MATH 301 Fall 2016 Homework Assignments:

Homework 1, due September 19

1) Anxiety disorders and symptoms can often be effectively treated with medications. The accompanying data on a receptor binding measure was read from a graph in a recent scientific paper on the subject. Use various methods from Chapter 1 to describe and summarize the data. In particular, we want to highlight the differences between the two groups of patients.

PTSD:	10	20	25	28	31	35	37	38	38	39
	39	42	46							
Healthy:	23	39	40	41	43	47	51	58	63	66
	67	69	72							

2) Consider a sample $x_1, x_2, \dots, x_n$ and suppose that the values of $\bar{x}$ and s have been calculated and are known. Let $y_i = x_i - \bar{x}$ and $z_i = y_i / s$ for all i's. (The y's have been "centered" and the z's have been "standardized".) Find the means and standard deviations for the two new lists, y and z.

(‡ Artifact 1 ‡) 3) Specimens of three different types of rope wire were selected, and the fatigue limit was determined for each specimen. Construct a comparative box plot and a plot with all three quantile plots superimposed. Comment on the information each display contains. Also explain which graphical display you prefer for comparing these data sets. Keep in mind your goal is to highlight the differences between the data sets.

Type 1	350	350	350	358	370	370	370	371	371
	372	372	384	391	391	392			
Type 2	350	354	359	363	365	368	369	371	373
	374	376	380	383	388	392			
Type 3	350	361	362	364	364	365	366	371	377
	377	377	379	380	380	392			

4) The sample data $x_1, x_2, \dots, x_n$ sometimes represents a time series, where x_t is the observed value of a response variable x at time t. Often the observed series shows a great deal of random variation, which makes it difficult to study long-term behavior. In such situations, we often prefer a smoothed version of the series. One technique for doing so involves **exponential smoothing**. A smoothing constant α is chosen $(0 < \alpha < 1)$ and then smoothed values y_t are calculated by the following rules: $y_1 = x_1$ and for $t = 2, 3, \dots, n$ $y_t = \alpha x_t + (1 - \alpha) y_{t-1}$.

a) Consider the following time series of the temperature of effluent at a sewage treatment plant on successive days: $47, 54, 53, 50, 46, 46, 47, 50, 51, 50, 46, 52, 50, 50$. Plot the data with time as the horizontal variable. Does there appear to be any pattern? Now calculate the y_t's using $\alpha = 0.1$. Repeat for $\alpha = 0.5$. Which value of α gives a smoother series?

b) Substitute $y_{t-1} = \alpha x_{t-1} + (1 - \alpha) y_{t-2}$ on the right-hand side of the expression for y_t, then substitute y_{t-2} in terms of x_{t-2} and y_{t-3}, and so on. On how many of the values $x_1, x_2, \ldots, x_n$ does y_t depend? What happens to the coefficient for y_{t-k} as k increases? If t is large, how sensitive is y_t to the initial condition $y_1 = x_1$?

Homework 2, due September 30

1) A utility company offers a lifeline rate to any household whose electricity usage falls below 240 kWh during a particular month. Let A denote the event that a randomly selected household in a certain community does not exceed the lifeline usage during January, and let B be the analogous event for the month of July (A and B refer to the same household). Suppose $P(A) = 0.8, P(B) = 0.7,$ and $P(A \cup B) = 0.9$. Compute a) $P(A \cap B)$, and b) the probability that the lifeline usage amount is exceeded in exactly one of the two months. Describe this last event in terms of A and B.

2) The route used by a motorist to get to work has two stoplights. The probability of signal 1 being red is 0.4 and for signal 2 it is 0.5. There is a 0.6 probability that at least one of the two is red. What is the probability that both signals are red? That the first is red, but not the second? That exactly one signal is red?

3) Three molecules of type A, three of type B, three of type C, and three of type D are to be linked together to form a chain molecule. An example of one such chain molecule is $ABCDABCDABCD$ and another is $BCDDAAABDBCC$.

a) How many such chain molecules are there? [Hint: If the three A's were distinguishable from one another, such as $A_1, A_2,$ and A_3, how many molecules would there be? How is this number reduced when the subscripts are removed from the A's?]

b) Suppose a chain molecule of the type described is randomly selected. What is the probability that all three molecules of each type end up next to one another (such as in $BBBAAADDDCCC$)?

4) Three married couples have purchased theater tickets and are seated in a row consisting of just six seats. If they take their seats in a completely random fashion (random order), what is the probability that Jim and Paula (husband and wife) sit in the two seats on the far left? What is the probability that Jim and Paula end up sitting next to one another? What is the probability that at least one of the wives ends up sitting next to her husband? [Note: if you cannot find the formula to answer this consider a brute force listing of the sample space.]

