Answers

Department of Mathematics (OU)

MATH 4753 Exam 1 Chapters 1-5 Due: Friday Feb. 27 2015

Name:
OUID:

Please answer **any 8 of the 10** questions. You can answer in any order you wish. Please hand in all the exam questions and answers at the end of the class.

Question	Marks earned	Out of
Q1		4
Q2		4
Q3		4
Q4		4
Q5		4
Q6		4
Q7		4
Q8		4
Q9		4
Q10		4
Total	Out of Best 8	32

Table 1: ANSWERS

Formulae you might need:

RV	$p(y)$	μ	σ^2	$m(t)=\sum_y e^{yt}p(y)$
Bernoulli	$p(y)=p^yq^{1-y}, y=0,1$	p	pq	pe^t+q
Binomial	$p(y)=\binom{n}{y}p^yq^{n-y}, y=0,1,\ldots,n$	np	npq	$(pe^t+q)^n$
Neg. Bin.	$p(y)=\binom{y-1}{r-1}p^rq^{y-r}, y=r,r+1,r+2,\ldots$	$\frac{r}{p}$	$\frac{rq}{p^2}$	Not given
Multinomial	$p(y_1,y_2,\ldots,y_k)=\frac{n!}{y_1!y_2!\ldots y_k!}(p_1)^{y_1}(p_2)^{y_2}\cdots(p_k)^{y_k}$	np_i	$np_i(1-p_i)$	Not given
Hyper-geom.	$p(y)=\frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}}$	$\frac{nr}{N}$	Not given	Not given
Poisson	$p(y)=\frac{\lambda^y e^{-\lambda}}{y!}$	λ	λ	$e^{\lambda(e^t-1)}$

Table 2: Formulae

Bayes rule:

