hat{\beta}_1$ must be negative, because the variance of Y decreases as X moves away from its mean.
- (c) $\hat{\beta}_1$ must be positive, because S_{XY} must be positive when X has positive variance.
- (d) $\hat{\beta}_1$ will be close to zero, because the covariance between X and Y is close to zero in this symmetric quadratic pattern.

25. The following table shows the weekly hours of study for a sample of students:

Hours	Relative Frequency (f_j/N)
0–5	0.10
5–10	0.16
10–20	0.24
20–30	0.30
30–40	0.20

What of the following options is the first quartile.

- (a) 7
- (b) 9.5
- (c) 10
- (d) 8.5

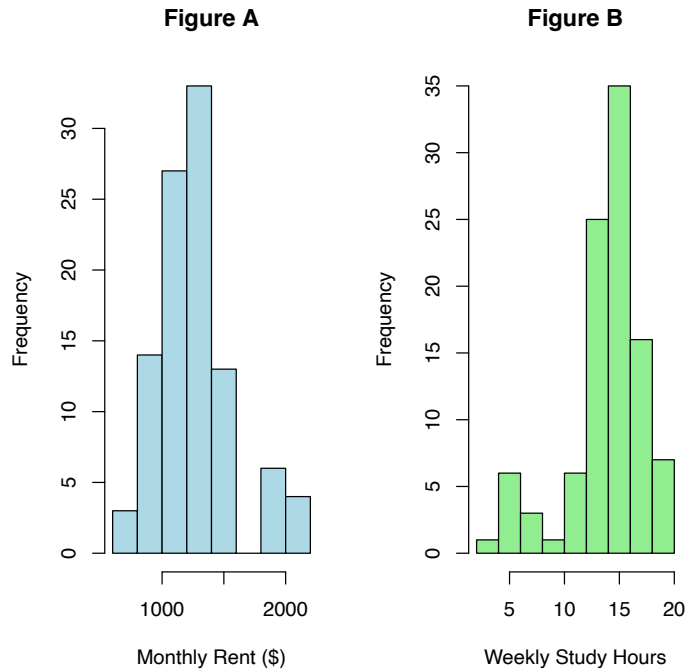
26. A grouped frequency table with equal-width classes has midpoints MP_j and relative frequencies f_j . The estimated mean is $(\bar{x} = \sum f_j MP_j)$. Which of the following changes will not affect this estimated mean?

- (a) Changing class boundaries slightly while keeping midpoints equal.
- (b) Adding a constant c to every individual observation.
- (c) Merging two adjacent classes into one wider class (combining frequencies and computing new midpoint).
- (d) Changing the width of every class but keeping the same midpoints and frequencies.

27. Which statement about the relationship between the Lorenz curve and the Gini index is correct?

- (a) If one Lorenz curve lies entirely below another, its Gini index must be higher.
- (b) A higher Gini index necessarily means the Lorenz curve is closer to the equality line.
- (c) Gini indices cannot be compared across samples of different sizes.
- (d) Two Lorenz curves that cross always imply identical Gini indices.

28. The following two figures show distributions of different variables for a sample of 100 observations each:



Which of the following statements is correct regarding the asymmetry of the distributions?

- (a) Figure A shows negative asymmetry; Figure B shows positive asymmetry.
- (b) Both figures show positive asymmetry.
- (c) Figure A shows positive asymmetry; Figure B shows negative asymmetry.
- (d) Both figures show negative asymmetry.

29. A variable Y has sample mean 69.62 and sample standard deviation 7.85. Define $Z = (Y - \text{mean}(Y))/3$. What is approximately the $\text{Var}(Z)$?

- (a) 6.844
- (b) $\text{Var}(Z) = 0$
- (c) 184.785
- (d) 61.595

30. A time series has the following chain components:

$$i_{2016,2015} = 1.02, \quad i_{2017,2016} = 0.99, \quad i_{2018,2017} = 1.05.$$

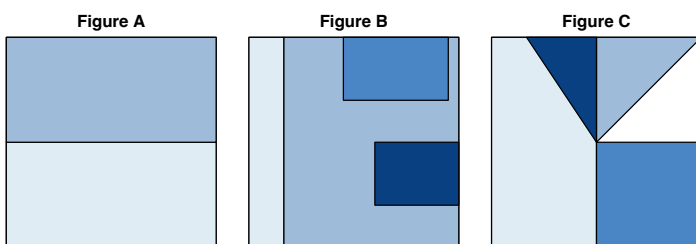
Compute the index $i_{2018,2015}$.

- (a) 1.059
- (b) 1.020
- (c) 1.040
- (d) 1.116

31. Below are three square diagrams representing land ownership patterns A, B and C. Each square has the same size and is divided into several land parcels of different sizes. Three Gini indices of land distribution are:

$$I_1 = 0.359, \quad I_2 = 0, \quad I_3 = 0.425.$$

Match each land pattern (A,B,C) to the correct index (I1,I2,I3).



- (a) $A \rightarrow I_3, B \rightarrow I_1, C \rightarrow I_2$
- (b) $A \rightarrow I_1, B \rightarrow I_2, C \rightarrow I_3$
- (c) $A \rightarrow I_2, B \rightarrow I_3, C \rightarrow I_1$

(d) $A \rightarrow I2, B \rightarrow I1, C \rightarrow I3$

32. In statistics, a population refers to

- (a) Only the largest observations in a dataset
- (b) The set of repeated measurements of the same individual
- (c) A subset extracted from a larger group
- (d) The entire group about which information is wanted

33. Aggregate growth for a group of countries is computed by combining individual growth rates using appropriate weights. Which of the following statements correctly describes this aggregation?

- (a) Aggregate growth must weight each country's growth rate by its share in the aggregate (e.g., GDP share) in the base period.
- (b) Aggregate growth must weight each country's growth rate by its share in the most recent period, because the new period reflects the updated composition of the group.
- (c) Aggregate growth is obtained by taking the simple average of individual growth rates whenever all countries have similar sizes.
- (d) Aggregate growth equals the weighted average of absolute changes divided by the weighted average of levels, so the weights cancel out and the result is the simple arithmetic average.

34. Suppose the exam scores (out of 100) of a class of 10 students are:

60, 65, 70, 72, 75, 78, 80, 85, 90, 95

Now, imagine that each student's score suddenly increases by 10 points due to a grading adjustment. Which of the following statements is correct?

- (a) The mean increases by 10, the variance remains the same, and the mode increases by 10
- (b) The mean increases by 10, the variance doubles, and the mode remains the same
- (c) The mean remains the same, variance increases, and the mode increases
- (d) The mean increases by 10, the variance increases by 10, and the mode remains the same

35. Which of the following statements correctly justifies the use of the average annual rate of growth over a multi-year period?

- (a) It replaces each annual growth rate with the relative change $\frac{X_{t+i}}{X_t} - 1$, ensuring the same proportional change every year.
- (b) It averages the annual rates by summing the k one-year growth rates and dividing the total by k .
- (c) It computes the multi-year change as $X_{t+k} - X_t$ and spreads this difference equally across years, assuming linear growth.
- (d) It finds the constant annual rate which, when compounded for k years, reproduces the total change from X_t to X_{t+k} .