

Lab 4 (100 pts.) - Interpretation of Confidence Intervals and Power Analysis for Z tests**Objectives: A Better Understanding of Confidence Intervals and Power Curves.**

A. (55 points) Interpretation of a Confidence Interval (no data file required). Use software to generate 40 observations from a normal distribution with $\mu = 10$ and $\sigma = 2$. Repeat this 30 times.

1. (30 points) From each set of observations, compute a 95% confidence interval. No data is required, however, you need to include all 30 confidence intervals.

Solution:**Sample Code 1:**

(You should run the following code 30 times. Each time, you should copy the table, record the confidence interval and determine by hand if this confidence interval contains the population mean: 10.)

```
n <- 40
RandomData <- rnorm(n, mean=10, sd=2)

ci <- t.test(RandomData, conf.level=0.95)

answer <- data.frame(i, ci$conf.int[1], ci$conf.int[2])
print(answer)}
```

You may choose to just print out 'ci' and then copy past the answers into another table.

Sample Code 2:

(You may consider the following code, which is more difficult to write, but generate the 30 tables simultaneously.)

```
res <- matrix(, nrow = 30, ncol = 3)
colnames(res) <- c("95% lower limit", "95% upper limit", "contain mu?")
for (i in 1:30){
  RandomSample <- rnorm(40, 10, 2)
  t <- t.test(RandomSample, conf.level=0.95)
  res[i, 1:2] <- t$conf.int[1:2]
  res[i, 3] <- (10 > t$conf.int[1]) & (10 < t$conf.int[2])
}
res
```

	95% lower limit	95% upper limit	contain mu?
[1,]	9.317943	10.675286	1
[2,]	9.503307	11.054551	1
[3,]	9.072149	10.264357	1
[4,]	8.834118	9.993473	0
[5,]	8.457915	9.828416	0
[6,]	9.562716	10.851474	1
[7,]	10.076547	11.217275	0
[8,]	9.668782	11.268103	1
[9,]	9.209580	10.808303	1
[10,]	9.144098	10.535114	1
[11,]	9.754032	10.904465	1
[12,]	8.910961	10.596295	1
[13,]	9.465037	10.603788	1
[14,]	9.267612	10.664503	1
[15,]	9.361587	10.483777	1
[16,]	9.385869	10.929714	1
[17,]	9.522930	11.157805	1
[18,]	9.000750	10.559026	1
[19,]	9.438331	10.792515	1
[20,]	9.775602	10.829115	1
[21,]	9.167850	10.469815	1
[22,]	9.333890	10.411812	1
[23,]	9.349506	10.688227	1
[24,]	9.360632	10.702535	1
[25,]	9.296839	10.573370	1
[26,]	9.032976	10.292199	1
[27,]	9.649248	10.966749	1
[28,]	9.174581	10.511815	1
[29,]	9.555404	10.749094	1
[30,]	8.909977	10.120440	1

2. (10 points) Determine how many of these intervals contain the population mean, $\mu = 10$. Please indicate for each confidence interval if it contains the value or not. Is this number that you would expect? Why or why not?

Solution:

We can see that, in this case, there are 3 confidence intervals that do not contain the true population mean, 10. This is roughly what we would expect. The confidence level is 95%, so we would expect $(0.05)(30) = 1.5$ of the confidence intervals to **NOT** include the population mean, 10.

3. (15 points) GROUP PART: This is a group assignment and is due on Blackboard at Midnight on FRIDAY, March 13. Be sure that the names and sections of each person at the top of the page. Combine your data with 3 or 4 other students (in any of my sections) and answer the following questions (no data is required for this part):
- Is the number of intervals that contain the mean what you would expect for the combined data?
 - How are the results from part 2 and part 3 different?

Solution:

You will need to show your work to the number of intervals that contain the mean (just add up the numbers from each student) and then calculate the percentage by dividing by the total number. Then compare the results from each of the students and the combined number of intervals

I would expect that this number would be more accurate because this is for a larger sample size. This percentage is only valid at large numbers.

B. (45 points) Water quality testing (no data file required). The Deely Laboratory is a drinking-water testing and analysis service. One of the common contaminants it tests for is lead. Lead enters drinking water through corrosion of plumbing materials, such as lead pipes, fixtures, and solder. The service knows that their analysis procedure is unbiased but not perfectly precise, so the laboratory analyzes each water sample three times and reports the mean result. The repeated measurements follow a Normal distribution quite closely. The standard deviation of this distribution is a property of the analytic procedure and is known to be $\sigma = 0.25$ parts per billion (ppb).

The Deely Laboratory has been asked by the university to evaluate a claim that the drinking water in the Student Union has a lead concentration of 6 ppb, well below the Environmental Protection Agency's action level of 15 ppb. Since the true concentration of the sample is the mean μ of the population of repeated analyses, the hypotheses are

$$H_0: \mu = 6$$

$$H_a: \mu \neq 6$$

The lab chooses the 1% level of significance, $\alpha = 0.01$. They plan to perform three analyses of one specimen ($n=3$).

