

## Stat 242 Quiz – Topics Drawn from Chapters 9 and 10

### *R Output Appendix*

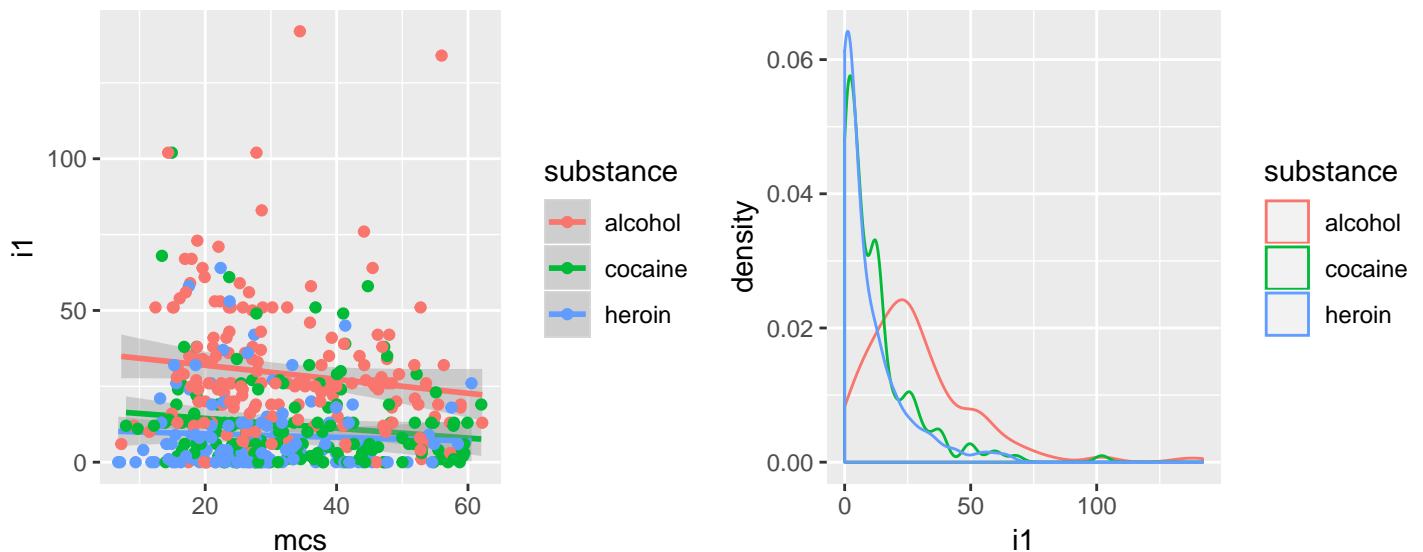
The HELP study was a clinical trial for adult inpatients recruited from a detoxification unit. Patients with no primary care physician were randomized to receive a multidisciplinary assessment and a brief motivational intervention or usual care, with the goal of linking them to primary medical care. As part of the study, a number of measurements were taken on the participants at baseline. Here we will look at the following three variables:

- **substance** is the primary substance of abuse: a variable with three levels (**alcohol**, **cocaine**, and **heroin**)
- **mcs** is the SF-36 Mental Component Score (measured at baseline, lower scores indicate worse mental health status)
- **i1** is the average number of drinks (standard units) consumed per day, in the past 30 days (measured at baseline)

We will use **i1** as the response and **substance** and **mcs** as explanatory variables.

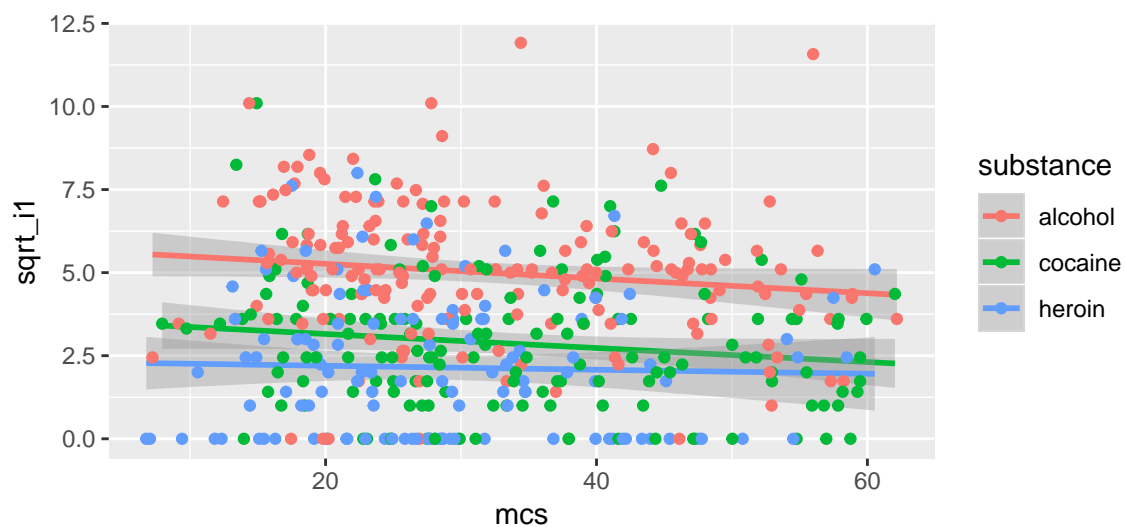
### Initial plots

```
p1 <- ggplot(data = HELPrct, mapping = aes(x = mcs, y = i1, color = substance)) +  
  geom_smooth(method = "lm") +  
  geom_point()  
  
p2 <- ggplot(data = HELPrct, mapping = aes(x = i1, color = substance)) +  
  geom_density()  
  
grid.arrange(p1, p2, ncol = 2)
```

