Lab Test 2

Sean Park

2021 8 23

1.

a) The probability is 0.15

```
pnorm(2500,3020,502)

## [1] 0.1501345
```

b) The probability is 0.75.

```
pnorm(2500,3020,502)*5

## [1] 0.7506726
```

i) H0 = The probability of underweight baby born is not changed.
(Pr(X < 3020) = 0.5)
H1 = The probability of underweight baby born is decreased.
(Pr(X < 3020) < 0.5)

ii) p-value is 0.026.

```
pbinom(180,400,0.5)

## [1] 0.02552011

prop.test(180,400,0.5, alternative="less", conf.level = 0.95)

##
##  1-sample proportions test with continuity correction
##
## data:  180 out of 400, null probability 0.5
## X-squared = 3.8025, df = 1, p-value = 0.02559
## alternative hypothesis: true p is less than 0.5
## 95 percent confidence interval:
##  0.0000000 0.4923665
## sample estimates:
##    p
## 0.45
```

iii) Since the p-value is less than 0.05. We reject null hypotheses. Therefore the proportion of underweight baby born statistically decreased in 95% confidence level.

iv) Based on the sample, the proportion of underweight babies should be between 0.401 and 0.499 in 95% confidence level. Confidence interval does not contain null hypothesis value 0.5.

```
t.score = qt(0.025,399,lower.tail = FALSE)
pi.hat.obs = 180/400
pi.hat.se = sqrt((pi.hat.obs*(1-pi.hat.obs))/400)
lower.95CI.bound = pi.hat.obs-(t.score*pi.hat.se)
upper.95CI.bound = pi.hat.obs+(t.score*pi.hat.se)
print(lower.95CI.bound)

## [1] 0.4010982

print(upper.95CI.bound)

## [1] 0.4989018
```

2.

a) observed odds of diabetes for males is 0.05.

```
pi.male = 4000/84000
odds.male = pi.male/(1-pi.male)
odds.male

## [1] 0.05
```

b) observed odds of diabetes for female is 0.04

```
pi.female = 3000/78000
odds.female = pi.female/(1-pi.female)
odds.female

## [1] 0.04
```

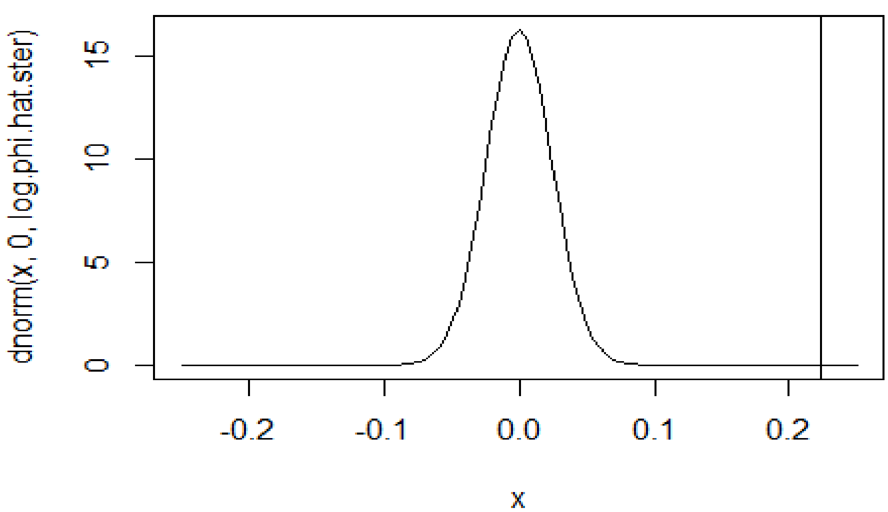
c) odds ratio of diabetes for males to female is 1.25

```
odds.ratio = odds.male/odds.female
odds.ratio

## [1] 1.25
```

d) For testing the odds of getting diabetes is greater for male and female.
Our H0 = ϕ = odds.male/odds.female = 1,
H1 = ϕ = odds.male/odds.female > 1
This equivalent to
H0 = log(ϕ) = log(odds.male/odds.female) = 0,
H1 = log(ϕ) = log(odds.male/odds.female) > 0
In this case if we reject null hypothesis and take alternative than it will prove that odds for male is greater.

Assume the natural logarithm of the estimator of the odds ratio is approximate Gaussian distribution.

```
phi.hat.obs = odds.ratio
number.male = 84000
number.female = 78000
pi.hat.obs = (4000+3000)/(84000+78000)
log(phi.hat.obs)

## [1] 0.2231436

log.phi.hat.ster = sqrt(1/(number.male*pi.hat.obs*(1-pi.hat.obs))+1/(number.f
emale*pi.hat.obs*(1-pi.hat.obs)))
log.phi.hat.ster

## [1] 0.02445517

log.phi.hat.obs = log(phi.hat.obs)
p.value = pnorm(log.phi.hat.obs,0,log.phi.hat.ster,lower.tail = FALSE)
p.value

## [1] 3.600224e-20

curve(dnorm(x,0,log.phi.hat.ster), from=-0.25, to=0.25)
abline(v = log.phi.hat.obs)
```



Since we have almost zero p-value, we reject our H0 at conventional significance level (5%). The data provides evidence that the true odds ratio of getting diabetes for males to compared to females is not one. Given the point estimate for odds ratio, it may tell us that males has higher chance to get diabetes than female. It needs further investigation.

e)

```
log.phi.hat.se.ci = sqrt(1/4000 + 1/80000 + 1/3000 + 1/75000)
UL = log.phi.hat.obs + 1.96*log.phi.hat.se.ci
LL = log.phi.hat.obs - 1.96*log.phi.hat.se.ci
exp(c(LL,UL))
```

```
## [1] 1.190970 1.311956
```

```
Calculating the standard error using the short-cut formula and apply 95% conf
idence interval gives us (1.19,1.31). We are 95% confident that this interval
 contains the true odds-ratio of diabetes for males to females.
Therefore, males are more likely to get diabetes than females.
```