Economics and Business Information

Formulae

Descriptive Statistics

Frequency tables

Suppose the values observed for a given variable X for statistical units *i* (i=1,…, n) are represented by: X_1, X_2, X_3,…, X_n

a. Discrete (ungrouped) data:

- Absolute frequency (F_j) – Number of times each value of variable X is observed (j=1,2,…k).
- Cumulative absolute frequency - $cum\ F_j = \sum_j F_j = N$
- Relative frequency (f_j) – Proportion of times each value of the variable X is observed: $f_j=F_j/N$
- Cumulative relative frequency - $cum\ f_j = \sum_j f_j = \sum_j \frac{F_j}{N} = 1$

b. Grouped data:

- *j* denominates class j^{th}, with j=1,2,3…,m
- Class width: $a_j=l_j - l_{j-1}$
- Class midpoint: $MP_j= (l_j + l_{j-1})/2$ or $MP_j= l_{j-1} + (l_j - l_{j-1})/2$
- Frequency density: $h_j = F_j/a_j$ or $h_j = f_j/a_j$

Measures of location

Arithmetic mean:	
Discrete data	$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$ $\bar{x} = \frac{1}{n}\sum_{j=1}^{m} F_j x_j = \sum_{j=1}^{m} f_j x_j$
Grouped continuous data	$\bar{x} = \frac{1}{n}\sum_{j=1}^{m} F_j MP_j = \sum_{j=1}^{m} f_j MP_j$
Median:	
Discrete data	Uneven number of observations: $x_{Me} = x_{\frac{n+1}{2}}$ Even number of observations: $x_{Me} = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2}$
Grouped continuous data	$x_{Me} = l_{j-1}(Me) + \frac{0.5 - cum\, f(Me-1)}{f(Me)} a(Me)$ where: - $l_{j-1}(Me)$: lower limit of the median class

	- *cum f(Me-1):* cumulative frequency of the class before the median class - *f(Me)*: relative frequency of the median class - *a(Me)*: width of the median class
Mode:	
Grouped continuous data	$x_{Mo} = l_{j-1}(Mo) + \frac{f(Mo+1)}{f(Mo-1)+f(Mo+1)} a(Mo)$
Quantiles: Q_1, Q_3	
Grouped continuous data	$x_{Q_1} = l_{j-1}(Q_1) + \frac{0.25 - cum\, f(Q_1 - 1)}{f(Q_1)} a(Q_1)$ $x_{Q_3} = l_{j-1}(Q_3) + \frac{0.75 - cum\, f(Q_3 - 1)}{f(Q_3)} a(Q_3)$

Measures of dispersion

Interquartile range:	IRQ=Q_3-Q_1
Mean deviation:	
Discrete data	$MD_x = \frac{\sum_{i=1}^{n} \lvert x_i - \bar{x} \rvert}{n}$
Grouped continuous data	$MD_x = \frac{\sum_{j=1}^{m} n_j \lvert MP_j - \bar{x} \rvert}{n} = \sum_{j=1}^{m} f_j \lvert MP_j - \bar{x} \rvert$
Standard deviation:	
Discrete data	$S_x = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}}$
Grouped continuous data	$S_x = \sqrt{\frac{\sum_{j=1}^{m} n_j (MP_j - \bar{x})^2}{n}} = \sqrt{\sum_{j=1}^{m} f_j (MP_j - \bar{x})^2}$
Variance:	
Discrete data	$S_x^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}$
Grouped continuous data	$S_x^2 = \frac{\sum_{j=1}^{m} n_j (MP_j - \bar{x})^2}{n} = \sum_{j=1}^{m} f_j (MP_j - \bar{x})^2$
Relative interquartile range:	$RIQR = \frac{IQR}{Q_2} = \frac{Q_3 - Q_1}{Q_2} = \frac{Q_3 - Q_1}{x_{ME}}$
Coefficient of variation (CV):	$CV_x = \frac{S_x}{\bar{x}}$

Measures of concentration

Gini Index	$GI = \frac{\sum_{j=1}^{m-1} (cumf_j(x) - cum\, f_j(y))}{\sum_{j=1}^{m-1} cumf_j(x)} = \frac{\sum_{j=1}^{m-1} (p_j - q_j)}{\sum_{j=1}^{m-1} p_j} = 1 - \frac{\sum_{j=1}^{m-1} q_j}{\sum_{j=1}^{m-1} p_j}$

Symmetry / Asymmetry of the distribution

- mean=median=mode: symmetrical distribution
- mean>median>mode: positive asymmetry – skewed to the left
- mode>median>mean: negative asymmetry – skewed to the right

