homework_ch4_JimmyNg

Jimmy Ng

October 9, 2018

Question 4.4

(a) The point estimate is 171.1, whereas the median is 170.3.
(b) The point estimate for the SD is 9.4, whereas the IQR (Q3 - Q1) is 14.
(c) In this sample, a person who is 180cm tall is not considered to be unusally tall. We should consider a one-tail test, (180 - 171.1) / 9.4 = 0.95. This is within one SD. In order to be considered unusually significantly taller than the rest, the person should be in the top 5 or 1 %, and that would give a Z-score at least 1.64 or above (for one-tail 95% significant level). On the other hand, for a person who is 155cm tall, he/she is considered to be unusally shorter than others, i.e. (155 - 171.1) / 9.4 = -1.71. That would position the person in the bottom 4 to 5% in the sample.
(d) The mean and SD would not be identical, but highly similar as there's always variation when drawing a random sample from the population.
(e) We measure the standard error, in this case, that is equal to 9.4 / sqrt(507) = 0.4174687.

Question 4.14

(a) False. A correct interpretation is an average spending for an American in the general adult population would be between the lower and upper intervals and that's true 95% of the time, in other words, there's a 5% chance that the true mean of the population lied outside the intervals.
(b) False (in terms of the statment only; the confidence interval (CI) is actually valid). As long as the sample is randomly and independently selected, and we have a large enough sample (n >= 30, less than 10% of the population), a minor skew can be tolerated.
(c) True. The true mean would lie between the lower and upper intervals 95% of the time. 95% of the sampling means would be covered within this range in the sampling distribution.
(d) True. Same reasoning as (a).
(e) True.
(f) False. The margin of error is equal to the critical value multiply by the standard error, i.e. SD / sqrt(N). Increasing the sample size by 3 times as many would not decrease the margin of error to a third. It's because the critical value (z-score 95% CI = 1.96) is a constant and sqrt(3N) is not equal to 3 * sqrt(N).
(g) True. 84.71 +/- 4.4 is equal to the 95% confidence intervals (80.31, 89.11).

Question 4.24

(a) Yes, samples are randomly and independently selected. Sample size is sufficiently large enough (n >= 30 and less than 10% of the population) for ignoring slight skew in the distribution.
(b) This is a one-tail test and the alpha level is 0.1. We can calculate the z-score as (30.69 - 32) / 4.31 = -0.3039443. The critical value of 90% CI for this one-tail test is -1.28. Since the result is larger than the critical value (and that is associated with the p-value), i.e. -0.3039443 > -1.28, we cannot discard the null hypothesis.
(c) The hypothesis is that gifted children can count to 10 statistically significantly younger than the rest. The null hypothesis is that there's no statistical difference. The alternative hypothesis is that there is a difference and we set the alpha level to 10%. The critical value for this one-tail test is -1.28, and the p-value associated with our result (-0.3039443) is 0.380945 or 38%. Since the p-value (38%) is larger than the alpha level that we have set (10%), we cannot discard the null hypothesis. As a result, there is no statistical evidence to demonstrate the difference between gifted children from the rest.

Question 4.26

(a) This is a one-tail test. Z-score = (118.2 - 100) / 6.5 = 2.8, and the p-value = 0.002555. Alpha level = 10% and the p-value is far less than 1%. As a result, we can discard the null hypothesis, i.e. mother of gifted children are not different from other mothers in IQ.

(b) x +/- (z * standard of error). Lower interval is equal to 118.2 - 1.64*(6.5/sqrt(36)), whereas the upper interval is equal to 118.2 + 1.64*(6.5/sqrt(36)), i.e. CI(116.4, 120.0).

(c) Yes, using 90% confidence interval, we see that the point estimate for gifted mothers have their IQ between 116.4 and 120, which is far above the average IQ of the general female mother population (100). The CI concur with the alternative hypothesis, which states that there's a statistical significant difference between gifted mothers' and normal mothers' IQ.

Question 4.34

CLT: The sampling distribution of the mean is a distribution of means that are collected through random sampling. However, the shape of the sampling distribution would vary by the size of each random sample that make up the distribution, such as a sample of size 10 would look different (flatter) than a sample of size 100. When the sample size increases, a sampling distribution would approach a normal distribution which the density curve looks symmetrical and centered about its mean with its spead determined by its SD.

