Charles Hwang
Dr. Xi
STAT 408-001
18 September 2022

Homework 1

1 A) $E(X) = \sum_{x \in D} x \cdot p(x) =$

$1 \cdot 0.05 + 2 \cdot 0.1 + 4 \cdot 0.35 + 8 \cdot 0.4 + 16 \cdot 0.1 =$

$0.05 + 0.2 + 1.4 + 3.2 + 1.6 = \boxed{6.45}$

B) $V(X) = \sum_{x \in D} (x-\mu)^2 \cdot p(x) =$

$(1-6.45)^2 \cdot 0.05 + (2-6.45)^2 \cdot 0.1 + (4-6.45)^2 \cdot 0.35 + (8-6.45)^2 \cdot 0.4 + (16-6.45)^2 \cdot 0.1 =$

$(-5.45)^2 \cdot 0.05 + (-4.45)^2 \cdot 0.1 + (-2.45)^2 \cdot 0.35 + (1.55)^2 \cdot 0.4 + (9.55)^2 \cdot 0.1 =$

$29.7025 \cdot 0.05 + 19.8025 \cdot 0.1 + 6.0025 \cdot 0.35 + 2.4025 \cdot 0.4 + 91.2025 \cdot 0.1 =$

$1.485125 + 1.98025 + 2.100875 + 0.961 + 9.12025 =$

$\boxed{15.6475}$

c) $SD(X) = \sqrt{V(X)} = \sqrt{15.6475} = \sqrt{\frac{6259}{400}} = \frac{\sqrt{6259}}{20} \approx \boxed{3.955692}$

2 A) Sample mean: $\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{0.83+0.88+0.88+1.04+1.09+1.12+1.29+1.31+1.48+1.49+1.59+1.62+1.65+1.71+1.76+1.83}{16} =$

$\frac{21.57}{16} = \boxed{1.348125}$

B) Sample variance: $\hat{\sigma}^2 = S^2 =$

$\frac{\sum (x_i - \bar{x})^2}{n-1} \approx \frac{1.719044}{16-1} = \frac{1.719044}{15} \approx \boxed{0.1146029}$

3 A) $\bar{x} \pm z^* \cdot \frac{\sigma}{\sqrt{n}} =$

$58.3 \pm 1.96 \cdot \frac{3}{\sqrt{25}} =$

$58.3 \pm 1.96 \cdot \frac{3}{5} =$

$58.3 \pm 1.176 =$

$\boxed{(57.124, 59.476)}$

B) $\bar{x} \pm z^* \cdot \frac{\sigma}{\sqrt{n}} =$

$58.3 \pm 1.96 \cdot \frac{3}{\sqrt{100}} =$

$58.3 \pm 1.96 \cdot \frac{3}{10} =$

$58.3 \pm 0.588 =$

$\boxed{(57.712, 58.888)}$

c) $\bar{x} \pm z^* \cdot \frac{\sigma}{\sqrt{n}} =$

$58.3 \pm 2.58 \cdot \frac{3}{\sqrt{25}} =$

$58.3 \pm 2.58 \cdot \frac{3}{5} =$

$58.3 \pm 1.548 =$

$\boxed{(56.752, 59.848)}$

4 A) $H_0: \mu = 100$ pounds-per-~~inch~~square-inch

$H_A: \mu > 100$ pounds-per-square-inch

There needs to be statistically significant evidence that the mean strength of welds exceeds 100 pounds-per-square-inch.

B) Type I error: Declaring the mean strength of welds exceeds 100 pounds-per-square-inch when it does not in reality

Type II error: Declaring the mean strength of welds does not exceed 100 pounds-per-square-inch when it does in reality

Homework 1

Charles Hwang

9/18/2022

Problem 5

Problem 5(a)

```
rm(list=ls())
NCbirths<-read.csv("/Users/newuser/Desktop/Notes/Graduate/STAT 408 - Applied Regression Analysis/births
```

Problem 5(b)

```
weights<-NCbirths$weight
```

Given these data are for newborn babies, it appears these weights are listed in ounces.