1. (30 points, 6 points each part) Using computer software, calculate the following powers:
 - a. At the 1% level of significance, what is the power of this test against the specific alternative $\mu = 6.5$?

Solution:

```
# variables to change:
# n, alpha, mu0, sigma
# muprime: the real value of mu
# by 0.05 - this needs to be fine enough so that the curve
# looks smooth

n <- 3
alpha <- 0.01
mu0 <- 6
sigma <- 0.25
sigman <- sigma/sqrt(n) #standard error
#z is from alpha/2 for a 2-tailed test
z <- qnorm(1 - alpha/2)
muprime <- 6.5
x1 <- mu0 - z*sigman #Value for which x < x1 => H_0 will be rejected
x2 <- mu0 + z*sigman #Value for which x > x2 => H_0 will be rejected
px1 <- pnorm(x1, muprime, sigman) #CDF up to x1 for various muprime
px2 <- pnorm(x2, muprime, sigman, lower.tail = FALSE) #upper tail
power <- px1 + px2 #Left Tail + Right Tail (Tails NOT symmetric!)
power
[1] 0.8128029
```

The power is 0.8128029

- b. At the 5% level of significance, what is the power of this test against the specific alternative $\mu = 6.5$?

Solution:

```
# variables to change:
# n, alpha, mu0, sigma
# muprime: the real value of mu
#   by 0.05 - this needs to be fine enough so that the curve
#             looks smooth

n <- 3
alpha <- 0.05
mu0 <- 6
sigma <- 0.25
sigman <- sigma/sqrt(n) #standard error
#z is from alpha/2 for a 2-tailed test
z <- qnorm(1 - alpha/2)
muprime <- 6.5
x1 <- mu0 - z*sigman #Value for which x < x1 => H_0 will be rejected
x2 <- mu0 + z*sigman #Value for which x > x2 => H_0 will be rejected
px1 <- pnorm(x1, muprime, sigman) #CDF up to x1 for various muprime
px2 <- pnorm(x2, muprime, sigman, lower.tail = FALSE) #upper tail
power <- px1 + px2 #Left Tail + Right Tail (Tails NOT symmetric!)
power
```

```
[1] 0.9337271
```

The power is 0.9337271

- c. At the 1% level of significance, what is the power of this test against the specific alternative $\mu = 6.75$?

Solution:

```
# variables to change:
# n, alpha, mu0, sigma
# muprime: the real value of mu
#   by 0.05 - this needs to be fine enough so that the curve
#             looks smooth

n <- 3
alpha <- 0.01
mu0 <- 6
sigma <- 0.25
sigman <- sigma/sqrt(n) #standard error
#z is from alpha/2 for a 2-tailed test
z <- qnorm(1 - alpha/2)
muprime <- 6.75
x1 <- mu0 - z*sigman #Value for which x < x1 => H_0 will be rejected
x2 <- mu0 + z*sigman #Value for which x > x2 => H_0 will be rejected
px1 <- pnorm(x1, muprime, sigman) #CDF up to x1 for various muprime
px2 <- pnorm(x2, muprime, sigman, lower.tail = FALSE) #upper tail
power <- px1 + px2 #Left Tail + Right Tail (Tails NOT symmetric!)
power
```

```
[1] 0.9956077
```

The power is 0.9956077

- d. If the lab performs five analyses of one specimen ($n=5$), what is the power of this test against the specific alternative $\mu = 6.5$?

Solution:

```
# variables to change:
# n, alpha, mu0, sigma
# muprime: the real value of mu
#   by 0.05 - this needs to be fine enough so that the curve
#             looks smooth

n <- 5
alpha <- 0.01
mu0 <- 6
sigma <- 0.25
sigman <- sigma/sqrt(n) #standard error
#z is from alpha/2 for a 2-tailed test
z <- qnorm(1 - alpha/2)
muprime <- 6.5
x1 <- mu0 - z*sigman #Value for which x < x1 => H_0 will be rejected
x2 <- mu0 + z*sigman #Value for which x > x2 => H_0 will be rejected
px1 <- pnorm(x1, muprime, sigman) #CDF up to x1 for various muprime
px2 <- pnorm(x2, muprime, sigman, lower.tail = FALSE) #upper tail
power <- px1 + px2 #Left Tail + Right Tail (Tails NOT symmetric!)
power

[1] 0.9710402
```

The power is 0.9710402

- e. Write a short paragraph explaining the consequences of changing the significance level, alternative μ and sample size on the power.

Solution:

We have the following conclusion:

The larger the value of α (or equivalently the higher the significance level), the greater the power is.

The larger the distance between the μ' and μ_0 is, the greater the power.

