Assignment_4_Solution

Jie Wang

3/6/2024

Solution to Question 1

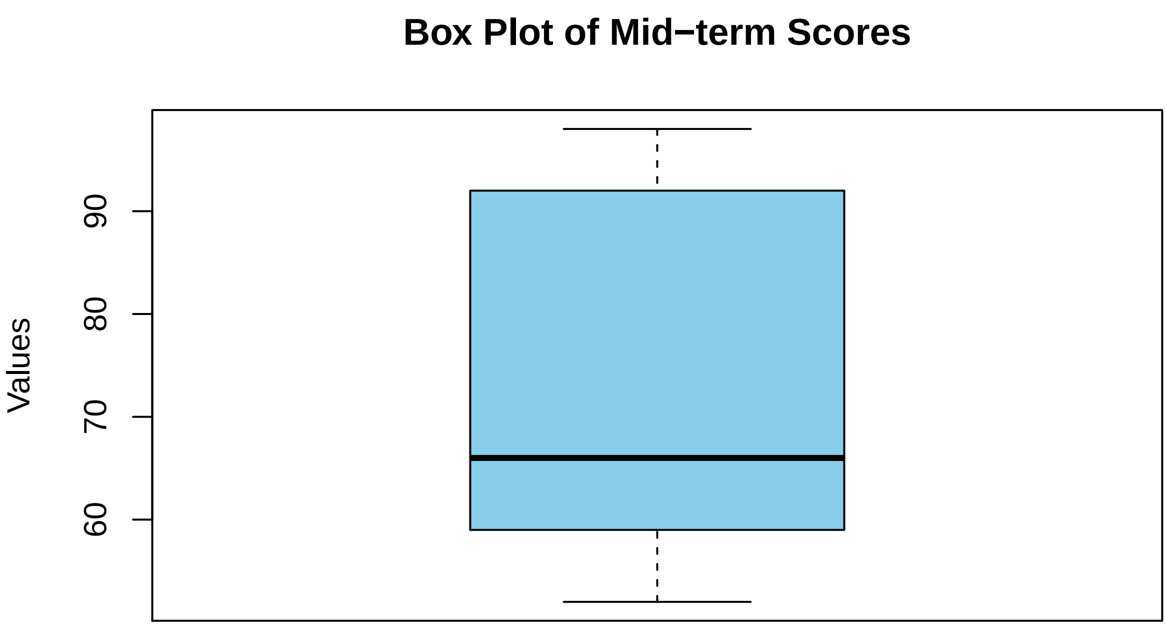
```
x <- c(92, 57, 92, 61, 94, 98, 95, 52, 66, 63, 61, 71, 52, 56, 86)
sample_mean <- mean(x)
sample_var  <- var(x)
sample_median <- median(x)
sample_quantile <- quantile(x, probs=c(0.25, 0.75))

print(c(sample_mean, sample_var, sample_median))
```

```
## [1]  73.06667 307.78095  66.00000
```

```
print(sample_quantile[2] - sample_quantile[1])
```

```
## 75%
##  33
```

```
x <- c(92, 57, 92, 61, 94, 98, 95, 52, 66, 63, 61, 71, 52, 56, 86)
boxplot(x, main = "Box Plot of Mid-term Scores", ylab = "Values", col = "skyblue")
```