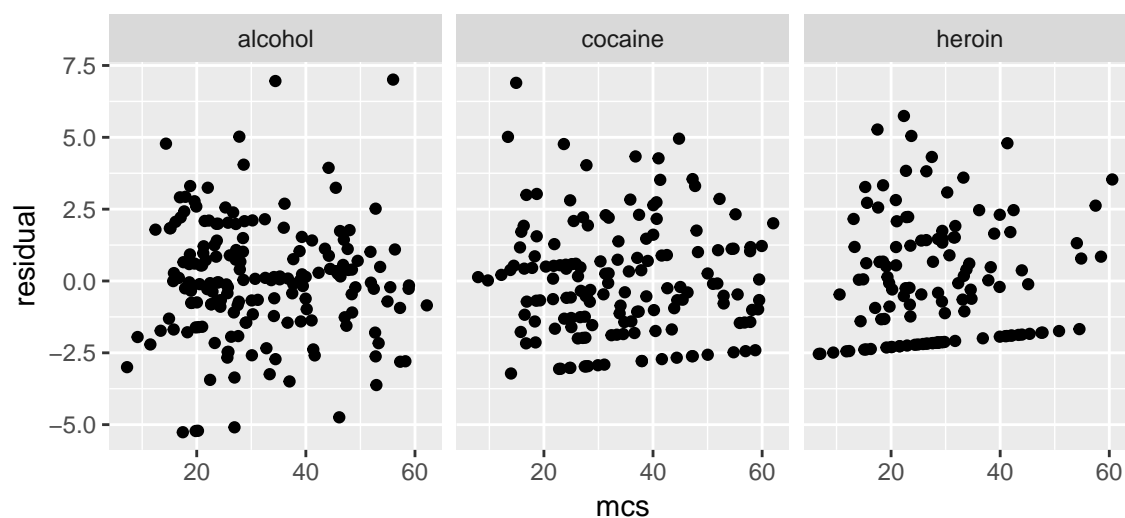


## Plots with transformed response variable

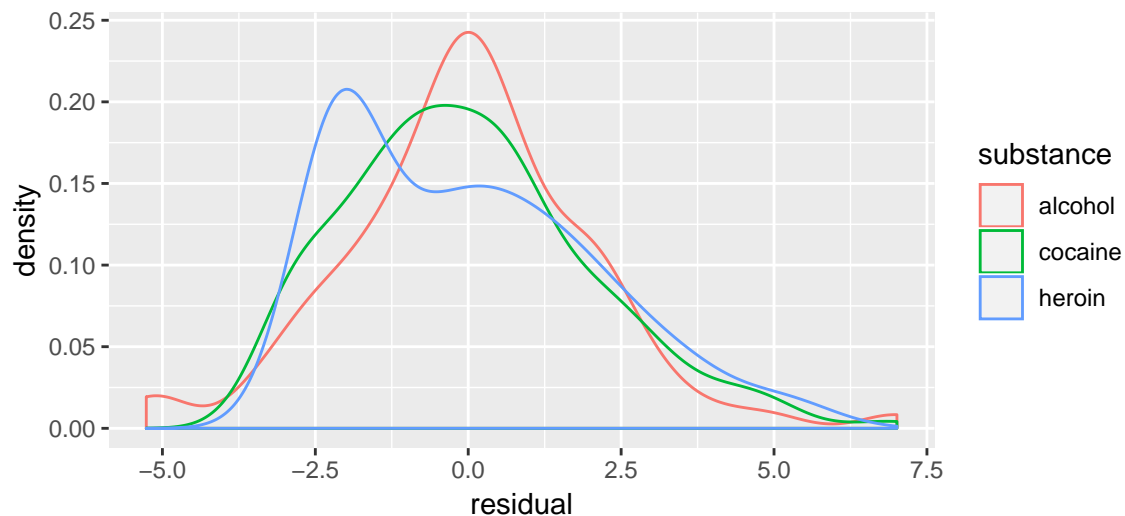
```
HELPrct <- HELPrct %>%  
  mutate(sqrt_i1 = sqrt(i1))  
  
ggplot(data = HELPrct, mapping = aes(x = mcs, y = sqrt_i1, color = substance)) +  
  geom_smooth(method = "lm") +  
  geom_point()
```



```
lm_fit <- lm(sqrt_i1 ~ substance + mcs, data = HELPrct)  
  
HELPrct <- HELPrct %>%  
  mutate(  
    residual = residuals(lm_fit)  
  )  
  
ggplot(data = HELPrct, mapping = aes(x = mcs, y = residual)) +  
  geom_point() +  
  facet_wrap(~ substance)
```



```
ggplot(data = HELPrct, mapping = aes(x = residual, color = substance)) +  
  geom_density()
```



