1) a)

(1) Diagnosis	(J) 1 Drugs	2 No Drugs	
1 Schizophrenia	105	8	113
2 Affective disorder	12	2	14
3 Neurosis	18	19	37
4 Personality disorder	47	52	99
5 Special symptoms	0	13	13
	182	94	276

```r
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)
pat.res$residuals

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

=> The p-value obtained based on the above codes of χ^2-test is nearly 0, which indicates that diagnoses and drug presciption are not independent.

b)

	Drugs	No Drugs
Schizophrenia	105	8
Affective Disorder	12	2

	Drugs	No Drugs
Neurosis	18	19
Personality Disorder	47	52

	Drugs	No Drugs
Schizophrenia + Affective Disorder	117	10
Neurosis + Personality Disorder	65	71
Special symptoms	0	13

R code 파일 첨부

i) The p-value is roughly 0.06, which is not significant to conclude that diagnoses and drug prescription are not independent.

ii) The p-value is roughly 0.8, which is not significant to conclude that diagnoses and drug prescription are not independent.

iii) The p-value is nearly 0, which is significant to conclude that diagnoses and drug prescription are not independent.

2)

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

fit2 <- glm(st.yes/total ~ cig, family=binomial (link=logit), weights=total,data=smoke)
summary(fit2)
```

```
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
```

Assumed model : $\log\left(\frac{\hat{\pi}}{1-\hat{\pi}}\right) = \alpha + \beta X = -1.8 + 0.29X$

=> The residual deviance is 0.56865, which is less than the degrees of freedom, 1. This indicates that the assumed model does successfully describe the data.

0.29 log-odds increases as one unit of X increases, which means for each number of smoking parent, the log-odds of their child smoking increase by 0.29.

3) a) $f(y) = \frac{e^{-\lambda}\lambda^{y}}{y!}$, $f(\lambda \mid \mu, k) = \frac{1}{\Gamma(k)\left(\frac{\mu}{k}\right)^{k}} \lambda^{k-1} e^{-\frac{k}{\mu}\lambda}$

$$P(y; k, \mu) = \int_0^\infty \frac{e^{-\lambda}\lambda^{y}}{y!} \cdot \frac{1}{\Gamma(k)\left(\frac{\mu}{k}\right)^{k}} \lambda^{k-1} e^{-\frac{k}{\mu}\lambda} \, d\lambda$$

$$= \frac{1}{y!\,\Gamma(k)\left(\frac{\mu}{k}\right)^{k}} \int_0^\infty \lambda^{y+k-1} e^{-\lambda\left(1+\frac{k}{\mu}\right)} \, d\lambda$$

$$= \frac{1}{y!\,\Gamma(k)\left(\frac{\mu}{k}\right)^{k}} \left(\frac{\mu}{\mu+k}\right)^{y+k} \Gamma(y+k)$$

$$= \frac{\Gamma(y+k)}{y!\,\Gamma(k)} \left(\frac{k}{\mu+k}\right)^{k} \left(\frac{\mu}{\mu+k}\right)^{y}$$

$$= \frac{\Gamma(y+k)}{\Gamma(k)\Gamma(y+1)} \left(\frac{k}{\mu+k}\right)^{k} \left(1 - \frac{k}{\mu+k}\right)^{y}$$

b) $Y \sim NB\left(k, \frac{k}{\mu+k}\right)$

$$\Rightarrow E(Y) = \frac{r(1-p)}{p} = \frac{(k)\left(1 - \frac{k}{\mu+k}\right)}{\left(\frac{k}{\mu+k}\right)} \quad , \; r = k \; , \; p = \frac{k}{\mu+r}$$

$$= \cancel{k}\left(\frac{\mu}{\cancel{\mu+k}}\right)\left(\frac{\cancel{\mu+k}}{\cancel{k}}\right)$$

$$= \mu$$

$$\Rightarrow Var(Y) = \frac{r(1-p)}{p^2} = k\left(\frac{\mu}{\mu+k}\right)\left(\frac{\mu+k}{k}\right)^2$$

$$= \frac{1}{k}\mu(\mu+k)$$

$$= \mu + \frac{1}{k}\mu^2$$