MAT 1372 Statistics with Probability Practice Final Exam\_answers Spring 2012

*Exam will be done in 2 parts. Once you have submitted part 1, you may work on part 2.*

**Part1: Please have your computer screens turned away. You may use a graphing calculator on part 1 but you may not rely on any of the statistical/probabilistic functions/tools. You should make use of the formula sheet from the Mann book.**

1. The probability that an unbalanced coin lands on head is 1/5. Coin is tossed 3 times a. List the sample space. The sample space is {HHH,HHT,HTH,HTT,THH,THT,TTH,TTT}

b. Construct a discrete probability distribution using X as the number of heads of the three-coin toss.

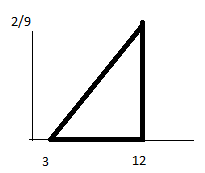
 or 

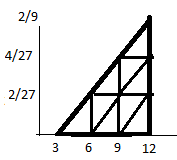
c. Find P(X<1), E(X) and E(1/(X+1))

P(X<1)=P(X=0)=.512, E(X)=.384+2(.096)+3(.008)=.6=3(.2) and

 , so E(1/(X+1))=.512+.384/2+.096/3+.008/4=.738

1. The probability density function is in the form of a right triangle which begins at 3 and rises to a peak at 12. Find P(X>9).

 b is 12-3=9, so ½ bh=1 so h=2/b=2/9

Points of line are (3,0) and (12,2/9)

Slope is m=(2/9-0)/(12-3)=2/81

y-y1=m(x-x1) or y-0=2/81(x-3)

or y=2/81(x-3)

Area of trapezoid is ½ (b1+b2)h

b1=y(9)= 2/81(9-3)=4/27, b2=2/9

A=½ (4/27+2/9)3=½ (4/27+6/27)3=5/9

1. a. Use graph paper to make a scatter plot.
2. Find the correlation coefficient and the regression line of the points.
3. If the x variable is the number of years in existence for a corner store and the y variable is profit in $10k, explain the meaning of the y-intercept and slope.
4. Predict the profit of a corner store after 9 years.



a. SSxy=∑xy-∑x∑y/n=146-25\*5/5=121

SSxx=∑x^2-(∑x)^2/n=199-(25)^2/5=74

SSyy=∑y^2-(∑y)^2/n=255-(5)^2/5=250

r=SSxy/Sqrt(SxxSyy)=121/sqrt(74\*250)

= 0.8896≈.89

b=SSxy/SSxx=121/74=1.635

a=y(bar)-bx(bar)=1-(121/74)(5)=74/74-605/74=-531/74=-7.176

b. Every year, the profit will increase by about $16k. The initial cost to start up the store is about $72k.

c. After 9 years, the store is predicted to make a profit of -7.2+1.6(9)=7.54 =$75k

1. If X is a random variable with E(X)= 2 and E(X^2)= 9, use Chebyshev’s inequality to find a lower bound for P(-1<X<5). V(X)=E(X^2)-E(X)^2=9-4=5 Hence σ=√5. Standardizing we find -3<X-2<3 or -3/√5<(X-2)/√5 <3/√5 or -3/√5<Z<3/√5. Hence we are looking at within 3/√5 standard deviations of the mean.

At least (1-1/k^2) lie within k std devs of mean. (1-1/(3/√5)^2)=(1-/(9/5))= 1-5/9)=4/9. So at least 44% of the values must lie within this interval, regardless of the distribution. In contrast, since 3/√5>1, the empirical rule says that at least 68% of the area lies within if it is approximately normal. A more precise measurement in that case is normsdist(3/√5)- normsdist(-3/√5)=.82 or 82%.

1. The distribution of grades on an quiz taken by 100 students is given in the table



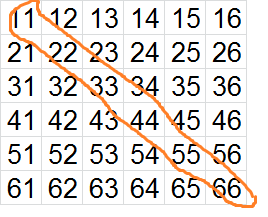
1. Compute the mean and standard deviation of this distribution. Use definition of standard deviation (weighted sum of deviations from mean square).

mean=10+80+90+80/100=2.6

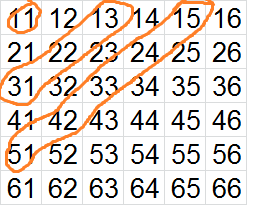


1. Describe mean & standard deviation of sample mean for random sample of size 16.

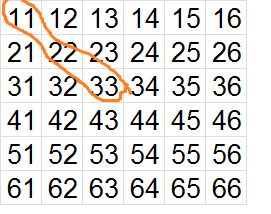
Random sample will have mean 2.6 and std dev= 0.9165/4= 0.2291

1. Suppose you roll two fair dice

(a) Determine each of these probabilities

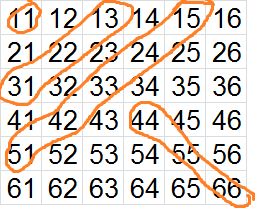
i. getting doubles: count 6 so 1/6

ii. getting an even sum of at most 6:

 count (36-6)/2=15 or 1+3+5=9 so 9/36=1/4

iii. getting doubles and an even sum of at most 6:

(1,1),(2,2),(3,3) so 3/36=1/12

iv. getting doubles or an even sum of at most 6

(use (i), (ii) and (iii))

P(A or B)=P(A)+P(B)-P(A and B)=1/6+1/4-1/12=(2+3-1)/12=1/3

(b) Which of the following pairs of events are disjoint? No proof needed if true.

