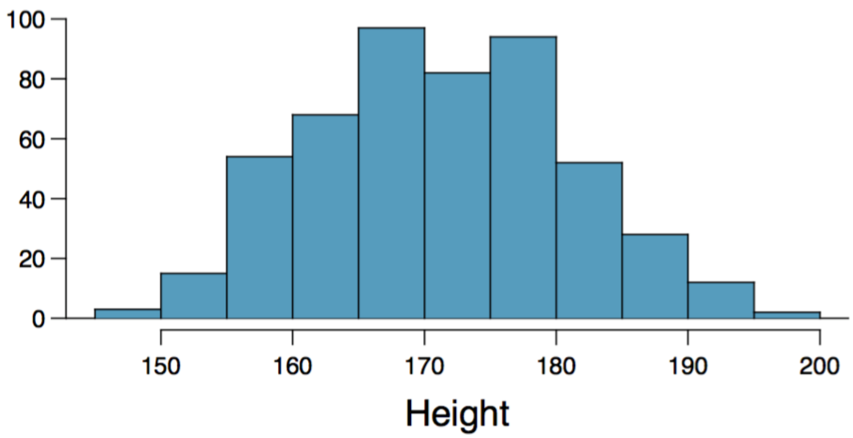
HW-4- 606
Dhananjay Kumar

4.4 Heights of adults. Researchers studying anthropometry collected body girth measurements and skeletal diameter measurements, as well as age, weight, height and gender, for 507 physically active individuals. The histogram below shows the sample distribution of heights in centimeters.



Min	147.2
Q1	163.8
Median	170.3
Mean	171.1
SD	9.4
Q3	177.8
Max	198.1

(a) What is the point estimate for the average height of active individuals? What about the median?

(b) What is the point estimate for the standard deviation of the heights of active individuals? What about the IQR?

(c) Is a person who is 1m 80cm (180 cm) tall considered unusually tall? And is a person who is 1m 55cm (155cm) considered unusually short? Explain your reasoning.

(d) The researchers take another random sample of physically active individuals. Would you expect the mean and the standard deviation of this new sample to be the ones given above? Explain your reasoning.

(e) The sample means obtained are point estimates for the mean height of all active individuals, if the sample of individuals is equivalent to a simple random sample. What measure do we use to quantify the variability of such an estimate (Hint: recall that SD_{x}-= p_n)? Compute this quantity using the data from the original sample under the condition that the data are a simple random sample.

Solution:

a) Point estimate of average height is the sample mean of height, that is: 171.1.
Point estimate of median is the sample median of height, that is: 170.3

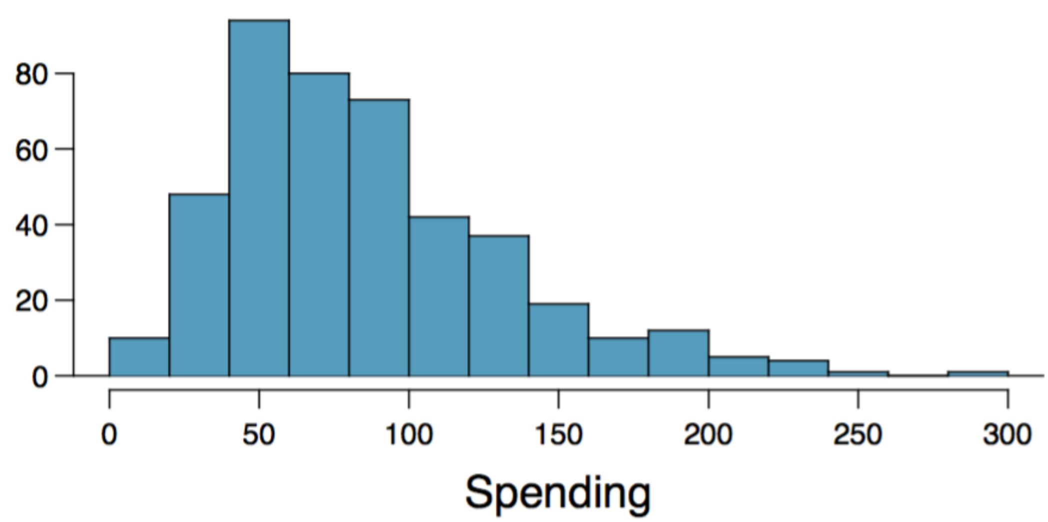
b) Point estimate of the standard deviation of the height is: 9.4
Point estimate of IQR: Q3-Q1 = 177.8-163.8 = 14