Problem 5(c)

```
weights_in_pounds<-weights/16
```

Problem 5(d)

```
head(weights_in_pounds,20)
```

```
##  [1]  7.7500 11.0625  6.6875  9.0000  7.3125  6.1250  9.1875  8.6250  6.5000
## [10]  7.6875  9.5625  8.0625  7.4375  6.7500  6.6250  7.8125  7.1875  8.0000
## [19]  8.2500  5.1875
```

Problem 5(e)

```
mean(weights_in_pounds)
```

```
## [1] 7.2532
```

We can see the mean weight of all babies is 7.2532003 pounds.

Problem 5(f)

```
table(NCbirths$Habit)
```

```
##
## NonSmoker    Smoker
##      1805       187
```

We can see that approximately $\frac{187}{1805+187} = \frac{187}{1992} \approx 9.3875502$ percent of mothers in the sample smoke.

Problem 5(g)

```
abs(14-as.numeric(100*table(NCbirths$Habit)["Smoker"]/sum(table(NCbirths$Habit))))
```

```
## [1] 4.61245
```

```
100*(1-as.numeric(100*table(NCbirths$Habit)["Smoker"]/sum(table(NCbirths$Habit)))/14)
```

```
## [1] 32.94607
```

The percentage I found in problem 5(f) is approximately 4.6124498 percentage points off from the percentage cited in the CDC report. Alternatively, we can see the percentage from 5(f) is approximately 32.94607 percent less than the percentage cited in the CDC report.

Problem 6

Problem 6(a)

```
flint<-read.csv("/Users/newuser/Desktop/Notes/Graduate/STAT 408 - Applied Regression Analysis/flint.csv
```

Problem 6(b)

```
mean(flint$Pb>=15)
```

```
## [1] 0.04436229
```

We can see that approximately 4.4362292 percent of the locations tested were found to have dangerous lead levels.

Problem 6(c)

```
mean(flint[flint$Region=="North","Cu"])
```

```
## [1] 44.6424
```

Problem 6(d)

```
mean(flint[flint$Region=="North" & flint$Pb>=15,"Cu"])
```

```
## [1] 255.7143
```

Problem 6(e)

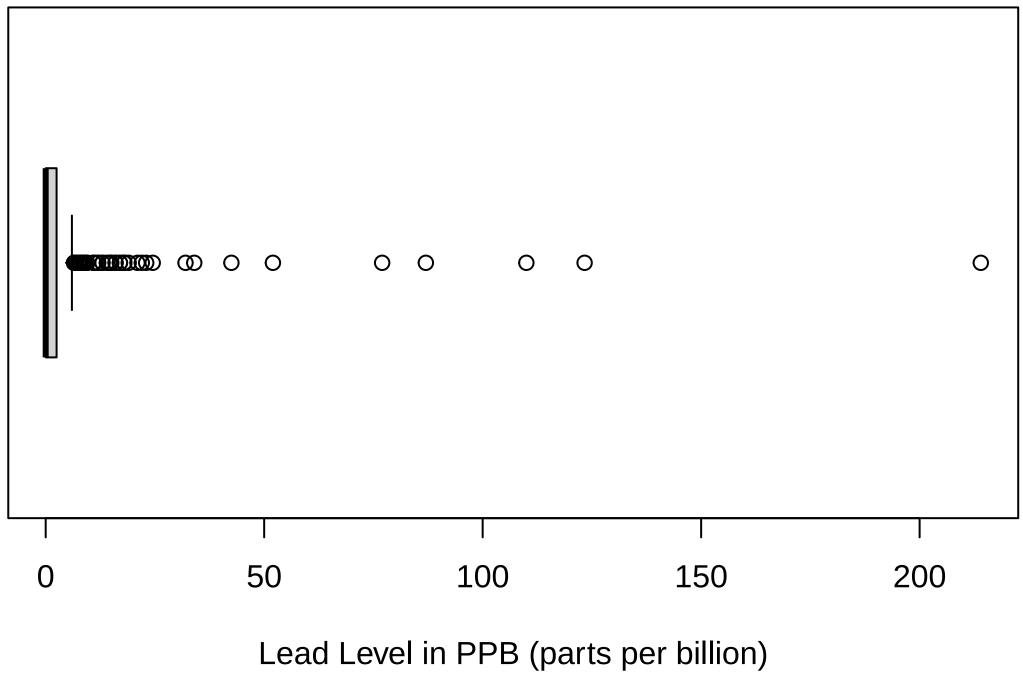
```
apply(flint[,c("Pb","Cu")],2,mean)
```

```
##        Pb        Cu
##  3.383272 54.581023
```