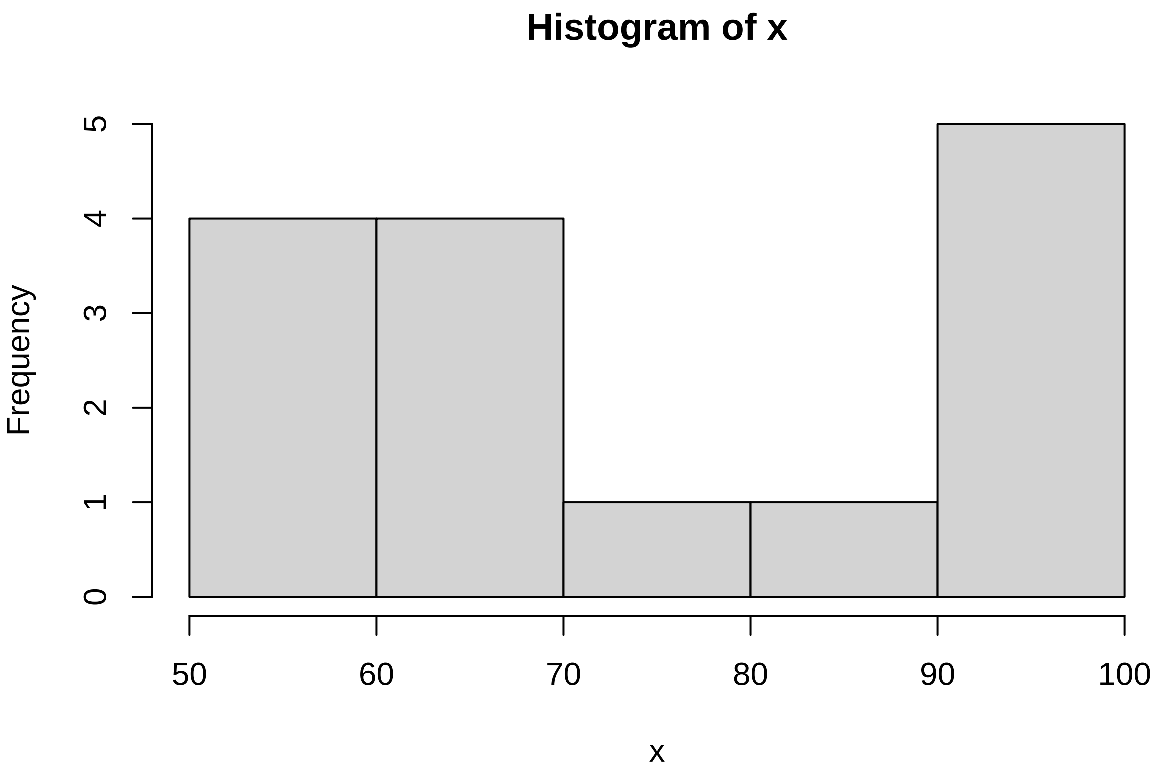



```
x <- c(92, 57, 92, 61, 94, 98, 95, 52, 66, 63, 61, 71, 52, 56, 86)
stem(x)
```

```
##
```

```
##   The decimal point is 1 digit(s) to the right of the |
##
##   5 | 2267
##   6 | 1136
##   7 | 1
##   8 | 6
##   9 | 22458
```

```r
x <- c(92, 57, 92, 61, 94, 98, 95, 52, 66, 63, 61, 71, 52, 56, 86)
hist(x, breaks = 5)
```



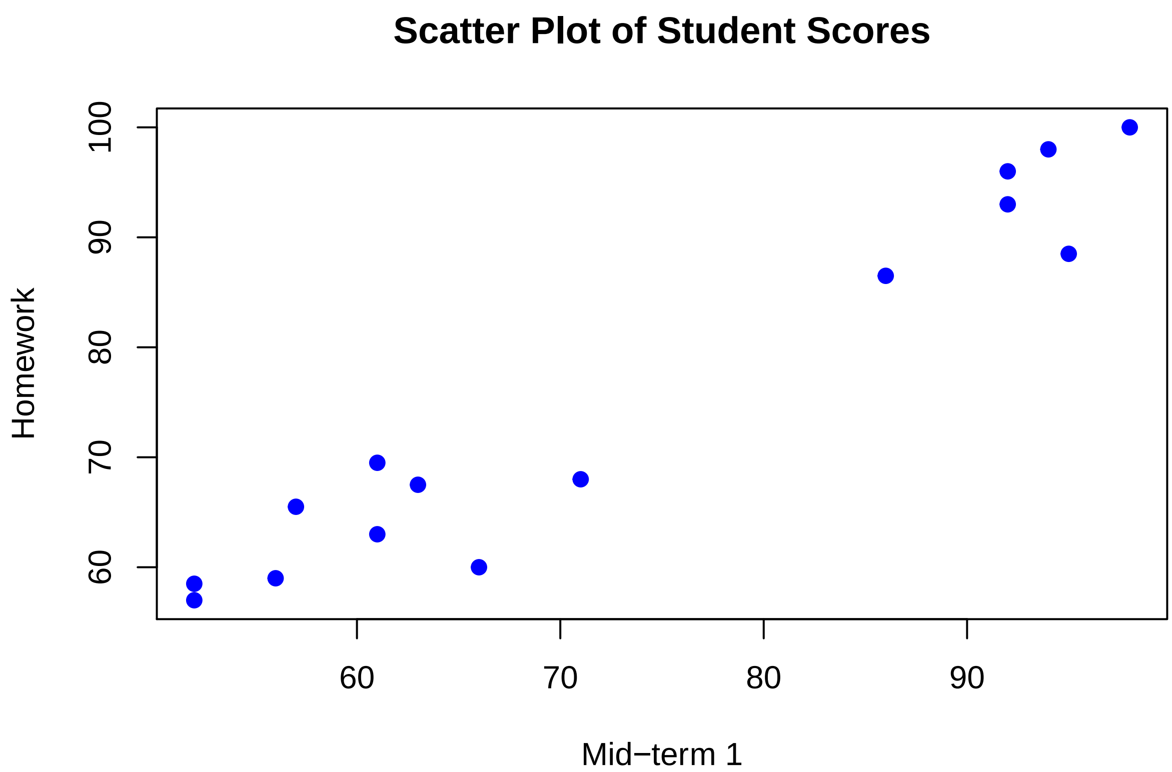

```r
x_new <- c(92, 57, 92, 61, 94, 98, 95, 52, 66, 63, 61, 71, 52, 56, 86, 0)

sample_mean <- mean(x_new)
sample_var  <- var(x_new)
sample_median <- median(x_new)
sample_quantile <- quantile(x_new, probs=c(0.25, 0.75))

print(c(sample_mean, sample_var, sample_median))
```

```
## [1]  68.5000 620.9333  64.5000
```

```r
print(sample_quantile[2] - sample_quantile[1])
```

```
##   75%
## 35.25
```

```r
x <- c(92, 57, 92, 61, 94, 98, 95, 52, 66, 63, 61, 71, 52, 56, 86)
y <- c(96, 65.5, 93, 69.5, 98, 100, 88.5, 57, 60, 67.5, 63, 68, 58.5, 59, 86.5)
plot(x, y, main = "Scatter Plot of Student Scores", xlab = "Mid-term 1", ylab = "Homework", col = "blue
```




```
print(cov(x,y))
```

```
## [1] 274.8333
```

```
print(cor(x,y))
```

```
## [1] 0.9664013
```

