Test 3 Study Guide
Math 160 Section 3

0.1 Chapter 13 - The Binomial Distribution

1. What are the characteristics of the Binomial setting?

   (a) A fixed number n of trials.
   (b) All trials are independent.
   (c) Each trial has only two possible outcomes - called "success" or "failure".
   (d) A fixed probability p of success in each trial.

   Such a sequence of trials is sometimes called a **Binomial experiment**.

2. What kind of random variable has a Binomial distribution?

   A count of successes in a Binomial experiment.

3. Say $X \sim B(n, p)$, in other words X has the Binomial distribution with parameters n and p.

   (a) What do n and p stand for?
   X is the number of successes in n independent trials, each of which has probability of success p. Thus, n is the number of trials and p is the probability of success.
   (b) What is the mean of X? What is the standard deviation of X?
   The mean of X is $\mu = np$. The standard deviation of X is $\sigma = \sqrt{np(1-p)}$.
   (c) What is the probability that $X = k$, if $0 \leq k \leq n$?
   $P(X = k) = \binom{n}{k} p^k (1-p)^{(n-k)}$.
   (d) What is the command on your calculator that gives the probability that $X \leq k$?
   binomcdf(n, p, k).

4. What is the binomial coefficient $\binom{n}{k}$? Compute $\binom{6}{3}$ and $\binom{7}{2}$

$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n*(n-1)*(n-2)*...*(n-k+1)}{k*(k-1)*(k-2)*...*1}$.

$\binom{6}{3} = \frac{6!}{3!3!} = \frac{6*5*4}{3*2*1} = 20.$

$\binom{7}{2} = \frac{7!}{2!5!} = \frac{7*6}{2*1} = 21.$

5. Martha plays poker with 5 friends every week. Martha is good at poker, and each week, she has probability .35 of winning. Let X be the number of times a year Martha wins at poker. (There are 52 weeks in a year.)

   (a) What is the formula giving the probability that Martha wins one-quarter of the poker games in a given year?
   One-quarter of 52 is 13.
   $P(X = 13) = \binom{52}{13} (.35)^{13} (1 - .35)^{(52-13)} = \binom{52}{13} (.35)^{13} (.65)^{(39)}$.
   (b) What is the probability that Martha wins one-quarter or fewer of the poker games in a given year?
   binomcdf$(52, .35, 13)$
   (c) On average, how many games a year does Martha win?
   The mean of $B(52, .35)$ is $\mu = 52 * .35 = 18.2$.

0.2 Chapter 3 - The Normal Distribution

1. What is the area under a density curve and above the x-axis?

2. If a density curve is left-skewed, which is larger: the mean or the median?

   The mean is not resistant, which in this case means it is more affected by skewness than the median. So, the mean is smaller than the median for a density curve which is left-skewed. The mean is larger than the median for a density curve which is right-skewed.

3. What does the $68 - 95 - 99.7$ Rule for normal distributions say?

   Over many, many observations from a normal distribution with mean μ and standard deviation σ,

   (a) Approximately 68% of observations should fall in the interval $(\mu - \sigma, \mu + \sigma)$.

   (b) Approximately 95% of observations should fall in the interval $(\mu - 2\sigma, \mu + 2\sigma)$.

   (c) Approximately 99.7% of observations should fall in the interval $(\mu - \sigma, \mu + \sigma)$.

4. Suppose $X \sim N(25, 25)$. Draw pictures of the following values as areas under a density curve. Find the values using Table A, and give the z-values used.

   (a) $P(X < 22)$.
   The z-score for 22 is
   $$\frac{22 - 25}{25} = -0.12.$$

   (b) $P(X > 32)$.
   The z-score for 32 is
   $$\frac{32 - 25}{25} = 0.28.$$

   (c) $P(22 \leq X \leq 32)$.

5. Let L be the random variable which is the length of human pregnancies from conception to birth. Then $L \sim N(266, 16)$, where L is given in days. Draw pictures of the following values as areas under a density curve. Use your calculator to find the following values, and write out the commands used.

   (a) The proportion of pregnancies lasting less than 240 days.
   .5-normalcdf(240, 266, 266, 16)

   (b) The proportion of pregnancies lasting between 240 and 260 days.
   normalcdf(240, 260, 266, 16)

   (c) The lengths of the longest one-fifth of pregnancies.

0.3 Chapter 11 - Sampling Distributions

1. Define the term *parameter*.

   A parameter is a number that describes a population. (A population is a group of individuals we would like to know about.)

2. Define the term *statistic*.

   A statistic is a number that can be calculated from the data we obtain from a sample. (A sample is a subset of the population that we actually measure.)