c) A person 180 cm tall is above the sample mean height and is unusually tall because it is greater than the third quartile range. Q1 and Q3 tells us where the middle 50 percent of the data lies. Since 180cm is greater than the Q3, it suggests that this is an unusually tall height in the given sample.
A person with 155 cm is below the sample mean and an unusually short height because it is less than the first quartile of the data(Q1).

d) I don't expect the mean and sample deviation of the new sample to be the same as in the given sample because sample mean is just a point estimate of the population mean and estimates vary. We can expect the point estimates to be close but not exactly same.

e) We would use the Standard Error for the sample mean to quantify the variability of such an estimate.
171.1/sqrt(507) = 7.5988

4.14 Thanksgiving spending, Part I. The 2009 holiday retail season, which kicked on November 27, 2009 (the day after Thanksgiving), had been marked by somewhat lower self-reported consumer spending than was seen during the comparable period in 2008. To get an estimate of consumer spending, 436 randomly sampled American adults were surveyed. Daily consumer spending for the six-day period after Thanksgiving, spanning the Black Friday weekend and Cyber Monday, averaged \$84.71. A 95% confidence interval based on this sample is (\$80.31, \$89.11). Determine whether the following statements are true or false, and explain your reasoning.



(a) We are 95% confident that the average spending of these 436 American adults is between \$80.31 and \$89.11.

(b) This confidence interval is not valid since the distribution of spending in the sample is right skewed.

(c) 95% of random samples have a sample mean between \$80.31 and \$89.11.

(d) We are 95% confident that the average spending of all American adults is between \$80.31 and \$89.11.

(e) A 90% confidence interval would be narrower than the 95% confidence interval since we don't need to be as sure about our estimate.

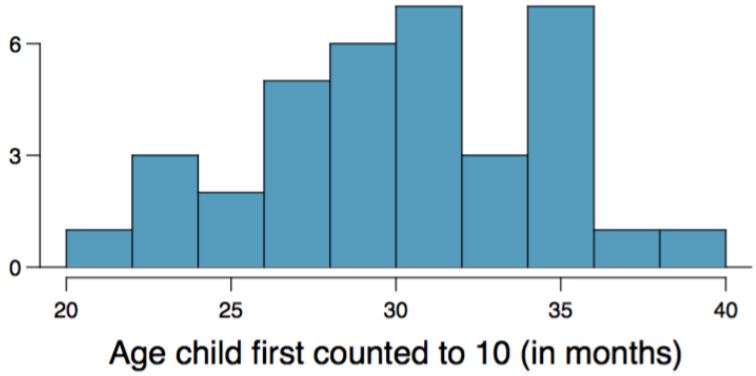
(f) In order to decrease the margin of error of a 95% confidence interval to a third of what it is now, we would need to use a sample 3 times larger.

(g) The margin of error is 4.4.

<u>Solution:</u>

a) False. I would rather say that we are 95% confident that the average spending of the American adult population is between \$80.31 and \$89.11. The confidence interval tells us the range of true population parameter from a sample.
b) False. The sample distribution is slightly right skewed but since the sample size is large (n=436) and is random we can say the confidence interval is still valid.
c) False. 95% of the samples would contain the population means within this confidence interval range.
d) True. This confidence interval means we are 95% confident that the average spending of all American adults is between \$80.31 and \$89.11.
e) True. 90% confidence interval would have narrower than 95% confidence interval because in this case the standard deviation would become less compared to 95% confidence interval, therefore the range would become narrow.
f) FALSE. We need sample size 9 times in order to reduce margin error to one third.
g) True. Margin error = 2 * SE
   84.71 +(1.96 * SE) = 89.11
   SE= 2.244
   So, Margin error = 2 * 2.244 = 4.4

4.24 Gifted children, Part I. Researchers investigating characteristics of gifted children collected data from schools in a large city on a random sample of thirty-six children who were identified as gifted children soon after they reached the age of four. The following histogram shows the distribution of the ages (in months) at which these children first counted to 10 successfully. Also provided are some sample statistics.[43]



n	36
min	21
mean	30.69
sd	4.31
max	39

(a) Are conditions for inference satisfied?

