Suggested Answer for Extra Practice

1) F, In a right-skewed distribution, extreme high values pull the arithmetic mean to the right, making the mean typically greater than the median.
2) T
3) F, Variance depends on the squared measurement unit and scale; relative dispersion across different units requires a scale-free measure like the Coefficient of Variation
4) T
5) F. Laspeyres indices hold weights fixed at *base-period* values; Paasche indices hold weights fixed at current-period values.

Part B

(20) a) Distinguish between nominal variables and ordinal variables. Provide one concrete business or economic example of each type.

Qualitative variables whose categories possess a natural, meaningful ordering or ranking. Example: prices. Nominal variables have no inherent order or ranking. Example: customer satisfaction levels.

(20) b) Define Pearson's Linear Correlation Coefficient (r_{xy}) between two quantitative variables. State its mathematical range of values and explain what a value of $r_{xy} = -0.85$ indicates regarding the strength and direction of the linear relationship.

Pearson's Linear Correlation Coefficient (r_{xy}) measures the strength and direction of the linear relationship between two quantitative variables x and y: $r_{xy} = -0.85$ means strong negative linear relationship.

(30) c) Starting from the definition of the sample variance of y, $s_y^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$, and using the simple linear regression decomposition $y_i - \bar{y} = (\hat{y}_i - \bar{y}) + e_i$:

Show step-by-step how total sample variance splits into explained variance ($s_{\hat{y}}^2$) and residual variance (s_e^2):

$$s_y^2 = s_{\hat{y}}^2 + s_e^2$$

(Hint: Use the key mathematical property that $\sum_{i=1}^{n}(\hat{y}_i - \bar{y})\, e_i = 0$.)

Substitute the decomposition $y_i - \bar{y} = (\hat{y}_i - \bar{y}) + e_i$ into the sum:

$$s_y^2 = \frac{1}{n}\sum_{i=1}^{n}[(\hat{y}_i - \bar{y}) + e_i]^2$$

Expand the squared term:

$$s_y^2 = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \frac{2}{n}\sum_{i=1}^{n}(\hat{y}_i - \bar{y})\, e_i + \frac{1}{n}\sum_{i=1}^{n} e_i^2$$

Use the Hint to remove the middle term:

$$s_y^2 = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \frac{1}{n}\sum_{i=1}^{n} e_i^2$$

$$s_y^2 = s_{\hat{y}}^2 + s_e^2$$

Part C

(40) **a)** A human resources management team studied the distribution of employee weekly working hours (in hours) across 100 staff members, summarized in the following grouped frequency table:

Working Hours (Interval)	**Number of Employees (nj)**
$[20,30)$	15
$[30,40)$	60
$[40,50)$	25
Total	**100**

Compute the sample mean ($\bar{x}$) and the sample median (Me) for these grouped data. Using these two values, comment briefly on the symmetry of the distribution of working hours.

Sample Mean ($\bar{x}$)

1. Midpoints (m_j) and Sample Mean ($\bar{x}$)

- $[20,30)$: $m_1 = 25, n_1 = 15 \Longrightarrow n_1 m_1 = 375$
- $[30,40)$: $m_2 = 35, n_2 = 60 \Longrightarrow n_2 m_2 = 2100$
- $[40,50)$: $m_3 = 45, n_3 = 25 \Longrightarrow n_3 m_3 = 1125$
- Total: $\sum n_j\, m_j = 375 + 2100 + 1125 = 3600$, $N = 100$

$$\bar{x} = \frac{\sum n_j\, m_j}{N} = \frac{3600}{100} = \mathbf{36.0} \text{ hours}$$

Sample Median (Me)

The median position is $\frac{N}{2} = 50$. Looking at cumulative frequencies:

- Class $[20,30)$: $F_1 = 15$
- Class $[30,40)$: $F_2 = 75 \Longrightarrow$ Median lies in interval $[30,40)$.

$$\text{Me} = l + \frac{0.5 - F^*(l)}{F^*(L) - F^*(l)} \cdot h$$

Class $[20,30)$: $f_1 = \frac{15}{100} = 0.15 \Longrightarrow F^*(30) = 0.15$

Class $[30,40)$: $f_2 = \frac{60}{100} = 0.60 \Longrightarrow F^*(40) = 0.15 + 0.60 = 0.75$

$$\text{Me} = 30 + \frac{0.5 - 0.15}{0.75 - 0.15} \cdot 10 = 30 + \frac{0.35}{0.60} \cdot 10 = 30 + \frac{3.5}{0.6} \approx \mathbf{35.83} \text{ hours}$$

(40) **b)** An economic research institute tracks an industrial production index using historical data. Due to an institutional update, the base year of the index series needs to be re-referenced from Year 0 to Year 2 ($b = 2$). The original index series with base Year 0 ($I_{t|0}$) is given below:

Period (t)	**Year 0**	**Year 1**	**Year 2**	**Year 3**	**Year 4**
Original Index ($I_{t\|0}$)	100.0	115.0	125.0	140.0	150.0

Shift the base of the entire index series to Year 2 ($b = 2$) so that $I_{2|2} = 100.0$. Calculate the new re-referenced index values ($I_{t|2}$) for all periods $t = 0,1,2,3,4$. Compute the percentage change in industrial production between Year 2 and Year 4 using both the original index ($I_{t|0}$) and your newly re-referenced index ($I_{t|2}$). Comment briefly on whether changing the base year affects the implied growth rate between these two periods.

Given $I_{2|0} = 125.0$:

- **Year 0 ($t = 0$):** $\frac{100.0}{125.0} \times 100 = \mathbf{80.00}$
- **Year 1 ($t = 1$):** $\frac{115.0}{125.0} \times 100 = \mathbf{92.00}$
- **Year 2 ($t = 2$):** $\frac{125.0}{125.0} \times 100 = \mathbf{100.00}$
- **Year 3 ($t = 3$):** $\frac{140.0}{125.0} \times 100 = \mathbf{112.00}$
- **Year 4 ($t = 4$):** $\frac{150.0}{125.0} \times 100 = \mathbf{120.00}$

. Growth Rate Comparison (Year 2 to Year 4)

- **Using $I_{t|0}$:** $\frac{150.0-125.0}{125.0} \times 100\% = \frac{25}{125} \times 100\% = \mathbf{20.00}\%$
- **Using $I_{t|2}$:** $\frac{120.00-100.00}{100.00} \times 100\% = \frac{20}{100} \times 100\% = \mathbf{20.00}\%$

Changing the base year rescales the entire index series proportionally, so it **does not affect** the growth rate between any two given periods.