Solution to Question 2

Here $X \sim \mathcal{N}(1.05, 0.005^2)$, and $\overline{X} \sim \mathcal{N}(1.05, 0.001^2)$. Therefore, the probability becomes

```
pnorm(1.051, 1.05, sd=0.001) - pnorm(1.049, 1.05, sd=0.001)
```

```
## [1] 0.6826895
```

Solution to Question 3

```
x = rnorm(100,mean=100,sd=3)
print(c(mean(x), var(x)))
```

```
## [1] 100.100775   9.080894
```

```
n = 100 #number of observations in one sample
S = 500 #number of simulations
X = matrix(0,nrow=S, ncol=n)
for(i in 1:S){
  X[i,] = rnorm(n,mean=100,sd=3)
}
```

```r
means = apply(X,1,mean)
variances = apply(X,1,var)
```

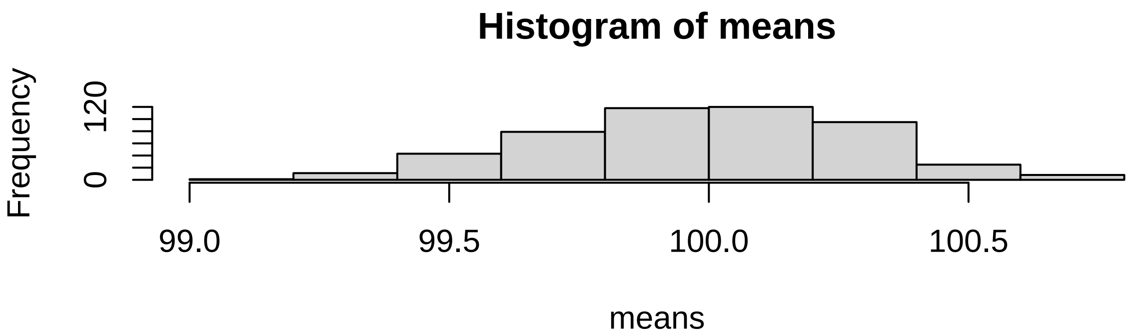
```r
summary(means)
```

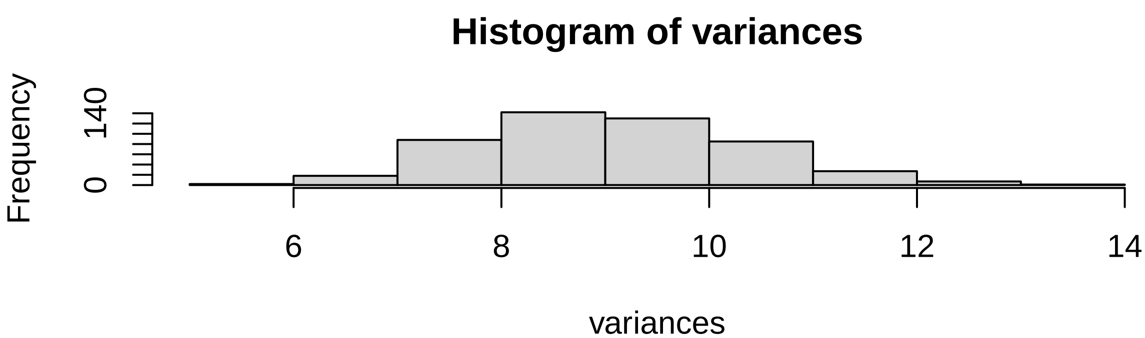
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   99.19   99.78  100.00   99.98  100.21  100.79
```

```r
summary(variances)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   5.711   8.130   9.003   9.077   9.916  13.275
```

```r
par(mfrow=c(2,1))
hist(means)
hist(variances)
```





The theoretical sampling distribution of $\hat{\mu}$ is $N(100, 0.3^2)$.