

# Multilevel Modeling Homework Reference Answers

## Exercise 2.8.2

```
setwd("c:/WashU/09Fall/CAS430/Exercise/01")

girl.pro <- read.table("girls.txt", header=TRUE)

n <- 3903
m <- mean(girl.pro$proportion)
o.sd <- sd(girl.pro$proportion)
e.sd <- sqrt(m*(1-m)/n)
e.var <- m*(1-m)/n
ci <- c(qchisq(p=0.025, df=23), qchisq(p=0.975, df=23))
sqrt(ci*e.var/23) }
```

## Exercise 2.8.4

```
G <- 1000                                # 1000 simulations
n <- 20                                   # 20 sample size

out <- c()
for (i in 1:G){
  u <- runif(n, min=0, max=1)
  draw <- sum(u)
  out[i] <- draw
}

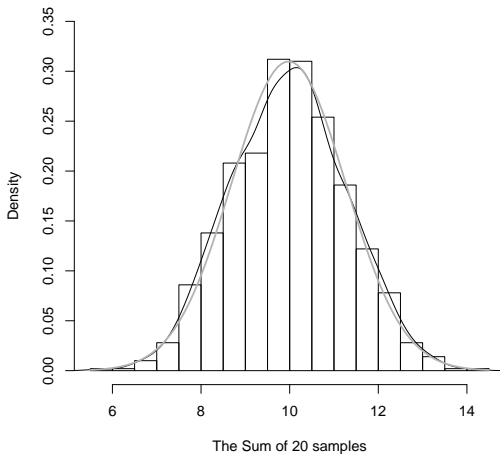
hist(out, probability=TRUE, breaks=20, ylim=c(0, 0.35),
main="1000 Draws from Uniform(0,1)",
xlab="The Sum of 20 samples")

lines(density(out))    ## Add the density line of data

curve(dnorm(x, mean=mean(out), sd=sd(out)),
add=T, col="grey70", lwd=2)
```

# Histogram

1000 Draws from Uniform(0,1)



## Exercise 3.9.4

```
library(foreign)
iq.data <- read.dta("child.iq.dta")

library(arm)
iq.m1 <- lm(ppvt ~ momage, data=iq.data)
display(iq.m1)
iq.m2 <- lm(ppvt ~ momage + educ_cat, data=iq.data)
display(iq.m2)

par(mfrow=c(1,2))
plot(ppvt ~ momage, data=iq.data, xlim=c(16, 32), ylim=c(20, 160),
     xlab="Momther's age at child's birth", ylab="Child's score", type="n")
colors <- ifelse(iq.data$educ_cat==1, "black",
                 ifelse(iq.data$educ_cat==2, "gray30",
                        ifelse(iq.data$educ_cat==3, "gray60", "gray90")))
points(ppvt ~ momage, data=iq.data, col=colors, pch=20)

curve(cbind(1,x) %*% coef(iq.m1), add=TRUE, lwd=2, col="blue")
```

## Exercise 3.9.4

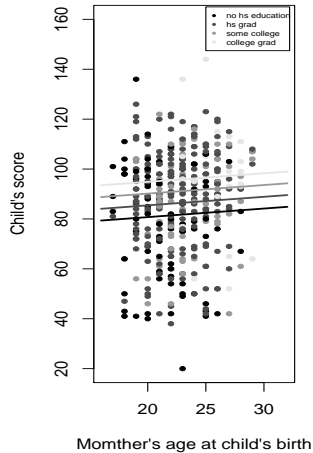
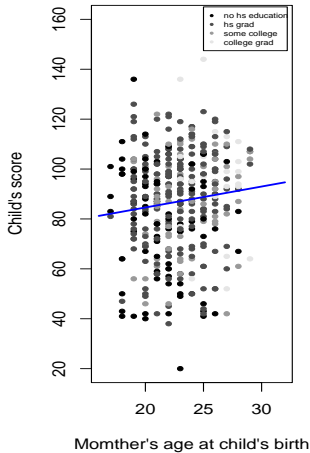
```
legend(x=25, y=165, # position of the legend
c("no hs education", "hs grad", "some college", "college grad"),
col=c("black", "gray30", "gray60", "gray90"),
pch=20, cex=0.5)
```

```
plot(ppvt ~ momage, data=iq.data, xlim=c(16, 32), ylim=c(20, 160),
xlab="Momther's age at child's birth", ylab="Child's score", type="n")
points(ppvt ~ momage, data=iq.data, col=colors, pch=20)
```

```
curve(cbind(1,x,1)%*%coef(iq.m2), add=TRUE, col="black", lwd=2)
curve(cbind(1,x,2)%*%coef(iq.m2), add=TRUE, col="gray30", lwd=2)
curve(cbind(1,x,3)%*%coef(iq.m2), add=TRUE, col="gray60", lwd=2)
curve(cbind(1,x,4)%*%coef(iq.m2), add=TRUE, col="gray90", lwd=2)
```

```
legend(x=25, y=165, # position of the legend
c("no hs education", "hs grad", "some college", "college grad"),
col=c("black", "gray30", "gray60", "gray90"),
pch=20, cex=0.5)
```

