MATH 3339 Statistics for the Sciences

Test 1 Fomula Sheet

Descriptive Statistics

- $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_1$, sample mean
- $\tilde{x} = \begin{cases} \text{middle value of ordered data} & \text{if } n \text{ is odd} \\ \text{mean of two middle values of ordered data} & \text{if } n \text{ is even} \end{cases}$, sample median
- $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$, sample variance
- $s = \sqrt{s^2}$, sample standard deviation
- $IQR = Q_3 - Q_1$, interquartile range where Q_3 is the 75^{th} percentile and Q_1 is the 25^{th} percentile.
- $z = \frac{x-\bar{x}}{s}$, standard score
- $CV = \frac{s}{\bar{x}}$, coefficient of variation
- $cov(x, y) = \frac{1}{n-1}\sum_{i=1}^{n}(x_1 - \bar{x})(y_i - \bar{y})$, sample covariance
- $r = \frac{cov(x,y)}{s_x s_y}$, sample correlation
- $R^2 = r^2$, coefficient of determination
- $\hat{\beta}_1 = r\frac{s_y}{s_x}$, sample slope of least-squares regression line
- $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$, sample y-intercept of least-squares regression line
- $y_i - \hat{y}_i$, residual of i^{th} observation

Counting and Probability

- $n! = n \cdot (n-1) \cdot (n-2) \cdots 1$, factorial
- $P(n, r) = \frac{n!}{(n-r)!}$, permutation
- $C(n, r) = \frac{n!}{r!(n-r)!}$, combination

- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, general addition rule for probability
- $P(A|B) = \frac{P(A \cap B)}{P(B)}$, conditional probability
- Given that events A_i, $i = 1, 2, \ldots, n$, are disjoint and exhaustive of the sample space S, then $P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$, law of total probability.

Discrete Probabilities

- $E(X) = \mu_x = \sum_{i=1}^{n} x_i p_i$, expected value of the distribution
- $Var(X) = \sigma_x^2 = \sum_{i=1}^{n} (x_i - \mu_x)^2 p_i = E(X^2) - E(X)^2$, variance of the distribution
- $SD(X) = \sigma_x = \sqrt{\sigma_x^2}$, standard deviation of the distribution
- **Binomial Distribution**, n = number of trials, p = probability of success
  - $P(X = k) = C(n, k)p^k(1 - p)^{n-k}$, probability of x successes in n independent trials
  - $E(X) = np$, expected value of binomial distribution
  - $Var(X) =$ np(1-p) , variance of binomial distribution
- **Geometric Distribution**, p = probability of success
  - $P(X = x) = (1 - p)^(x - 1)p$, probability that the x^{th} trial is the first success
  - $P(X > x) = (1 - p)^x$, probability that the first success is more than the x^{th} trial
  - $E(X) = \frac{1}{p}$ expected value of geometric distribution
  - $Var(X) = \frac{1-p}{p^2}$, variance of geometric distribution
- **Hypergeometric Distribution**, m = number of successes in the population, n = number of failures in the population, k = number of trails
  - $P(X = x) = \frac{C(m,x) \times C(n,k-x)}{C(m+n,k)}$, probability of x successes
  - $E(X) = \frac{km}{m+n}$, expected value of hypergeometric distribution
  - $Var(X) = kp(1 - p)\left(1 - \frac{k-1}{m=n-1}\right)$, variance of hypergeometric distribution, where $p = \frac{k}{m+n}$
- **Poisson Distribution**, $\mu =$ mean number of successes in a unit
  - $P(X = x) = \frac{e^{-\mu}\mu^x}{x!}$ probability of x successes
  - $E(X) = \mu$, expected value for Poisson distribution
  - $Var(X) = \mu$, variance for Poisson distribution

R Commands

- `factorial(n)`, factorial
- `factorial(n)/factorial(n-r)`, permutation
- `choose(n,r)`, combination
- `x = c(1, 2, ...)`, to input a vector
- `mean(x)`, mean
- `median(x)`, median
- `sd(x)`, standard deviation
- `cov(x,y)`, covariance
- `lm(y ~ x)`, least-squares regression line
- **Probability Distributions** "-" filled with d, p or q
  - `-binom(x,n,p)` binomial
  - `-gamma(x-1,p)` gamma
  - `-hyper(x,m,n,k)` hypergeometric
  - `-pois(x,mu)` Poisson