

STATISTICS WORKSHEET 1

- 1) A (True)
- 2) A (Central Limit Theorem)
- 3) C (Modeling bounded count data)
- 4) D (All of the mentioned)
- 5) C (Poisson)
- 6) B (False)
- 7) B (Hypothesis)
- 8) A (0)
- 9) C (Outliers cannot conform to the regression relationship)
- 10) The normal distribution is the most widely known and used of all distributions. Because the normal distribution approximates many natural phenomena so well, it has developed into a standard of reference for many probability problems.
- 11) a: mean or median imputation b: multivariate imputation by chained equation c: random forest
- 12) A/B testing, also known as split testing, refers to a randomized experimentation process wherein two or more versions of a variable (web page, page element, etc.) are shown to different segments of website visitors at the same time to determine which version leaves the maximum impact and drive business metrics.
- 13) Bad practice in general. If just estimating means: mean imputation preserves the mean of the observed data. Leads to an underestimate of the standard deviation. Distorts relationships between variables by “pulling” estimates of the correlation toward zero
- 14) If we want to use a variable x to draw conclusions concerning a variable y : y is called dependent or response variable. x is called independent, predictor, or explanatory variables. If a relationship between two variables is linear it can be summarized by a straight line A straight line can be describe by an equation ' $a+bx$ '
- 15) Descriptive statistics and inferential statistics