BIOS: 4120 Lab 11 Answers

April 14-15, 2018

Practice Problems

1. We are interested in testing whether a certain at-risk population for diabetes has a daily sugar intake that is equal to the general population, which is equal to 77 grams/day. A sample of size 37 was taken from this at-risk population, and we obtained a sample mean of 80 and sample standard deviation of 11 grams.
   Perform a hypothesis test to test whether this population has a significantly different mean sugar intake from 77 grams.

2. The distribution of LDL cholesterol levels in a certain population is approximately normal with mean 90 mg/dl and standard deviation 8 mg/dl.

(a) What is the probability an individual will have a LDL cholesterol level above 95 mg/dl?

(b) Suppose we have a sample of 10 people from this population. What is the probability of exactly 3 of them being above 95 mg/dl?

(c) Take the sample of size 10, as in part b. What is the probability that the sample mean will be above 95 mg/dl?

(d) Suppose we take 5 samples of size 10 from the population. What is the probability that at least one of the sample means will be greater than 95 mg/dl?

3. In the following scenarios, identify what will happen to the power of a hypothesis test:

(a) We increase the sample size.

(b) The standard deviation of the sample is larger than what we expected.

(c) Our effect size moves from 5 units to 10 units.

Solutions

Problem 1

```
mu <- 77
xbar <- 80
s <- 11
n <- 37

# Use a t-test for the sample

t <- (xbar-mu)/(s/sqrt(n))

2*pt(abs(t),n-1,lower.tail=FALSE)
```

```
## [1] 0.1058177
```

We do not have significant evidence to conclude that the at-risk mean sugar intake is different from the general population mean sugar intake (p = 0.11).

Problem 2

```
## Part 2a
mu <- 90
sigma <- 8

(p <- pnorm(95,mu,sigma,lower.tail=FALSE))
```

```
## [1] 0.2659855
```

```
## Part 2b
dbinom(3,10,p)
```

```
## [1] 0.2592314
```

```
## Part 2c
z <- (95-90)/(sigma/sqrt(10))

(p <- pnorm(z,lower.tail=FALSE))
```

```
## [1] 0.02405341
```

```
## Part 2d
1-dbinom(0,5,p)
```

```
## [1] 0.1146189
```

Problem 3

$H_0 : \pi = 0.5$

$H_A : \pi \neq 0.5$

$\pi = 0.5$
$\hat{\pi} = \frac{155}{235} = 0.66$
$n = 235$

$SE = \sqrt{\frac{\pi(1-\pi)}{n}}$

$SE = \sqrt{\frac{0.5(1-0.5)}{235}}$

$SE = 0.0326$

Remember that to compute a test statistic we use:

$z = \frac{\hat{\pi}-\pi}{SE}$

$z = \frac{0.66-0.5}{0.0326}$

$z = 4.91$

Find 2-tailed probability by looking up this z-score on the z-table or in R:

```
(pnorm(4.91, lower.tail = F)*2)
```

```
## [1] 9.107639e-07
```

Interpretation: Based on this data, there is stong evidence to suggest that there is a difference in the true proportion of fans who cheer for the Cubs vs the Sox.

Creating a confidence interval (z)

Now we want to create a 95% confidence interval for π. *Interpret the interval.*

Remember that now standard error is based on $\hat{\pi}$ and becomes:

$SE = \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}$

$SE = \sqrt{\frac{0.66(1-0.66)}{235}}$

$SE = 0.0309$

We will have to find $z_{\alpha/2}$ using the z-table. *What is our α for a 95% confidence interval?*

$z_{\alpha/2} = 1.96$ (from table)

Remember that the equation for the confidence interval is:

$\hat{\pi} \pm z_{\alpha/2} * SE$
$0.66 \pm 1.96 * 0.0309$
(0.599, 0.721)

```
0.66 + qnorm(c(0.025,0.975))* sqrt((0.66*(1-0.66))/235)
```

```
## [1] 0.5994345 0.7205655
```

Interpretation: We can say with 95% confidence that this interval (0.599, 0.721) contains the true proportion of Chicago fans who cheer for the Cubs. Note that we reject the null hypothesis that there is no difference as 0.5 does not fall within our confidence interval. We can conclude that in fact there are more Chicago fans cheering for the Cubs.

Problem 4

a)

The variability of the sample means will be lower than the variability of the population by a factor of $\frac{1}{\sqrt{20}}$.

b)

The probability that our sample mean will lie between 29 and 31 is given by $P(29 < X < 31)$, where $X \sim N(29.5, (\frac{9.25}{\sqrt{20}})^2)$. We then observe $P(29 < X < 31)$ can be given by the following R code:

```
pnorm(31, mean = 29.5, sd = 9.25/sqrt(20)) - pnorm(29, mean = 29.5, sd = 9.25/sqrt(20))
```

```
## [1] 0.3613468
```

```
#or#

pnorm((31 - 29.5)/(9.25/sqrt(20))) - pnorm((29 - 29.5)/(9.25/sqrt(20)))
```

```
## [1] 0.3613468
```

c)

In order to find the values that contain the middle 50% of our data, we need to find the 25th and 75th percentile of the data. That can be done with the following code:

```
qnorm(.25, mean = 29.5, sd = 9.25/sqrt(20))
```

```
## [1] 28.10491
```

```
qnorm(.75, mean = 29.5, sd = 9.25/sqrt(20))
```

```
## [1] 30.89509
```

```
#or#

(9.25/sqrt(20))*qnorm(.25) + 29.5
```

```
## [1] 28.10491
```

```
(9.25/sqrt(20))*qnorm(.75) + 29.5
```

```
## [1] 30.89509
```

d)

Now since we do not know our standard deviation for the population, we need to use a t-test. Thus, our 95% confidence interval will be given by $\bar{X} \pm t_{19,.975} \times SE(\bar{X}) = 30.1 \pm 2.093 \times \frac{9.25}{\sqrt{20}}$. This is given in R with:

```
xbar <- 30.1
se <- 8.95/sqrt(20)

xbar + c(-1, 1)*qt(.975, df = 19)*se
```

```
## [1] 25.91127 34.28873
```

e)

The confidence interval in part d) is wider because we are using a t-test instead of a z-test because we are not aware of the true population variaibility of albumin levels in part d). To account for our uncertainty, we needed to use the t-test which allows for the possibility that the true variance is larger than we have observed. The variables that affect the width of our confidence interval are our test statistic, which was given in this case to be the 97.5th percentile of the t-distribution, and our standard error which is our sample standard

deviation divided by the square root of n. ### Problem 5
Part A: Power increases
Part B: Power decreases
Part C: Power increases