Question 4.40

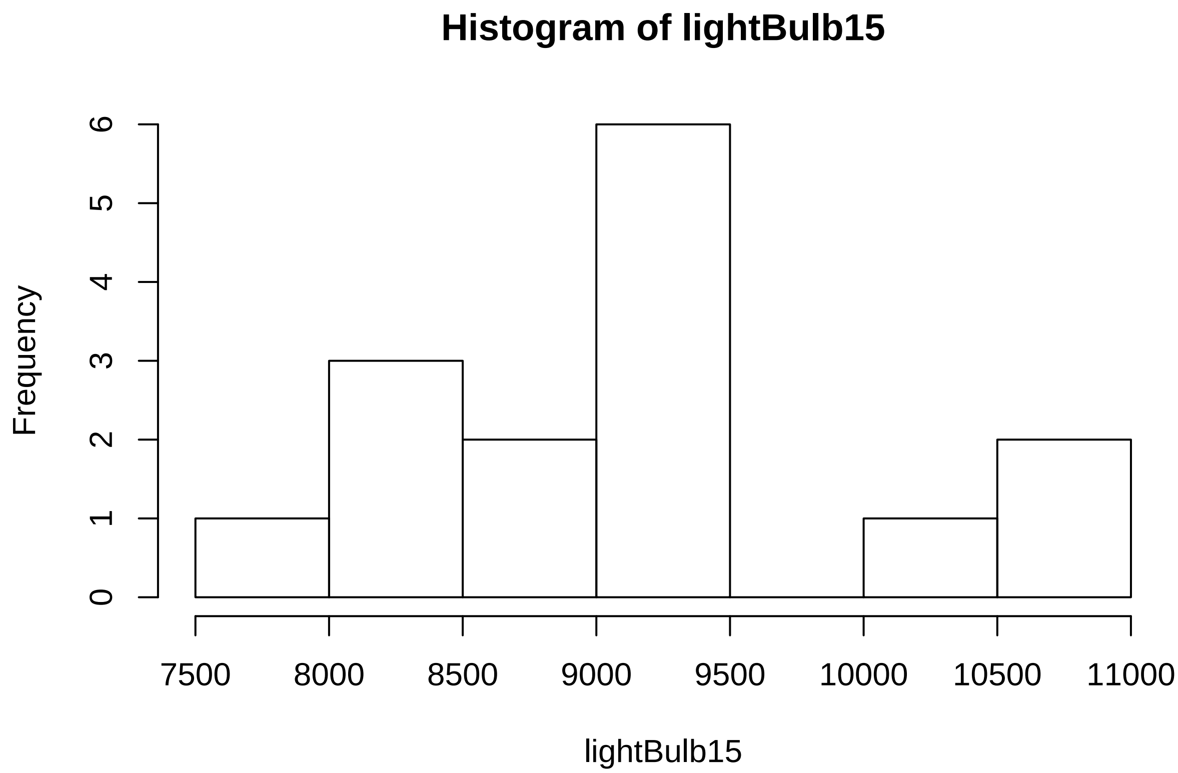
```
# (a)
mu = 9000; sd = 1000
print(paste("(4.40a) The probability of a randomly chosen light bulb that lasts more than 10500 hours i
            round(100 * pnorm(q = 10500, mu, sd, lower.tail = F), 2),
            "%",
            sep = ""))
```

```
## [1] "(4.40a) The probability of a randomly chosen light bulb that lasts more than 10500 hours is appr
```

(4.40a) The probability of a randomly chosen light bulb that lasts more than 10500 hours is approximately 6.68%

```
# (b)
set.seed(123)
lightBulb15 <- rnorm(n = 15, mu, sd)
summary(lightBulb15); hist(lightBulb15)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    7735    8499    9111    9152    9431   10715
```



