STAT 242 Midterm Exam - R output

Potentially relevant R code - Problem 1

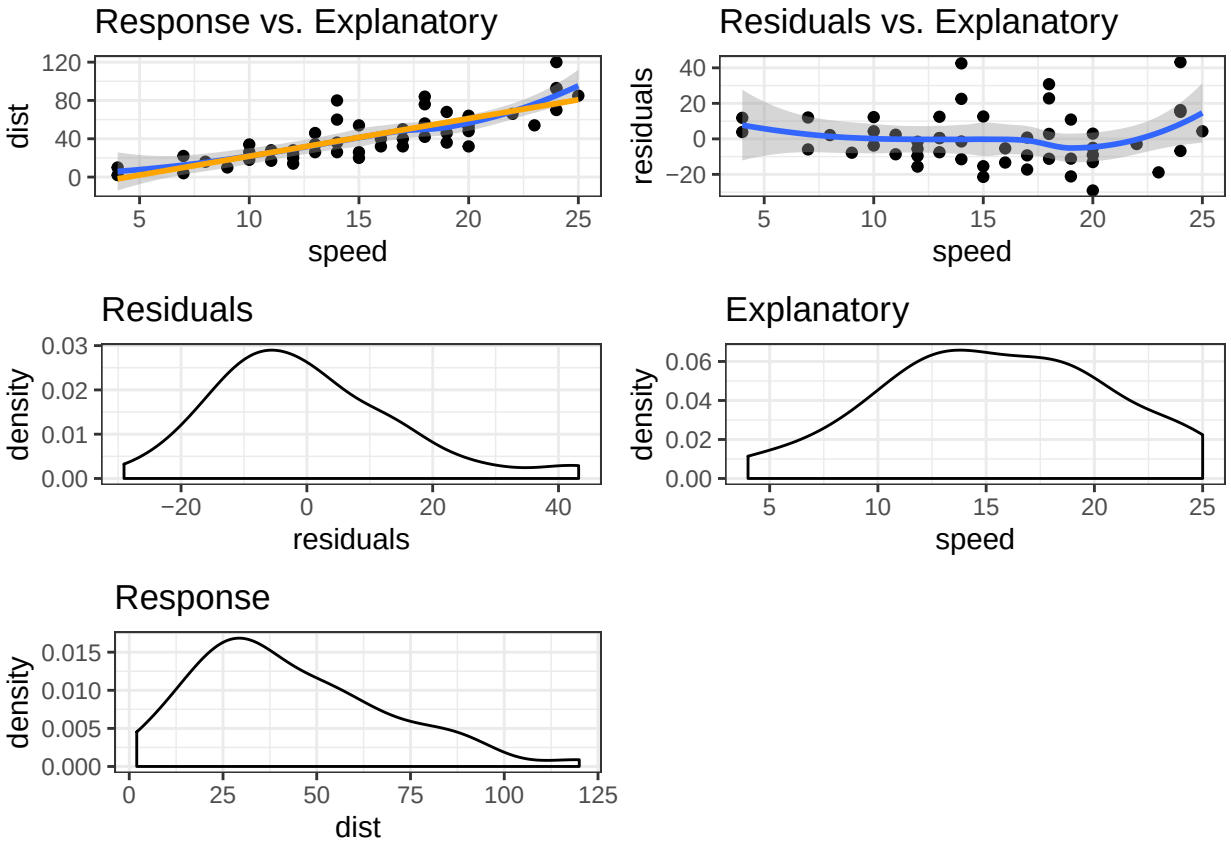
```
head(cars)
```

```
##   speed dist
## 1     4    2
## 2     4   10
## 3     7    4
## 4     7   22
## 5     8   16
## 6     9   10
```

```
fit <- lm(dist ~ speed, data = cars)
cars <- cars %>%
  mutate(
  residuals = residuals(fit),
  fitted = predict(fit)
)

p1 <- ggplot(data = cars, mapping = aes(x = speed, y = dist)) +
  geom_point() +
  geom_smooth() +
  theme_bw() +
  geom_smooth(method = "lm", color = "orange", se = FALSE) +
  ggtitle("Response vs. Explanatory")
p2 <- ggplot(data = cars, mapping = aes(x = speed, y = residuals)) +
  geom_point() +
  theme_bw() +
  geom_smooth() +
  ggtitle("Residuals vs. Explanatory")
p3 <- ggplot(data = cars, mapping = aes(x = residuals)) +
  geom_density() +
  theme_bw() +
  ggtitle("Residuals")
p4 <- ggplot(data = cars, mapping = aes(x = speed)) +
  geom_density() +
  theme_bw() +
  ggtitle("Explanatory")
p5 <- ggplot(data = cars, mapping = aes(x = dist)) +
  geom_density() +
  theme_bw() +
  ggtitle("Response")
grid.arrange(p1, p2, p3, p4, p5, ncol = 2)
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```




