401 Discussion Boards

Week 1

1 The Gestation period is the closest physiological relationships a mother and baby will have in their relationship. During Gestation, the nutrient intake for a mother is shared with the baby through the placenta. As a result, when a mother ingests illegal drugs the baby also ingests the drugs. A baby’s development is much smaller on a scale compared to its mother and prone to nutrient risks. Illegal drugs are often filled with toxic nutrients for a baby and can often have a fatal reaction on the developing baby (John Hopkins Medicine).

         Drugs are created with different substances, and some ingredients are more toxic than other ingredients depending on the drug. Mothers have different reactions to drugs and likewise, the baby will have reactions. Overall, given the negative effects drugs have on the Gestation period, professionals see a correlation between SIDS and drug use especially cocaine.

 Reference

Johns Hopkins Medicine, based in Baltimore, Maryland. (n.d.). *Johns Hopkins Medicine, based in Baltimore, Maryland*. Retrieved June 20, 2012, from http://www.hopkinsmedicine.org

General Website Inquiry

   2. There are correlating factors that could link SIDS and drug abuse if it is found that ingesting drugs does not directly cause SIDS. According to John Hopkins Medicine, the healthier a mother is during her pregnancy the greater chances are her baby will be healthy. The inverse to this statement is the deterioration of a woman’s health increases the chances of a SIDS baby. While drug use might not directly cause SIDS, it could in fact lead to deteriorated health which is a factor in the causation of SIDS. Also, drug use is often associated with certain diseases because of ingesting techniques such as sharing needles. As a result, Hepatitis and immune deficiency diseases are associated with drug users. If a mother contracted one of the diseases and became pregnant it would greatly increase her chances for SIDS due to poor health and the suspected health of the developing baby. Drug abuse and poor health are often correlated. Likewise, poor health and SIDS are correlated. Drug abuse might not directly cause SIDS, but given the correlating relationship between drugs, poor health, and SIDS one could draw a connecting relationship.

Johns Hopkins Medicine, based in Baltimore, Maryland. (n.d.). *Johns Hopkins Medicine, based in Baltimore, Maryland*. Retrieved June 20, 2012, from http://www.hopkinsmedicine.org

3.  The example of mothers, drug use, and SIDS demonstrates a relationship. In predictive analytics, experiments are conducted to analyze and understand relationships. Our assigned reading taught how to form a hypothesis. The definition of a hypothesis according to R. Mark Sirkin is, “A statement that name the variables that appear to be related and indicates the nature of that relationship.(Page 10, 2006).” A hypothesis should have a dependent variable that is dependent or affected on an independent variable. In the case of the example for Session 1, the hypothesis would be worded, “there is a relationship between using drugs and SIDS such that when drugs are consumed by a mother there is a greater chance for SIDS.” From the hypothesis, an experiment can be conducted and analytics would help ascertain whether the statement is scientifically accurate. Analytics, relevant research and experiments help reveal pertinent information to issues like drug use and SIDS.

Clearly articulated inputs and outputs are pivotal for the decision making process and communication plan for a firm. Precisely defined inputs and outputs create “a single version of the truth” for a firm (Davenport and Harris). Statistics seeks to understand and answer complex questions in a scientific manner in order to be profitable. In order to accurately gather information, the parameters of the experiment must be clearly defined and understood. Take for example, customer satisfaction as a desired inquiry. There are many different perspectives on the meaning of customer satisfaction. Management might perceive it to mean overall buying experience. Marketing might understand the term to mean likeliness to purchase in the next six months. If a costly experiment is conducted and the overall definition is not defined, major errors will result. Using statistics, one would have to clearly identify the meaning of customer satisfaction. Once established, an experiment would communicate input questions based on the definition to derive output driven results to help answer the defined inquiry. The result of such an experiment would prove beneficial to the overall success of a firm.