(b) Suppose you read online that children first count to 10 successfully when they are 32 months old, on average. Perform a hypothesis test to evaluate if these data provide convincing evidence that the average age at which gifted children fist count to 10 successfully is less than the general average of 32 months. Use a significance level of 0.10.

(c) Interpret the p-value in context of the hypothesis test and the data.

(d) Calculate a 90% confidence interval for the average age at which gifted children first count to 10 successfully.

(e) Do your results from the hypothesis test and the confidence interval agree? Explain.

<u>Solution:</u>

a) Not all the conditions of inference are satisfied
   1. Sample is not independent, as it is collected from only schools in large cities.
   2. Sample size should be large (>30), this condition is met (sample size is 36).
   3. Distribution should not be skewed, which is true in this case.

b) H0 (Null hypothesis): Average age, $\mu = 32$ months
   Ha (Alternative hypothesis): Average age, $\mu < 32$ months
   $Z = (30.69 - 32)/0.7183 = -1.82$
   pnorm(-1.82) = 0.0343
   At significance level 0.10 we reject the null hypothesis

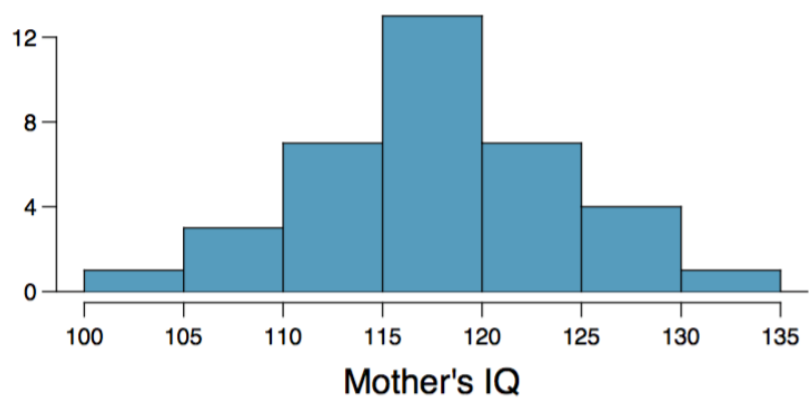
c) $P = 0.0343$
   The probability of null hypothesis to be true is 3.43% which is less than the

significance level , 10%(0.10). Therefore, we reject the null hypothesis.

d) 90 % confidence interval:
S.E = 4.31/sqrt 36 = 0.7183
30.69 +- 1.645*0.7183
30.69 +- 1.182
(29.508, 31.872)

e) Yes, the results match. We had rejected the null hypothesis, which claimed that the mean is 32 months and at 90% confidence interval also, 32 is not in this range.

4.26 Gifted children, Part II. Exercise 4.24 describes a study on gifted children. In this study, along with variables on the children, the researchers also collected data on the mother's and father's IQ of the 36 randomly sampled gifted children. The histogram below shows the distribution of mother's IQ. Also provided are some sample statistics.



n	36
min	101
mean	118.2
sd	6.5
max	131

(a) Perform a hypothesis test to evaluate if these data provide convincing evidence that the average IQ of mothers of gifted children is different than the average IQ for the population at large, which is 100. Use a significance level of 0.10.

(b) Calculate a 90% confidence interval for the average IQ of mothers of gifted children.

(c) Do your results from the hypothesis test and the confidence interval agree? Explain.

<u>Solution:</u>

a) H0 (Null Hypothesis): Average IQ, $\mu = 100$
Ha (Alternative hypothesis): Average IQ, $\mu \neq 100$

Z = 118.2-100/(6.5/sqrt1.083) = 16.8
This is a two-tailed test.
2*(1- Pnorm(16.8)) = 0
Since the probability is almost 0, we reject the null hypothesis.

b) S.E = 1.083
118.2 +- 1.65*1.083
(116.2+- 1.78695)
(116.41,119.99)

c) Yes, the results match. In hypothesis testing we had failed the null hypothesis which claimed that average IQ was 100. The 90% confidence interval range also does not have 100, it is much greater than 100.