```
summary(fit)
```

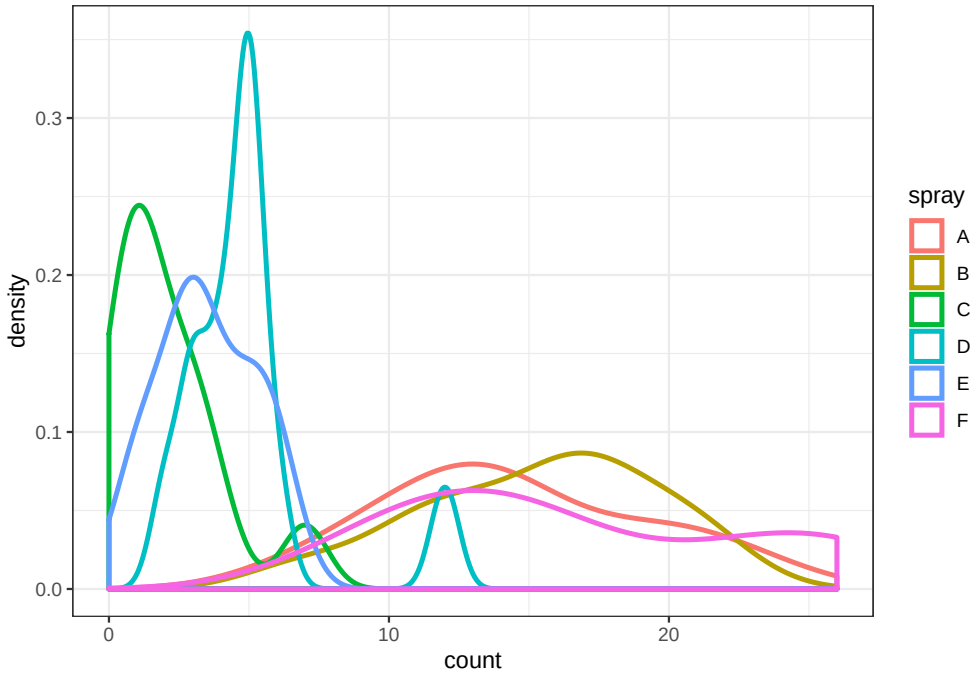
```
##
## Call:
## lm(formula = dist ~ speed, data = cars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -29.069  -9.525  -2.272   9.215  43.201
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -17.5791     6.7584  -2.601   0.0123 *
## speed         3.9324     0.4155   9.464 1.49e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.38 on 48 degrees of freedom
## Multiple R-squared:  0.6511, Adjusted R-squared:  0.6438
## F-statistic: 89.57 on 1 and 48 DF,  p-value: 1.49e-12
```

Potentially relevant R code - Problem 2

```
head(InsectSprays, 4)
```

```
##   count spray spray_grouped
## 1    10     A       grouped
## 2     7     A       grouped
## 3    20     A       grouped
## 4    14     A       grouped
```

```
ggplot(data=InsectSprays, aes(x=count, color=spray)) +
  geom_density(size=1.1) +
  theme_bw()
```