$$p(A_i|B)=\frac{p(A_i)p(B|A_i)}{p(B)}$$

where $p(B)=\sum_{j=1}^n p(A_j)p(B|A_j)$

$$E((X-\mu)^2)=E(X^2)-\mu^2$$

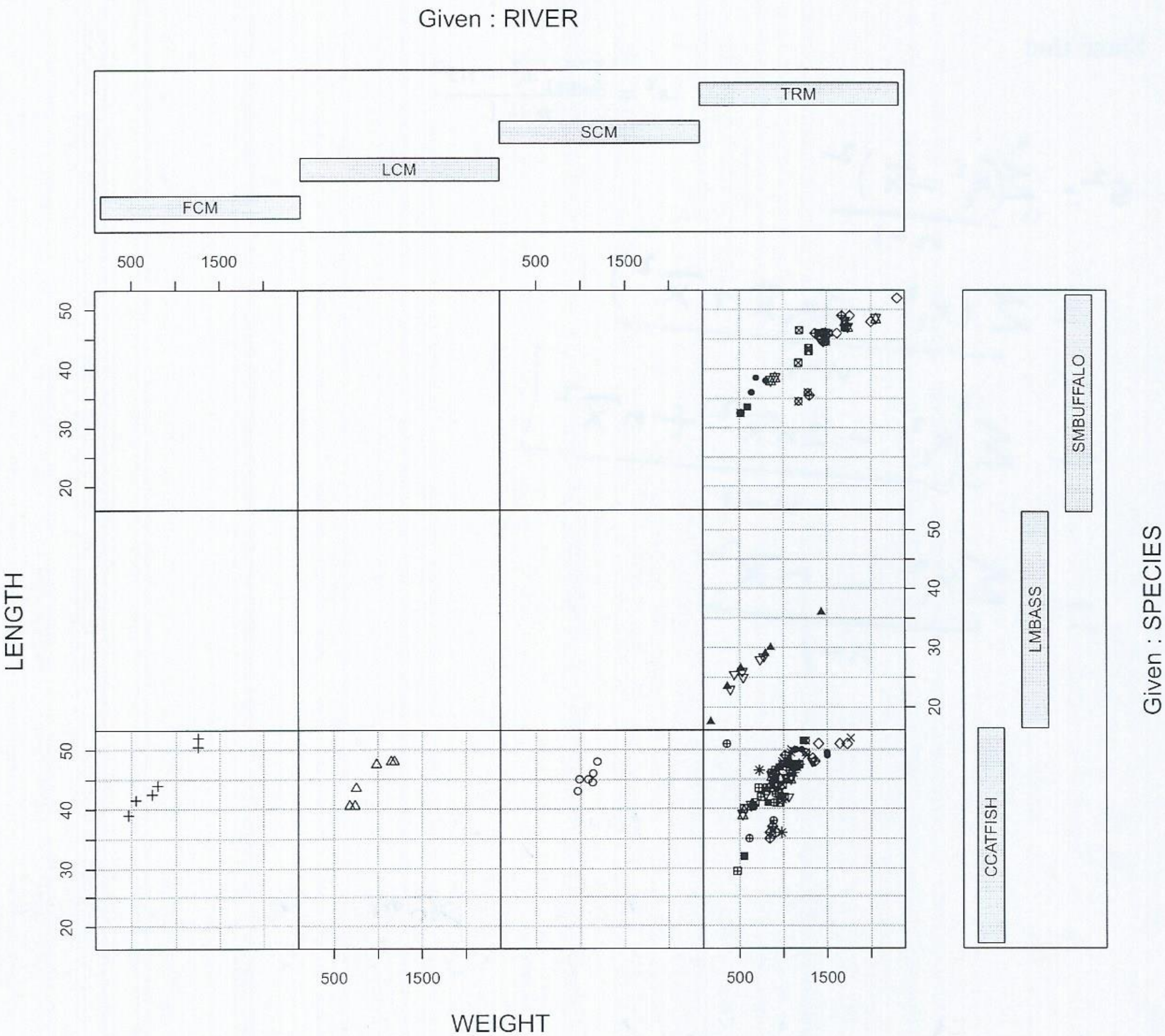


Figure 1: Coplot of Length Vs weight

3. To standardize a data set, x, we often use a z transformation, where

$$z_i = \frac{x_i - \overline{x}}{s}$$

To standardize a population with variable X we often use a Z transformation, where

$$Z = \frac{X - \mu}{\sigma}$$

Show the following

(a) For a sample: $\overline{z} = 0$

(b) For a sample: $sd(z) = 1$

(c) For a population $E(Z) = 0$

(d) For a population $V(Z) = 1$

(a) $\bar{z} = \frac{\sum z_i}{n} = \frac{\sum \frac{(x_i - \bar{x})}{s}}{n} = \frac{1}{ns}\left[\sum x_i - n\bar{x}\right]$

$= \frac{1}{ns}\left[n\bar{x} - n\bar{x}\right]$

$= 0$

(b) $sd(z) = \frac{\sum_{i=1}^{n} (z_i - \bar{z})^2}{n-1}$

$= \frac{\sum z_i^2}{n-1} = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{s^2(n-1)} = \frac{s^2}{s^2} = 1$

(c) $E(Z) = E\left(\frac{X - \mu}{\sigma}\right) = \frac{1}{\sigma}E(X) - \frac{\mu}{\sigma} = \frac{\mu}{\sigma} - \frac{\mu}{\sigma} = 0$

(d) $V(Z) = V\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma^2}V(X) = \frac{\sigma^2}{\sigma^2} = 1$

5. If $S_k = \{i, 1 \le i \le n : |x_i - \overline{x}| < ks\}$ where $k > 1$ and s is the standard deviation of the sample, prove that

$$\boxed{\frac{N(S_k)}{n} > 1 - \frac{1}{k^2}}$$

$N(S_k)$ is the number of elements in S_k.

See D2L.

