Homework 1

Miles Tweed

2/7/2021

#Problem 1

1.1

```r
prop.sample.size <- function(m, alpha) {
  (qnorm(1-(alpha/2))^2 * 0.25) / m^2
}
```

The worst case scenario is the one that leads to the largest standard error and would require the largest sample size to achieve the desired margin of error and confidence level. This occurs when the proportion are equivalent $\hat{p} = 0.5$ and $(1 - \hat{p}) = 0.5$

1.2

```r
# Exercise 8.50
# a)
prop.sample.size(m = 0.10, alpha = 0.05)
```

```
## [1] 96.03647
```

```r
# b)
prop.sample.size(m = 0.05, alpha = 0.05)
```

```
## [1] 384.1459
```

```r
# c)
prop.sample.size(m = 0.05, alpha = 0.01)
```

```
## [1] 663.4897
```

d)

By comparing parts a and b, decreasing the margin of error by half required the sample size to increase by almost four times. By comparing parts b and c, increasing the confidence interval from 0.95 to 0.99 doubled the required sample size.

1.3

```r
mean.sample.size <- function(m, alpha, s) {
  (s^2 * qnorm(1-alpha/2)^2)/m^2
}

# 8.53
# Since the range of incomes fall between $0 and $120,000
#and the distribution is approximately normal, we can assume
#that six standard deviations will capture most of the distribution.
```

```
#This means that a good approximation for one standard deviation
#would be s = 120000/6 = 20000
mean.sample.size(m=1000, alpha=0.01, s=20000)
```

[1] 2653.959

1.4 For a two-sample proportion test the formula for the standard deviation is:

$$se = \sqrt{\frac{\hat{p_1} \cdot (1 - \hat{p_1})}{n_1} + \frac{\hat{p_2} \cdot (1 - \hat{p_2})}{n_2}}$$

Using equally sized samples ($n_1 = n_2 \equiv n$):

$$m = z_{(1-\frac{\alpha}{2})} \cdot \sqrt{\frac{\hat{p_1} \cdot (1 - \hat{p_1})}{n} + \frac{\hat{p_2} \cdot (1 - \hat{p_2})}{n}}$$
$$m^2 = z^2_{(1-\frac{\alpha}{2})} \cdot \frac{\hat{p_1} \cdot (1 - \hat{p_1}) + \hat{p_2} \cdot (1 - \hat{p_2})}{n}$$
$$n = \frac{z^2_{(1-\frac{\alpha}{2})} \cdot (\hat{p_1} \cdot (1 - \hat{p_1}) + \hat{p_2} \cdot (1 - \hat{p_2}))}{m^2}$$

Assuming the worse case scenario where the sample size will be maximal ($p_1 = p_2 = 0.5$):

$$n = \frac{z_{(1-\frac{\alpha}{2})} \cdot (0.25 + 0.25)}{m^2}$$
$$= \frac{z_{(1-\frac{\alpha}{2})} \cdot 0.50}{m^2}$$

This equation differs from the one-sample proportion test in that the proportion component is double the size (0.50 versus 0.25).

Problem 2

2.1

Problem 9.66 from Agresti

a) Increasing the sample size reduces the probability of a Type II error because it causes the distributions of the null and alternative hypothesis to become more narrow which reduces their overlap.

b) The minimum sample size that achieved a power of 90% was $n = 83$.

c) The probability of a Type II error will increase if the true p is 0.40 instead of 0.50. This is because 0.40 is closer to, and harder to distinguish from, the null hypothesis of 0.33. This was confirmed using the app. The probability of a Type II error went from 10% to 72% ($n = 83$).

d) The power will increase if the significance level is changed from 0.05 to 0.10 because it is less probable to commit a Type II error. In the app the probability of a Type II error went from 10% to 6% and therefore the power went from 90% to 94% ($n = 83$)

2.2

a) $H_0 : p = 0.28 \ H_a : p \neq 0.28$

b) This means that, the probability of a Type II error will be 0.23 with a confidence level of 95% when the true proportion differs from the null hypothesis by 10% and an appropriate sample size is used.

c)
```
n <- 100
p.0 <- 0.28
p.a <- 0.38
alpha <- 0.05

rej.q <- qnorm(1-alpha/2,
               mean=p.0,
               sd= sqrt(p.0*(1-p.0)/n))

rej.p <-
pnorm(rej.q, mean = p.a, sd = sqrt(p.a*(1-p.a)/n)) -
  pnorm(-rej.q, mean = p.a, sd = sqrt(p.a*(1-p.a)/n))

power <- 1-rej.p
power
```

```
## [1] 0.597616
```

d)
```
beta <- 0.15

z.score.100 <-
(sqrt(n)*(p.a-p.0)+qnorm(beta)*sqrt(p.a*(1-p.a)))/sqrt(p.0*(1-p.0))

(1-pnorm(z.score.100))*2
```

```
## [1] 0.2684023
```

In this case, $\alpha = 0.26$ indicating that in order to have a power of 85% when the sample size is 100 the significance level of this study would be 0.74.

e)
```
sample.sz <-
  (qnorm(1-alpha/2)*sqrt(p.0*(1-p.0)) -
     qnorm(beta)*sqrt(p.a*(1-p.a)))^2 /
  (p.a-p.0)^2
sample.sz
```

```
## [1] 191.2944
```

