**I. Basic probability formulas**

* P(A ⋃ B) = P(A) + P(B) - P(A ⋂ B)
* P(A | B) =
* P(A | B) =
* If A, B independent: P(A ⋂ B) = P(A) . P(B)

**II. Discrete random variables**

* = E(x) = xi . P(x=xi)
* 2 = V(x) = (xi - )2 . P(x=xi)

= xi2 . P(x=xi) - 2

* E(ax + by) = a.E(x) + b.E(y)
* V(ax + by) = a2 . V(x) + b2 . V(y)
* Probability mass function: f(xi) = P(x=xi)
* Cumulative distribution function: F(xi) = P(xxi)
* Some special distribution:

1. Discrete uniform distribution
   * P(x=Xi) =
   * =
   * 2 =
2. Binomial distribution
   * P(x=k) = nCk . pk . (1-p)n-k
   * = n.p
   * 2 = n.p . (1-p)
3. Poisson distribution
   * P(x=k) = (.T)k
   * = .T
   * 2 = .T
4. Hypergeometric distribution
   * P(x=k) =
   * = n.p
   * 2 = n.p.(1-p).
5. Geometric distribution
   * P(x=k) = (1-p)k-1 . p
   * =
   * 2 =
6. Negative binomial distribution
   * P(x=k) = (k-1)C(r-1) . pr . (1-p)k-r
   * =
   * 2 =

**III. Continuous random variable**

* Probability density function f(x): P(a<x<b) = f(x) dx
* Cumulative distribution function F(x):
  + F(xi) = P(xxi)
  + F(xi)’ = f(xi)
* = E(x) =
* E(xn) =
* 2 = V(x) = -
* Some special distribution:

1. Continuous uniform distribution
   * f(x) = , a x b

= 0 , elsewhere

* + =
  + 2 =

1. Normal distribution N(, 2)
   * z =
   * f(z) =
   * (x) = p(z<xi)
   * (-x) = 1 - (x)
2. Normal distribution approximate binomial and poisson distribution
3. Binomial (np > 5 and n(1-p) > 5)
   * + z =
     + P(XBINORM  a) = P(XNORMAL  a+0.5)
     + P(XBINORM  a) = P(XNORMAL  a-0.5)
4. Poisson
   * + z =
     + P(XPOISSON  a) = P(XNORMAL  a+0.5)
     + P(XPOISSON  a) = P(XNORMAL  a-0.5)
5. Exponential distribution
   * f(x) = . , x > 0
   * = 0 , elsewhere
   * P(x a) = ,(a > 0)
   * =
   * 2 =

**IV. Descriptive statistic** (Take a sample of size n from population N)

* Sample mean: =
* Sample median: L = so Median =
* Mode: Số phần tử xuất hiện nhiều nhất
* Range: max - min
* Sample variance: s2 =
* Quatiles:
  + L1 = so Q1 =
  + L2 = so Q2 =
  + L3 = so Q3 =

**V. Sampling distribution**

* Population mean , variance 2. Sample size n. *(Normal distribution or n > 30)*:
  + Phân phối của có dạng: N( , )
  + Phân phối của - có dạng: N(1 - 2 , + )
* For proportion of population p, sample size n. *(np ≧ 5 or n.(1-p) ≧ 5)*:
  + Phân phối của có dạng: N( , )
  + Phân phối của - có dạng: N( - , + )

**VI. Statistical intervals - Test claims for one sample**

* (l, u) = ( - E, + E)
* width = 2E
* P-value = 2 . P(Z > |Z0|)

1. Population variance known
   * E = .
   * =
2. Population variance unknown
   * n > 30:
     + E = .
     + =
   * n 30:
     + E = .
     + =

* For propotion:
  + (l, u) = ( - E, + E)
  + E = .
  + =
  + Nếu đề không cho , mặc định = 0.5
* Nếu là one-side thì tương tự nhưng thay thành

**VII. Test claims for 2 samples** (2 population independent, normal distribution or both n1, n2 > 30)

* (l, u) = (- - E , - + E)

1. Population variance known
   * E = .
   * =
2. Population variance unknown
   * Assume 2 = 2
     + Degree of freedom: df = + + 2
     + =
     + E = .
     + =
   * Not assume 2 = 2
     + Degree of freedom: df =
     + E = .
     + =

* For propotion:
  + (l, u) = ( - - E , - + E)
  + E = .
  + = (trong đó xi = n . )
  + =

**VIII. Linear Regression**

* SXY = =
* SXX = =
* SYY = =
* Slope: = =
* Intercept: = - .
* Error sum of square: SSE =
* Regression sum of square: SSR =
* Total sum of square: SST =
* SSE + SSR = SST
* Standard error: =
* Coefficient of correlation: R = =
* Test claims about the slope *(df = n-2)*:
  + se() =
  + =
* Test claims about the intercept *(df = n-2)*:
  + se() =
  + =
* Test claims about the coefficient of correlation *(df = n-2)*: =