## Summary Output – Model 1

```
lm_fit1 <- lm(sqrt_i1 ~ substance + mcs, data = HELPrct)
summary(lm_fit1)
```

```
##
## Call:
## lm(formula = sqrt_i1 ~ substance + mcs, data = HELPrct)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.2641 -1.5367 -0.0864  1.1850  7.0094
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    5.579993   0.284606  19.606  <2e-16 ***
## substancecocaine -2.106686   0.224915  -9.367  <2e-16 ***
## substanceheroin  -2.918426   0.238913 -12.215  <2e-16 ***
## mcs             -0.018102   0.007545  -2.399   0.0168 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.028 on 449 degrees of freedom
## Multiple R-squared:  0.2788, Adjusted R-squared:  0.274
## F-statistic: 57.85 on 3 and 449 DF,  p-value: < 2.2e-16
```

```
confint(lm_fit1)
```

```
##              2.5 %      97.5 %
## (Intercept)    5.02066734  6.13931833
## substancecocaine -2.54870229 -1.66467032
## substanceheroin  -3.38795270 -2.44889857
## mcs             -0.03292972 -0.00327382
```

Define the following notation for the coefficients in Model 1:

- $\beta_0^{M1}$  labeled (Intercept)
- $\beta_1^{M1}$  labeled **substancecocaine**
- $\beta_2^{M1}$  labeled **substanceheroin**
- $\beta_3^{M1}$  labeled **mcs**

## Summary Output – Model 2

```
lm_fit2 <- lm(sqrt_i1 ~ substance * mcs, data = HELPrct)
summary(lm_fit2)

##
## Call:
## lm(formula = sqrt_i1 ~ substance * mcs, data = HELPrct)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.3214 -1.4969 -0.0909  1.2096  7.1054
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    5.7067029   0.4104223   13.904 < 2e-16 ***
## substancocaine -2.1310403   0.6102941   -3.492 0.000527 ***
## substanceheroin -3.3885455   0.6376873   -5.314 1.7e-07 ***
## mcs             -0.0220794   0.0119598   -1.846 0.065533 .
## substancocaine:mcs 0.0009804   0.0171670    0.057 0.954481
## substanceheroin:mcs 0.0160741   0.0199329    0.806 0.420436
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.031 on 447 degrees of freedom
## Multiple R-squared:  0.28, Adjusted R-squared:  0.2719
## F-statistic: 34.76 on 5 and 447 DF, p-value: < 2.2e-16

confint(lm_fit2)

##              2.5 %      97.5 %
## (Intercept)    4.90010597  6.513299881
## substancocaine -3.33044232 -0.931638201
## substanceheroin -4.64178295 -2.135308101
## mcs             -0.04558386  0.001424983
## substancocaine:mcs -0.03275757  0.034718470
## substanceheroin:mcs -0.02309980  0.055247949
```

Define the following notation for the coefficients in Model 2:

- $\beta_0^{M2}$  labeled (Intercept)
- $\beta_1^{M2}$  labeled substancocaine
- $\beta_2^{M2}$  labeled substanceheroin
- $\beta_3^{M2}$  labeled mcs
- $\beta_4^{M2}$  labeled substancocaine:mcs
- $\beta_5^{M2}$  labeled substanceheroin:mcs

## ANOVA comparison of Model 1 and Model 2

```
anova(lm_fit1, lm_fit2)

## Analysis of Variance Table
##
## Model 1: sqrt_i1 ~ substance + mcs
## Model 2: sqrt_i1 ~ substance * mcs
##   Res.Df  RSS Df Sum of Sq    F Pr(>F)
## 1      449 1846.4
## 2      447 1843.3  2     3.0733 0.3726 0.6891
```