(4.40b) Above is a simulation of normal distribution of 15 samples based on the given parameters.

```
# (c)
simulation <- function(var, iteration = 5000, size = 50, seed = 1234){
        if(!require(purrr)){install.packages("purrr"); require(purrr)}
        if(!require(plyr)){install.packages("plyr"); require(plyr)}

        set.seed(seed)
        sim <- as.vector(rep(NA, iteration))
        sim <- map(1:iteration, function(x) sim[x] <- mean(sample(var, size)))
        sim <- sim %>%
                plyr::ldply(., data.frame)
        sim <- as.vector(sim[,1])
}

sample_means15 <- simulation(rnorm(n = 1000, mu, sd), iteration = 1000, size = 15, seed = 123)
```

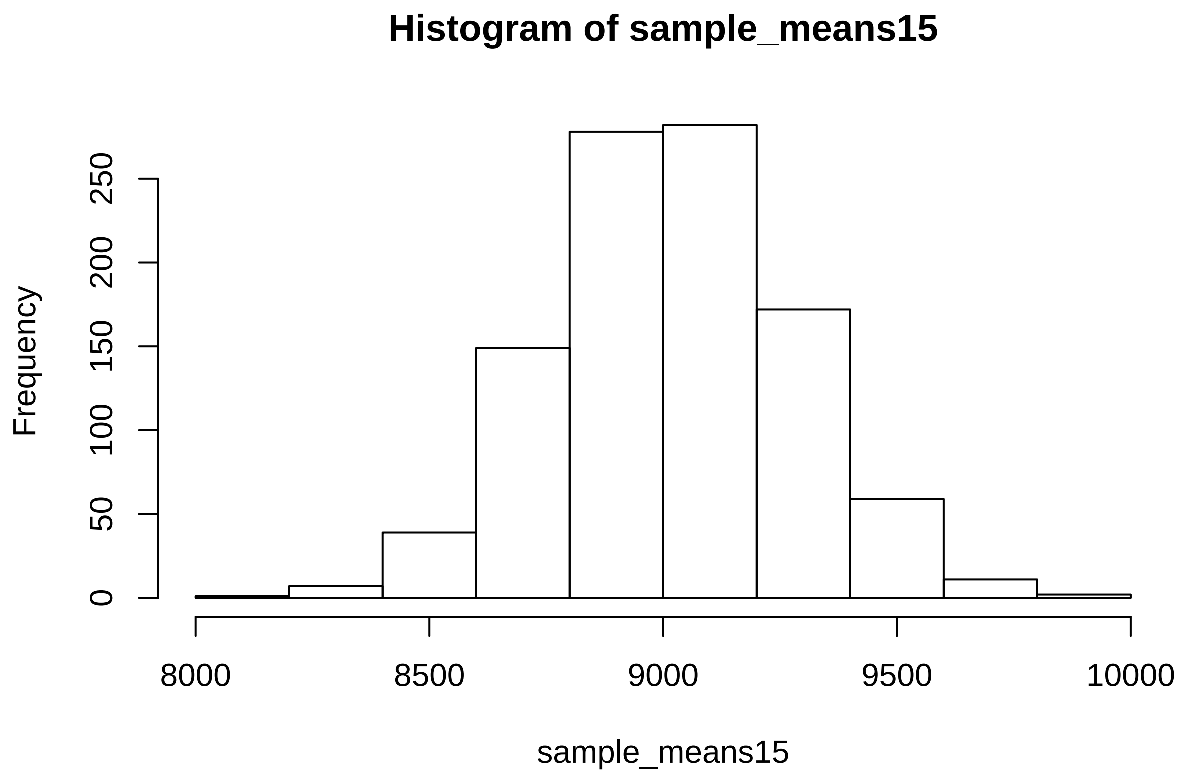
```
## Loading required package: purrr

## Loading required package: plyr

##
## Attaching package: 'plyr'

## The following object is masked from 'package:purrr':
##
##     compact
```

```r
above10500 <- sum(sample_means15 > 10500) / 1000
hist(sample_means15)
```