Homework 3, due October 7

(‡ Artifact 2 ‡) 1) In a Little League baseball game, suppose the pitcher has a 50% chance of throwing a strike and a 50% chance of throwing a ball, and that successive pitches are independent of one another. Knowing this, the opposing team manager has instructed his hitters to not swing at anything. What is the chance that the batter walks on four pitches? What is the chance that the batter walks on the sixth pitch? What is the chance that the batter walks (not necessarily on four pitches)? Note: in baseball, if a batter gets three strikes he is out, and if he gets four balls he walks.

2) A car insurance company classifies each driver as good risk, medium risk, or poor risk. Of their current customers, 30% are good risks, 50% are medium risks, and 20% are poor risks. In any given year, the chance that a driver will have at least one citation is 10% for good risk drivers, 30% for medium risk drivers, and 50% for poor risk drivers. If a randomly selected driver insured by this company has at least one citation during the next year, what is the chance that the driver was a good risk? A medium risk?

3) An insurance company offers its policyholders a number of different premium payment options. For a randomly selected policyholder, let X be the number of months between successive payments. The cdf of X is:

$$F(x) = \begin{cases} 0 & x < 1 \\ 0.3 & 1 \le x < 3 \\ 0.4 & 3 \le x < 4 \\ 0.45 & 4 \le x < 6 \\ 0.6 & 6 \le x < 12 \\ 1 & 12 \le x \end{cases}$$

What is the pmf of X? Using just the cdf, compute $P(3 \le X \le 6)$ and $P(4 \le X)$. [Of course you can use your pmf to check your work.]

4) The pmf for X, the number of major defects on a randomly selected appliance in our warehouse, is

x	0	1	2	3	4
$p(x)$	0.08	0.15	0.45	0.27	0.05

Compute $E(X)$, $V(X)$ using the definition, and $V(X)$ using the shortcut.

<u>Homework 4, due October 24</u>

1) The time it takes a read/write head to locate a desired record on a computer disk once positioned on the right track can be reasonably modeled with a uniform distribution. If the disk rotates once every 25 msec, then assume $X \sim \text{Unif}[0,25]$. Compute $P(10 \le X \le 20)$, $P(10 \le X)$, the cdf $F(x)$, $E(X)$, and $V(X)$.

(‡ Artifact 3 ‡) 2) Use the following pdf and find a) the cdf b) the mean and c) the median of the distribution.

$$f(x) = \begin{cases} (4 - x^2)/9 & -1 < x < 2 \\ 0 & \text{otherwise} \end{cases}.$$

3) There are two machines available for cutting corks intended for use in wine bottles. The first produces corks with diameters that are normally distributed with mean 3 cm and standard deviation 0.1 cm. The second machine produces corks with diameters that have a normal distribution with mean 3.04 cm and standard deviation 0.02 cm. Acceptable corks have diameters between 2.9 cm and 3.1 cm. Which machine is more likely to produce an acceptable cork?

4) Suppose the time it takes for Jed to mow his lawn can be modeled with a gamma distribution using $\alpha = 2$ and $\beta = 0.5$. What is the chance that it takes at most 1 hour for Jed to mow his lawn? At least 2 hours? Between 0.5 and 1.5 hours?

<u>Homework 5, due October 31</u>

1) The data below are precipitation values during March over a 30-year period in Minneapolis-St. Paul.

0.77	1.20	3.00	1.62	2.81	2.48	1.74	0.47	3.09	1.31
1.87	0.96	0.81	1.43	1.51	0.32	1.18	1.89	1.20	3.37
2.10	0.59	1.35	0.90	1.95	2.20	0.52	0.81	4.75	2.05

Construct and interpret a normal probability plot for this data set. The large outliers should make the data look non-normal. Can a transformation make the data more normal looking? Calculate the square root and the cube root of each observation, and construct and interpret normal probability plots. What do you conclude is the best choice: leave the data along, use the square root, or use the cube root?

(‡ Artifact 4 ‡) 2) A particular type of tennis racket comes in a midsize version and an oversize version. Sixty percent of all customers who shop at a certain store want the oversize version. Assume the next ten customers that come to the store are a random sample of all customers. (This assumption is sometimes difficult to justify in practice, but is almost always made to facilitate our calculations. Whether this is good practice or not is a worthy discussion.) What is the chance that at least six of the next ten customers want the oversize racket? What is the chance that the number of customers out of the next ten who want the oversize racket is within one standard deviation of the mean? If the store currently has only seven rackets of each version, what is the chance that ***all*** of the next ten customers can get the version they want? [Hint: It might be easier to calculate the last part by examining the complement.]

3) Suppose n in a binomial experiment is known and fixed. Are there any values of p for which the variance is zero? Explain this result in words. For what value of p is the variance maximized? [Hint: Either graph variance as a function of p or try to minimize using calculus.]

<u>Homework 6, due November 9</u>

1) A second stage smog alert has been called in a certain area of Los Angeles county in which there are 50 industrial firms. An inspector will visit 10 randomly selected firms to check for violations of regulations. If 15 of the firms are actually violating at least one

regulation, what is the pmf of the number of firms visited by the inspector that are in violation of at least one regulation? Find the Expected Value and Variance for your pmf.

