|  |  |
| --- | --- |
| Basic definitions  Coding For Classes Class interval: Midpoint of the class  Class boundary: Midpoint between the upper and lower boundaries of 2 different classes Interpolation to find quartiles Find the nth value using  Find the quartile using  A&B = top & bottom class boundary for class that pos lies on  Af & Bf = Accumulated frequency of A & B | Basic CQT |
| Charts and graphsHistograms ***Chart*** {w: class unit, h: fd}  ***Bar*** {w: class interval, h: fd}  Be careful about units and scale! Skewness Lean left → +ve skew  Lean right → -ve skew  The formulas of skewness are often given in the question  🡪 skew arrow direction (+ve) Comparing data  1. Measure of location    * Median Q2    * Skewness 2. Measure of spread    * IQR Q3 – Q1    * Standard deviation    * Range Max – Min   **Use Q2 & IQR if data is skewed!** (More resistant to outliers) | Chart CQT |
| Probability Sum of probabilities must equal to 1   |  |  |  |  | | --- | --- | --- | --- | |  |  | A & B | Intersect | |  |  | A or B | Union | |  |  | not A | Complement |   Probability for something to occur twice = Conditional probabilities For probability of B given that A occurred: Event types Independent:  Mutually exclusive: Tree diagrams   When calculating probabilities of events, make sure to consider each repetition of cases very carefully. For example, when selecting objects, the total number of objects reduces by one. Make sure to read questions very carefully and consider if the case even exists. | Probability CQT |
| Data correlation & predictions For a scatter diagram:  If data is linearly displaced → Strong  If data displaced with +ve slope → Positive  Positive correlation means that x & y increase & decrease together Product moment correlation coefficient Magnitude → Linear characteristic  +veness → Positive correlation  Practical meaning of r: when x increases y increase / decrease  Value of r is not affected by coding, but PMCC between coded data and value changes depending on the slope of the code.  An absolute relationship have a PMCC of Regression lines The regression line y on x predicts values for y  Interpolation predicts values within data range, otherwise it’s extrapolation and the prediction are less reliable  General formula: | Correlation CQT Find the estimated value of y when x = …  Comment on the validity of your answer   * Check if is extrapolation * If not, check if the model applies to the estimated quantity |
| Random variablesProbability distribution Random variables 🡪 X  Possible outcomes of X 🡪 x  Probability that X has value of x:   |  |  |  |  | | --- | --- | --- | --- | |  | 1 | 2 | 3 | |  | 0.1 | 0.3 | 0.6 |  Expected values Expected value is the mean outcome of infinite observations      Variance of X:   |  |  |  |  | | --- | --- | --- | --- | |  | 1 | 2 | 3 | |  | 0.1 | 0.3 | 0.6 | |  | 0.1 | 0.6 | 1.8 | |  |  | | | |  | 0.1 | 1.2 | 5.4 | |  |  | | | |  |  | | |  Cumulative distribution Cumulative distribution has the probabilities carried on   * First possible value of x P(x) = F(x) * Final possible value of x F(x) = 1  |  |  |  |  | | --- | --- | --- | --- | |  | 1 | 2 | 3 | |  | 0.1 | 0.3 | 0.6 | |  | 0.1 | 0.1 + 0.3 = 0.4 | 1 |  Function of X For   |  |  |  |  | | --- | --- | --- | --- | |  | 1 | 2 | 3 | |  | 1a + b | 2a + b | 3a + b | |  | 0.1 | 0.3 | 0.6 | |  | 0.1 | 0.3 | 0.6 |  Nested conditions For   |  |  |  |  | | --- | --- | --- | --- | |  | 1 | 2 | 3 | |  | 1 | 4 | 9 | |  | 0.1 | 0.3 | 0.6 | |  |  | | |   *It always helps to draw the distribution table!* | RandVars CQT Find the probability distribution of X given F(x)   * Find values and draw table:  |  |  |  |  | | --- | --- | --- | --- | |  | … | … | … | |  | … | … | … | |
| Normal distribution  * Symmetrical * Total area = 1  Standardizing normal distribution to ZNested cases Draw a graph of the standard distribution (with calculator or not) to best understand what area of value the question is asking for or considering. | NormDist CQT |