

CSC 315, Fall 2018

Exam II Format and Outline

The Exam II format will be as follows:

- Exam II will be a written exam
- You will not have access to a computer during the exam
- You may bring one page of notes, front and back, any size font
- The formula sheet will be provided for you
- Question format:
 - You may be asked to interpret output from *R* code, such as results from a hypothesis test,
 - You may be asked to add to *R* code (such as adding code to extract or calculate a *p*-value)
 - You may be asked to write *R* code, as required on previous Labs 4 – 7, *but*
 - You will not be asked to write *R* code for generating any graphs, but
 - You may be asked to interpret an *R* graph that is provided to you
 - You are responsible for all material in Lab #5

Exam II Outline

1. Empirical probability

- a. The empirical probability of an event *E* is the long run proportion of times that the event occurs

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

- 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* =

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

when all possible outcomes are equally likely

- b. Can be calculated in *R* using the *permutations* function (when order matters) from the library *gtools* or the *combinations* function (when order does not matter).

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

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

- 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 $P(a < X < b)$.

4. The normal probability distribution

- a. $X \sim N(\mu, \sigma)$ if it is a unimodal, symmetric, bell-shaped distribution with mean μ and standard deviation σ
- b. Empirical 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* – calculates probabilities of the form $P(X < k)$, equivalent to the area under the curve to the *left* of k
 - ii. *dnorm* – calculates the probability density (e.g., if plotting the curve)
 - iii. *qnorm* – calculates 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 σ/\sqrt{n} . Furthermore, the distribution of \bar{X}_n is normal if X is normally distributed, and approximately normal for other distributions if $n > 30$. The larger the sample size, the closer the distribution is to normal.
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.
 - 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