Qualifying Exam 2019

Exam #7

6/1/2019

Question 1

a) Number of meals involving fish as a positive test

```r
# epiR package for calculating sensitivity and specificity
sensspec0 <- epi.tests(ctable0)
sensspec1 <- epi.tests(ctable1)
sensspec2 <- epi.tests(ctable2)
sensspec3 <- epi.tests(ctable3)
sensspec4 <- epi.tests(ctable4)
sensspec7 <- epi.tests(ctable7)
sensspec14 <- epi.tests(ctable14)
sensspec21 <- epi.tests(ctable21)
```

	Sensitivity	Specificity
>=0	100	0.0
>=1	100	8.0
>=2	100	19.2
>=3	100	28.0
>=4	90	28.8
>=7	70	36.8
>=14	30	89.6
>=21	30	93.6

b) Appropriate thresholds

Sensitivity refers to the true positive rate, or the probability that a test will rule in disease correctly. Specificity indicates the true negative rate, or the probability that a test will correctly rule out disease. Therefore, the probability of a false negative is 100 - sensitivity and the the false positive rate is 100 - specificity.

i. True positives

If we want to maximize true positives while minimizing false positives, the optimal threshold is the one with the highest sensitivity and lowest 100 - specificity. A threshold of >= 3 meals per week including fish would provide a 100% true positive rate and a 72% false negative rate.

ii. True negatives

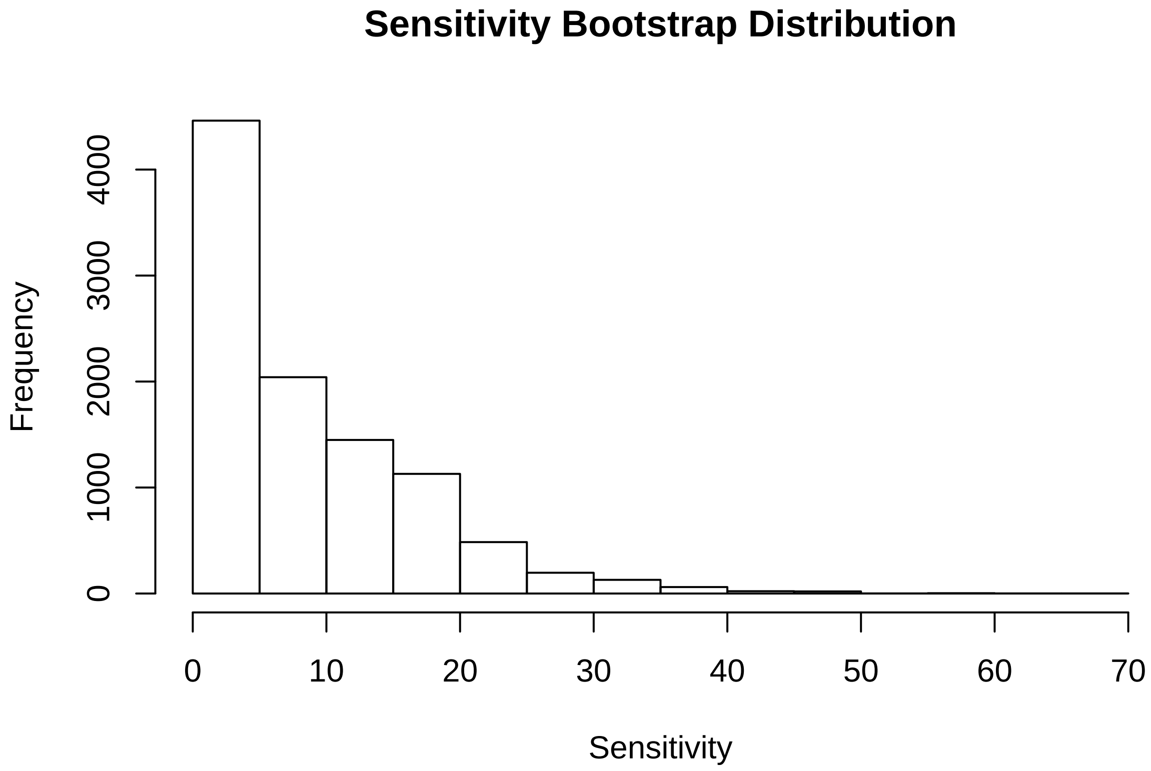
Maximizing true negatives first and then true positives requires choosing the test with highest specificity and highest sensitivity. In this case a threshold of >= 21 meals including fish per week would provide a true negative detection rate of 93.6% and a true positive rate of 30%.