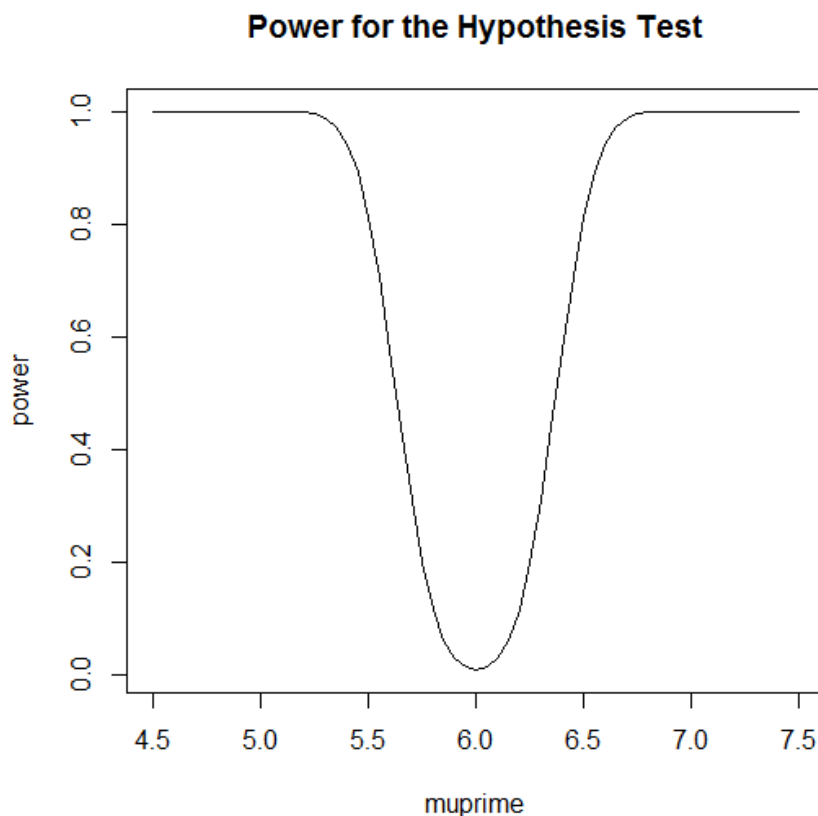
The larger the sample size, the greater the power.

2. (10 points) Generate a power curve when $n=3$ at a 1% significance level. Please use an interval length of 3.

Solution:

```
# variables to change:
# n, alpha, mu0, sigma
# muprime: the real value of mu
# by 0.05 - this needs to be fine enough so that the curve
# looks smooth

n <- 3
alpha <- 0.01
mu0 <- 6
sigma <- 0.25
sigman <- sigma/sqrt(n) #standard error
#z is from alpha/2 for a 2-tailed test
z <- qnorm(1 - alpha/2)
muprime <- seq (from=4.5,to=7.5, by=0.05)
x1 <- mu0 - z*sigman #Value for which  $x < x_1 \Rightarrow H_0$  will be rejected
x2 <- mu0 + z*sigman #Value for which  $x > x_2 \Rightarrow H_0$  will be rejected
px1 <- pnorm(x1, muprime, sigman) #CDF up to  $x_1$  for various muprime
px2 <- pnorm(x2, muprime, sigman, lower.tail = FALSE) #upper tail
power <- px1 + px2 #Left Tail + Right Tail (Tails NOT symmetric!)
plot(muprime,power,main="Power for the Hypothesis Test",type="l")
```



3. (5 points) What sample size would be required for the power to be at least 0.90 at the 1% level of significance against the specific alternative $\mu = 6.3$?

Solution:

```
#this time, I will fix muprime and change n;
n <- 1:20 #these values may be changed as appropriate
muprime <- 6.3
alpha <- .01
mu0 <- 6
sigma <- 0.25
sigman <- sigma/sqrt(n)
z <- qnorm(1 - alpha/2)
x1 <- mu0 - z*sigman #Value for which x < x1 => H_0 will be rejected
x2 <- mu0 + z*sigman #Value for which x > x2 => H_0 will be rejected
px1 <- pnorm(x1, muprime, sigman) #CDF up to x1 for various muprime
px2 <- pnorm(x2, muprime, sigman, lower.tail = FALSE) #S_X(x2): muprime
power <- px1 + px2 #Left Tail + Right Tail (Tails NOT symmetric!)
beta <- 1 - power #When obtaining by 'hand' beta is easier to find first
```

```
answer <- data.frame(n, power)
```

```
answer
```

	n	power
1	1	0.08451698
2	2	0.18977182
3	3	0.30946629
4	4	0.43021435
5	5	0.54278498
6	6	0.64190611
7	7	0.72543766
8	8	0.79340227
9	9	0.84712267
10	10	0.88855966
11	11	0.91985851
12	12	0.94307157
13	13	0.96001259
14	14	0.97220011
15	15	0.98085565
16	16	0.98693152
17	17	0.99115150
18	18	0.99405411
19	19	0.99603281
20	20	0.99737057

As seen above, we need the sample size to be at least 11.