Homework 8

Gabriel Lahman

3/22/2019

Question 1

```
rbinorm = function(mu, var, p) {
  mu1 = mu[1]
  mu2 = mu[2]
  var1 = var[1]
  var2 = var[2]

  beta = (p * sqrt(var2)) / sqrt(var1)
  alpha = mu2 - beta * mu1
  var21 = var2 * (1 - p^2)

  x1 = rnorm(1, mean=mu1, sqrt(var1))
  x2 = alpha + (beta * x1) + (sqrt(var21) * rnorm(1))

  return (c(x1, x2))
}
```

Question 2

The paramaters needed are: $\mu_1, \mu_2, \mu_3, \sigma_1^2, \sigma_2^2, \sigma_3^2, p_{2,3}, p_{1,3}$. Using these parameters, the regression formulations are as follows:

$Z_3 \sim N(\mu_3, \sigma_3^2)$

$Z_2|Z_3 \sim N(\alpha + \beta Z_1, \sigma_{2|3}^2)$ w/ $\beta = \frac{p_{2,3}\sigma_2}{\sigma_3}$, $\alpha = \mu_2 - \beta\mu_3$, and $\sigma_{2|3}^2 = \sigma_2^2(1 - p_{2,3}^2)$

$Z_1|Z_2, Z_3 = Z_1|Z_3 \sim N(\gamma + \delta Z_3, \sigma_{1|3}^2)$ w $\delta = \frac{p_{1,3}\sigma_1}{\sigma 3}$, $\gamma = \mu_1 - \delta\mu_3$, and $\sigma_{1|3}^2 = \sigma_1^2(1 - p_{1,3}^2)$.

Since we know that $Z_1 \perp Z_2|Z_3$, $p_{1,2}$ is not needed if we formulate the variables starting with Z_3.

```
rtrinorm = function(mu, var, p23, p13) {
  sd = sqrt(var)

  z3 = rnorm(mu[3], sd[3])

  beta = (p23 * sd[2]) / sd[3]
  alpha = mu[2] - beta * mu[3]
  var23 = var[2] * (1 - p23^2)
  z2 = alpha + (beta * z3) + (sqrt(var23) * rnorm(1))

  delta = (p13 * sd[1]) / sd[3]
  gamma = mu[1] - delta * mu[3]
  var13 = var[1] * (1 - p13^2)
  z1 = gamma + (delta * z3) + (sqrt(var13) * rnorm(1))

  return (c(z1, z2, z3))
}
```

Question 3

The total number of parameters required for the model when it is assumed that $A \perp B|C$ is five: $p_A, p_B, p_C, cor_{b,c}, cor_{a,c}$. Using these five parameters, the distribution can be defined as following:

$C \sim Bern(p_C)$

$B|C \sim Bern(p_B|C)$ where $p_B|C = p_B + cor_{b,c} * (\sigma_B)\frac{C-p_C}{\sigma_C}$. Since these are bernouli, the standard deviation is $p * (1-p)$.

$A|B, C = A|C \sim Bern(p_A|C)$ where $p_A|C = p_A + cor_{a,c} * (\sigma_A)\frac{C-p_C}{\sigma_C}$.

```
tribern_MLE = function(data) {
  n = dim(data)[[1]]
  pA = sum(data[,1]) / n
  pB = sum(data[,2]) / n
  pC = sum(data[,3]) / n

  corMat = matrix(nrow = 3, ncol = 3)
  corMat[1,2] = cor(data[,1], data[,2])
  corMat[2,1] = corMat[1,2]
  corMat[1,3] = cor(data[,1], data[,3])
  corMat[3,1] = corMat[1,3]
  corMat[2,3] = cor(data[,2], data[,3])
  corMat[3,2] = corMat[2,3]

  out = list(p = list(pA, pB, pC),
             cor = corMat)

  return (out)
}

rtribern = function(p, cor_bc, cor_ac) {
  pa = p[1]
  pb = p[2]
  pc = p[3]

  sdC = sqrt(pc * (1 - pc))
  sdB = sqrt(pb * (1 - pb))
  sdA = sqrt(pa * (1 - pa))

  c = rbinom(1,1,pc)

  pb_c = pb + (cor_bc * (sdB / sdC) * (c - pc))
  b = rbinom(1,1,pb_c)

  pa_c = pa + (cor_ac * (sdA / sdC) * (c - pc))
  a = rbinom(1,1,pa_c)

  return(c(a,b,c))
}
```

```
p = c(.25, .5, .75)
out = matrix(nrow = 10000, ncol = 3)
for (i in 1:10000) {
  out[i,] = rtribern(p, .3, .1)
}

print(tribern_MLE(out))
```

```
## $p
## $p[[1]]
## [1] 0.2543
##
## $p[[2]]
## [1] 0.4964
##
## $p[[3]]
## [1] 0.7473
##
##
## $cor
##            [,1]       [,2]      [,3]
## [1,]         NA 0.03428813 0.1007289
## [2,] 0.03428813         NA 0.2988858
## [3,] 0.10072893 0.29888581        NA
```

Question 4

To specify this model, a total of 19 parameters would be needed. First, we need $P(Z = z)$ for $z \in (A, B, C, D, E, F)$, which is 6 different probabilities. However, I concede that if you were given 5 out of the 6, you could simply subtract the sum of those 5 from 1 to get the 6th. Then, for each department, $P(X = admit|Z = z)$ and $P(Y = Male|Z = z)$ are both needed, giving twelve more. Finally, since X and Y are not independent when $Z = A$, we would need the correlation between X and Y when $Z = A$. That brings the total parameters to 19 (6 + 12 + 1).