CSC 315, Fall 2016
Exam II Outline

1. Empirical probability
    a. Formula: empirical probability of an event E = its long run proportion

$$\mathrm{P}(E) = \frac{\#\ of\ times\ event\ occurs}{\#\ of\ trials\ or\ experiments}$$

    b. Can be calculated in *R* using the *replicate* function to replicate an event (implemented in a function) such as flipping a coin.

2. Classical probability
    a. Formula: classical probability of an event E =

$$\mathrm{P}(E) = \frac{\#\ of\ ways\ in\ which\ E\ occurs}{\#\ of\ possible\ outcomes\ in\ the\ sample\ space}$$

when all possible outcomes are equally likely

    b. Can be calculated in *R* using the *permutations* function from the library *gtools* .Note: in some cases, the *combinations* function can also be used (for example, for the sample space of possible poker hands, where the order of cards does not matter). I will give you the code if *combinations* is appropriate.

3. Probability distributions
    a. The probability that a random variable X is less than the value k can be calculated based on classical probability and is

$$\mathrm{P}(X < k) = \frac{\#\ of\ oservations < k}{total\ \#\ of\ observations\ (sample\ size)}$$

    b. This probability is equivalent to the sum of histogram densities for the bars corresponding to $X < k$.
    c. This probability is equivalent to the area under the curve between $-\infty$ and k for a probability density function.
    d. For any probability density function, the area under the curve between points a and b is equal to the $\mathrm{P}(a < X < b)$.

4. The normal probability distribution.
    a. $\mathrm{X} \sim \mathrm{N}(\mu, \sigma)$ if it is a unimodal, symmetric, bell-shaped distribution with mean μ and standard deviation σ

b. Empircal rule: for a normal distribution, approximately 68%, 95% and 99% of observations are within 1, 2, and 3 standard deviations of the mean.
c. The standard normal distribution is a normal distribution with $\mu = 0$ and $\sigma = 1$.
d. *R* functions:
    i. *pnorm* – to calculate probabilities (area under the curve and to the *left* of the specified value)
    ii. *dnorm* – to calculate the probability density (e.g., if plotting the curve)
    iii. *qnorm* – to calculate percentiles of the normal distribution
    iv. *rnorm* – random number generation

5. Sampling distribution of the sample mean
    a. The sample mean is a random variable!
    b. Central Limit Theorem: If a distribution *X* has mean μ and standard deviation σ then the sample distribution of the sample mean $\bar{X}_n$ from a sample of size *n* has mean μ and standard deviation $\sigma/\sqrt{n}$. Furthermore, the distribution of $\bar{X}_n$ is normal if *X* is normally distributed, and approximately normal if $n > 30$. The larger the sample size, the closer to normal the distribution.

6. Hypothesis testing based on a population proportion, a population mean, the difference between population means, or difference between two proportions.
    a. State the null and alternative hypotheses
    b. Specify the distribution of the sample statistic (i.e., a sample proportion or sample mean) under the null hypothesis, and graph this distribution.
    c. Calculate/find the test statistic (using the *prop.test* or *t.test* functions), or manually using the appropriate formula
    d. Find the *p*-value
        i. From *prop.test* or *t.test*
        ii. Based on a z- or t- test statistic and using the *pnorm* or *pt* function with appropriate degrees of freedom.
    e. State the conclusion regarding the null hypothesis in the context of the problem, and justify your conclusion based on the *p*-value.
    f. Interpretation of Type I or Type II errors