7. If $Y = 3X - 2$ and $X \sim N(\mu = 2, \sigma = 1)$ evaluate:

(a) $E(Y)$

(b) $V(Y)$

If $H = \frac{4D-\mu}{\sigma}$ and $D \sim N(\mu, \sigma)$ give simplified expressions for (working *MUST* be shown):

(a) $E(H)$

(b) $V(H)$

$E(Y) = E(3X-2) = 3E(X) - 2 = 3\times 2 - 2 = 4$

$V(Y) = V(3X-2) = 3^2 V(X) = 9 \times 1^2 = 9$

$E(H) = E\left(\frac{4D-\mu}{\sigma}\right) = \frac{4E(D)}{\sigma} - \frac{\mu}{\sigma} = \frac{4\mu-\mu}{\sigma} = \frac{3\mu}{\sigma}$

$V(H) = V\left(\frac{4D-\mu}{\sigma}\right) = \frac{16}{\sigma^2} V(D) = \frac{16}{}$

9. The following is code used to make plots of normal densities, create a polygonal area, calculate the area and plot it on the graph.

```
norma<-function(mn=0,std=1,a=0,b=2,colo="Green",xlimits=c(-4,4),xlab="x"){
curve(dnorm(x,mean=mn,sd=std),xlim=xlimits,xlab=xlab,ylab="Density",
main=paste("Normal with mean=",mn, "sd=",std),las=1)
curvex=seq(a,b,length=1000) #A
curvey=dnorm(curvex,mean=mn,sd=std)  #B
polygon(c(a,curvex,b),c(0,curvey,0),col=colo)
av=(a+b)/2   # halfway beteen a and b
A=pnorm(b,mean=mn,sd=std)-pnorm(a,mean=mn,sd=std)
if(abs(a)>abs(b)){
text(b,0.5*dnorm(b,mean=mn,sd=std),paste("Area=",round(A,4)),cex=1.5)
}
else{
text(a,0.5*dnorm(a,mean=mn,sd=std),paste("Area=",round(A,4)),cex=1.5)
}
}
norma(xlimits=c(-4,15),mn=4,std=3,a=5,b=10)
```

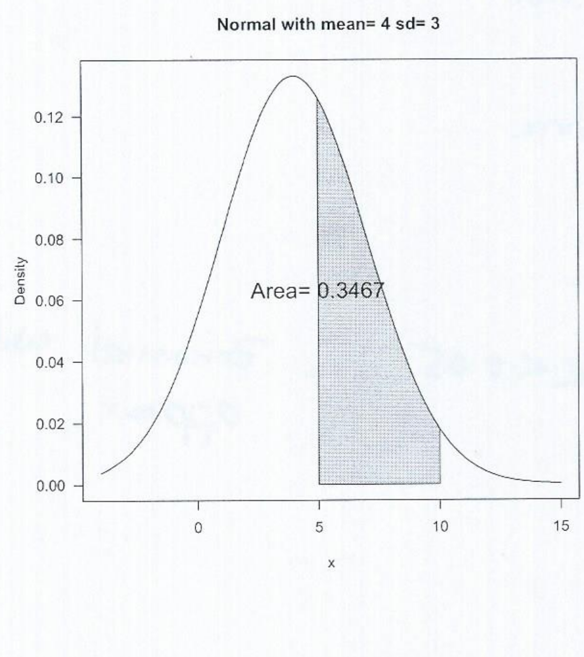
(a) `norma(xlimits=c(-4,15),mn=4,std=3,a=5,b=10)` is used to make the plot below. Find $P(5 \leq X < 10)$ 0.3467

(b) How many x values are created in line A? 1000

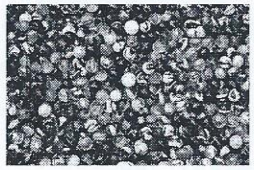
(c) How many y values are created in line B? 1000

(d) What color would the area in the plot below be seen if created and viewed on the computer?