3. What is the difference between $\bar{x}$ and μ?

   The population men is denoted μ - this is the mean value of a variable for all the individuals in a population, and it is usually unknown. The sample mean is denoted $\bar{x}$ - this is the mean of a variable for the sample we've selected.

4. What does the "Law of Large Numbers" say?

5. Explain what the "Sampling Distribution of the Mean" is.

6. Suppose the time between nerve impulses (interspike interval) in humans follows an exponential distribution with mean 0.219 seconds, and standard deviation 0.148 seconds. As part of a test of the effects of nicotine on the nervous system, the times between 101 nerve impulses are recorded from the arms of control subjects. The result are 100 interspike intervals for each subject. These 100 intervals are then averaged, to give an average interspike interval for each subject.

   (a) What is the distribution of the average interspike intervals?
   The average interspike intervals are sample means for a sample of size 100. Thus, they are distributed normally, with mean 0.219 seconds, and standard deviation $0.148/\sqrt{100} = .00148$ seconds.

   (b) What is the probability that a given subject has an interspike interval longer than 0.25 seconds?

   (c) Within what range do the interspike intervals of 90% of subjects lie?

0.4 Chapter 14 - Confidence Intervals

1. What is the general definition of a confidence interval at level C for a parameter θ?

   A confidence interval at level C is an interval of the form

   $$\text{Estimate} \pm \text{Margin of Error}$$

   along with a confidence level C. For C% of a very large number of samples drawn from the population, the interval will contain the parameter θ.

2. Which of the following is true about a 99% confidence interval?

   (a) 99% of the possible sample means will be included in this interval.

   (b) 99% of the intervals constructed using this process based on repeated samples from this population will include the population mean.

   (c) 99% of the time the interval will include the sample mean.

   (d) 99% of the time the interval will include the population mean.

   (e) 99% of the possible population means will be included in the interval.

   (B) is true - 99% of the intervals constructed using this process based on repeated samples from this population will include the population mean.

3. Given the same sample size and the same population, which is longer: a 95% confidence interval or a 90% confidence interval?

   A 95% confidence interval.

4. How can we decrease the margin of error of a confidence interval?

   By decreasing the confidence level or increasing the sample size.

5. What are the assumptions (the book calls them "Simple Conditions") we have made so far when calculating confidence intervals?

   (a) The population is normally distributed.

   (b) The sample is a simple random sample.

   (c) We know the population standard deviation, σ.

6. Say we have a population which is normally distributed with $\sigma = 3.75$. We take a simple random sample of 25 individuals from the population, and find $\bar{x} = 10$.

   (a) Find a 92% confidence interval for the population mean.

   (b) How many more individuals must we sample to obtain a 96% confidence interval with the same margin of error?

0.5 Chapter 15 - Tests of Significance

1. What is the result of a test of significance? What does it tell you?

   We obtain a p-value, and by comparing it to α, we conclude either to reject the null hypothesis H_0, or that the evidence is consistend with the null hypothesis. If we reject the null hypothesis, we have concluded that the evidence we gathered is so unlikely, given the null hypothesis, that the null hypothesis is probably untrue.

2. In hypothesis testing, what is H_a? What is H_0?

   H_0 is the null hypothesis, the hypothesis we are gathering evidence against.

   H_a is the alternative hypothesis. It tells us what kind of evidence against H_0 we will accept.

3. Suppose $H_0 : \mu = \mu_0$. Set up a one-sided H_a.

   Either $\mu < \mu_0$, or $\mu > \mu_0$.

4. What's an α, in the context of significance?

   It is a significance level, set before the test of significance is performed. The significance level is a measure of how much evidence against H_0 we require before rejecting it.

5. True or false: a p-value of 0.05 is stronger evidence against the null hypothesis than a p-value of 0.1.

   True.

6. Experiments on learning in animals sometimes measure how long it takes mice to find their way through a maze. The mean time is 18.4 seconds, with standard deviation 2.3 seconds, for one particular maze. A researcher thinks that a loud noise will cause the mice to complete the maze in less time. She measures how long each of 10 mice takes with a noise as stimulus. She assumes that the standard deviation for maze times under these conditions will also be 2.3 seconds, and chooses a significance level of $\alpha = .05$. On average, it takes the 10 mice 19.2 seconds to find their way through the maze.

   (a) What are the null and alternative hypotheses for this test?
   The null hypothesis is $\mu = 18.4$.
   The alternative hypothesis is that $\mu < 18.4$.

   (b) What is the p-value for this test?

   (c) What conclusion can be drawn?

7. Explain, by drawing a suitable diagram, why a significance test that is significant at the 1% level must also be significant at the 5% level.