Davenport, T. H., & Harris, J. G. (2007). Analytics and Business Performance. *Competing on analytics the new science of winning* (p. 46). Boston, Mass: Harvard Business School Press.

References

Sirkin, R. M. (2006). How We Reason. *Statistics for the social sciences* (3rd ed., p. 10). Thousand Oaks, Calif.: Sage Publications.

Week 2

Find an ad or article in a current newspaper or magazine (*USA Today* is a fertile source for this) in which data are being described.  Excerpt and cite the article for the class and determine whether or not the data have been fairly represented. If you were writing that same ad or article, what would you have done differently and why?

Nancy Hellmich of USA Today wrote an article citing the results of a small study conducted by the National Institutes of Health (Hellmich 2012). The weight loss experiment had the following structure:

* 21 obese people ranging in age from 18-40 were studied that had lost 10-15% percent of their bodyweight.
* The 21 participants were split into three smaller groups and were given specific diets.
* Their health was monitored from a medical standpoint.

The data in this experiment is very misleading given the size of the experiment. According to the law of large numbers, the sample size should be no less than 30 for the central limit theorem to be applicable (Sirkin, 2006 p.245). In this experiment, n=7 which means that the results will not conform to a normal distribution. According to data for 2010, 37.5% percent of American adults are obese (Ogden 2012). An obese American who reads this article might conclude the results from the experiment would apply to the greater population. Sadly, based on how small the sample size is compared to the adult obese population ranging from 18-40, which is 113.502 million, the z-score would be far below the necessary 1.65 (one tailed) or 1.96 (two-tailed) z score needed to reject the null hypothesis (Census.gov 2010).

If I were writing this article, I would make it explicitly clear that the results do not pertain in any way to the US population. Specifically, before stating the results of the experiment I would make it clear that the group was too small for the results to pertain to the general population. Also, at the end of the article I would state the same disclaimer.

References

Hellmich, N. (n.d.). Low-carb diet burns the most calories in small study - “ USATODAY.com. *News, Travel, Weather, Entertainment, Sports, Technology, U.S. & World - USATODAY.com*. Retrieved June 28, 2012, from <http://www.usatoday.com/news/health/story/2012-06-27/calories-low-carb-weight-loss/55843134/1>

Ogden, C., Carroll, M., Kit, B., & Flegal, K. (2012). Prevalence of Obesity in the United States, 2009-“2010. *NCHS Data Brief*, *82*, 1-8.

Sirkin, R. M. (2006). *Statistics for the social sciences* (3. ed.). Thousand Oaks, Calif.: Sage Publications.

Table 7. Resident Population by Sex and Age: 1980 to 2010. (n.d.). *Census*. Retrieved June 28, 2012, from <http://www.census.gov/compendia/statab/2012/tables/12s0007.pdf>

Drawing from your own professional experience, give an example of a variable for which the mean would be a more meaningful or appropriate descriptor than the median.   Also give an example of a variable for which the median would be a more meaningful or appropriate descriptor than the mean. Discuss why this would be the case for each of them.

One of the programs I oversee in my refugee resettlement position is a tutoring program. In this program, I utilize the benefits of using mean and median when communicating the mean age of students and median attendance.

    The mean is applicable when communicating information about the background of our students in the program. For example, we have 135 students in the program, and the average age is 10. Given that our program age ranges from 5-18, there is not a huge gap between ages to skew the mean.     
    The median is applicable when communicating the median amount of students that show up on a given night of tutoring. For example, during Ramadan our program might have 1-3 students but during registration there will be 60-80 students. The large swings in attendance would skew a mean, as a result using the median attendance more accurately reflects the number of students we have on a given night.

Week 3

Determining the probability of an event is often a difficult process for which our intuition is often not very reliable. Consider the two questions below:

1. How many people have to be in the class so there is a 50% chance that at least two people in the class have the same birthday?
2. How many people have to be in the class so there is a 50% chance of having the same birthday as the instructor?