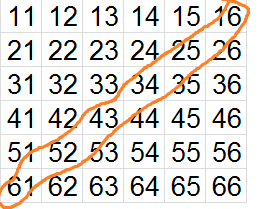
i. getting doubles, getting an even sum of at most 6 no (2,2)

ii. getting doubles, getting a sum of 7 yes

iii. getting doubles, getting a sum of 6 no (3,3)

iv. getting 2 on the ﬁrst die, getting 2 on the second die no (2,2)

(c) Which of these pairs of events are independent? Proof needed.

i. getting doubles, getting an even sum of at most 6:

P(A)P(B)=?P(A and B) 1/6 \*1/4=?1/12 NO!

ii. getting doubles, getting a sum of 7: 1/6\*1/6=?0 NO!

iii. getting doubles, getting a sum of 6: 1/6\*5/36=?1/36 NO!

iv. getting 2 on the ﬁrst die, getting 2 on the second die: 1/6\*1/6=?1/36 Yes

1. 4% of a clinic’s patients are known to have Lyme’s disease. A test is developed that is positive in 98% of patients with Lyme’s disease, but it is also positive in 3% of patients who do not have disease. A patient is chosen at random from clinic. Use this information to ﬁll in the following table. Do not to round your answers (use all available digits).



1. What is the probability that a random patient’s test comes out positive for Lyme disease? 6.8%
2. What is the probability that a person actually has Lyme’s disease given that the test comes out positive?3.92%/6.8%=0.576470588=58%

**Part 2: Use Excel to do your analysis and get your answers. Use a separate blue book. Please send via email the excel file. Note, the file will only be used for verification purposes and will not be directly graded. Put all that is asked for into the blue book.**

1. Is TV watching among children related to being overweight?

(col A: lbs overwt (or underwt), col B: hr/wk of TV watching).

1. Find the trend line and interpret the slope and y-intercept. y = 0.787x - 17.456
2. Every additional hour of television that a child watches per week translates to a gain of .79 pound. A child who does not watch any television will be about 17.5 lbs underweight.
3. What is the correlation coefficient? How closely are the variables correlated?

The correlation coefficient r is 0.69098, r^2 is .4775 or slightly less than 50% of the child’s weight increase might be explained by a television viewing increase. This is considered a moderate correlation.

1. What is the average number of lbs the children are overweight? 30.56 How many hours of TV do they view per week on average? 6.596
2. Verify that the coordinate corresponding to the answer to c. lies on the trend line.

y = 0.787\*30.56 - 17.456=6.595

1. A president’s Job Approval rating is one of the best indicators for assessing his/her chances of reelection. On May 5, 43% of potential voters say they at least somewhat approve of the president's job performance. 53% at least somewhat disapprove. You would like to see if the national opinion is the same as that of New York State. You do a poll of 1000 potential voters randomly sampled from the entire state and would like to find out if New York’s approval of the President differs from that of the nation as a whole. (See the excel file for detailed data.)
2. What are the null and alternative hypotheses?

H0=NYS support for Obama is same as national

H1= NYS support for Obama is not same as national

1. Calculate the Test Statistic and the degree of freedom.

TS=423 and df=4

1. Calculate the p value=chidist(423,4)=0 (3E-90)

At the α=5% significance level, what can you say? Clearly, Obama has more support in NYS compared with the rest of the nation, at any significance level.

1. On average, Rona’s computer crashes once every 2 days and the distribution is Poisson.
2. On average, how many times will her computer crash in one day? 1/2
3. Find the probability that her computer will not crash on that day. =poisson(0,1/2,true)=.607
4. Find the probability that her computer will crash at least once on that day (Hint: it is the complement event to (b))=.393
5. Calculate the chance of her first crash happening on day 1, day 2,…, day 10.

Hint: for day 5, for instance, multiply the probability of not having a crash for 4 days and the probability of having a crash on the next day. See spreadsheet.

1. Use the chart tool to display the bar chart of probabilities from d. You should have 10 columns (1 to 10). See spreadsheet. Transfer a rough sketch to your exam.
2. The FDA has decided to measure 8-ounce cans of Hershey Cocoa. The weights of 50 containers are found in the accompanying excel file.
3. What are the null and alternative hypotheses?

H0: µ=8

H1: µ<8

1. Use a critical value approach to the t-test at the α = 5% significance level.

Xbar=7.9882, s=.04034, sxbar=s/sqrt(50)=.0057

Note that tinv assumes a two-tailed test. As the Tinv function help file explains, you can use Tinv for a one-tailed test by doubling α. In other words, our input into tinv, will be .10,49. .10=2\*.05 and df=n-1=50-1

Crit val= -.0057\*tinv(.10,49)+8=7.9904

Hence xbar falls within our rejection region so we reject the null hypothesis. In other words, we have strong evidence that Hershey is not putting enough Cocoa into its cans.

(c) Make a histogram to verify that the weights are approximately normally distributed.

Transfer a rough sketch to your exam. We use .02 as our bin widths. (see spreadsheet). The distribution is roughly a normal distribution with its right tail cut off. While not a normal distribution, our sample size is high enough that our sample mean will have a distribution which is very close to normal.

1. Do people spend more time reading articles that agree with their disposition? An antismoking group conducted a survey asking those who saw an advertisement against smoking if (1) they read headline only, (2) some detail, or (3) most of the advertisement. The questionnaire asks respondents to identify themselves as heavy smokers (1), moderate smokers (2), light smokers (3) or nonsmokers (4). The result is found in the accompanying excel file (col A: type of smoker, col B: amt of advertisement read). Does how much you smoke affect how much you read of an antismoking advertisement? Let the null hypothesis be that it does not. Use countifs to create the contingency table. Use a pvalue approach to decide whether to reject the null hypothesis at the 1% and 5% levels. The pvalue is .00002 which means that the null hypothesis will be rejected at any level. In other words, smokers less likely to read ads attacking smoking than nonsmokers.