```
InsectSprays %>%
  group_by(spray) %>%
  summarise(
    sd_count = sd(count)
  )
```

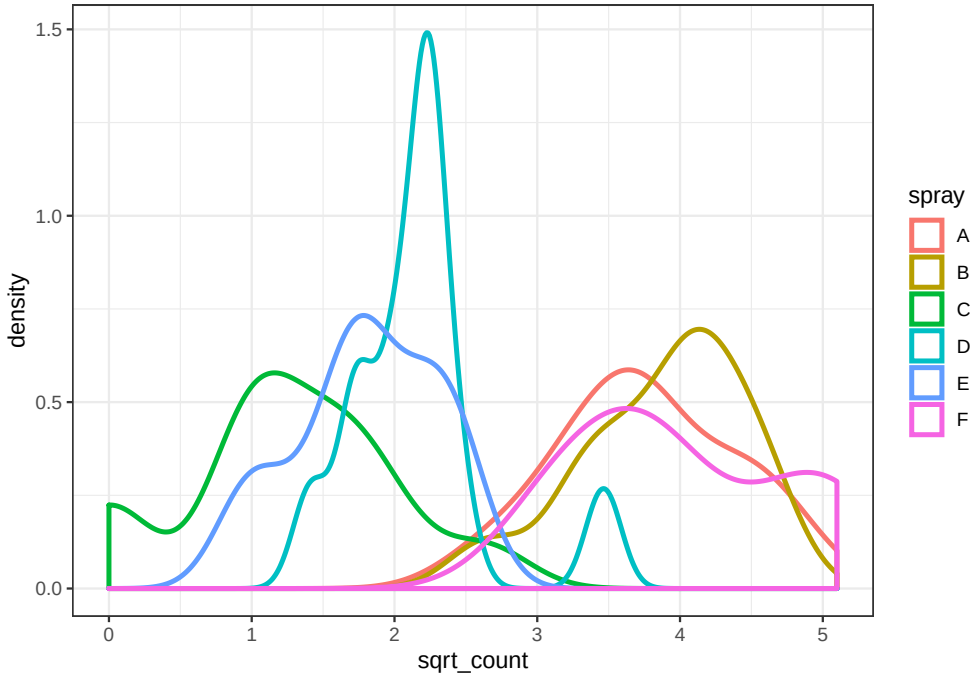
```
## # A tibble: 6 x 2
##   spray sd_count
##   <fct>    <dbl>
## 1 A         4.72
## 2 B         4.27
## 3 C         1.98
## 4 D         2.50
## 5 E         1.73
## 6 F         6.21
```

```r
model_fit1 <- lm(count ~ spray, data=InsectSprays)
summary(model_fit1)
```

```
##
## Call:
## lm(formula = count ~ spray, data = InsectSprays)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -8.333 -1.958 -0.500  1.667  9.333
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  14.5000     1.1322  12.807  < 2e-16 ***
## sprayB        0.8333     1.6011   0.520    0.604
## sprayC      -12.4167     1.6011  -7.755 7.27e-11 ***
## sprayD       -9.5833     1.6011  -5.985 9.82e-08 ***
## sprayE      -11.0000     1.6011  -6.870 2.75e-09 ***
## sprayF        2.1667     1.6011   1.353    0.181
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.922 on 66 degrees of freedom
## Multiple R-squared:  0.7244, Adjusted R-squared:  0.7036
## F-statistic:  34.7 on 5 and 66 DF,  p-value: < 2.2e-16
```

```r
InsectSprays <- InsectSprays %>%
  mutate(
    sqrt_count = sqrt(count)
  )

ggplot(data=InsectSprays, aes(x=sqrt_count, color=spray)) +
  geom_density(size=1.1) +
  theme_bw()
```




```r
InsectSprays %>%
  group_by(spray) %>%
  summarise(
    sd_sqrt_count = sd(sqrt_count)
  )
```

```
## # A tibble: 6 x 2
##   spray sd_sqrt_count
##   <fct>         <dbl>
## 1 A             0.624
## 2 B             0.577
## 3 C             0.763
## 4 D             0.503
## 5 E             0.496
## 6 F             0.751
```

```r
model_fit2 <- lm(sqrt_count ~ spray, data=InsectSprays)
summary(model_fit2)
```

