MTH304 (Statistics) Summary Sheet Jeff Mair

$\bar{X}$, S are mean & std dev of a SAMPLE, while...
μ, σ are mean & std dev of the POPULATION

Section 1 - Descriptive Statistics

$n = \sum f$

Sample Mean (avg): $\bar{x} = \dfrac{\sum f \cdot x}{n}$

Sample Variance: $s^2 = \dfrac{\sum f \cdot (x_i - \bar{x})^2}{n-1}$

Sample Std Dev: $\sqrt{s^2}$

The mean of a given freq. distribution is: x-bar +/- std dev

NOTE: $\sum f \cdot (x_i - \bar{x})^2 = \sum f \cdot (x_i^2) - \dfrac{(\sum f \cdot x_i)^2}{n}$

Coding: $u = \dfrac{x - A}{B}$, $\bar{u} = \dfrac{\bar{x} - A}{B}$, $s_x^2 = B^2 \cdot s_u^2$

Note: when converting variance back from coded, need to mult by B^2, but no need to translate by A

Section 2 - Probability

$P(A) = \dfrac{n(A)}{n(s)}$

Complimentary event A: 1 - P(A)

Or => $P(A \cup B)$ ***** two cases*****

If A & B are mutually exclusive events:

$P(A \cup B) = P(A) + P(B)$

Else:

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

And => $P(A \cap B)$ ***** two cases*****

If A & B occur at the same time **INDEPENDENT:**

$P(A \cap B) = P(A) \cdot P(B)$

Else:

Conditional Probability

$P(B/A) = \dfrac{P(A \cap B)}{P(A)}$ $\qquad P(A/B) = \dfrac{P(A \cap B)}{P(B)}$

$\Rightarrow P(A) \cdot P(B/A) = P(B) \cdot P(A/B)$

Methods of Enumeration

1. Listing
2. Tree diagram
3. Permutations (order is significant)... **P -> Postion!!**
4. Combinations (order is not significant)
5. Generalized Permutations

Permutation ${}_nP_k = \dfrac{n!}{(n-k)!}$

Combination ${}_nC_k = \dfrac{n!}{k!(n-k)!}$

Generalized Permutation ${}_nP_{n1\ n2\ n3 \ldots} = \dfrac{n!}{n_1! \cdot n_2! \cdot n_3! \ldots \cdot n_k!}$

Discrete Distributions

If N is given, can choose HYPERGEOMETRIC or BINOMIAL distributions. If no N, can only use BINOMIAL
(n is sample size, N is population size)

$\mu = \sum x \cdot Pr(x)$, $\qquad \sigma^2 = \sum x^2 \cdot Pr(x) - \mu^2$

Types:

1. 'No name' (coins)
2. Hypergeometric $P(x=k) = \dfrac{\binom{X}{k} \cdot \binom{N-X}{n-k}}{\binom{N}{n}}$
3. Binomial $P(x=k) = \binom{n}{k} \cdot p^k \cdot q^{n-k}$

   Also: $\mu = n \cdot p$, and $\sigma^2 = n \cdot p \cdot q$

   If n > 30, approximate with Normal or Poisson

   if $n \cdot p \geq 5$ use Normal, else use Poisson

   Normal $\rightarrow$ $\binom{n}{k} \cdot p^k \cdot q^{n-k} \approx \int \dfrac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{\frac{-(x-\mu)^2}{2\sigma^2}}$

   Poisson $\rightarrow$ $\lambda t = \mu = n \cdot p$
4. Poisson $P(x=k) = \dfrac{(\lambda t)^k \cdot e^{-\lambda t}}{k!}$ where λt = rate

   Also: $\mu = \lambda t$ and $\sigma^2 = \lambda t$
5. Geometric

Continuous Distributions

Where f(x) is PDF:

$\mu = \int x \cdot f(x)\, dx$

$\sigma = \int x^2 \cdot f(x)\, dx - \mu^2$

1. Normal Distribution, 'z' coding: $z = \dfrac{x - \mu}{\sigma}$
2. Exponential Distribution: $\lambda e^{-\lambda x}$

   When the rand var 'x' represents a continuous quantity (like time), must use Exponential, NOT Poisson.
3. T-Distribution
4. F-Distribution

Confidence Interval of 'True mean'

Confidence interval: % probability **(1 - α)** that the population mean lies within a particular interval on the distribution curve

$\pm Z_{\alpha/2} = \dfrac{\bar{x} - \mu}{\sigma}$, where $Z_{\alpha/2}$ is the z-value of the prob/2.

Large Sample (>= 30):

=> Central Limit Theorem applies, so the samples are said to be Normally distributed

Memorize variations!

Binomial:

$z = \dfrac{\hat{p} - P}{\sqrt{\dfrac{P \cdot Q}{n}}}$; $\quad z = \dfrac{\hat{p}_1 - \hat{p}_2 - (P_1 - P_2)}{\sqrt{\dfrac{P_1 \cdot Q_1}{n} + \dfrac{P_2 \cdot Q_2}{n}}}$

Small Sample (< 30)
CLT not valid, but sample follows distribution of pop
If pop dist not given, must state assumption that it is conts!! (and ND?)
Case I: σ known: => sample is ND, so use Z dist, same as LS
Case II: σ unknown: => approx with S, sample is T-dist, with d.f. n-1
See cases I, II, III of $\mu_1 - \mu_2$

$$\pm t_{\alpha/2} = \frac{\bar{x} - \mu}{S/\sqrt{n}}$$

Two Populations
If σ ? $\sigma_a = \sigma_2$, and can't guess:
Use F-Dist, with Hypothesis Test
If h_0 -> $\sigma_1 = \sigma_2$
Then use Pooled Variance
Else: use T-Dist with long fmla

Chi-Squared Distribution dist of variances

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$

Hypothesis Testing
Step 1 - Decision Rule
Step 2 - Statistic (find where 'z' lands on the curve)
Step 3 - Conclusion
If random var needs to be 'equal' to a value, that must be part of the hypothesis (H_0). The opposite direction is H_1.

F-Distribution

$$F = \frac{S_1^2}{S_2^2} \quad \text{s.t. } S_1 > S_2$$

In reading the table, take $\mathbf{1 - \alpha}$, and find $\boldsymbol{v_1}$ in horizontal and $\boldsymbol{v_2}$ in the vertical