STAT 2593 Formula Sheet

Measures of Location and Variability

Mean: $\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1+x_2+\ldots+x_n}{n}$

Median: If n is even then $\tilde{x} = \frac{(\frac{n}{2})^{\text{th}} \text{ obs.} + (\frac{n+1}{2})^{\text{th}} \text{ obs.}}{2}$

If n is odd then $\tilde{x} = \frac{n+1}{2}^{\text{th}}$ obs.

Range: $x_{\text{max}} - x_{\text{min}}$

Variance: $s^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1} = \frac{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2}{n-1}$

IQR: IQR = Q3 − Q1

Discrete Random variables

Probability Mass Function (pmf): $p(x) = P(X = x)$

1. $1 \geq p(x) \geq 0$ for all x in X
2. $\sum_x p(x) = 1$

Cumulative Distributive Function (cdf): $F(x) = P(X \leq x) = \sum_{k \leq x} f(k)$

Mean (μ): $E(X) = \sum_x x f(x)$

Expected value: $E(g(X)) = \sum_x g(x) f(x)$

Variance (σ^2): $\text{var}(X) = \sum_x (x-\mu)^2 f(x) = E(X^2) - [E(X)]^2$

SD (σ): $SD(X) = \sqrt{\text{var}(X)}$

Basic Probability Rules

- *General Addition Rule:* $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- *Complement Rule:* $P(A^c) = 1 - P(A)$
- $P(\emptyset) = 0$ and $P(S) = 1$
- $0 \leq P(A) \leq 1$ for all A
- $P(A|B) = \frac{P(A\cap B)}{P(B)}$ and $P(B|A) = \frac{P(A\cap B)}{P(A)}$
- *Multiplication Rule:* $P(A \cap B) = P(B) \times P(A|B) = P(A) \times P(B|A)$
- *Law of Total Probability:* $P(B) = P(B|A_1)P(A_1) + \cdots + P(B|A_k)P(A_k)$ where $A_1, \ldots, A_k$ form a partition of the sample space.
- $A \perp B$ if and only if $P(A \cap B) = P(A)P(B)$
- *Bayes' Theorem:* $P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{i=1}^{n} P(A_i)P(B|A_i)}$

Counting Rules

- *Product rule for counting:* n_j choices for decision j; $n_1 \times \cdots \times n_k$ total choices.
- Permutations are given by $P_{k,n} = \frac{n!}{(n-k)!}$.
- Combinations are given by $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.

Continuous Random variables

Probability Density Function (pdf): $P(a \leq X \leq b) = \int_a^b f(x)dx = F(b) - F(a)$

1. $f(x) \geq 0$ for all x
2. $\int_{-\infty}^{\infty} f(x)dx = 1$

Cumulative Distributive Function (cdf): $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$

Median: x such that $F(x) = 0.5$

Q_1 and Q_3: x such that $F(x) = 0.25$ and x such that $F(x) = 0.75$

Mean (μ): $E(X) = \int_{-\infty}^{\infty} x f(x)dx$

Expected value: $E(g(X)) = \int_{-\infty}^{\infty} g(x)f(x)dx$

Variance (σ^2): $\text{var}(X) = \int_{-\infty}^{\infty}(x-\mu)^2 f(x)dx = E(X^2) - [E(X)]^2$

SD (σ): $SD(X) = \sqrt{\text{var}(X)}$

Expectation and Variance Rules

- $E(aX + b) = aE(X) + b$, for $a, b \in \mathbb{R}$
- $E(X + Y) = E(X) + E(Y)$, for all pairs of X and Y
- $\text{var}(aX + b) = a^2 \text{var}(X)$, for $a, b \in \mathbb{R}$
- $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y)$, for independent X and Y

Uniform and Exponential

$X \sim \text{Unif}(a, b)$

Mean: $\mu = E(X) = \frac{a+b}{2}$

Variance: $\sigma^2 = \text{var}(X) = \frac{(b-a)^2}{12}$

pdf of X

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

cdf of X

$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases}$$

$X \sim \text{Exp}(\lambda)$

Mean: $\mu = E(X) = \frac{1}{\lambda}$

Variance: $\sigma^2 = \text{var}(X) = \frac{1}{\lambda^2}$

pdf of X

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

cdf of X

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Normal Distribution

$X \sim N(\mu, \sigma^2)$
Standard Normal: $Z \sim N(0,1)$ where $Z = \frac{X-\mu}{\sigma}$; cdf denoted as $\Phi(z)$
Empirical Rule:

- approximately 68% of observations fall within σ of μ
- approximately 95% of observations fall within 2σ of μ
- approximately 99.7% of observations fall within 3σ of μ

Bernoulli and Binomial Random variables

$X \sim \text{Bernoulli}(p)$
pmf: $P(X=x) = p^x(1-p)^{1-x}$ for $x = 0, 1$

Mean: $\mu = E(X) = p$
Variance: $\sigma^2 = \text{var}(X) = p(1-p)$

$X \sim \text{Bin}(n, p)$
pmf: $P(X=x) = \binom{n}{x} p^x (1-p)^{n-x}$ for $x = 0, 1, 2, ..., n$

Mean: $\mu = E(X) = np$
Variance: $\sigma^2 = \text{var}(X) = np(1-p)$

Geometric Distribution

$X \sim \text{Geo}(p)$
pmf: $P(X=x) = p(1-p)^{x-1}$ for $x = 1, 2, 3, ...$

Mean: $\mu = E(X) = \frac{1}{p}$
Variance: $\sigma^2 = \text{var}(X) = \frac{1-p}{p^2}$

Hypergeometric Distribution

$X \sim \text{Hypgeo}(N, M, n)$
pmf: $P(X=x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$ for $x = \max\{0, n-N+M\}, \ldots, \min\{n, M\}$.

Mean: $\mu = E(X) = n\frac{M}{N}$
Variance: $\sigma^2 = \text{var}(X) = \frac{N-n}{N-1} \cdot n \cdot \frac{M}{N} \cdot \left(1 - \frac{M}{N}\right)$

Negative Binomial Distribution

$X \sim \text{NB}(r, p)$
pmf: $P(X=x) = \binom{x+r-1}{r-1} p^r (1-p)^x$ for $x = 0, 1, 2, 3, ...$
Mean: $\mu = E(X) = \frac{r(1-p)}{p}$
Variance: $\sigma^2 = \text{var}(X) = \frac{r(1-p)}{p^2}$

Poisson Distribution

$X \sim \text{Poi}(\lambda)$ pmf: $P(X=x) = \frac{\lambda^x e^{-\lambda}}{x!}$ for $x = 0, 1, 2, 3, ...$
cdf: $P(X \le x) = \sum_{i=0}^{x} \frac{\lambda^i e^{-\lambda}}{i!}$ for $x = 0, 1, 2, 3, ...$

Mean: $\mu = E(X) = \lambda$
Variance: $\sigma^2 = \text{var}(X) = \lambda$

Central Limit Theorem

Let X_1, X_2,..., X_n be a random sample from an arbitrary population/distribution with mean μ and Variance σ^2. When n is large ($n \ge 30$) then

$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{n} \sim N(\mu, \frac{\sigma^2}{n})$, approx.

Point Estimators

- Bias $= E[\widehat{\theta}] - \theta$.
- $\text{MSE}(\widehat{\theta}) = E[(\widehat{\theta} - \theta)^2]$
- $\text{MSE}(\widehat{\theta}) = \text{Bias}(\widehat{\theta})^2 + \text{var}(\widehat{\theta})$

Confidence Interval for Mean

General Form: point estimate $\pm$ margin of error
When σ^2 is **known**: $\overline{x} \pm z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$
When σ^2 is **unknown** (Z): $\overline{x} \pm Z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$
When σ^2 is **unknown** (t): $\overline{x} \pm t_{\frac{\alpha}{2}, n-1} \frac{s}{\sqrt{n}}$

Pooled Variance

Requires assumptions that population Variances are equal: $\sigma_1^2 = \sigma_2^2 = \sigma^2$
$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$

Critical Values

- $\Phi(-2.58) = 0.005$;
- $\Phi(-1.96) = 0.025$;
- $\Phi(-1.645) = 0.05$;
- $\Phi(1.645) = 0.95$;
- $\Phi(1.96) = 0.975$;
- $\Phi(2.58) = 0.995$