# Scatterplot



## Exercise 4.9.4 (a)

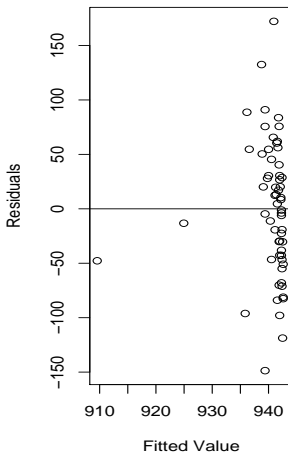
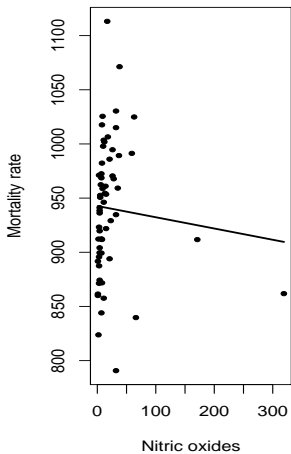
```
library(foreign)
pollution <- read.dta("pollution.dta")

library(arm)
pol.m1 <- lm(mort ~ nox, data=pollution)
display(pol.m1)

par(mfrow=c(1,2))
plot(mort ~ nox, data=pollution, xlab="Nitric oxides",
      ylab="Mortality rate", pch=20)

plot(residuals(pol.m1) ~ fitted(pol.m1),
      xlab="Fitted Value", ylab="Residuals")
abline(h=0)
```

# Scatterplot

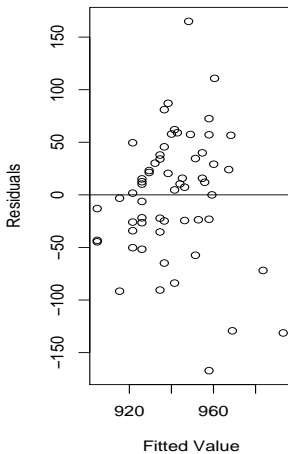
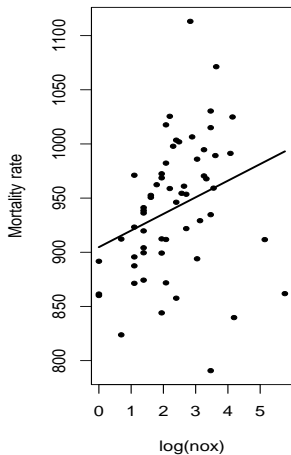


## Exercise 4.9.4 (b)

```
pol.m2 <- lm(mort ~ log(nox), data=pollution)
summary(pol.m2)
```

```
par(mfrow=c(1,2))
plot(mort ~ log(nox), ylab="Mortality rate",
     data=pollution, pch=20)
abline(pol.m2, lwd=2)
plot(residuals(pol.m2) ~ fitted(pol.m2),
     xlab="Fitted Value", ylab="Residuals")
abline(h=0)
```

# Scatterplot



## Exercise 4.9.4 (d)

```
par(mfrow=c(2,2))
plot(mort ~ hc, data=pollution)
plot(mort ~ so2, data=pollution)
plot(mort ~ log(hc), data=pollution)
plot(mort ~ sqrt(so2), data=pollution)

pol.m3 <- lm(mort ~ log(nox) + log(hc) + sqrt(so2),
data=pollution)
summary(pol.m3)

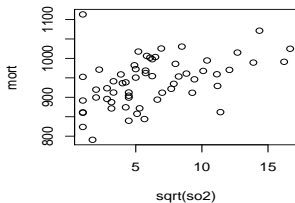
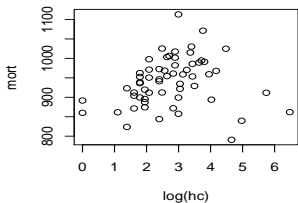
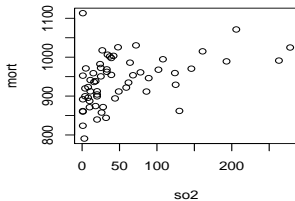
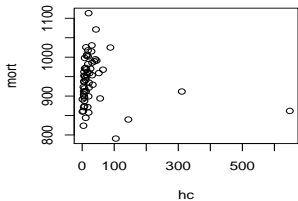
par(mfrow=c(2,2))
plot(mort ~ log(nox), data=pollution, xlab="ln(Nitric oxides)",
ylab="Mortality rate", pch=20)
curve(cbind(1, x, mean(log(pollution$hc)),
mean(sqrt(pollution$so2)))%*%coef(pol.m3), add=TRUE, lwd=2)
```