Problem 6(f)

```
boxplot(flint$Pb,horizontal=TRUE,main="Lead (Pb) Levels at Various Locations in Flint, Michigan",xlab="
```

Lead (Pb) Levels at Various Locations in Flint, Michigan



Problem 6(g)

```
mean(flint$Pb>0)
```

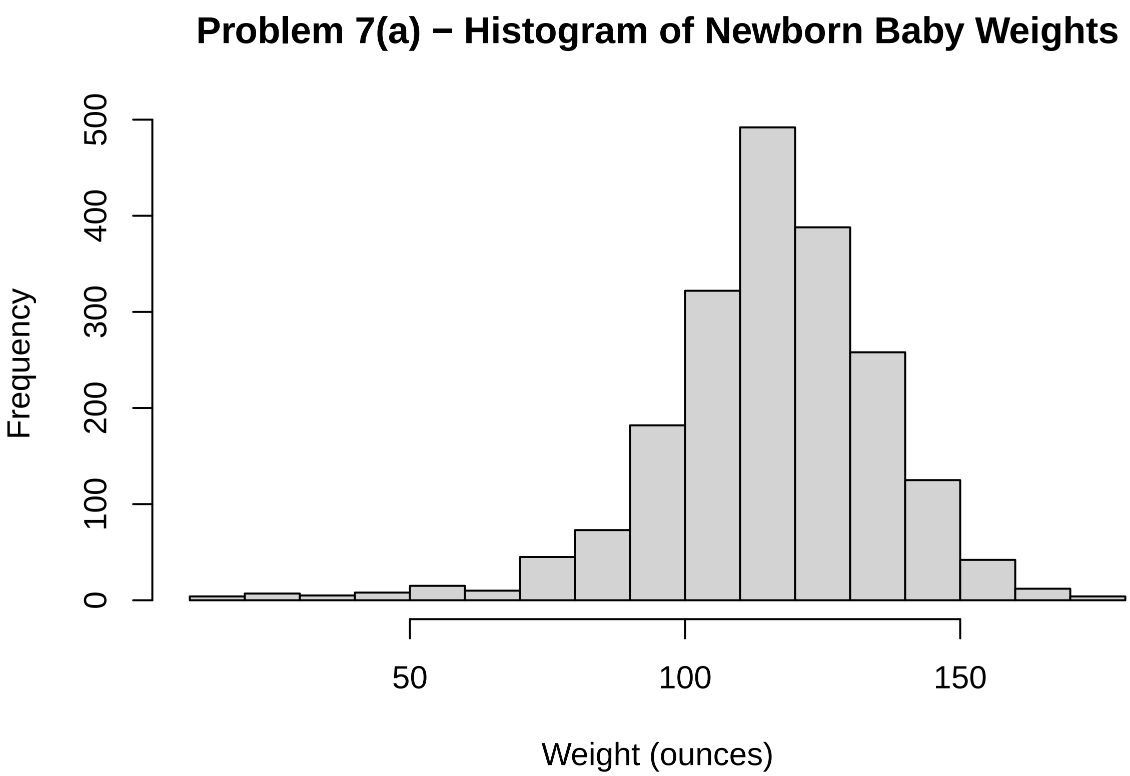
```
## [1] 0.4454713
```

No, the mean does not seem to be a good measure of center for these data. A more useful statistic for these data could be the proportion of locations at which *any* level of lead was measured (0.4454713), or the proportion found in problem 6(b).

