(Hopefully) Helpful Notes

Note: This material is purely supplemental, and is not a substitute for the actual course material. This does not include all course material, and contains additional material not necessary for the purposes of this course. If you find any typos please let me know at sherrard@ucsb.edu.

Section 1:

Parameter: Parameters in statistics are an important component of any statistical analysis. In simple words, a parameter is any numerical quantity that characterizes a given population or some aspect of it.

Statistic: Any function of our data i.e. mean, median, mode, count, proportion of the data.

Alternative Hypothesis: The alternative hypothesis is usually our research hypothesis. For example, we may want to test if kissing couples tend to lean their heads to the right more often than by random chance.

Null Hypothesis: The statement being tested. It is usually our previous knowledge, or the complement of the alternative hypothesis. For example if our research hypothesis is that couples turn their head to the right while kissing more often than not, then our null hypothesis may be that they turn their head towards each side equally.

P-value: The p-value, or calculated probability, is the probability of finding the observed, or more extreme, results when the null hypothesis of a study question is true. For example, suppose we are interested in testing whether a certain policy has been effective. In this case our null hypotheis will be that there is no effect, and our alternative hypothesis will be that there was an effect. We donote this as follows: Let β be the parameter representing the effect of the policy, H_0 represent the null hypothesis, and H_a be the alternative hypothesis. We write:

$$H_0 : \beta = 0$$

$$H_a : \beta \neq 0$$

The p-value tells us the probability of drawing the sample in question given that the null hypothesis is true (the policy had no effect). A low p-value (typically we use 0.05 as a threshold) tells us that it would be incredibly unlikely to see the sample we are seeing if the null hypothesis were true, and as such we have reason to reject the null hypothesis. A high p-value (say anything above 0.1) tells us that it is not sufficiently unlikely that we would see our data under the null hypothesis, and thus we cannot reject the possibility that the null is true.

Section 2:

Probability of Some Event:
$$P(\text{Some Event}) = \frac{\text{\# of outcomes in event}}{\text{total \# of outcomes}}$$

Probability of All outcomes:
$$P(\text{All Outcomes}) = 1$$

Probability of Impossible Outcomes:
$$P(\text{Impossible Outcome}) = 0$$

Probability Range:
$$0 \leq P(A) \leq 1$$

Rule of Complementary Events:
$$P(A') + P(A) = 1$$

Independent Events:
$$P(A \cap B) = P(A) \cdot P(B)$$

Disjoint (Mutually Exclusive) Events:
$$P(A \cap B) = 0$$

Probability of the Union of Two Sets:
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Conditional Probability:

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

Bayes Theorem:

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

Relationship between Conditional/Joint Probabilities:

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

De Morgan's Laws:

$$(A \cup B)' = A' \cap B'$$
$$(A \cap B)' = A' \cup B'$$

Section 3:

Expected Value (Mean) of X:

$$\mathbb{E}[X] = \mu = \sum_i x_i f(x_i) = \sum_i x_i P(X = x_i)$$

Expectation Properties:

Let X and Y be random variables with finite expected values:

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$$

Let a be any constant:

$$\mathbb{E}[aX] = a\,\mathbb{E}[X]$$

Variance of X:

$$Var(X) = \sigma^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}[X^2] - \mu^2$$

Note: You may also see the following formula:

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

To see how these formulas are all equivalent check this link: https://en.wikipedia.org/wiki/Variance#Definition

Variance Properties:

$$Var(X) \geq 0$$

Let a be any constant:

$$Var(X + a) = Var(X)$$
$$Var(aX) = a^2 Var(X)$$

Standard Deviation of X:

$$\sigma = \sqrt{Var(X)} = \sqrt{\mathbb{E}[X^2] - \mu^2}$$

Combinatorics Table:

	Without Replacement	With Replacement
Ordered	$\frac{n!}{(n-k)!}$	n^k
Unordered	$\binom{n}{k} = \frac{n!}{(n-k)!k!}$	Don't worry about this one

For the purposes of this class we will mostly just worry about combinations $\binom{n}{k}$ (Unordered/Without Replacement) as it appears in a variety of distributions. While we may at some point use the other two formulas, they are mostly here merely for future reference. The following are examples which may help you better understand when to use each one:

Combinations (Unorderd/Without Replacement): Denoted $\binom{n}{k}$, combinations simply describe the number of ways we can arrange k chosen objects from a pool of size n when order does not matter and chosen objects are not replaced. The simplest example would be a lottery. Suppose for example that in a lottery 5 numbers are chosen from a pool of 10 possible numbers. Each number, once chosen, is not placed back in the pool, so there can be no duplicates in our group of 5. The order the numbers are chosen, however, does not matter. 1,2,3,4,5 is equivalent to 5,4,3,2,1. In this example the calculation would be:

$$\binom{10}{5} = \frac{10!}{(10-5)!5!} = 252$$

Another example would be the following: Suppose a group of 8 friends were playing checkers, and everyone wanted to play eachother once. We could calculate the total number of games played:

$$\binom{8}{2} = \frac{8!}{(8-2)!2!} = 28$$

Permutations (Ordered/Without Replacement) Whereas with combinations order does not matter, with permutations the order in which the objects are chosen matters. As such, the number of permutations will be larger than the number of combinations for a given group of size n choosing k objects. One example of a permutation is the following: Suppose in a given Olympic event there are 10 competitors. How many different permutations of Gold/Silver/Bronze medals can there be? Here it is clear that the order in which competitors are "chosen" matters. Being "chosen" first gives you a gold medal, second a silver, and third a bronze. The calculation would be as follows:

$$\frac{10!}{(10-3)!} = 720$$

Ordered/With Replacement: This example can be a little less intuitive. Here we are looking at the amount of ways n possible choices can be arranged in k spots when we allow for duplicates. Perhaps the easiest example would be that of a coin flip. Suppose you want to know that number of different outcomes you can get from flipping a fair coin 3 times. In this case $n = 2$, as there are two possible outcomes (Head and Tails). $k = 3$ as we are placing these 2 possibilities into 3 spots (note this is only possible as we allow replacement). In this case we see that there are 8 total outcomes:

$$2^3 = 8$$

But to see this more clearly, note that the following is the set of outcomes:

$$\{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

Binomial Distribution: The binomial distribution describes the probability distribution of k successes over n independent bernoulli (binary) trials, each with probability π of success.

$$PMF : f_X(x) = \binom{n}{k}\pi^k(1-\pi)^{n-k}$$

Example: Suppose you want to know the probability of getting 5 heads over 10 total flips of a rigged coin. Let $\pi = .25$ be the probability of heads in a given trial. Then:

$$P(X = 5) = \binom{10}{5}0.25^5(1-0.25)^{10-5} \approx 0.0584$$

Properties of the Binomial Distribution:

$$\mathbb{E}[X] = n\pi$$
$$Var(X) = n\pi(1-\pi)$$

Section 4:

Hypergeometric Distribution: The Hypergeometric distribution describes the probability of k successes in n draws from a population of size N where there are N_1 possible success states. This is similar to the binomial distribution, but now we are drawing without replacement.

$$PMF : F_K(k) = \frac{\binom{N_1}{k}\binom{N-N_1}{n-k}}{\binom{N}{n}}$$

Example: Suppose you have a standard deck of cards, and would like to know the probability of drawing 2 aces in a hand of 5. In this case we have $N = 52$ (Total number of cards in the deck), $N_1 = 4$ (Total number of success states, in this

case aces, in the deck), $n = 5$ (There are 5 trials or draws), and $k = 2$ (the number of successes we are interested in). Plugging this into our formula we get:

$$P(K = 2) = \frac{\binom{4}{2}\binom{48}{3}}{\binom{52}{5}} \approx 0.0399$$

Section 5:

Properties of PDFs (Continuous Distributions):

Let X be a random variable, and let a and b be constants:

$$P(a \leq X \leq b) = \int_a^b f_X(x)dx$$

When turning a PDF into a CDF:

$$F_X(x) = \int_{-\infty}^x f_X(t)dt$$

When turning a CDF into a PDF:

$$f_X(x) = \frac{d}{dx}F_X(x)$$

The probability of any given point is zero. This is a property of integrals. Put mathematically:

$$P(X = a) = \int_a^a f_X(x)dx = 0$$

Expectation of a Continuous Random Variable:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} xf(x)dx$$

Consequently:

$$\mathbb{E}[X^2] = \int_{-\infty}^{\infty} x^2 f(x)dx$$

The Continuous Uniform Distribution:

For a continuous uniform random variable X with bounds a and b:

$$PDF : f_X(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq x \leq b \\ 0 & \text{else} \end{cases}$$

$$CDF : F_X(x) = \begin{cases} 0 & \text{for } x < a \\ \frac{x-a}{b-a} & \text{for } a \leq x < b \\ 1 & \text{for } x \geq b \end{cases}$$

Properties of the Continuous Uniform Distribution:

$$\mathbb{E}[X] = \frac{a+b}{2}$$

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

The Normal (Gaussian) Distribution:

For a normal random variable X:

$$PDF : f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{\frac{1}{2\sigma^2}(x-\mu)^2} \quad \text{for } -\infty \leq x \leq \infty$$

Properties of the Normal (Gaussian) Distribution:

$$\mathbb{E}[X] = \mu$$

$$Var(X) = \sigma^2$$

$$SD(X) = \sigma$$

For a normal random variable X with mean μ and variance σ^2 we often write:

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

Note that the distribution is symmetric about the mean. We also designate the "standard" normal distribution as a normal distribution with mean 0 and variance 1.