Just when I thought I was starting to get the hang of stats I get hit with this seemingly simple questions. In my mind, the equation for this question would read x/365=.5. “x” equals the number of classmates needed to have a probability equal to 50%. “365” is the number of days in a year, which I am not figuring twins, leap year or other anomalies. “.5” equals the probability we desire which is 50% After running the calculation, I came up with 183.3 or 184. I thought to myself, “self, that was pretty easy.” I then ran the problem in excel to double check and sure enough basic math works in excel.

After feeling really smart, I “googled” the question and everything changed. It turns out this problem is very complicated and in reality I don’t understand probability like I thought. It turns out “23” is the answer, but after scrolling through all the fancy formulas I still did not understand (pleacher.com). Did anyone else in our class come up with my answer? Hopefully I am not an extreme outlier for this class.

I checked the work on the problem from pleacher.com and it turns out it is right, but seeing we can't attach anything in this discussion board you will have to take my word for it.

In regard to the second question. My thought process is that it is larger than the first question based on the fact that the Birthday is on specific date. Other than that, I do not have a clue as to how to solve it. When researching this problem, I stumbled onto this website (math.harvard.edu) which also presents other mind numbing probability conundrums.

References

McMullen, C. (2011, May 4). Probability Theory. *Math Harvard*. Retrieved July 5, 2012, from www.math.harvard.edu/~ctm/papers/home/text/class/harvard/154/course/course.pdf

Pleacher, D. (n.d.). The Birthday Problem. *The Pleacher Page*. Retrieved July 5, 2012, from <http://www.pleacher.com/mp/mlessons/stat/birthday.html>

Provide a scenario from your professional or personal experience in which the underlying data are likely to assume a normal distribution, and a scenario in which the underlying data are unlikely to assume a normal distribution.

I currently work as a program director leading initiatives that serve the refugee population in Eden Prairie, Minnesota. The family size we serve follows a normal distribution of the greater Somali refugee community in Minnesota. For example, the average size is around six people in a household with a standard deviation of four. This follows the same normal distribution for Somali families in and around Minnesota. While this example serves as an example of a sample group compared to the actual population, both groups follow a normal distribution in regard to family size.

The volunteers that serve in our programs do not follow a normal distribution in regard to the average age compared to the average age in Eden Prairie, Minnesota. Seeing that the majority of our volunteers are retired or their kids have left for college, our average age is negatively skewed and does not follow a normal distribution relative to age.

Week 4

Consider these two research statements: (1) “ We are 90 percent confident that the population mean is between 5 and 15.” (2)  
“90 percent of the confidence intervals formed in this way will contain  
the population mean, and this particular interval extends from a lower   
limit of 5 to an upper limit of 15.” Are these statements the same? If not, is one statement more accurate or more appropriate than the other?

           From my short time of being a student of statistics, I have learned communicating information in a pithy fashion is important. Both answers could be communicated in a more professional sense conveying the information more accurately. The first statement is communicating confidence from the statistician, but we do not know what that confidence is based on, thus I find the statement to be erroneous. The following statement is loquacious, but makes an attempt to communicate in a statistical fashion.

A confidence interval is a mathematical measurement. The definition of a measurement is a concrete indicator that can be gathered (Groves...2009) The second statement is more accurate from a statistical standpoint, but the wording is confusing. Wording the findings of a statistical proof should reflect the indicator in a concise manner.

Let’s face it, statistics is a confusing field, which is why it is important to communicate as clearly as possible. Sirkin would communicate the above situation as, “90% of the sample means lie between 5 and 15 (2010 p 254). As statisticians, it is important that “our” confidence does not get wrapped up into the work we are doing and not be communicated in our findings as the first statement communicates. The second statement is more accurate in communicating confidence intervals.

Groves, R. M. (2009). *Survey methodology* (2nd ed.). Hoboken, N.J.: Wiley.

