DSC 462: Assignment 4

Shouman Das

December 5, 2017

Q1

(a)

$$n = 2030, \quad \hat{p} = 1050/n$$
$$CI_{95\%} = \hat{p} \pm z_{0.025}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$
$$\approx 0.52 \pm 0.02$$

```
n = 2030
p = 1050/n
p
```

```
## [1] 0.5172414
```

```
1.96* sqrt(p*(1-p)/n)
```

```
## [1] 0.02173801
```

(b) Note that

$$\alpha = 0.05$$
$$Z_{\alpha/2} = 1.96$$
$$E_o = 0.01$$

So we have $n = (\frac{Z_{\alpha/2}}{2E_o})^2 = 9604$.

```
(1.96/2/0.01)^2
```

```
## [1] 9604
```

Q2

(a)

```
n = 56
pHat = 20.5/n #for continuity correction
p0 = 0.25
Zobs = (pHat-p0)/sqrt(p0*(1-p0)/n)
alpha_obs = 1-pnorm(Zobs)
alpha_obs
```

```
## [1] 0.02243114
```

So p-value $= 0.0224$ and we can reject H_o at a significance level $\alpha = 0.05$.

(b) We can check the previous result by using R.

```
prop.test(x = 21, n = 56, p = .25, alternative = "greater", correct = TRUE)
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  21 out of 56, null probability 0.25
## X-squared = 4.0238, df = 1, p-value = 0.02243
## alternative hypothesis: true p is greater than 0.25
## 95 percent confidence interval:
##  0.268643 1.000000
## sample estimates:
##     p
## 0.375
```

Q3

We can do this problem by using R.

```
n1 = 250
n2 = 500
x1 = 65
x2 = 100
prop.test(x =c(x2,x1), n = c(n2,n1), correct = F)
```

```
##
##  2-sample test for equality of proportions without continuity
##  correction
##
## data:  c(x2, x1) out of c(n2, n1)
## X-squared = 3.4965, df = 1, p-value = 0.0615
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  -0.124696627  0.004696627
## sample estimates:
## prop 1 prop 2
##   0.20   0.26
```

(1) So the 0.95 confidence interval for $p_2 - p_1$ is $(-0.12469, 0.00469)$.

(2) Since the P-value is 0.0615, the null hypothesis cannot be rejected.

Q4

```
n11 = 3
n12 = 31
n21 = 12
n22 = 45
OR = (n11*n22)/(n12*n21)
SE.log.OR = sqrt(1/n11+1/n12+1/n21+1/n22)
log(OR)
```

```
## [1] -1.013619
```

```
1.96*SE.log.OR
```

```
## [1] 1.345347
```

So 95% CI for the log odds ratis is -1.10136 ± 1.3453. This also means that we cannot reject the null hypothesis, since 0 is contained in the interval.

Q5

(a) To construct a 90% upper confidence bound for σ, we need

$$\chi^2_{n-1,1-\alpha} = \chi^2_{5,0.90} = 1.61$$

which gives

$$\sigma < \frac{S_n}{\sqrt{\chi^2_{n-1,1-\alpha}/(n-1)}} = \frac{0.088}{\sqrt{1.61/5}} = 0.1550796$$

```
x = c(9.43, 9.85, 10.12, 9.89, 9.81, 10.3)
Sn = var(x)
Sn
```

```
## [1] 0.088
```

```
Sn/sqrt(1.61/5)
```

```
## [1] 0.1550796
```

(b) Using this upper bound for sample size estimation we get $n = 10$

```
alpha = 0.05
z.alpha.half = qnorm(1-alpha/2)
E.o = 0.1
sigma = 0.1550796
(z.alpha.half*sigma/E.o)^2
```

```
## [1] 9.238586
```

Q6

```
n1 = 14
n2 = 18
S1.sq = 645.16
S2.sq = 1413.76
alpha = 0.1
F = S1.sq/S2.sq
crit.up = qf(1-0.05, 13,17)
crit.low = qf(0.05, 13,17)
F
```

```
## [1] 0.4563434
```

```
crit.low
```

```
## [1] 0.4002126
```

```
crit.up
```

```
## [1] 2.353063
```

So the rejection region is $F \leq 0.4$ or $F \geq 2.35$ So we cannot reject the null hypothesis. Also, we can get our p-value $= 0.1569$.

```
2*min(pf(F, 13,17), 1- pf(F,13,17))
```

```
## [1] 0.156946
```

Q7

First calculate the variance for null hypothesis.