Problem 7

Problem 7(a)

```
set.seed(2022)
hist(NCbirths$weight,main="Problem 7(a) - Histogram of Newborn Baby Weights",xlab="Weight (ounces)")
```



The weight variable appears to be approximately normal, but probably does not *exactly* follow a normal distribution. We can see the shape is symmetric and approximately normal, but the long left tail is a notable sign of non-normality.

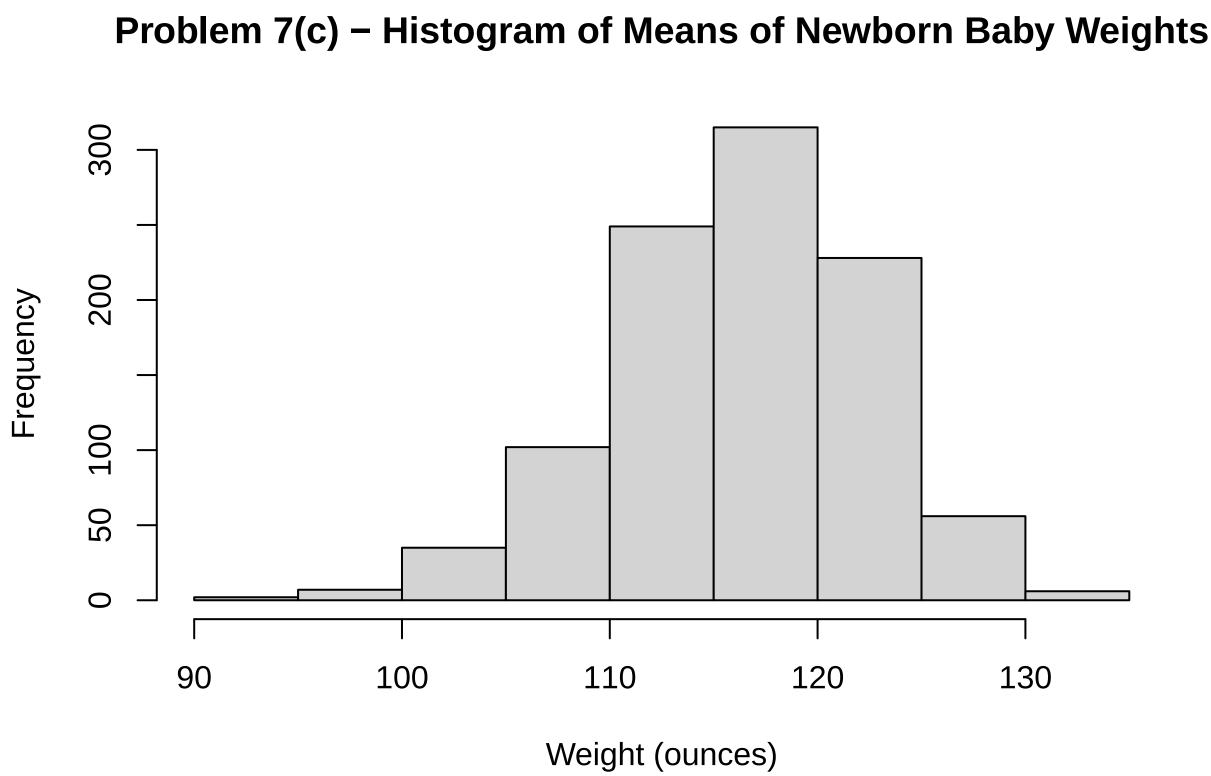
Problem 7(b)

```
set.seed(2022)
mean(sample(NCbirths$weight,10))
```

```
## [1] 119
```

Problem 7(c)