Sirkin, R. M. (2006). *Statistics for the social sciences* (3rd ed.). Thousand Oaks, Calif.: Sage Publications.

We hear a lot about the results of opinion surveys or election polls in the news. Such  
surveys are essentially attempts to infer information for a general   
population based on information from a sample of that population. Assuming  
normal distributions, discuss the relationship between basic   
statistical quantities - such as the mean, standard deviation, and   
confidence interval - and reported polling results such as "margin of   
error".   
  
Discuss some of the following questions:  What do these statistical quantities correspond to in the polling results? Which item in this list is never mentioned by any name in connection   
with the results (unless you read the fine print about survey   
methodology)? Why do you suppose that is? What implications does this have for interpreting survey and poll results?

In order to best answer the questions, I have the latest details from the Gallup Poll on the Presidential Election.

Gallup Poll Information -   
            As of today, Barack Obama has 47% and Mitt Romney has 45%.

Each seven-day rolling average is based on telephone interviews with approximately 3,050 registered voters; Margin of error is ±2 percentage points. Results from April 15 through May 6 are based on five-day rolling averages with approximately 2,200 registered voters each; Margin of error is ±3 percentage points. Editorial note: Due to a technical issue, the May 1-5 data point is not displayed at this time. Information gathered from Gallup.com

The Gallup example demonstrates the mean relationships of 47% and 45% relating to one standard deviation equalling 2 percentage points. What I do not see listed in any fashion is the confidence interval. This might be as a result of the sample being too small in comparison to the greater population it is inferring. The last time I posted on a sample size being too small, I was told ANOVA would be relevant (Sirkin 2006). I do not know enough about ANOVA to comment. If the confidence interval was communicated outright, I doubt individuals would take the poll seriously. A 90% confidence interval in the sample mean would probably have such a large range that the poll findings would be irrelevant.

Gallup Presidential Election Trial Heat Results: Barack Obama vs. Mitt Romney. (n.d.). *Gallup.Com - Daily News, Polls, Public Opinion on Politics, Economy, Wellbeing, and World*. Retrieved July 12, 2012, from <http://www.gallup.com/poll/150743/Obama-Romney.aspx>

Sirkin, R. M. (2006). *Statistics for the social sciences* (3rd ed.). Thousand Oaks, Calif.: Sage Publications.

Week 5

In a jury trial in the US judicial system, everyone is innocent until proven guilty.

1. In  
   this scenario, there are several different questions we could discuss,   
   such as:  Do you feel it is more preferable for the jury to make a Type I  
   or Type II error?  Does the nature of the crime, or the possible   
   sentence upon conviction, affect your answer?  Which kind of error is   
   more costly if it is committed?  As a society, do we feel that one kind   
   of jury error is more acceptable than the other?
2. In  
   general (that is, outside of this particular context), is it more   
   preferable to make a Type I or Type II error?  How do you make that   
   determination?

     If one were to structure the statement into the format used in Sirkin it would look like the following:

* H1= The person on trial is guilty
* Ho= The person on trial is not guilty
  + Both examples assume an action of proof

    According to Sirkin a type 1 error is, “The probability of falsely rejecting a true null hypothesis (2006, p 207).” A type 2 error is rejecting H1, when one should have accepted it, based on the methods used for calculation (2006).  In our culture, the idea of wrongly committing someone is atrocious. But, given the ability of certain lawyers, the jury can easily be swayed that making a type 2 error is more preferable. From a cost standpoint, committing a type 1 error generally costs more on the taxpayer. A type 2 error results in a guilty man going free, but in the end it will still cost the average taxpayer if the guilty commits another crime. Both errors are costly to the average person, one simply has a time delay.    
   More important than the nature of the crime or sentence is the context of the accused. Take for example someone who stands convicted of aggravated assault with a deadly weapon. If the accused has no known prior incidence the jury will be more likely to make a type 2 error. Likewise, if the accused has quite a long track record the jury is far more likely to make a type 1 error. Based on context rather than the type of a crime, it is more preferable to make a type 1 or 2 error.