2) A couple wants to have exactly two girls and they will have children until they have two girls. What is the chance that they have x boys? What is the chance they have 4 children altogether? How many children would you expect this couple to have? (Find the Expected Value.)

3) Let X have a binomial distribution with $n = 25$. For $p = 0.5, 0.6$, and 0.9, calculate the following probabilities **both** exactly and with the normal approximation to the binomial. a) $P(15 \leq X \leq 20)$. b) $P(X \leq 15)$. c) $P(20 \leq X)$. Comment on the accuracy of the normal approximation for these parameter choices.

Homework 7, due November 30

1) There are 40 students in a statistics class, and from past experience, the instructor knows that grading exams will take an average of 6 minutes, with a standard deviation of 6 minutes. If grading times are independent of one another, and the instructor begins grading at 5:50 p.m., what is the chance that grading will be done before the 10 p.m. news begins?

(‡ Artifact 5 ‡) 2) A student has a class that is supposed to end at 9:00 a.m. and another that is supposed to begin at 9:10 a.m. Suppose the actual ending time of the first class is normally distributed with mean 9:02 and standard deviation 1.5 minutes. Suppose the starting time of the second class is also normally distributed, with mean 9:10 and standard deviation 1 minute. Suppose also that the time it takes to walk between the classes is a normally distributed random variable with mean 6 minutes and standard deviation 1 minute. If we assume independence between all three variables, what is the chance the student makes it to the second class before the lecture begins? [Hint: Consider the quantity $x_4 = x_2 - x_1 - x_3$. Positive values of x_4 correspond to making it to class on time.]

3) A 90% confidence interval for the true average IQ of a group of 100 people is (114.4, 115.6). Deduce the sample mean and population standard deviation used to calculate this interval, and then produce a 99% interval from the same data.

4) An experimenter would like to construct a 99% confidence interval with a length of no more than 0.2 ohms, for the average resistance of a segment of copper cable of a certain length. If the experimenter is willing to assume that the true standard deviation is no larger than 0.15 ohms, what sample size would you recommend?

Homework 8, due December 7

1) Fifteen samples of soil were tested for the presence of a compound, yielding these data values: 26.7, 25.8, 24.0, 24.9, 26.4, 25.9, 24.4, 21.7, 24.1, 25.9, 27.3, 26.9, 27.3, 24.8, 23.6. Is it plausible that these data came from a normal curve? Support your answer. Now calculate a 95% confidence interval for the true average amount of compound present. Comment on any assumptions you had to make.

2) A hot tub manufacturer advertises that with its heating equipment, a temperature of $100°$ F can be achieved in at most 15 minutes. A random sample of 32 tubs is selected, and the time necessary to achieve $100°$ F is determined for each tub. The sample average time and sample standard deviation are 17.5 minutes and 2.2 minutes, respectively. Does this data cast doubt on the company's claim? Calculate a P-value, and comment on any assumptions you had to make.

3) A sample of 50 lenses used in eyeglasses yields a sample mean thickness of 3.05 mm and a population standard deviation of 0.30 mm. The desired true average thickness of such lenses is 3.20 mm. Does the data strongly suggest that the true average thickness of such lenses is undesirable? Use $\alpha = 0.05$. Now suppose the experimenter wished the probability of a Type II error to be 0.05 when $\mu = 3.00$. Was a sample of size 50 unnecessarily large?

4) Suppose that the true average viscosity should be 3,000 in a certain process. Do the following measurements support that standard? State and test the appropriate hypotheses.

2,781	2,900	3,013	2,856	2,888

Homework 9, due December 12

1) A random-number generator is supposed to produce a sequence of 0s and 1s with each value being equally likely to be a 0 or a 1 and with all values being independent. In an examination of the random-number generator, a sequence of 50,000 values is obtained of which 25,264 are 0s.

a) Formulate a set of hypotheses to test whether there is any evidence that the random-number generator is producing 0s and 1s with unequal probabilities, and calculate the corresponding P-value.

b) Compute a 99% confidence interval for the probability p that a value produced by the random-number generator is a 0.

c) If a two-sided 99% confidence interval for this probability is required with a total length no larger than 0.005, how many additional values need to be investigated?

2) In a survey of 4,722 American youngsters, 15% were seriously overweight, as measured by BMI. Calculate and interpret a 99% confidence interval for the proportion of all American youngsters who are seriously overweight. Discuss whether the Associated Press (who reported this data) actually took or could have taken a random sample of American youngsters.

3) In is known that roughly 2/3 of all human beings have a dominant right foot or eye. Is there also right-sided dominance in kissing behavior? One scientific article reported that in a random sample of 124 kissing couples, both people in 80 of the couple tended to lean more to the right than to the left. If 2/3 of all kissing couples exhibit this right-leaning behavior, what is the probability that the number in a sample of 124 who do so differs from the expected value by at least as much as what was actually observed? (i.e.

calculate a P-value.) Does the result of the experiment suggest that the $2/3$ figure is plausible or implausible? State and test the appropriate hypotheses.