Sanity check for the above calculation (should return $\alpha = 0.05$):
```
beta <- 0.15

z.score.adj <-
(sqrt(sample.sz)*(p.a-p.0)+qnorm(beta)*sqrt(p.a*(1-p.a)))/sqrt(p.0*(1-p.0))

(1-pnorm(z.score.adj))*2
```

```
## [1] 0.05
```

f)

I think it would be better to have a higher power with this type of clinical trial. It would be best to have a
high probability of detecting a difference if one exists because these studies are expensive and it becomes
un-ethical to subject large numbers of subjects to an experimental drug.

g)

If you could not increase the sample size, you could accept a lower significance level. The example below
shows that increasing the power to 0.90 could be achieved by relaxing the significance to $\alpha = 0.40$.

```
n <- 100
beta <- 0.10

z.score.100 <-
(sqrt(n)*(p.a-p.0)+qnorm(beta)*sqrt(p.a*(1-p.a)))/sqrt(p.0*(1-p.0))

(1-pnorm(z.score.100))*2
```

```
## [1] 0.3999186
```

#Problem 3

To introduce the term β I considered the relationship between the distribution and defined a term m' that
represents the margin of the Type II error and is defined as:

$$m' = z_\beta \sqrt{\frac{p_a(1-p_a)}{n}}$$

Consider figure 1 on the following page (not necessarily accurate to scale).

The difference in the proportion is equal to the difference of these two margins Therefore:

$$p_a - p_0 = m - m'$$

$$p_a - p_a = z_{(1-\frac{\alpha}{2})} \cdot \sqrt{\frac{p_0(1-p_0)}{n}} - z_\beta \cdot \sqrt{\frac{p_a(1-p_a)}{n}}$$

$$p_a - p_a = \frac{z_{(1-\frac{\alpha}{2})} \cdot \sqrt{p_0(1-p_0)} - z_\beta \cdot \sqrt{p_a(1-p_a)}}{\sqrt{n}}$$

$$(p_a - p_a)^2 = \frac{z_{(1-\frac{\alpha}{2})}^2 \cdot p_0(1-p_0) - z_\beta^2 \cdot p_a(1-p_a)}{n}$$

$$n = \frac{z_{(1-\frac{\alpha}{2})}^2 \cdot p_0(1-p_0) - z_\beta^2 \cdot p_a(1-p_a)}{(p_a - p_a)^2}$$

$$n = \left(\frac{z_{(1-\frac{\alpha}{2})} \cdot \sqrt{p_0(1-p_0)} - z_\beta \cdot \sqrt{p_a(1-p_a)}}{p_a - p_a} \right)^2$$
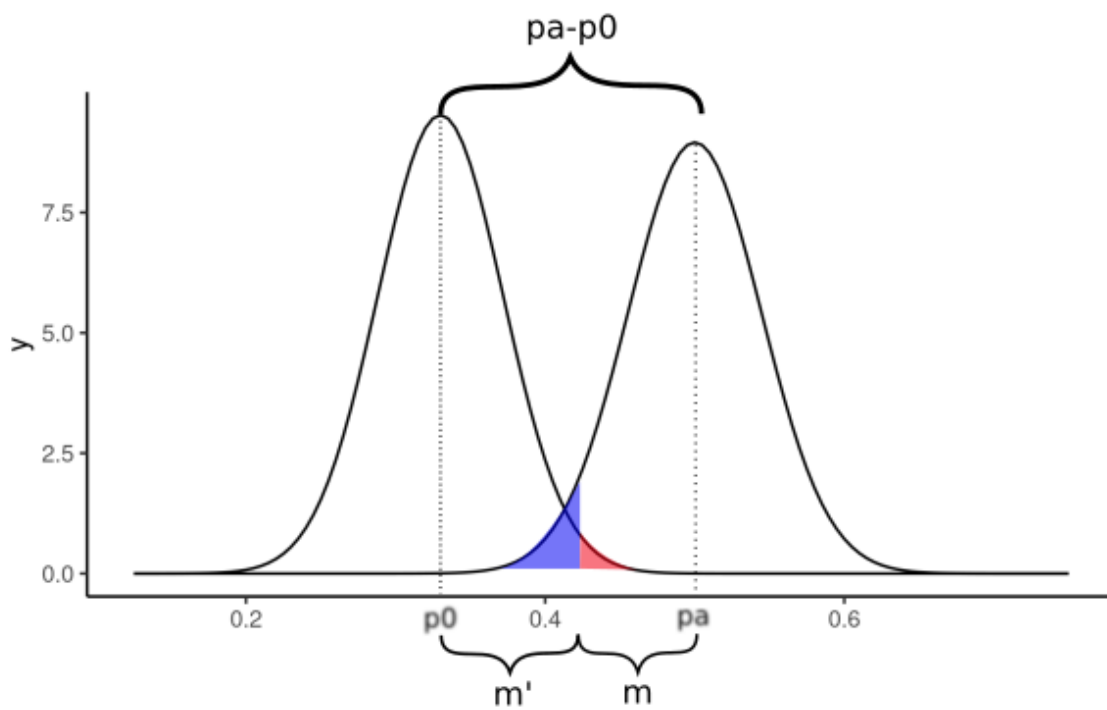
Figure 1: Relationship between the distribution of H_a and the distribution of H_0.