Green



10. A barrel contains 3000 marbles, 1000 black marbles and 2000 white marbles. A random sample of 5 marbles is drawn without replacement. Y is the number of black marbles drawn.



(a) Classify this problem in terms of the *precise* underlying distribution of Y.

(b) Using the output below, what is the probability that three of the five marbles are black? You must give the probability to 5 dec places

(c) If the marbles are taken one at a time and each marble is replaced and the marbles randomized before the next marble is taken and replaced - what is the probability that after 5 such trials there are three black marbles?

(d) Why do the binomial and hyper-geometric distributions seem to give close answers? When do they appreciably differ?

```
> dhyper(3,m=1000,n=2000,k=5)
[1] 0.1645815
> dbinom(3,prob=1/3,size=5)
[1] 0.1646091
> phyper(3,m=1000,n=2000,k=5)
[1] 0.9548697
> pbinom(3,prob=1/3,size=5)
[1] 0.9547325
> dhyper(3,m=10,n=20,k=5)
[1] 0.1599933
```

a) hypergeometric

b) 0.16458

c) 0.16461

d) $\frac{n}{N} = \frac{5}{3000} << 0.05$ $\therefore$ Binomial will be a good approx.

8. To attach the housing on a motor, a production line assembler must use an electrical hand tool to set and tighten *four* bolts. Suppose that the probability of setting and tightening a bolt in any 1-second time interval is $p = 0.8$. If the assembler fails in the first second, the probability of success during the next 1-second interval is 0.8, and so on.

(a) Find the probability distribution of Y, the length of time until a complete housing is attached.

(b) Find $p(6)$.

(c) Find the mean Y.

(d) Find the variance of Y.

a) $Y \sim \text{Neg Bin}(p=0.8, r=4)$

b) $p(6) = \binom{6-1}{4-1} 0.8^4 0.2^2$

$= \binom{5}{3} 0.8^4 0.2^2$

$= \frac{5 \times 4 \times \cancel{3!}}{2! \, \cancel{3!}} 0.8^4 0.2^2$

$= 10 \cdot 0.8^4 0.2^2$

$=$

c) $E(Y) = \frac{r}{p} = \frac{4}{0.8} = 5$

d) $V(Y) = \frac{rq}{p^2} = \frac{4 \times 0.2}{0.8^2} =$

6. Suppose a drug test is 99% sensitive and 98% specific. That is, the test will produce 99% true positive results for drug users and 98% true negative results for non-drug users. Suppose that 0.4% of people are users of the drug. If a randomly selected individual tests positive, what is the probability he or she is a user?

Use good notation.

Notice what we want! $p(u|+)$

Use Bayes Rule

$$p(u|+) = \frac{p(u)p(+|u)}{p(u)p(+|u) + p(\bar{u})p(+|\bar{u})} \quad (*)$$

$$p(u) = 0.4\% = \frac{0.4}{100} = \frac{4}{1000}$$

$$p(\bar{u}) = 1 - \frac{4}{1000} = \frac{996}{1000}$$

$$p(+|\bar{u}) + p(-|\bar{u}) = 1$$

$$p(+|\bar{u}) + 0.98 = 1$$

$$p(+|\bar{u}) = 0.02$$

Now fill in the rest of $(*)$

$$p(u|+) = \frac{\frac{4}{1000} \times 0.99}{\frac{4}{1000} \times 0.99 + \frac{996}{1000} \times 0.02}$$

$$= 0.1658291$$

$$= \underline{16.6\%}$$

4. Using the definition of a Bernoulli random variable $X \sim Bern(p)$, derive the moment generating function for

(a) The Bernoulli, $M_X(t) = E(e^{Xt})$

(b) If $Y = X_1 + X_2 + \ldots + X_n$ where $X_i \sim Bern(p)$, derive the moment generating function for the Binomial.

(c) Using $M_Y(t)$ find the population mean.

