Assignment 7
Modern Applied Statistics
Shaheed Shihan

1. Using basic statistical properties of the variance, as well as single variable calculus, derive (5.6). In other words, prove that α given by (5.6) does indeed minimize Var(αX + (1 − α)Y).

$$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$$
$$Var(aX) = a^2 Var(X)$$
$$Cov(aX, bY) = abCov(X,Y)$$

$$Var(aX + (1-a)Y) = Var(aX) + Var\big((1-a)Y\big) + 2Cov(aX, (1-a)Y)$$
$$= a^2 Var(X) + (1-a)^2 Var(Y) + 2a(1-a)Cov(X,Y)$$
$$f(a) = \sigma_X^2 a^2 + \sigma_Y^2 (1-a)^2 + 2\sigma_{XY}(-a^2 + a)$$

Taking the first derivative:

$$\frac{d}{da} f(a) = 0$$
$$2\sigma_X^2 a + 2\sigma_Y^2 (1-a)(-1) + 2\sigma_{XY}(-2a+1) = 0$$
$$\sigma_X^2 a + \sigma_Y^2 (1-a) + \sigma_{XY}(-2a+1) = 0$$
$$a(\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}) + \sigma_{XY} - \sigma_Y^2 = 0$$
$$\frac{\sigma_Y^2 - \sigma_{XY}}{\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}}$$

6.

a.

```
Call:
glm(formula = default ~ balance + income, family = "binomial",
   data = D)

Deviance Residuals:
   Min       1Q   Median       3Q      Max
-2.4725  -0.1444  -0.0574  -0.0211   3.7245

Coefficients:
            Estimate Std.   Error       z value     Pr(>|z|)
(Intercept) -1.154e+01      4.348e-01   -26.545     < 2e-16 ***
balance      5.647e-03      2.274e-04   24.836      < 2e-16 ***
income       2.081e-05      4.985e-06   4.174       2.99e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2920.6  on 9999  degrees of freedom
Residual deviance: 1579.0  on 9997  degrees of freedom
AIC: 1585

Number of Fisher Scoring iterations: 8
```

c.

```
ORDINARY NONPARAMETRIC BOOTSTRAP


Call:
boot(data = D, statistic = boot_fn, R = 1000)


Bootstrap Statistics :
        original          bias         std. error
t1* -1.154047e+01   -1.929224e-02   4.260635e-01
t2*  2.080898e-05   -4.043053e-09   4.824970e-06
t3*  5.647103e-03    1.261121e-05   2.212360e-04
```

d. The standard error for both the bootstrapping and standard glm seem to similar to one another.

9.

a.

```
> mu <- mean(medv)
> mu
[1] 22.53281
```

b.

```
> se <- sd(medv)/sqrt(dim(Boston)[1])
> se
[1] 0.4088611
```

c.

```
ORDINARY NONPARAMETRIC BOOTSTRAP
```

```
Call:
boot(data = medv, statistic = boot_f2, R = 1000)


Bootstrap Statistics :
    original        bias       std. error
t1* 22.53281     0.00801581    0.408103
```

The error rate computed by both methods are very similar.

d.

```
One Sample t-test

data:  medv
t = 55.111, df = 505, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
21.72953 23.33608
sample estimates:
mean of x
22.53281
```

```
> Confidence_interval <- c(22.53 - 2 * 0.4119, 22.53 + 2 * 0.4119)
> Confidence_interval
[1] 21.7062 23.3538
```

Once again, both confidence intervals are very similar.

e.

```
> med <- median(medv)
> med
[1] 21.2
```

f.

```
ORDINARY NONPARAMETRIC BOOTSTRAP

Call:
boot(data = medv, statistic = boot_f3, R = 1000)


Bootstrap Statistics :
```

```
   original   bias      std. error
t1*   21.2   -0.0264   0.3731921
```

The estimated median value using bootstrapping is 21.2 which is the value we have from e. Also, the standard error in computing this value is small as compared to the true median value.

g.

```
> perc
  10%
12.75
```

h.

```
ORDINARY NONPARAMETRIC BOOTSTRAP


Call:
boot(data = medv, statistic = boot_f4, R = 1000)


Bootstrap Statistics :
      original   bias      std. error
t1*   12.75      0.00355   0.4978984
```

The values obtained in both g and h are the same for the 10th percentile. The standard error is also relatively small.