

범주형자료분석 과제 2 - 정희철

(1) (1 - a)

```
diagnosis <- c('schizophrenia','affective disorder','neurosis','personality
disorder','special symptoms')
drugs <- c(105,12,18,47,0)
nodrugs <- c(8,2,19,52,13)
patients = data.frame(drugs, nodrugs)
pat.res = chisq.test(patients)

## Warning in chisq.test(patients): Chi-squared approximation may be
incorrect

pat.res$residuals

##           drugs   nodrugs
## [1,]  3.5316125 -4.914109
## [2,]  0.9110435 -1.267684
## [3,] -1.2953872  1.802484
## [4,] -2.2627648  3.148554
## [5,] -2.9278770  4.074034

st.res = pat.res$stdres
pchisq(sum(st.res^2), (5-1)*(2-1), lower.tail = F )

## [1] 8.224131e-51

gtest = GTest(patients)
```

(1 - b)

```
diagnosis <- c('schizophrenia','affective disorder')
drugs <- c(105,12)
nodrugs <- c(8,2)
patients.g1 = data.frame(drugs, nodrugs)
pat.res.g1 = chisq.test(patients.g1)

## Warning in chisq.test(patients.g1): Chi-squared approximation may be
incorrect

st.res.g1 = pat.res.g1$stdres
pchisq(sum(st.res.g1^2), (2-1)*(2-1), lower.tail = F )

## [1] 0.05894513

gtest.g1 = GTest(patients.g1)

diagnosis <- c('neurosis','personality disorder')
drugs <- c(18,47)
```

```

nodrugs <- c(19,52)
patients.g2 = data.frame(drugs, nodrugs)
pat.res.g2 = chisq.test(patients.g2)
st.res.g2 = pat.res.g2$stdres
pchisq(sum(st.res.g2^2), (2-1)*(2-1), lower.tail = F )

## [1] 0.8072857

gtest.g2 = GTest(patients.g2)

diagnosis <- c('schizophrenia + affective disorder', 'neurosis + personality
disorder', 'special symptoms')
drugs <- c(117,65,0)
nodrugs <- c(10,71,13)
patients.g3 = data.frame(drugs, nodrugs)
pat.res.g3 = chisq.test(patients.g3)

## Warning in chisq.test(patients.g3): Chi-squared approximation may be
incorrect

st.res.g3 = pat.res.g3$stdres
pchisq(sum(st.res.g3^2), (3-1)*(2-1), lower.tail = F )

## [1] 1.835906e-60

gtest.g3 = GTest(patients.g3)

gtest$statistic == gtest.g1$statistic + gtest.g2$statistic +
gtest.g3$statistic # G2가 동일한 것을 확인

##      G
## TRUE

```

2

```

smoking <- c('both parents smoke', 'one parent smokes', 'neither parent
smokes')
st.yes <- c(400,416,188)
st.no <- c(1380,1823,1168)
total <- st.yes+st.no
smoke <- data.frame(smoking, st.yes, st.no, total)
smoke

##           smoking st.yes st.no total
## 1 both parents smoke   400 1380 1780
## 2  one parent smokes   416 1823 2239
## 3 neither parent smokes  188 1168 1356

smoke$cig <- c(2, 1, 0)

fit2 <- glm(st.yes/total ~ cig, family=binomial (link=logit),

```

```

weights=total,data=smoke)
summary(fit2)

##
## Call:
## glm(formula = st.yes/total ~ cig, family = binomial(link = logit),
##      data = smoke, weights = total)
##
## Deviance Residuals:
##      1      2      3
## -0.2929  0.5658 -0.4034
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.79502    0.06576 -27.299  < 2e-16 ***
## cig          0.28663    0.04704   6.093 1.11e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 38.36582  on 2  degrees of freedom
## Residual deviance:  0.56865  on 1  degrees of freedom
## AIC: 26.733
##
## Number of Fisher Scoring iterations: 3

```