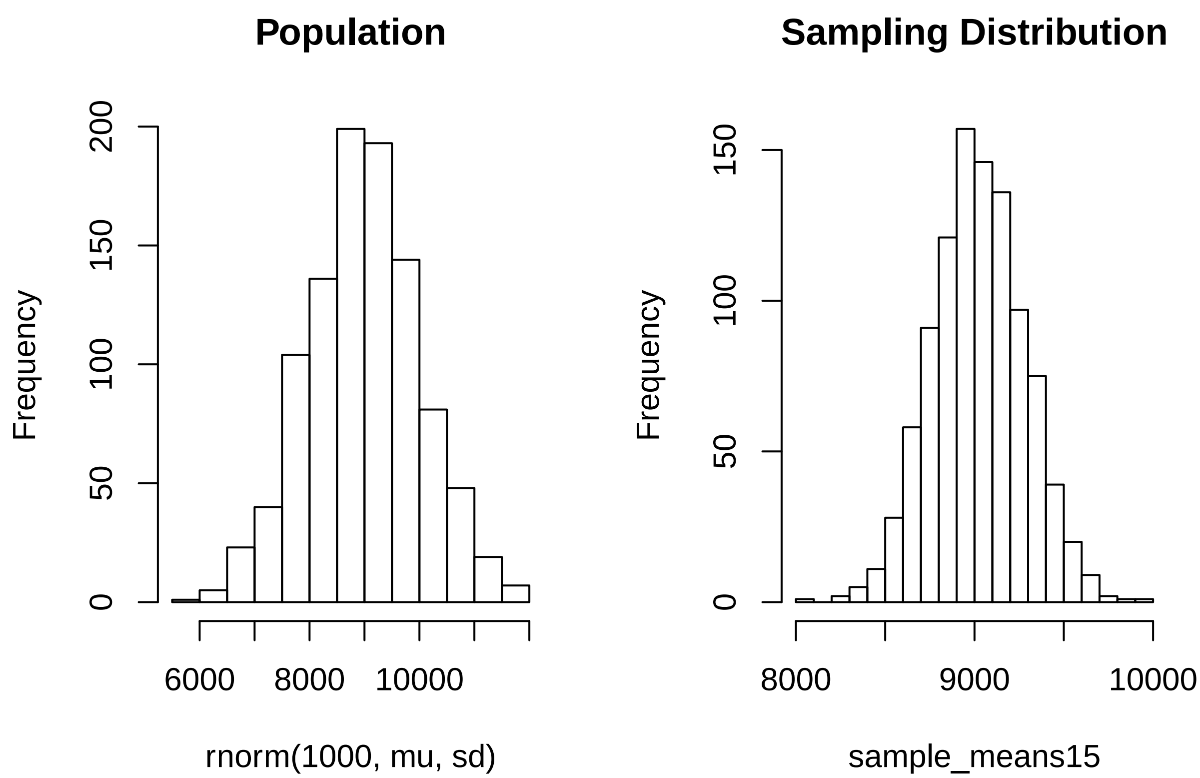



```r
print(paste("(4.40c) The probability that the mean lifespan of 15 randomly chosen life bulbs last more
            above10500,
            ".",
            sep = ""))
```

```
## [1] "(4.40c) The probability that the mean lifespan of 15 randomly chosen life bulbs last more than
```

(4.40c) The probability that the mean lifespan of 15 randomly chosen life bulbs last more than 10500 hours is approximately 0. Above sampling distribution is based on 1000 iterations/simulations using sample size of 15 each and the given parameters.

```r
# (d)
par(mfrow = c(1, 2))
hist(rnorm(1000, mu, sd), main = "Population", breaks = 20)
hist(sample_means15, main = "Sampling Distribution", breaks = 20)
```



(4.40d) Above left histogram displays 1000 random samples based on the given parameters, whereas the right histogram is based on sampling distribution using the same set of parameters with 1000 iterations and sample size of 15 in each sampling.

(4.40e) Yes, according to the Central Limit Theorem, the sampling distribution of the mean approaches normal distribution when sample size increases and each draw is randomly and independently carried out. Even if the underlying distribution is skewed, We can still use the sampling distribution to perform a normal approximation in order to estimate the mean.

Question 4.48

p-value would likely decrease when the sample size increases and that's because the variance of the sample mean (standard error) would decrease and hence precision increases. When standard error decreases, a small (even a tiny bit) difference between two or multiple comparisons would be easily (mis)interpreted as significant. In this case, p-value can be very misleading without concerning the effect size of the study.