$$\sigma_0^2 = (1/2^2 + 1/6^2 + 1/3^2)\sigma_X^2 = 9.722$$

Also the variance for the alternative hypothesis,

$$\sigma_1^2 = \sigma_0^2 + 2/25 * (0.56/12 + 0.18/6 + 0.21/18) = 14.13889$$

So our hypothesis would be

$$H_o \ : \ \sigma^2 = \sigma_0^2, \quad H_a : \sigma^2 = \sigma_1^2$$

Now our test statistic

$$W = \frac{(n-1)S_n^2}{\sigma_0^2}$$

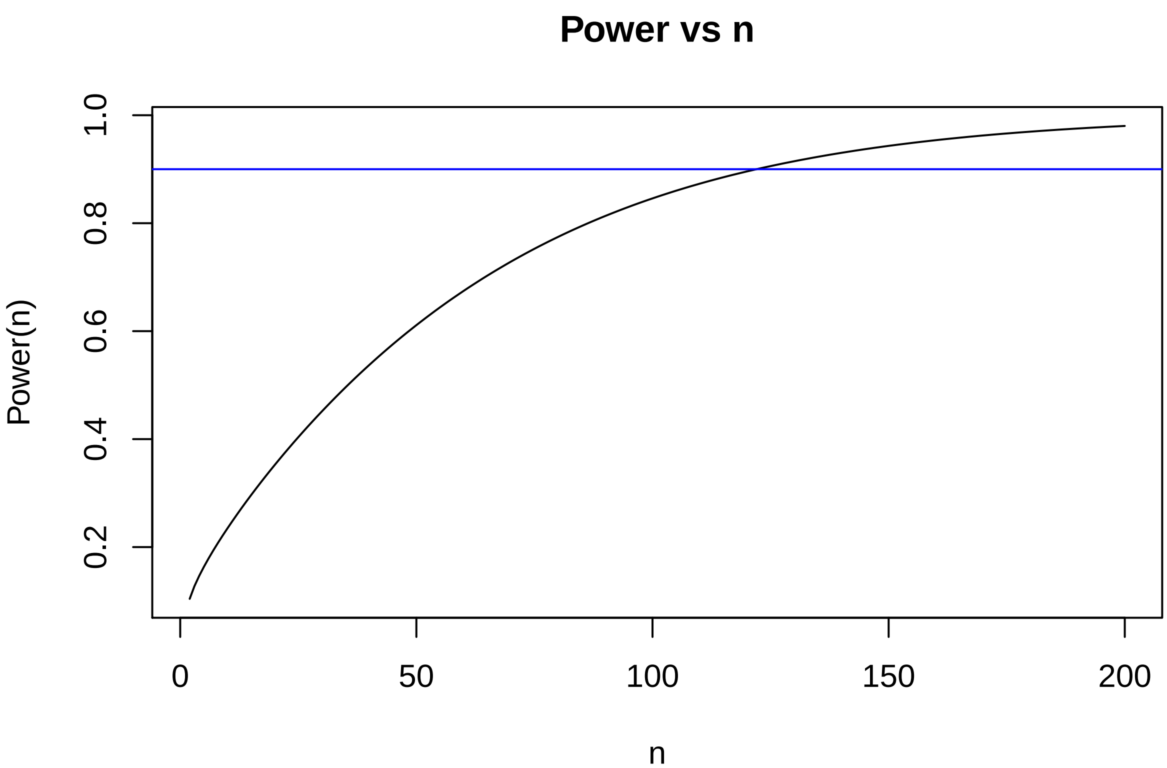
Now, if H_o is true then $W \sim \chi^2_{n-1}$. If H_a is true then $\frac{\sigma_0^2}{\sigma_1^2}W \sim \chi^2_{n-1}$. So our power would be

$$\text{Power}(n) = P(\text{reject } H_0 | H_a)$$

$$= P(W > \chi^2_{n-1,0.025}) = P(\frac{\sigma_0^2}{\sigma_1^2} \ W > \frac{\sigma_0^2}{\sigma_1^2} \ \chi^2_{n-1,0.025})$$

where $\frac{\sigma_0^2}{\sigma_1^2}W \sim \chi^2_{n-1}$.

```r
sigma0.sq = (1/4+1/36+1/9)*25
sigma1.sq = sigma0.sq + 2*25*(0.56/12+0.18/6+0.21/18)
Power = function(n){
    crit = sigma0.sq/sigma1.sq*qchisq(p = 0.05, df = n-1, lower.tail = FALSE)
    pchisq(q = crit, df = n-1, lower.tail = FALSE)
}
n = 2:200
plot(n, Power(n), type = 'l', main = "Power vs n")
abline(h = 0.9, col = 'blue')
```




```r
n[Power(n)>=0.9][1]
```

```
## [1] 123
```

We can see that 123 is the minimum number of sample required to have 90% power.