STAT 211 – Midterm 1 Formula Cheat Sheet

Chapter 1: Statistical Concepts

Population vs Sample

Population Parameter (true value, unknown)

$$\mu, \quad \sigma^2, \quad \sigma, \quad p$$

Sample Statistic (estimate of parameter)

$$\bar{x}, \quad s^2, \quad s, \quad \hat{p}$$

Population Mean

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$$

Sample Mean

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Population Variance

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$$

Population Standard Deviation

$$\sigma = \sqrt{\sigma^2}$$

Sample Variance (Unbiased)

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

Sample Standard Deviation

$$s = \sqrt{s^2}$$

Linear Transformation Property

$$\mathrm{Var}(\alpha + \beta X) = \beta^2 \mathrm{Var}(X)$$
$$\mathrm{SD}(\alpha + \beta X) = |\beta| \mathrm{SD}(X)$$

Median & Quartiles

Sample Median

Ordered data:

$$\text{If } n \text{ odd: } x_{\frac{n+1}{2}}$$

$$\text{If } n \text{ even: } \frac{x_{n/2} + x_{n/2+1}}{2}$$

Quartile Position Formula

$$h = (n-1)\frac{k}{4} + 1$$

Interpolation:

$$Q_k = x_{\lfloor h \rfloor} + (h - \lfloor h \rfloor)(x_{\lceil h \rceil} - x_{\lfloor h \rfloor})$$

Interquartile Range

$$IQR = Q_3 - Q_1$$

Outlier Rules

Mild Outlier:

$$x < Q_1 - 1.5IQR \quad \text{or} \quad x > Q_3 + 1.5IQR$$

Extreme Outlier:

$$x < Q_1 - 3IQR \quad \text{or} \quad x > Q_3 + 3IQR$$

Proportions

Population Proportion

$$p = \frac{N_i}{N}$$

Sample Proportion

$$\hat{p} = \frac{n_i}{n}$$

Histograms – Bin Rules

Square Root Rule (No. Bins)

$$k \approx \sqrt{n}$$

Freedman–Diaconis Rule

$$\text{Bin Width} = 2 \times IQR \times n^{-1/3}$$

Chapter 2: Probability

Probability Axioms

$$0 \le P(A) \le 1$$

$$P(S) = 1$$

$$P(A_1 \cup A_2 \cup \ldots) = \sum P(A_i) \quad \text{(disjoint)}$$

Complement Rule

$$P(A^c) = 1 - P(A)$$

Addition Rule (General)

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

Addition Rule (Disjoint)

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

Difference Rule

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

Subset Rule

$$A \subset B \Rightarrow P(A) \le P(B)$$

Classical Probability

If outcomes equally likely:

$$P(A) = \frac{\text{number of favorable outcomes}}{\text{number of total outcomes}}$$

Conditional Probability

Definition

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (\text{if } P(B) > 0)$$

Bayes' Theorem

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Multiplication Rule

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

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

Independence

Definition

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

$$P(A|B) = P(A)$$

Set Operations

$$A \cup B \quad A \cap B \quad A^c$$

$$(A \cup B)^c = A^c \cap B^c$$

$$(A \cap B)^c = A^c \cup B^c$$

Chapter 3: Distributions

Random Variables

Discrete or Continuous

$$X = \text{random variable}$$

Discrete

$$f(x) = P(X = x) \quad \sum_x f(x) = 1$$

$$F(x) = P(X \le x)$$

$$E(X) = \sum_x x f(x)$$

$$\text{Var}(X) = E(X^2) - [E(X)]^2$$

Binomial

$$X \sim \text{Binomial}(n, p)$$

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$E(X) = np$$

$$\text{Var}(X) = np(1-p)$$

Continuous

$$f(x) = \frac{d}{dx} F(x)$$

$$f(x) \ge 0$$

$$\int_{-\infty}^{\infty} f(x)\, dx = 1$$

$$P(a < X < b) = \int_a^b f(x)\, dx$$

$$F(x) = P(X \le x)$$

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

$$\text{Var}(X) = E(X^2) - [E(X)]^2$$

Uniform

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

$$f(x) = \frac{1}{b-a}$$

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

$$\text{Var}(X) = \frac{(b-a)^2}{12}$$

Normal

$$X \sim N(\mu, \sigma^2)$$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$Z = \frac{X - \mu}{\sigma}$$

$$x_p = \mu + z_p \sigma$$

$$IQR = 2(0.67449)\sigma \approx 1.349\sigma$$

Empirical Rule (Normal Distribution)

Approximately:

68% within 1σ

95% within 2σ

99.7% within 3σ

Effect of Linear Transformation on Expectation and Variance

$$E(\alpha + \beta X) = \alpha + \beta E(X)$$

$$\text{Var}(\alpha + \beta X) = \beta^2 \text{Var}(X)$$