Stat 151 Midterm Formula Sheet

Descriptive Statistics

- Sample Mean:
  $$\bar{y} = \frac{y_1 + y_2 + \ldots + y_n}{n} = \frac{\sum y_i}{n}$$

- Sample Variance
  $$s^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1} = \frac{\sum_{i=1}^{n} y_i^{\,2} - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}}{n-1}$$

- Sample Standard Deviation:
  $$s = \sqrt{\text{sample variance}} = \sqrt{s^2}$$

- Range = max – min

- Interquartile Range (IQR) = $Q_3 - Q_1$

- 5 # Summary: min, Q_1, median, Q_3, max

- Outliers:
  Lower fence = $Q_1 - 1.5\times$IQR
  Upper fence = $Q_3 + 1.5\times$IQR

- z-score: $z = \frac{y - \mu}{\sigma}$

- Empirical Rule:
  Empirical rule: For a relatively bell-shaped data set, approximately 68%, 95%, 99.7% of the measurements are within one, two and three standard deviations of the mean, respectively.

Probability Theory

- Addition Rule:
  P(A or B) = P(A U B) = P(A) + P(B) – P($A \cap B$)

- Complement Rule: P(A^c) = 1 – P(A)

- Multiplication Rule (general):
  P(A and B) = P($A \cap B$)
  = P(A) × P($B|A$)
  = P(B) × P($A|B$)

- Multiplication Rule for **Independent** Events:
  If A and B are independent, then
  P(A and B) = P(A ∩ B) = P(A) × P(B)

- Conditional Probability of A given B, if P(B) > 0 :
  $$\text{P}(A|B) = \frac{\text{P}(A \cap B)}{\text{P}(B)}$$

- Probability of A:
  P(A) = P($A \cap B$) + P($A \cap B^c$)

Population Distributions

- The mean (expected value) of a discrete random variable: $\mu = E(X) = \sum xp(x)$

- The variance of a discrete random variable:
  $$\sigma^2 = Var(X) = \sum (x-\mu)^2\, p(x)$$

- The standard deviation of a discrete random variable: $\sigma = \sqrt{\sigma^2}$

- For two random variables X and Y, and constants a, b, and c:
  - $E[aX + b] = aE[X] + b$
  - $V[aX + b] = a^2 V[X]$
  - $E[aX + bY + c] = aE[X] + bE[Y] + c$
  - $V[aX + bY + c] = a^2 V[X] + b^2 V[Y]$, if X and Y are independent.

Stat 151 Midterm Formula Sheet

Normal Distribution

$Y \sim Normal(\mu, \sigma), \quad -\infty < y < \infty$

- $P(a < Y < b) = P\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right)$

Sampling Distributions

- Sampling Distribution of a Sample Proportion, $\hat{p}$:
  - $Mean(\hat{p}) = \mu_{\hat{p}} = p$
  - $SD(\hat{p}) = \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$
  - For large n ($np \geq 10$ and $n(1-p) \geq 10$),

    $\hat{p} \dot{\sim} N\left(\mu_{\hat{p}} = p, \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}\right)$

- Sampling Distribution of a Sample Mean, $\bar{y}$:
  - $Mean(\bar{y}) = \mu_{\bar{y}} = \mu$
  - $SD(\bar{y}) = \sigma_{\bar{y}} = \frac{\sigma}{\sqrt{n}}$
  - If $y \sim N(\mu, \sigma)$, then

    $\bar{y} \dot{\sim} N\left(\mu_{\bar{y}} = \mu, \sigma_{\bar{y}} = \frac{\sigma}{\sqrt{n}}\right)$
  - CLT: If $y \sim ?(\mu, \sigma)$, then for $n \geq 30$

    $\bar{y} \approx N\left(\mu_{\bar{y}} = \mu, \sigma_{\bar{y}} = \frac{\sigma}{\sqrt{n}}\right)$