4.34 CLT. Define the term "sampling distribution" of the mean, and describe how the shape, center, and spread of the sampling distribution of the mean change as sample size increases.

Solution: When samples of size n are taken from a population that are random and independent, then the mean of each sample is calculated and plotted, then it is called sampling distribution of the mean.

As the sample size increases it's shape resembles **bell curve,** center is towards **mean** of the distribution and **taller** and spread becomes **narrower.**

4.40 CFLBs. A manufacturer of compact fluorescent light bulbs advertises that the distribution of the lifespans of these light bulbs is nearly normal with a mean of 9,000 hours and a standard deviation of 1,000 hours.

(a) What is the probability that a randomly chosen light bulb lasts more than 10,500 hours?

(b) Describe the distribution of the mean lifespan of 15 light bulbs.

(c) What is the probability that the mean lifespan of 15 randomly chosen light bulbs is more than 10,500 hours?

(d) Sketch the two distributions (population and sampling) on the same scale.

(e) Could you estimate the probabilities from parts (a) and (c) if the lifespans of light bulbs had a skewed distribution?

Solution:

a) Z = 10500-9000/1000 = 1.5
   1-pnorm(1.5) **= 0.067**

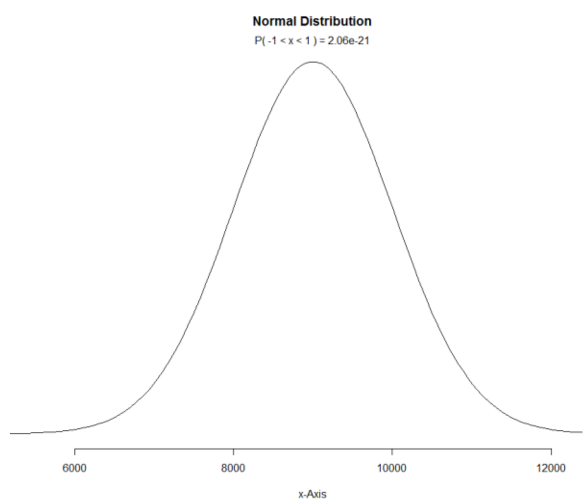
b) S.E = 1000/sqrt(15) = 258.199

c) Z = 1500-9000/S.E = 5.8094

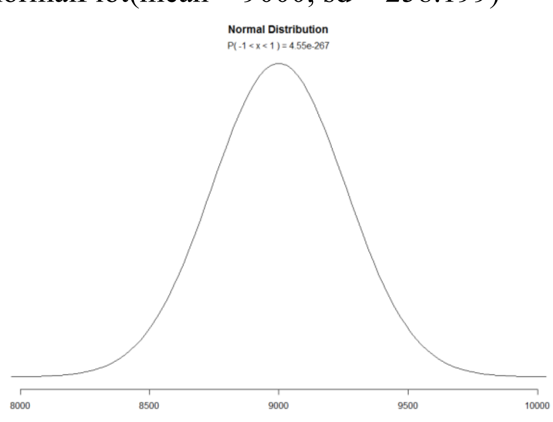
   1-pnorm(5.8094) = 3.134856e-09 (extremely small)

   There is almost 0 probability that the mean lifespan of 15 randomly chosen light bulb is more than 10,500 hours.

d) normalPlot(mean = 9000, sd = 1000)



normalPlot(mean = 9000, sd = 258.199)



e) No, because we assume that the distribution is normal, and for normal distribution data should not be skewed.

4.48 Same observation, different sample size. Suppose you conduct a hypothesis test based on a sample where the sample size is $n = 50$, and arrive at a p-value of 0.08. You then refer back to your notes and discover that you made a careless mistake, the sample size should have been $n = 500$. Will your p-value increase, decrease, or stay the same? Explain.

Solution: As the sample size increases the SD would decrease. This would result in larger Z score. Larger Z value means p value would decrease.