References

Sirkin, R. M. (2006). *Statistics for the social sciences* (3. ed.). Thousand Oaks, Calif.: Sage Publications.

**Subject:** Misinterpretation & Conditional Probabilities

The following scenario illustrates how easy it is to misinterpret   
medical test results. The results illustrate why some medical   
professionals argue against universal screening for certain diseases. A   
medical procedure that is used to test for cancer has the following   
characteristics:

* Type I error (false positive) = 0.01 (1%)
* Type II error (false negative) = 0.05 (5%)
* The power or accuracy of the test is defined as 1 – Type II error = 1 – 0.05 = 0.95 (95%).

The accuracy of the test means that the test will detect cancer 95% of   
the time IF a person has cancer. The type of cancer that the test is   
used for occurs in approximately 0.5% of the population (i.e., about 1   
in 200 people).  
  
**Question:** Since the test is 95%   
accurate, does this mean that if someone tests positive for cancer,   
there is a 95% chance that the person has cancer? Use the   
characteristics of the test and the incident rate of the cancer in the   
population to determine the probability that a   
person actually has cancer if the test is positive.  
  
If you're not sure where to start, one approach is to consider the number of people that would fall   
into each of the four possible outcomes if the test were administered to  
100,000 randomly selected persons. The four possible outcomes are:

1. Person does not have cancer, test is negative
2. Person does not have cancer, test is positive (Type I error)
3. Person has cancer, test is negative (Type II error)
4. Person has cancer, test is positive

Remember, we do not know ahead of time whether a person having the test  
has cancer or not. Therefore if a person tests positive, the result could  
be either an Outcome 2 or an Outcome 4. The probability that a person   
with a positive outcome has cancer is given by the ratio:   
Outcome 4/(Outcome 2 + Outcome 4). See if you can calculate the   
probability given by this ratio by determining the values for Outcomes 2  
and 4 based on testing 100,000 persons.

To answer this question I needed to work with the Math tutoring place. The easiest thing for me to do was put the question into a chart and extrapolate from it.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Positive Test | Negative Test | Total |
| Has Cancer | 475.00 | 25.00 | 500.00 |
| Does Not Have Cancer | 995.00 | 98,505.00 | 99,500.00 |
| Total | 1,470.00 | 98,530.00 | 100,000.00 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Outcome 4 | Outcome 2 | Outcome 4 | Sum | Num/Den |
| 475.00 | 995.00 | 475.00 | 1,470.00 | 0.32 |

32% of the people that test positive for cancer will in fact have cancer. Having a 95% confidence interval for this test does not make much sense at all given that 68% percent of the people who test positive will not have cancer but will be majorly emotionally engaged.

Week 6

How could two-sample testing be used to determine whether a company’s   
new proposed website is more effective than its existing website? (If   
you feel the need to define “effectiveness” of a website in order to   
answer this question, please do so.)

Merriam Webster’s definition of effective is, “producing a decided, decisive, or desired effect (merriam-webster.com).” For the stated example, the desired effect would be more hits on the web page as well as longer time spent on the web page after the new website is launched. For each desired effect, a previous mean covering perhaps a week or month would be recorded to establish one data set. After the launch of the website, new data would be recorded to establish a new mean for the website hits and time spent on page. One would conduct an independent t-test to analyze if there is a difference. The experiment would be non-directional given that a direction is unknown. The null hypothesis would be that page hits and page time has remained the same.

ANOVA can be used in a wide variety of areas of research. Describe  
a few concrete experimental situations in which ANOVA could be used to   
compare the effects of different treatments upon the population being   
studied.

Also discuss some of the assumptions and issues involved with both *t*-tests and ANOVA.  For example:  What assumptions underlie both procedures?  Which   
assumptions are the procedures robust with regard to?  What kind of   
research hypothesis Ha can be handled by *t*-testing but not ANOVA?  And so forth...