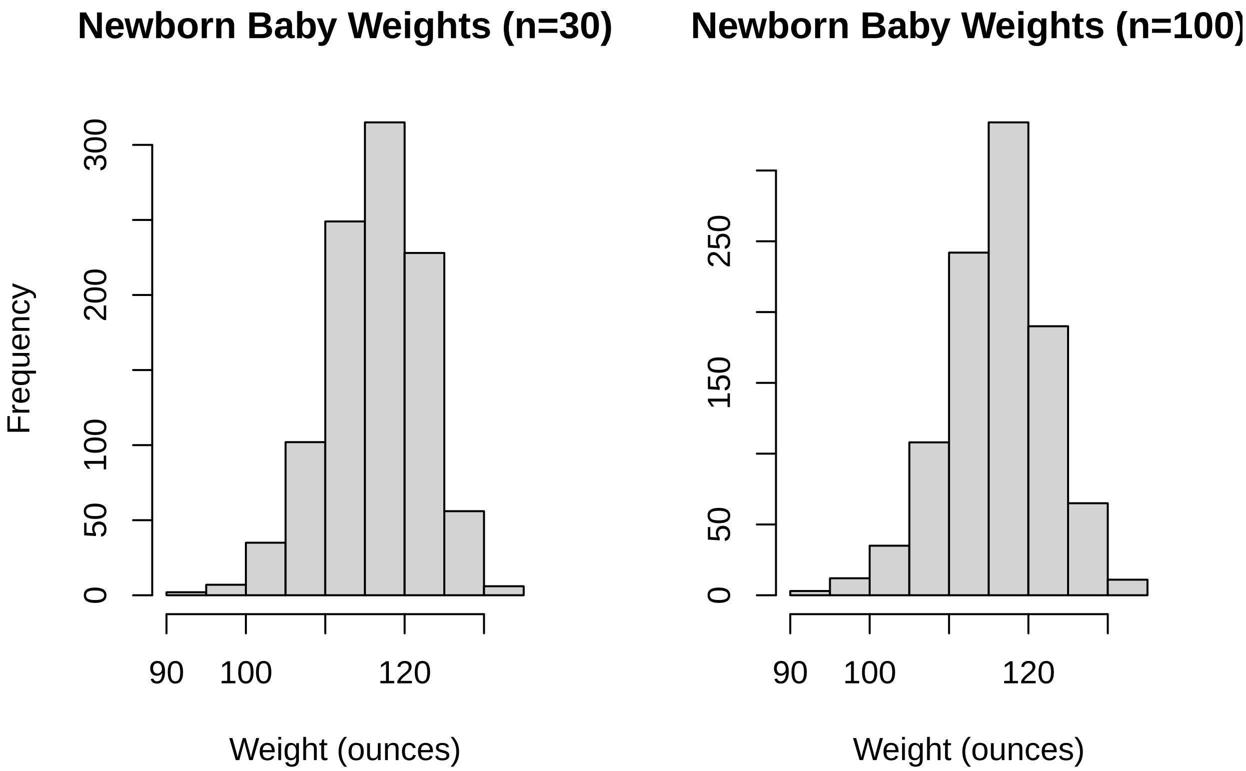
```
set.seed(2022)
m<-m30<-m100<-rep(NULL,1000)
for (i in 1:1000){m[i]<-mean(sample(NCbirths$weight,10))}
hist(m,main="Problem 7(c) - Histogram of Means of Newborn Baby Weights",xlab="Weight (ounces)")
```



Yes, this distribution is close to normal.

Problem 7(d)

```
set.seed(2022)
for (i in 1:1000){m30[i]<-mean(sample(NCbirths$weight,10))}
for (i in 1:1000){m100[i]<-mean(sample(NCbirths$weight,10))}
par(mfrow=c(1,2))
hist(m30,main="Newborn Baby Weights (n=30)",xlab="Weight (ounces)")
hist(m100,main="Newborn Baby Weights (n=100)",xlab="Weight (ounces)",ylab="")
```



Yes, these two distributions are both close to normal. We can see the shape is symmetric and consistent with distributions that are approximately normal.