Rates of change and indices

Absolute change	$\Delta X_{t+k,t} = X_{t+k} - X_t, \qquad with\ k = 1,\ 2,$ … (time periods)
Mean absolute change	$m\Delta X_{t+k,t} = \frac{X_{t+k} - X_t}{k}, \qquad with\ k = 1,\ 2,$ … (time periods)
Relative change – rate of change or rate of growth	$r_{t+k,t} = \frac{X_{t+k} - X_t}{X_t} = \frac{\Delta X_{t+k,t}}{X_t} = \frac{X_{t+k}}{X_t} - 1,$
Average relative change or rate of change	$mr_{t+k,t} = \left(\frac{X_{t+k}}{X_t}\right)^{\frac{1}{k}} - 1$ $= \left(1 + r_{t+k,t}\right)^{\frac{1}{k}} - 1$ $= \left[\left(1 + r_{t+1,t}\right)\left(1 + r_{t+2,t+1}\right) \ldots \left(1 + r_{t+k,t+k-1}\right)\right]^{\frac{1}{k}}$-1
Year-on-year rate of change	$h_{t,s} = \frac{x_{t,s} - x_{t-1,s}}{x_{t-1,s}}, \qquad with\ t\ the\ year\ and\ s\ the\ period$

Simple Indices

Consider a time series for variable X between years 0 and t, X_1, X_2, X_3,…, X_t:

Chain index	$i_{1,0} = \frac{X_1}{X_0}, i_{2,1} = \frac{X_2}{X_1}, i_{3,2} = \frac{X_3}{X_2} \ldots, i_{t,t-1} = \frac{X_t}{X_{t-1}}$
Fixed base index	$i_{1,0} = \frac{X_1}{X_0}, i_{2,0} = \frac{X_2}{X_0}, i_{3,0} = \frac{X_3}{X_0} \ldots, i_{t,0} = \frac{X_t}{X_0}$
Relationship between indices and rates of change	Chain index: $i_{t,t-1} = \left(1 + r_{t,t-1}\right)$ Fixed base index: $i_{t,0} = \left(1 + r_{t,0}\right)$
Properties of indices:	

Circularity	$i_{t,0} = i_{t,t-1} * \dots * i_{3,2} * i_{2,1} * i_{1,0}$
Rebasing	$i_{t,b} = \frac{i_{t,0}}{i_{b,0}}$ *because* $\frac{\frac{x_t}{x_0}}{\frac{x_b}{x_0}} = \frac{x_t}{x_b}$
Reversibility	$i_{t,0} = \frac{1}{i_{0,t}}$ *because* $\frac{x_t}{x_0} = \frac{1}{\frac{x_0}{x_t}}$

Composite or aggregate indices

Index of value	Value Index = Price Index*Quantity Index $I_{value} = \frac{\sum p_t.q_t}{\sum p_o.q_o} = I_{prices} * I_{quantity} = L_{t,0}^{P} * P_{t,0}^{Q} = P_{t,0}^{P} * L_{t,0}^{Q}$
Laspeyres Price Index $L_{t,0}^{P} = \frac{\sum_{j=1}^{m} p_t^j \cdot q_0^j}{\sum_{j=1}^{m} p_0^j \cdot q_0^j}$	**Laspeyres Quantity Index** $L_{t,0}^{Q} = \frac{\sum_{j=1}^{m} p_0^j \cdot q_t^j}{\sum_{j=1}^{m} p_0^j \cdot q_0^j}$
Laspeyres indices as the weighted average of simple indices $L_{t,0}^{P} = \sum_{j=1}^{m} w_0^j \frac{p_t^j}{p_0^j}$, $L_{t,0}^{Q} = \sum_{j=1}^{m} w_0^j \frac{q_t^j}{q_0^j}$, with $w_0^j = \frac{p_0^j \cdot q_0^j}{\sum_{j=1}^{m} p_0^j \cdot q_0^j}$	
Paasche Price Index $P_{t,0}^{P} = \frac{\sum_{j=1}^{m} p_t^j \cdot q_t^j}{\sum_{j=1}^{m} p_0^j \cdot q_t^j}$	**Paasche Quantity Index** $P_{t,0}^{Q} = \frac{\sum_{j=1}^{m} p_t^j \cdot q_t^j}{\sum_{j=1}^{m} p_t^j \cdot q_0^j}$

Association and linear relation between variables

Covariance between X and Y	$S_{YX} = \frac{\sum_{j=1}^{N} (x_j - \bar{X})(y_j - \bar{Y})}{n}$
Linear correlation coefficient between X and Y	$r_{YX} = \frac{S_{YX}}{S_X S_Y} = \frac{\frac{1}{n}\sum_{j=1}^{n}(x_j - \bar{x})(y_j - \bar{y})}{\sqrt{\frac{1}{n}\sum_{j=1}^{n}(x_j - \bar{x})^2 \frac{1}{n}\sum_{j=1}^{n}(y_j - \bar{y})^2}}$
Regression line	$Y_i=b_0+b_1X_i+\varepsilon_i$
Regression parameters	$b_0 = \bar{Y} - b_1\bar{X}$ $b_1 = \frac{S_{YX}}{S_X^2}$