|  |  |
| --- | --- |
| **Mean:**  Mean = n x p  **Median:**  1) Line up all the numbers.  2) Median = (n+1)/2 th position = 10.5 position = (85+85)/2 = 85%  First quartile: median of the first half of ordered data  Second Quartile: just typical median.  Third Quartile: median of the second half of the ordered data  Interquartile range (IQR) = upper quartile – lower quartile  negative skewness: median will exceed the mean.  positive skewness: median will be less than mean.  **Standard Deviation:**  Standard Deviation Formula    Use sample mode in calculator when dealing with standard deviation.  **Z-score – How many standard deviations away from mean**  Z = (y - µ)/ σ  Once z score is found you can look at percentages from the chart, always skew lower. We estimate conservatively.  Unstandardizing: y = zσ + µ  **Correlation:**  R2 – % of variation in y that can be attributed to x. Just square r.  Least squares regression:  ŷ = b0 + b1 x # x is the variable being measured. b0 = intercept, b1 = slope. Y = response, x = explanatory  Let y represent …. and x represent ….  **Residual:**  e = y - ŷ  residuals usally ploted against explanatory var. ???What is this exactly???  e – Residual – How much the prediction misses the actual observation by.  If a residual plots points are spread over graph, its good and indicates linear | **General:**  68-95-99.7  Response variable: variable that is affected.  Blocking: Occurs when giving causal thing, is a part of experiment design. Ie placebo vs real.  Stratifying: Occurs at sampling, ie male vs female.  Observational studies cannot determine cause and effect, only well designed experimental can.  **Probabilities:**  P(A or B) = P(a) + P(b)  P(A and B) = P(a) x P(b) #assuming a and b are independent  Expected Value: x1p1 + x2p2 + x3p3 …  **Binomial:**  P(X > 10) =  n = number  k= number of successes.  SPIN: Success condition, Prob of success, independence, number of trials fixed.  If the question asks for probability of greater than or less than, add up the probabilities from the binomial table.  **Central Limit Theorem:**  An SRS will have:   * Normal distribution * Sample Mean = Population Mean. * Std deviation = σ/sqrt(n)   **Confidence Intervals:**  Z for proportion, t for mean.  Margin of error will be smaller if confidence interval is??  df = n – 1  find critical value based on df and confidence interval.  **Sample Size:**  Grab z value from t-table for percentages. |

|  |  |
| --- | --- |
| **Hypothesis:**  H0 , HA  We use t for means and z for proportions. t is used when we don’t know population std deviation.  **Sign Test:**  Binomial situation.  Test statistic = number of plus  n = 9  p = probability of success = 0.5 for sign test, always. Cross off 0s.  Is a non-parametric test so spell out hypothesis in words.  p-value = probability of getting ‘x’ successes at 0.5 on the binomial chart.  If no level of significance, we tend to assume 5%.  **Testing Proportions:**  np ≥ 10 AND nq ≥ 10  1) Random sample  2) 10% condition (sample is < 10%)  3) Independence  4) np >= 10, nq >= 10 (If this is not met use non-parametric test for hypothesis). # these are values in null hypothesis.  Let p be the proportion …  **Testing Means:**  df = n-1  1) Random sample  2) 10% condition (sample is < 10%)  3) Independence  4) Population is nearly normally distributed. (If this is not met use non-parametric test for hypothesis).  **Unpaired Matching:**  H0 : µcurrent = µnew model Ha : µcurrent < µnew model  df = smaller (n-1)  1) SRS  2) 10% condition (sample is < 10%)  3) Independence (each individual is independent of another individual)  4) Nearly normal (we assume the population of differences are normally distributed).  5) Data is paired. IE same individual measured twice. | **Paired Matching:**  d̅ = difference between …. Average difference???  Define sub d separately.  df = nd - 1 # if there are 12 before and afters, its 12 - 1.  **Chi Squared:**  Expected count = (row total x column total)/table total  df = (r-1) x (c-1) where r is the no. of rows and c is the no. of columns  H0 = A and B are not associated  Ha = A and B are associated  Is a non-parametric test so spell out hypothesis in words.  1) SRS -  2) Independence -  3) 10% condition - sample is less than 10% of population  4) Expected counts >= 5 - expected counts at least 5.  **Type I vs Type II:**  In the case of a Type I error the null hypothesis is rejected, but is actually true.  In the case of a type II error the null hypothesis is accepted, but is actually false.  **P-value:**  After finding the test value search for the p-value. This will typically be representative of percentage probability. In general the following holds true for P-values:  >10% – insufficient evidence to support Ha  5-10% – slight evidence to support Ha  1-5% – moderate evidence to support Ha  0.1 – 1% – strong evidence to support Ha  < 0.1% – very strong evidence to support Ha  Level of Significance: if p-value is greater than level of significance we do not reject the null hypothesis.  “Sufficient and insufficient evidence to reject accept” |