```
##
```

```
## Call:
## lm(formula = sqrt_count ~ spray, data = InsectSprays)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.24486 -0.39970 -0.01902  0.42661  1.40089
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.7607     0.1814  20.733  < 2e-16 ***
## sprayB        0.1160     0.2565   0.452    0.653
## sprayC       -2.5158     0.2565  -9.807 1.64e-14 ***
## sprayD       -1.5963     0.2565  -6.223 3.80e-08 ***
## sprayE       -1.9512     0.2565  -7.606 1.34e-10 ***
## sprayF        0.2579     0.2565   1.006    0.318
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6283 on 66 degrees of freedom
## Multiple R-squared:  0.7724, Adjusted R-squared:  0.7552
## F-statistic:  44.8 on 5 and 66 DF,  p-value: < 2.2e-16
```

```
model_fit3 <- lm(count ~ spray_grouped, data=InsectSprays)
summary(model_fit3)
```

```
##
## Call:
## lm(formula = count ~ spray_grouped, data = InsectSprays)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -8.500 -2.083 -0.500  1.500 10.500
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)              2.083      1.131   1.842   0.0698 .
## spray_groupedD           2.833      1.600   1.771   0.0810 .
## spray_groupedE           1.417      1.600   0.886   0.3789
## spray_groupedgrouped    13.417      1.306  10.273 1.79e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.918 on 68 degrees of freedom
## Multiple R-squared:  0.7167, Adjusted R-squared:  0.7042
## F-statistic: 57.33 on 3 and 68 DF,  p-value: < 2.2e-16
```

```
model_fit4 <- lm(sqrt_count ~ spray_grouped, data=InsectSprays)
summary(model_fit4)
```

```
##
## Call:
## lm(formula = sqrt_count ~ spray_grouped, data = InsectSprays)
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.24486 -0.30863 -0.04487  0.42661  1.40089
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)            1.2449     0.1801   6.913 2.03e-09 ***
## spray_groupedD         0.9195     0.2547   3.611 0.000579 ***
## spray_groupedE         0.5646     0.2547   2.217 0.029961 *
## spray_groupedgrouped   2.6405     0.2079  12.699  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6238 on 68 degrees of freedom
## Multiple R-squared:  0.7689, Adjusted R-squared:  0.7587
## F-statistic: 75.42 on 3 and 68 DF,  p-value: < 2.2e-16
```

```
anova(model_fit1)
```

```
## Analysis of Variance Table
##
## Response: count
##           Df Sum Sq Mean Sq F value    Pr(>F)
## spray      5 2668.8  533.77  34.702 < 2.2e-16 ***
## Residuals 66 1015.2   15.38
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(model_fit2)
```

```
## Analysis of Variance Table
##
## Response: sqrt_count
##           Df Sum Sq Mean Sq F value    Pr(>F)
## spray      5 88.438 17.6876  44.799 < 2.2e-16 ***
## Residuals 66 26.058  0.3948
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(model_fit3)
```

```
## Analysis of Variance Table
##
## Response: count
##               Df Sum Sq Mean Sq F value    Pr(>F)
## spray_grouped  3 2640.2  880.06  57.331 < 2.2e-16 ***
## Residuals     68 1043.8   15.35
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(model_fit4)
```

```
## Analysis of Variance Table
##
## Response: sqrt_count
##               Df Sum Sq Mean Sq F value    Pr(>F)
## spray_grouped  3 88.037 29.3458   75.42 < 2.2e-16 ***
## Residuals     68 26.459  0.3891
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(model_fit3, model_fit1)
```

```
## Analysis of Variance Table
##
## Model 1: count ~ spray_grouped
## Model 2: count ~ spray
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     68 1043.8
## 2     66 1015.2  2    28.667 0.9319 0.3989
```

```r
anova(model_fit4, model_fit2)
```

```
## Analysis of Variance Table
##
## Model 1: sqrt_count ~ spray_grouped
## Model 2: sqrt_count ~ spray
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     68 26.459
## 2     66 26.058  2   0.40055 0.5073 0.6045
```