c) Bootstrap sampling for >= 21 meals threshold

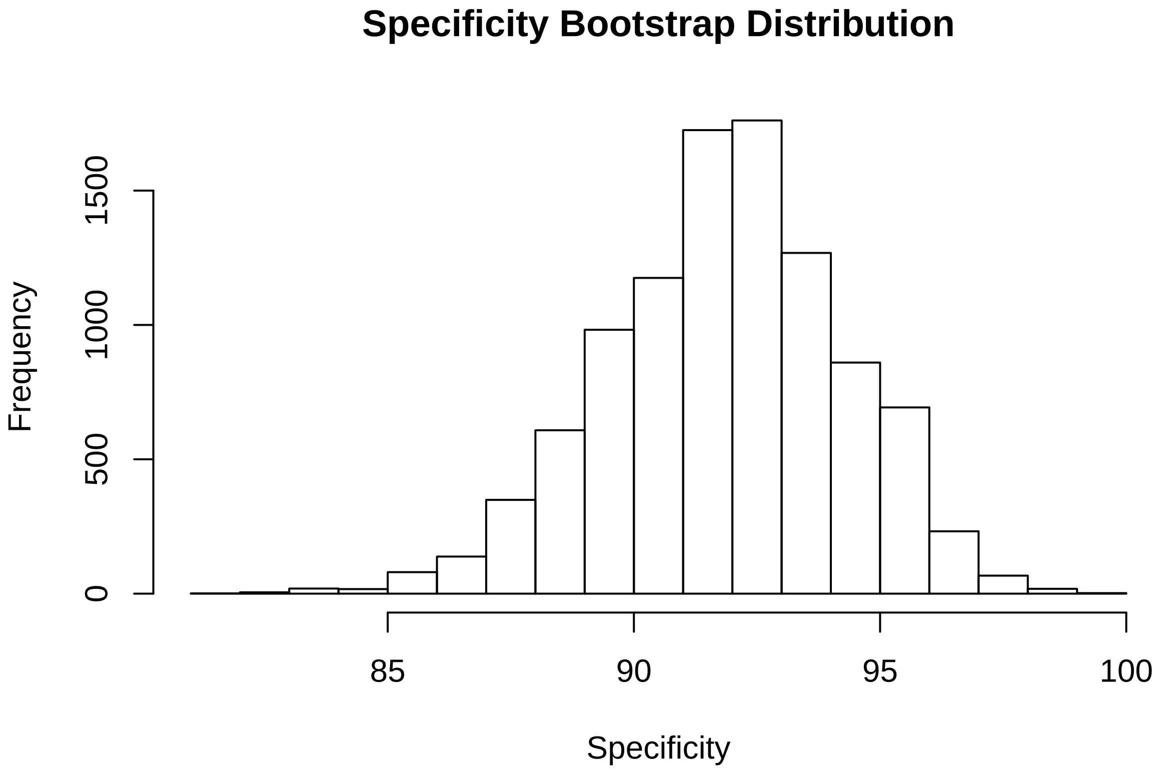
```
# Vector for storing results
set.seed(1234)
B <- 10000
sens_results <- numeric(B)
spec_results <- numeric(B)
# Loop
for (i in 1:B) {
  meals <- sample(fish$fishmlwk,replace = T)
  meals <- ifelse(meals >= 21,1,0)
  response <- sample(fish$MeHg,replace = T)
  response <- ifelse(response >= 8,1,0)
  table <- table(factor(meals,levels=1:0),factor(response,levels=1:0))
  sens_results[i] <- (table[1,1]/sum(table[,1])) * 100
  spec_results[i] <- (table[2,2]/sum(table[,2])) * 100
}
```

i. Plots

```
# Plots
hist(sens_results,main = "Sensitivity Bootstrap Distribution",xlab = "Sensitivity")
```



```
hist(spec_results,main = "Specificity Bootstrap Distribution",xlab = "Specificity")
```



ii. Mean, SE, and Bias From Bootstrap Distributions

	Mean	Standard Error	Bias
Sensitivity	8.16521	0.0922186	-21.83479
Specificity	91.87575	0.0240551	-1.72425