ANOVA would be helpful in discerning the effects of the following treatment experiments:

* Cholesterol drug that is supposed to lower cholesterol,
  + Compare the control group to the drug group
  + The hypothesis would be a lowering in cholesterol (Unidirectional).
  + Null hypothesis: there is no difference between the two groups.
* Erectile dysfunction drug, (results desired do not need explaining)
  + Compare the control group to the drug group
  + The hypothesis would be a change after taking the drug (Unidirectional).
  + Null hypothesis: there is no difference between the two groups.
* Acne treatment drug
  + Compare the control group to the drug group
  + The hypothesis would be a physical lessening of acne.
  + Null hypothesis: there is no difference between the two groups.

    The assumptions associated with ANOVA relate to the myopic nature of the test. One is assuming the change in one of the groups is caused by the experiment, not some other phenomenon. The procedure is robust in analyzing the precise changes from the groups being compared to each other. T-testing can analyze a wide variety of topics, it simply represents a sample population desiring to explain the greater population. Sales forecasting, polling, and research can all integrate t-tesing. ANOVA seems to be used when comparing one variable to two groups. While this is just our first session of ANOVA, I would expect to learn more techniques for ANOVA in the future.

Week 7

Imagine that we have collected the following data on each U.S. citizen   
over the age of 25: 1) level of education (expressed in years of   
schooling) and 2) amount of money donated to charities over the last   
year. How might the results of a Pearson’s correlation and of a linear   
regression for these two variables be useful?  What does regression tell   
us that correlation does not?

According to Sirkin, Pearson’s r is, “a measure of how close the point of distribution come to linearity” (p. 468, 2010). This is demonstrated through a 0-1 scale, expressing the degree of linear relationship. In the situation stated above, the Pearson’s correlation would be helpful in demonstrating how strong the linear relationship is between education (being the independent variable) and money donated to charities (dependent variable). Once Pearson’s r is calculated, one would better understand the linearity of the two variables, which leads to predictive modeling.

Linear Regression is a visual representation of the two variables being compared. Utilizing the Pearson’s r, one could create a linear regression model for any variety of points to predict giving trends based on education level. The regression is going to allow one to predict virtual variables. What is important to note is that the regression model accuracy is dependent on the Pearson’s r for the degree of reliability. While both tools are very helpful, Pearson’s r is key for an accurate regression model.

Reference

Sirkin, R. M. (2006). *Statistics for the social sciences* (3. ed.). Thousand Oaks, Calif.: Sage Publications.

Top of Form

n reviewing the last year’s data from a customer  
relationship management (CRM) system, a sales manager notes a 0.50 correlation  
between the number of sales calls made on major accounts (i.e. face-to-face  
meetings of more than thirty minutes) and account sales for the year. In the first quarter of the current year, the  
manager instructs sales representatives to double the number of sales calls on  
major accounts. They comply with her  
request, but there is no change in major account sales.  What are some possible explanations for this?

One thing is obvious about this sales manager, she did not take 401 Stats from NU. The situation is rather ambiguous with the correlation. Has the correlation been squared or has it not been squared already? I am assuming it has not been squared, thus the r squared equals .25. This correlative score does imply some amount of a relationship, so the sales manager is not completely inept. But, critical thinking is needed to better understand the relatively weak relationship.     
    Business sales is not as myopically focused as two variables. Given the r score, more variables should be added to the equation to strengthen the relationship. Perhaps exploring other variables in the CRM would reveal why no changes occurred in sales. The clients needs could already be satisfied, timing for purchasing could be different, and a myriad of other explanations could explain the stagnant sales. Also, one sales quarter is a relatively short time. Perhaps she needs to explore a larger timetable to better understand buying patterns and other variables.   

