

1) $\mu_0 = 15$ mg/day

Conjectured true population mean, $\mu_A = 17$ mg/day

Hypothesis:

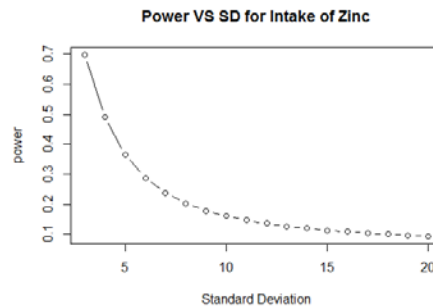
$$H_0 : \mu \leq \mu_0 : \mu \leq 15$$

$$H_A : \mu > \mu_0 : \mu > 15$$

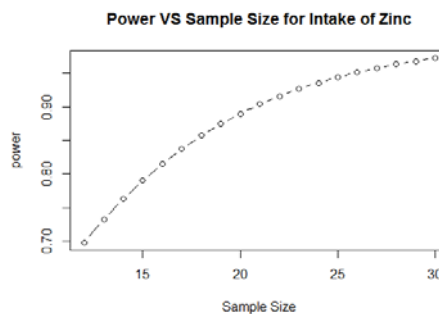
$n = 12$, $\alpha = 0.05$, conjectured s.d $\sigma = 3$ mg/day

A) power = 0.6981908 [See appendix for R code and Calculation]

B) If the sample deviation was larger, the power would be lower than that of the power that was calculated in part A. Following graph shows the Power ~ SD relation for the data.



C) If the sample size was larger, power would increase. The following graph shows the Power ~ Sample size relation for Zinc data.

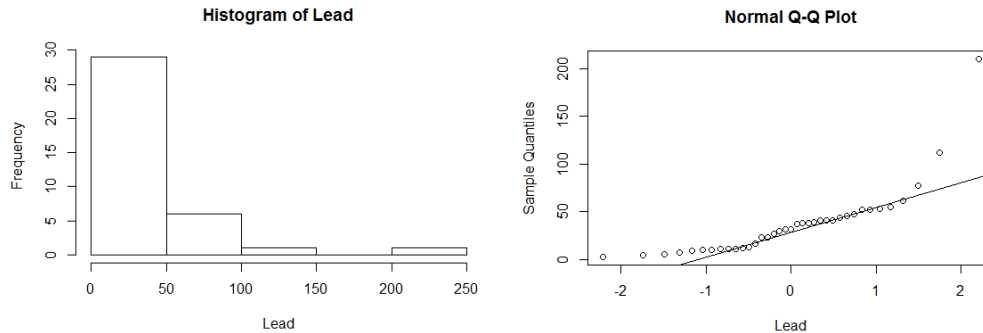


D) If $\alpha = 0.10$, power = 0.8260609 [See Appendix for calculation], which is higher than the power calculated in part A.

E) For $\mu_A = 16$ mg/day, power = 0.2874441 [See Appendix for calculation], which is lower than the power calculated in part A.

F) For Power = 0.9, sample size = 21 (rounded up) [See Appendix for calculation].

2) A)



Shapiro-Wilk normality test p-value = $1.928e-07 < 0.05$. [See appendix]
So, the null hypothesis (lead sample data is normally distributed) can be rejected.

The result of the shapiro-wilk test matches the histogram and normal Q-Q plot. From the Histogram we can see that the lead sample data is skewed right, and from the normal Q-Q plot we see that the plot is not a straight line.

B) Mean = 37.24324

Median = 32

C) $H_0 : M = 30$

$H_A : M \neq 30$

From the sign test, we obtain,
 s (#of values > 30) = 20

p-value = 0.6177

As P value $> \alpha = 0.05$, we cannot reject H_0 .

We do not have enough evidence to reject the population median = 30 with 95% confidence.

D)

95% confidence interval (using Upper Archieved CI) is (17.0000, 41)

E)

$H_0 : \mu = \mu_0 : \mu = 30$

$H_A : \mu \neq \mu_0 : \mu \neq 30$

p-value = 0.2431 $> \alpha = 0.05$. So we cannot reject H_0 .

We do not have enough evidence to reject that the population mean = 30 with 95% confidence.