iii. 90% Bootstrap and Normal Percentile Confidence Intervals

```
# Sensitivity
# Normal percentiles
L <- mean(sens_results) - (1.645 * sd(sens_results))
L
```

```
## [1] -7.004749
```

```
U <- mean(sens_results) + (1.645 * sd(sens_results))
U
```

```
## [1] 23.33517
```

```
# Coverage
sum(sens_results < L)/B
```

```
## [1] 0
```

```r
sum(sens_results > U)/B
```

```
## [1] 0.0671
```

```r
# Bootstrap percentiles
quantile(sens_results,c(0.05,0.95))
```

```
##  5% 95%
##   0  25
```

The bootstrap distribution for sensitivity is not at all normal. The 90% confidence interval for this distribution using normal percentiles is (-7.00%,23.34%), which does not make sense as sensitivity cannot be negative. Also, none of the bootstrap values were below the lower limit (again, because this is impossible), when we'd expect that 5% would be for a normal distribution. So in this case it would probably be better to use the bootstrap confidence interval (0%,25%).

```r
# Specificity
# Normal percentiles
Lc <- mean(spec_results) - (1.645 * sd(spec_results))
Lc
```

```
## [1] 87.91868
```

```r
Uc <- mean(spec_results) + (1.645 * sd(spec_results))
Uc
```

```
## [1] 95.83282
```

```r
# Coverage
sum(spec_results < Lc)/B
```

```
## [1] 0.0558
```

```r
sum(spec_results > Uc)/B
```

```
## [1] 0.049
```

```r
# Bootstrap percentiles
quantile(spec_results,c(0.05,0.95))
```

```
##       5%      95%
## 87.80488 95.79832
```

The bootstrap distribution for specificity appears to be much closer to normal than sensitivity. The 90% normal percentile confidence interval is (87.92%,95.83%), which matches the bootstrap confidence interval very closely (87.80%,95.80%). Also, approximately 5% percent of the bootstrap values were below the lower limit and above the upper limit, which is what we would expect from a normal distribution.

d. 90% Confidence Intervals Using Exact and Asymptotic Methods

i. Sensitivity

Clopper-Pearson Method

$$\hat{p} = \frac{x}{n} 0.3 = \frac{x}{135} x = 0.3 * 135 = 40.5$$

$$\text{CI} = (\frac{x}{x + (n - x + 1)F_{1-\frac{\alpha}{2};2(n-x+1),2x}}, \frac{(x+1)F_{1-\frac{\alpha}{2};2(x+1),2(n-x)}}{(n-x) + (x+1)F_{1-\frac{\alpha}{2};2(x+1),2(n-x)}})$$

```
n <- 135
x <- 0.3 * 135
L <- x/(x+((n-x+1)*qf(0.95,(2*(n-x+1)),2*x))) * 100
L
```

```
## [1] 23.52392
```

```
U <- (x+1)*qf(0.95,(2*(x+1)),2*(n-x))/((n-x)+(x+1)*qf(0.95,(2*(x+1)),2*(n-x))) * 100
U
```

```
## [1] 37.14843
```

The Clopper-Pearson CI for sensitivity is (23.52%,37.15%).

ii. Specificity

1. Clopper-Pearson Method

$$\hat{p} = \frac{x}{n}0.936 = \frac{x}{135}x = 0.936 * 135 = 126.36$$

$$\text{CI} = (\frac{x}{x + (n-x+1)F_{1-\frac{\alpha}{2};2(n-x+1),2x}}, \frac{(x+1)F_{1-\frac{\alpha}{2};2(x+1),2(n-x)}}{(n-x)+(x+1)F_{1-\frac{\alpha}{2};2(x+1),2(n-x)}})$$

```
n <- 135
x <- 0.936*135
L <- x/(x+((n-x+1)*qf(0.95,(2*(n-x+1)),2*x))) * 100
L
```

```
## [1] 88.98013
```

```
U <- (x+1)*qf(0.95,(2*(x+1)),2*(n-x))/((n-x)+(x+1)*qf(0.95,(2*(x+1)),2*(n-x))) * 100
U
```

```
## [1] 96.67401
```

The Clopper-Pearson CI for specificity is (88.98%,96.67%).