RE: Understanding Correlations

Week 8

Think of your favorite casino game (craps, black-jack, roulette, poker, keno, lottery etc.) and analyze the probability of a ***simple*** outcome.   
For example, what is the probability of getting three straight craps   
(sevens) in the game of craps, or the probability of being dealt a ‘21’   
in blackjack?  Be sure to state clearly what assumptions you're making   
to simplify your analysis.  You also need to find out what the payoff   
is, or assume a payoff value, for this bet.  Based on your analysis of   
this outcome, what  would be the expectation value for a single $100 bet  
on your outcome?  
  
Again, keep it ***simple***  
and straightforward so that you can explain it to others (and   
understand it yourself).  Especially with a game like poker,   
particularly Texas Hold'Em, the calculations can get very unwieldy for   
all but the simplest scenarios.  Don't get too ambitious!

Given that I am a tyro to this principle, I am going to keep it really simple and use the game of Roulette.  
  
Scenario 1: Bet on one of three colors to win, the following probabilities:  
Red: 18 out of 38: Probability (P) =  .474  
Black: 18 out of 38 P = .474  
Green: 2 out of 38 P = .052  
  
What is the probability that you will win four out of four bets on black wagering $25 each turn?   
  
P = .474  
Q = .526  
N = 4 Trials  
R = 4 Successes   
  
 Formula on Page 152 of The Binomial Probability Distribution   
  
Cnr = n!/ r!(n - r)!   
Work  4!/ 4!(4-4)! = 1   
  
CnrX(P^r)X(Q^n-r)  
Work 1X(.474^4)X(.526^0)= .05 Probability of winning all four spins.   
The Payoff value would be 100 dollars because the bet remains the same every time.     
  
The expectation value for this specific outcome would be very similar to a single bet for green with a probability of .052. Therefore the expected payout with a $100 bet on my outcome should be rewarded at a 8:1 ratio. Thus the payout expected would be $800.     
  
The payout for red or black is 1:1 & for green is 8:1

**Drawing on your own  
experience in work or life, discuss some examples of applications of Bernoulli processes, binomial random variables, and the probability models associated with them.**

Situation: What is the probability of being accepted into one, two, or all three graduate programs for which I applied.  
  
Binomial Random Variables:   
Either I get into the program or I get denied. (Succeed or Fail)  
  
For simplicity purposes: The probability of getting into each school is .3  
Each School’s decision is independent of the other schools.   
Each School allows one submission per year, thus my number of trials are 3.  
  
The Probability model can be used.  
  
CnrP^rQ^n-r  
  
N = 3  
R = 1,2,3  
P = .3  
Q = .7   
  
3!/(3-1)!\*1! = 3 |   3\*.3^1\*.7^2 = .441 Probability of being accepted at one of the three schools   
3!/(3-2)!\*2! = 1.5 | 1.5\*.3^2\*.7^1 = .09 Probability of being accepted at two of the three schools  
3!/(3-3)!\*3! = 1 | 1\*.3^3\*.7^0 = .027 Probability of being accepted at all three of the schools

Week 9

Looking over the results of a nationwide survey of retail customers who have shopped for clothing in the last six months (n =150,000), an apparel brand manufacturer finds that every categorical variable has a statistically significant relationship with every other categorical variable. What can the brand manager do to make sense out of this survey?

The Brand Manger has a number of options to make sense out of the survey. The starting point to better understanding the survey begins with defining "statistically significant relationship" as well as "variable". Depending on how many variables are in the survey, the manager will want to discern relationships between the variables. Specifically, finding which variables are independent and dependent would help prioritize which aspects to focus on to enhance customer experience. Given that we just learned about the Chi-test, I would recommend structuring the data to conduct such a test. The problem does not specifically state what kind of data the survey has collected. Is the data nominal, ordinal, ratio, or interval. Based on the answer to this question, the manager can conduct a specific test such as Gama, Lambda, Chi, or Phi to find specific relationships in the survey.