Week Three Review Answers

1. (a) On the interval (a,b) the uniform density function is $\frac{1}{b-a}$.
   Since b=1, and a=0, b-a=1 and
   $\int_{0.55}^{0.65} dx = x|_{0.55}^{0.65} = 0.65 - 0.55 = 0.1$

   (b) $\int_{0.55}^{0.65} dX_1 \int_{0.3}^{0.4} dX_2 = (0.1) \times (0.1) = 0.01$

   (c) $\int dX_1 ..100.. \int dX_{100} = 0.1^{100} = 10^{-100}$

   (d) It's impossible to get any observation within the prescribed distance.

   (e) Find c_1 and c_2 such that $\int_{c_1}^{c_2} dX = 0.1$
   Let the length of side be c_2-c_1=δ. Then when p=1, δ =0.1
   when p=2, δ^2=0.1 or δ=0.31
   when p=100, δ^{100}=0.1 or δ=0.977

   At large p the hypercube sides include the vast range of each variable, so you can't get very good predictions.
2. The probability that a student will receive an A can be determined from the logistic model, $p(X) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2}}{1+e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2}}$.
   (a) Substituting the parameter values gives $\frac{e^{-6+0.05(40)+1(3.5)}}{1+e^{-6+0.05(40)+1(3.5)}} = 0.378$
   (b) We know that the log of the odds ratio equals $\beta_0 + \beta_1 x_1 + \beta_2 x_2$ or
   $$\log\left(\frac{0.5}{1-0.5}\right) = -6 + 0.05x_1 + 1(3.5)$$
   $$\frac{6-3.5}{0.05} = 50 \text{ hrs} = X_1$$
3. Stock Market Data

   Can we predict the direction of stock market change based on the five previous days (lag 1-5) of change, today's volume and date (year)?

   Based on the previous logistic regression results only Lag1 and Lag2 will be used and the training data will come from year 2005.

```
> library(ISLR2)        attach(Smarket)
> unique(Year)

[1] 2001 2002 2003 2004 2005

> train <- (Year < 2005)
```

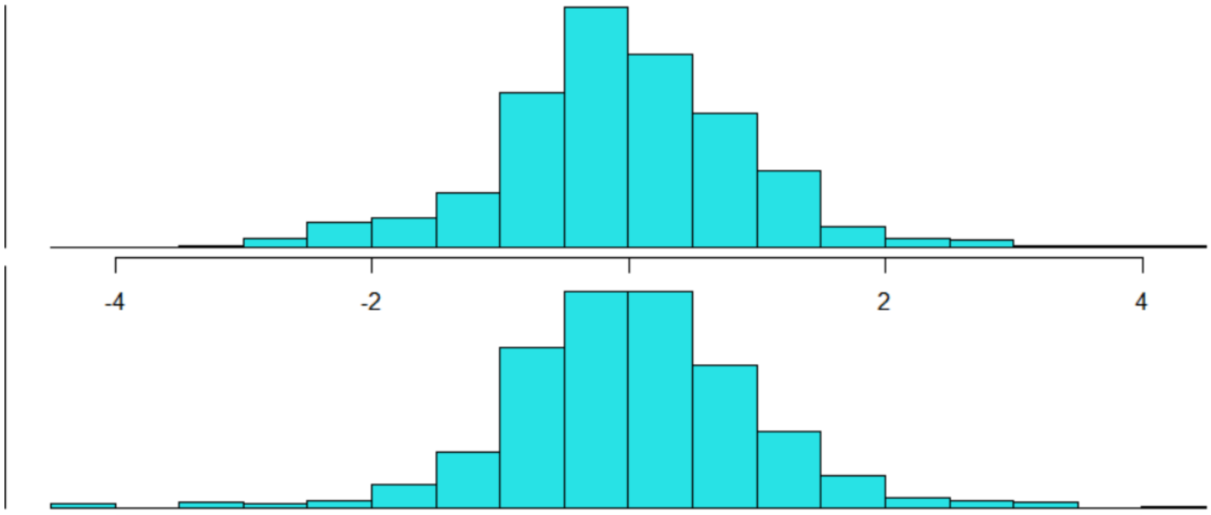
```
> Smarket.2005 <- Smarket[!train, ]# all the Smarket data
                   not equal to 2001, 2002, 2003, and 2004
> library(MASS)
> lda.fit <- lda(Direction ~ Lag1 + Lag2 , data = Smarket ,
+                subset = train)
> lda.fit
Call:
lda(Direction ~ Lag1 + Lag2, data = Smarket, subset =
train)

Prior probabilities of groups:
    Down       Up
0.491984 0.508016

Group means:
            Lag1        Lag2
Down  0.04279022  0.03389409
Up   -0.03954635 -0.03132544

Coefficients of linear discriminants:
            LD1
Lag1 -0.6420190
Lag2 -0.5135293
> contrasts(Direction)
     Up
Down  0
Up    1
#########
# To predict class membership Lag1×(-0.64)+ Lag2×(-0.51)
```

```
# If this is large -> predict Up, small -> predict down
> par(mar=c(0,0,0,0))
> plot(lda.fit)
```




```
> lda.pred <- predict(lda.fit , Smarket.2005) # Use the
                                                test set of 2005 data
> names(lda.pred) # lda.pred is a list of three items with
                    these names
[1] "class"     "posterior" "x"
# class are the lda predictions
# posterior has a column for each class with the posterior
# probability predicted from equation (4.15)
# and x is the discriminant function
> Direction.2005 <- Direction[!train] #Observed directions
                                                    in 2005


> table(lda.class, Direction.2005)
         Direction.2005
lda.class Down  Up
```

```
     Down   35  35
     Up     76 106
> mean(lda.class == Direction.2005) # This works since
          FALSE has a numerical value of 0 and TRUE, 1
[1] 0.5595238 #This is just a little better than guessing
               with the prior probabilities
```

 # The total number of samples predicted to be Down was 70. These correspond to posterior probabilities greater than 50%.

```
> sum(lda.pred$posterior[, 1] >= .5)
[1] 70
```

 # You can increase this probability to be more certain of a correct calculation although with these data none of the probabilities is much greater than 50%.