Section 6-7:

Normal CDF and Z-Scores:

For a given normal random variable we often want to find the $P(X < \text{some value})$. Whereas with other distributions we simply use the CDF, with the normal we are unable to do that as we cannot express the CDF with standard functional forms. Note that we often designate the CDF with $\Phi(x)$. Because we cannot use the CDF to estimate our probability, we must turn to using z-scores and a z-table, which distill the information from the normal CDF into a form we can work with easily. The z-score, defined thusly:

$$z = \frac{x - \mu}{\sigma}$$

describes the number of standard deviations a given point (x) is from the mean value. Suppose we want to calculate the $P(X < x)$ where x describes some point in our distribution. We will take the following steps:

1. Using the formula above we calculate the z-score.
2. We use the z-table to find the probability associated with the calculated z-score.

To give an example: Suppose our random variable X follows a normal distribution with $\mu = 35, \sigma = 20$, and we want to know $P(X < 40)$. We will first calculate our z-score:

$$z = \frac{40 - 35}{20} = 0.25$$

And then, using a z-table, we will find the probability associated with this value. In this case we get:

$$P(X < 40) = 0.5987$$

Continuity Correction of a Binomial Random Variable to a Normal Distribution:

Suppose we wish to know the probability of some range of values occuring from a binomial distribution. Given only the probability of success (π) and the number of trials (n) we would need to calculate the probability for each possible value separately, and then sum over them to find the probability of the range. Depending on the size of the range this could be a difficult task. Consider the following example: Suppose you are given a fair coin ($\pi = 0.5$) and are asked to calculate the probability of getting fewer than 40 heads over 100 flips (n = 100). To calculate this we would need to find:

$$P(X = 0) + P(X = 1) + P(X = 2) + ... + P(X = 39)$$

But that would require us to do 40 individual evaluations of the binomial distribution. No easy task. Thus what we will do instead is approximate $P(X < 40)$ using a normal distribution and employing a continuity correction. The continuity correction serves to make our estimate of the probability more accurate by correcting some of the error which occurs when we estimate a discrete distribution (Binomial) with a continous (Normal). To do the continuity correction we will take the follwing steps:

1. Determine the "included" and "excluded" sets. The included set will be the set of values that lies in the range we are seeking to calculate. The excluded set is the set of values outside of this range.
2. Change the value we are estimating to that which is halfway between the included and excluded sets, and calculate the probability using a normal distribution with mean and variance equal to that of the binomial.

Returning to our example, we know that the mean and standard deviation of the binomial distribution we are drawing from are:

$$\mu = n\pi = 50$$
$$\sigma = \sqrt{n\pi(1 - \pi)} = 5$$

Thus we will estimate it with:

$$X \sim N(\mu = 50, \sigma = 5)$$

Here the included set is all integers between 0 and 39, whereas the excluded set is all integers from 40 to 100. Thus, employing the continuity correction, we will estimate:

$$P(X < 39.5) = 0.0179$$

Let's consider a slightly different example. Suppose instead we are asked to find $P(X \leq 40)$. Now our included set goes from 0 to 40, whereas the excluded set will go from 41 to 100. Thus, with our continuity correction, we will estimate:

$$P(X < 40.5) = 0.0287$$

Central Limit Theorem:

Let X be a random variable with mean μ and variance σ^2, and let X_i designate draws from the distribution. Furthermore, let $\bar{X}$ be the sample mean of the draws. Mathematically:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

By the central limit theorem, we know that:

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Thus $\bar{X}$ will have mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$.

Consequently, let $SUM = \sum_{i=1}^{n} X_i$, then:

$$SUM \sim N\left(n\mu, \sqrt{n}\sigma\right)$$

Section 8:

Type 1 Error:

A type 1 error is the incorrect rejection of a true null hypothesis (False Positive). The probability of a type 1 error is equal to our significance level α. We can write this as:

$$P(\text{Type 1 Error}) = P(\text{Reject } H_0 \mid H_0 \text{ is True})$$

Type 2 Error:

A type 2 error occurs when we incorrectly retain a false null hypothesis (False Negative). We often denote the probability of a type 2 error with β. We can write this as:

$$P(\text{Type 2 Error}) = P(\text{Fail to Reject } H_0 \mid H_0 \text{ is False})$$

Power of a Test:

The power of a test is the probability of not committing a type 2 error. As such it is equal to $1 - \beta$. In other words, the power of a test is the probability of correctly rejecting a false null hypothesis. We can write this as:

$$\text{Power} = P(\text{Reject } H_0 \mid H_0 \text{ is False})$$

,

Sample Variance:

$$s^2 = \frac{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2}{n - 1}$$

Confidence Interval:

For a known standard deviation (σ):

$$CI = \bar{X} \pm z^* \cdot \frac{\sigma}{\sqrt{n}}$$

For an unknown standard deviation:

$$CI = \bar{X} \pm t^* \cdot \frac{s}{\sqrt{n}}$$

Chebyshev Inequality:

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

or

$$P(|x - \mu| \leq k\sigma) \geq 1 - \frac{1}{k^2}$$

Markov's Inequality:

$$P(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$$