## Exercise 4.9.4 (d)

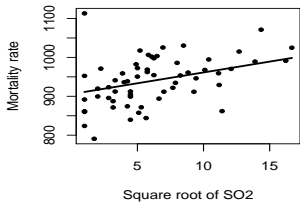
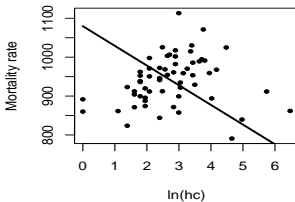
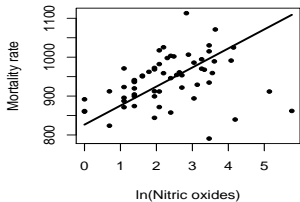
```
plot(mort ~ log(hc), data=pollution, xlab="ln(hc)",  
     ylab="Mortality rate", pch=20)  
curve(cbind(1, mean(log(pollution$nox)), x,  
            mean(sqrt(pollution$so2)))%*%coef(pol.m3), add=TRUE, lwd=2)
```

```
plot(mort ~ sqrt(so2), data=pollution,  
     xlab="Square root of SO2", ylab="Mortality rate", pch=20)  
curve(cbind(1, mean(log(pollution$nox)),  
            mean(log(pollution$hc)), x)%*%coef(pol.m3), add=TRUE, lwd=2)
```

# Scatterplot



# Scatterplot



## Exercise 4.9.4 (e)

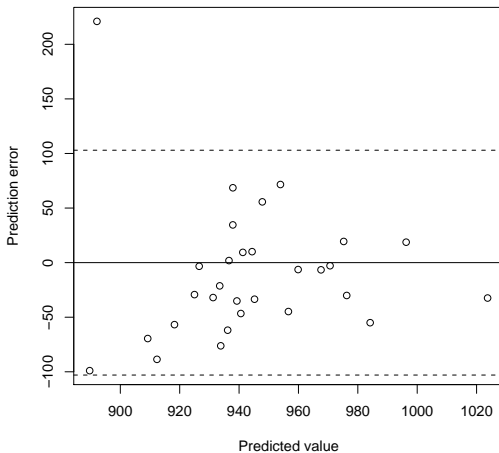
```
first.half <- pollution[1:30,]
sec.half <- pollution[31:60,]

pol.m4 <- lm(mort ~ log(nox) + log(hc) +
sqrt(so2), data=first.half)
summary(pol.m4)
m4 <- summary(pol.m4)
names(m4)

prediction <- predict(pol.m4, newdata=sec.half,
interval="prediction", level=0.95)

pre.err <- sec.half$mort - prediction[,1]
plot(pre.err ~ prediction[,1],
xlab="Predicted value", ylab="Prediction error")
abline(h=c(0, 2*m4$sigma, -2*m4$sigma), lty=c(1,2,2))
```

# Scatterplot



## Exercise 6.10.1 (a)

```
library(foreign)
hiv <- read.dta ("risky_behaviors.dta", convert.factors=T)
str(hiv)
levels(hiv$sex)
levels(hiv$bs_hiv)
names(hiv)[1] <- "male"
names(hiv)[4] <- "bs_hiv_pos"

n.act.1 <- glm(bupacts ~ couples + women_alone,
family=poisson(link="log"), data=hiv)
display(n.act.1)

dif.dev <- n.act.1$null.deviance-deviance(n.act.1)
dif.df <- n.act.1$df.null-df.residual(n.act.1)
pchisq(dif.dev, df=dif.df, lower.tail = FALSE)
```

## Exercise 6.10.1 (a)

```
n <- nrow(hiv)
k1 <- length(n.act.1$coef)
yhat1 <- predict(n.act.1, type="response")
z1 <- (hiv$bupacts-yhat1)/sqrt(yhat1)
cat("overdispersion ratio is ", sum(z1^2)/(n-k1), "\n")
cat("p-value of overdispersion test is ",
pchisq(sum(z1^2), n-k1), "\n")
```

## Exercise 6.10.1 (b)

```
n.act.2 <- glm(bupacts ~ couples + women_alone +  
bs_hiv_pos + male, family=poisson(link="log"), data=hiv)  
display(n.act.2)
```

```
dif.dev <- n.act.2$null.deviance-deviance(n.act.2)  
dif.df <- n.act.2$df.null-df.residual(n.act.2)  
pchisq(dif.dev, df=dif.df, lower.tail = FALSE)
```

```
k2 <- length(n.act.2$coef)  
yhat2 <- predict(n.act.2, type="response")  
z2 <- (hiv$bupacts-yhat2)/sqrt(yhat2)  
cat("overdispersion ratio is ", sum(z2^2)/(n-k2), "\n")  
cat("p-value of overdispersion test is ",  
pchisq(sum(z2^2), n-k2), "\n")
```

## Exercise 6.10.1 (c)

```
n.act.3 <- glm(bupacts ~ couples + women_alone +  
bs_hiv_pos + male, family=quasipoisson(link="log"), data=hiv)  
display(n.act.3)
```