(d) Using $M_Y(t)$ find the population variance

Remember that if $M_X(t)$ is a moment generating function then

$$\mu_k' = \left. \frac{d^k M_X(t)}{dt^k} \right|_{t=0}$$

(a) $M_X(t) = E(e^{Xt}) = \sum_x e^{xt} p(x) = e^{0t} p(0) + e^{t} p(1)$
$= q + pe^t.$

(b) $Y = X_1 + \ldots + X_n$; $X_i \sim Bern(p)$

$M_Y(t) = E(e^{Yt}) = E(e^{\Sigma X_i t}) = E(e^{X_1 t}) \ldots E(e^{X_n t})$
$= (M_X(t))^n$
$= (q + pe^t)^n.$

(c) $\mu_Y = \left. \frac{d}{dt} M_Y(t) \right|_{t=0} = np.$

(d) $\mu_2' = \left. \frac{d^2}{dt^2} M_Y(t) \right|_{t=0}$ = see note.

Then use $\sigma^2 = \mu_2' - (\mu_1')^2$

2. The formula for the sample variance is

$$s^2 = \frac{\sum_{i=1}^n (x_i - \overline{x})^2}{n-1}$$

Show that

$$s^2 = \frac{\sum_{i=1}^n x_i^2 - n\overline{x}^2}{n-1}$$

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$$

$$= \frac{\sum (x_i^2 - 2x_i\overline{x} + \overline{x}^2)}{n-1}$$

$$= \frac{\sum x_i^2 - 2n\overline{x}^2 + n\overline{x}^2}{n-1}$$

$$= \frac{\sum x_i^2 - n\overline{x}^2}{n-1}$$

1. A biologist wants to make a coplot of **LENGTH Vs WEIGHT** given **RIVER*SPECIES** for fish caught in rivers that feed into the Tennessee river and recorded in the **DDT.xls** data set, so that each point is plotted with a symbol according to the variable **MILE** which is treated as a factor (Qualitative variable). The biologist approaches a MATH 4753 student knowing that he or she is well instructed in statistics and capable of helping in this project.
The MATH 4753 student makes the plot (see figure 1) using the code below.

```
> head(ddt)
  RIVER MILE  SPECIES LENGTH WEIGHT DDT
1   FCM    5 CCATFISH   42.5    732  10
2   FCM    5 CCATFISH   44.0    795  16
3   FCM    5 CCATFISH   41.5    547  23
4   FCM    5 CCATFISH   39.0    465  21
5   FCM    5 CCATFISH   50.5   1252  50
6   FCM    5 CCATFISH   52.0   1255 150
# The following code may help
m=with(ddt, as.numeric(levels(factor(MILE))))
colm=c()
for(i in 1:length(ddt$MILE)){
 colm[i]=which(ddt$MILE[i]==m)
}
> colm
  [1]  3  3  3  3  3  3  2  2  2  2  2  2  1  1  1  1  1  1  4  4  4  4  4  4  5  5  5
 [30]  5  5  5  5  5  5  5  6  6  6  6  6  6  6  6  6  6  6  6  7  7  7  7  7  7  7  7
 [59]  7  7  8  8  8  8  8  8  9  9  9  9  9  9  9  9  9  9  9  9 10 10 10 10 10 10 11 1
 [88] 11 11 11 11 11 11 11 11 11 12 12 12 12 12 12 13 13 13 13 13 13 13 13 13 13 13 1
[117] 14 14 14 14 15 15 15 15 15 15 15 15 16 16 16 16 16 16 16 16 16 16 17 17 17 17 17 1

coplot(LENGTH~WEIGHT|RIVER*SPECIES,data=ddt,pch=colm) #A
```

(a) What are the categorical variables? River, Species

(b) MILE is a quantitative variable what are the other quantitative variables? Length, Weight, DDT

(c) What does pch=colm do in line A? Gives a plotting character to each mile level

(d) Why are the top six plots empty? No data there ie No Fish caught of Species LMBASS & SMBUFFALO in the NON TRM Rives.