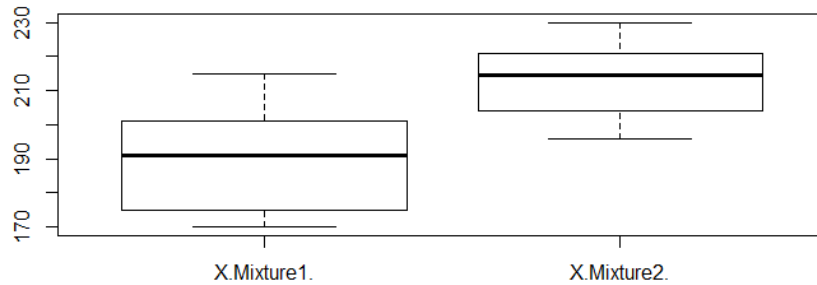
F)

95 percent confidence interval: (24.86550, 49.62099)

G) Studentized confidence interval mean= (27.46, 57.19)

H) Assuming the cumulative lead exposure is of interest, the mean would be of more interest.

3) A)



B) Sample Mean for Mixture 1 $\bar{y}_1 = 190$

Sample SD for Mixture 1 $s_1 = 14.65151$

Mixture 1 Sample size $n_1 = 10$

Sample Mean for Mixture 2 $\bar{y}_2 = 213.4$

Sample SD for Mixture 2 $s_2 = 10.62701$

Mixture 1 Sample size $n_1 = 10$

C) Assuming equal variances, the 95% confidence interval for the difference between the mean = (-35.423, -11.375). Based on this interval, we can conclude that there is a difference between the population mean as the 95% confidence interval does not cover 0.

d) The summary stats show that the sample sizes of mixture 1 and 2 are equal.

$s_1/s_2 = 14.65151/10 = 1.46 < 2$. So equal variances can be assumed. In this case, Pooled t-test is more appropriate here.

e) $H_0: \mu_1 - \mu_2 = 0$

$H_1: \mu_1 - \mu_2 \neq 0$

p-value = 0.0006899 < 0.05, so reject H_0 .

This means we do not have sufficient evidence to conclude with 95% confidence that there are no differences between the means.

APPENDIX

#QUESTION 1

```
> #1A
> power.t.test(n=12, delta=2, sd=3, sig.level = 0.05, type = "one.sample", alternative = "one.sided")
```

One-sample t test power calculation

```
      n = 12
  delta = 2
      sd = 3
sig.level = 0.05
  power = 0.6981908
alternative = one.sided
```

```
>
> #1B
> testSD<- seq(3, 20, 1)
> powerVal1B_DiffSD<- power.t.test(n=12, delta=2, sd=testSD, sig.level = 0.05, type = "one.sample", alternative = "one.sided")
> plot(powerVal1B_DiffSD$power ~ testSD, type = "b", xlab = "Standard Deviation", ylab = "power", main = "Power VS SD for Intake of Zinc")
>
> #1C
> testSampleSize <- seq(12, 30, 1)
> powerVal1C_Diffn <- power.t.test(n=testSampleSize, delta=2, sd=3, sig.level = 0.05, type = "one.sample", alternative = "one.sided")
> plot(powerVal1C_Diffn$power ~ testSampleSize, type = "b", xlab = "Sample Size", ylab = "power", main = "Power VS Sample Size for Intake of Zinc")
>
> #1D
> power.t.test(n=12, delta=2, sd=3, sig.level = 0.10, type = "one.sample", alternative = "one.sided")
```

One-sample t test power calculation

```
      n = 12
  delta = 2
      sd = 3
sig.level = 0.1
  power = 0.8260609
alternative = one.sided
```

```
>
> #1E
> power.t.test(n=12, delta=1, sd=3, sig.level = 0.05, type = "one.sample", alternative = "one.sided")
```

One-sample t test power calculation

```
      n = 12
  delta = 1
      sd = 3
```

```
sig.level = 0.05
power = 0.2874441
alternative = "one.sided"
```

```
>
> #1F
> power.t.test(delta=2, sd=3, p=0.9, sig.level = 0.05, type = "one.sample", alternative = "one.sided")
```

One-sample t test power calculation

```
      n = 20.69914
delta = 2
sd = 3
sig.level = 0.05
power = 0.9
alternative = "one.sided"
```

#QUESTION2

```
> hist(DataHW4_2$X.Lead, xlab = "Lead", main = "Histogram of Lead")
> qqnorm(DataHW4_2$X.Lead, xlab = "Lead")
> qqline(DataHW4_2$X.Lead, )
> shapiro.test(DataHW4_2$X.Lead.)
```

Shapiro-Wilk normality test

```
data: DataHW4_2$X.Lead.
W = 0.69693, p-value = 1.928e-07
```

```
> #B
> mean(DataHW4_2$X.Lead.)
[1] 37.24324
> median(DataHW4_2$X.Lead.)
[1] 32
>
> #C
> hist(DataHW4_2$X.Lead, xlab = "Lead", main = "Histogram of Lead")
> summary(DataHW4_2$X.Lead.)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  3.00  11.00   32.00   37.24  46.00  210.00
> sort(DataHW4_2$X.Lead.)
[1] 3 5 6 7 9 10 10 11 11 11 12 13 17 23 23 27 30 32 32
[20] 37 38 38 39 41 41 41 44 46 48 52 52 53 55 62 77 112 210
> library(BSDA)
> SIGN.test(DataHW4_2$X.Lead, md=30)
```

One-sample Sign-Test

```
data: DataHW4_2$X.Lead.
s = 20, p-value = 0.6177
alternative hypothesis: true median is not equal to 30
95 percent confidence interval:
 17.34363 41.00000
sample estimates:
median of x
      32
```

	Conf. Level	L. E. pt	U. E. pt
Lower Achieved CI	0.9011	23.0000	41
Interpolated CI	0.9500	17.3436	41
Upper Achieved CI	0.9530	17.0000	41

```
> t.test(DataHW4_2$X.Lead., mu=30)
```

One Sample t-test

data: DataHW4_2\$X.Lead.
t = 1.1868, df = 36, p-value = 0.2431
alternative hypothesis: true mean is not equal to 30
95 percent confidence interval:
24.86550 49.62099
sample estimates:
mean of x
37.24324

2G

```
> mean.fun <- function(d,i)
+ {
+   m <- mean(d[i])
+   n <- length(i)
+   v <- (n-1)*var(d[i])/n^2
+   c(m,v)
+ }
> set.seed(7255)
> resultsHW4_2F <- boot(data=DataHW4_2$X.Lead., mean.fun, R=1000)
> boot.ci(resultsHW4_2F, type="all")
BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 1000 bootstrap replicates
```

CALL :
boot.ci(boot.out = resultsHW4_2F, type = "all")

Intervals :

Level	Normal	Basic	Studentized
95%	(25.31, 48.78)	(23.66, 47.92)	(27.46, 57.19)

Level	Percentile	BCa
95%	(26.57, 50.83)	(28.02, 53.82)

Calculations and Intervals on Original Scale
Some BCa intervals may be unstable

```
> RocketPropellantData <- read.csv(file.choose())
> RocketPropellantData
  X.Mixture1. X.Mixture2.
1          185          221
2          192          210
3          201          215
```

4	215	202
5	170	204
6	190	196
7	175	225
8	172	230
9	198	214
10	202	217

```

> y1 <- mean(RocketPropel antData$X. Mi xture1. )
> s1 <- sd(RocketPropel antData$X. Mi xture1. )
> y2 <- mean(RocketPropel antData$X. Mi xture2. )
> s2 <- sd(RocketPropel antData$X. Mi xture2. )
> y1; s1; y2; s2
[1] 190
[1] 14.65151
[1] 213.4
[1] 10.62701
> n1 <- length(RocketPropel antData$X. Mi xture1. )
> n2 <- length(RocketPropel antData$X. Mi xture2. )
> t.test(RocketPropel antData$X. Mi xture1. , RocketPropel antData$X. Mi xture2. , var
.equal = TRUE)

```

Two Sample t-test

```

data: RocketPropel antData$X. Mi xture1. and RocketPropel antData$X. Mi xture2.
t = -4.0883, df = 18, p-value = 0.0006899
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -35.42491 - 11.37509
sample estimates:
mean of x mean of y
 190.0      213.4

```