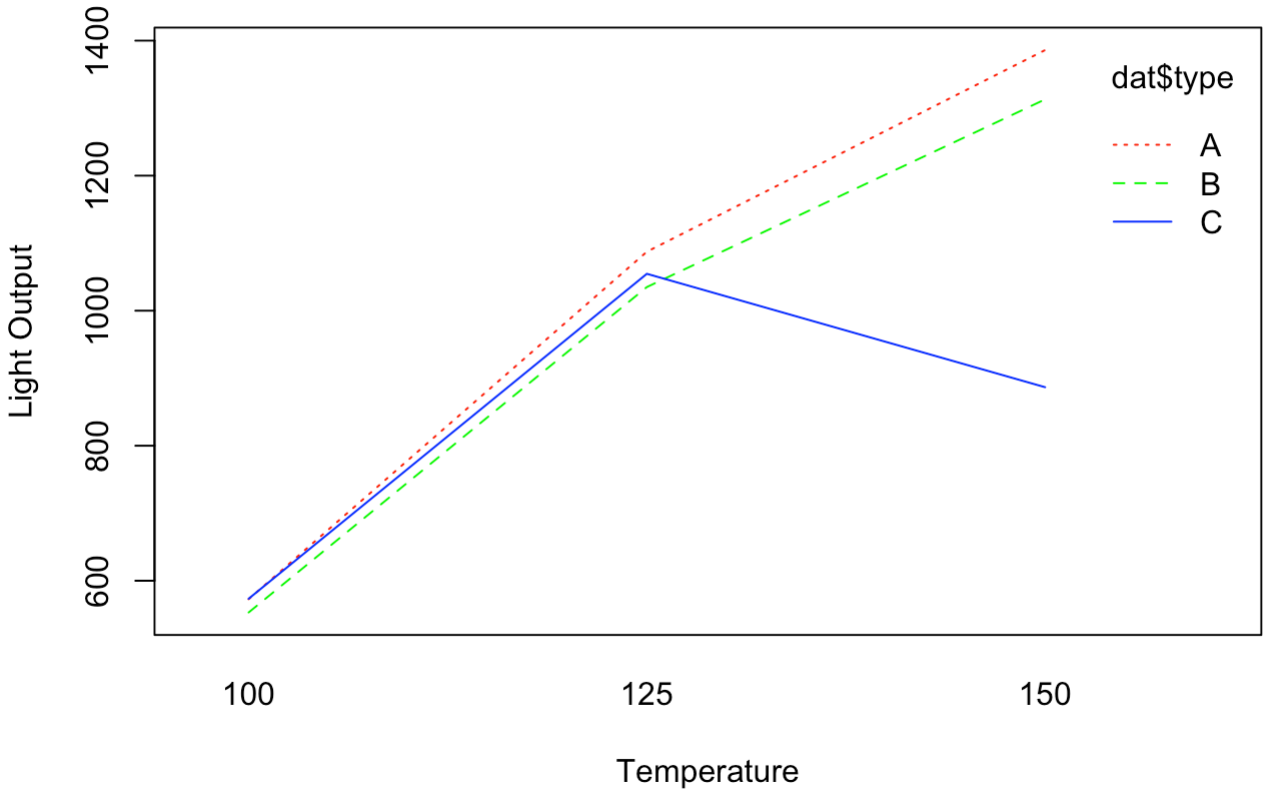
hw4

2024-10-19

```
dat = data.frame(type=c(rep("A",9), rep("B",9), rep("C",9)),
     temp=rep(c(100,125,150), 9),
     y=c(580,1090,1392,568,1087,1380,570,1085,1386,550,1070,1328,530,1035,1312,
       579,1000,1299,546,1045,867,575,1053,904,599,1066,889))
```

```
# Interaction plot of type and temperature
interaction.plot(dat$temp, dat$type, dat$y,
                 xlab = "Temperature",
                 ylab = "Light Output",
                 legend = TRUE,
                 col = c("red", "green", "blue"))
```




```
# Fit a linear model with main effects and interaction term
model <- lm(y ~ type * factor(temp), data = dat)

# Summary of the model
summary(model)
```

```
##
## Call:
## lm(formula = y ~ type * factor(temp), data = dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -35.000  -5.333  -0.333   6.667  35.000
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)             572.6667    11.0381  51.881  < 2e-16 ***
## typeB                   -19.6667    15.6102  -1.260   0.2238
## typeC                     0.6667    15.6102   0.043   0.9664
## factor(temp)125         514.6667    15.6102  32.970  < 2e-16 ***
## factor(temp)150         813.3333    15.6102  52.103  < 2e-16 ***
## typeB:factor(temp)125   -32.6667    22.0762  -1.480   0.1562
## typeC:factor(temp)125   -33.3333    22.0762  -1.510   0.1484
## typeB:factor(temp)150   -53.3333    22.0762  -2.416   0.0265 *
## typeC:factor(temp)150 -500.0000    22.0762 -22.649 1.11e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 19.12 on 18 degrees of freedom
## Multiple R-squared:  0.9973, Adjusted R-squared:  0.9961
## F-statistic: 824.8 on 8 and 18 DF,  p-value: < 2.2e-16
```

c.

H0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = 0$;

H1: at least one $\beta_j \neq 0$, for j = 1, 2, 3, 4, 5, 6, 7, 8

The overall model is significant. From summary data, we find the f-statistic is 824.8 with the p value being < 2.2e-16, which is less than 0.05. Therefore, our decision is to reject H0. It suggests that the overall model is statistically significant, indicating that at least one predictor (type, temperature, or their interaction) has a significant effect on the light output of the oscilloscope tube.

```
drop1(model, test='F')
```

```
## Single term deletions
##
## Model:
## y ~ type * factor(temp)
##                   Df Sum of Sq    RSS    AIC F value    Pr(>F)
## <none>                          6579 166.39
## type:factor(temp)  4    290552 297131 261.26  198.73 1.254e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

d.

H0: $\beta_5 = \beta_6 = \beta_7 = \beta_8 = 0$;

H1: at least one $\beta_j \neq 0$, for j = 5, 6, 7, 8

The interaction is significant. From drop1, we find the f-statistic is 198.73, with p value being 1.254e-14 < 0.05, so our decision is to reject H0 and conclude that there is an interaction.

e.

The results indicate that both the main effects of temperature and glass type, as well as their interaction, significantly affect light output. Type A glass shows the most consistent and favorable performance, exhibiting steady increases in light output as temperature rises from 100°C to 150°C, reaching the highest output levels among all types. Type C glass behaves markedly differently from the others - while it initially follows the upward trend until 125°C, it shows a dramatic decrease in light output at 150°C (as evidenced by the significant -500 unit interaction effect), suggesting that Type C glass may be unsuitable for high-temperature operations. Based on these findings, Type A glass appears to be the optimal choice for oscilloscope tubes, particularly when operating at higher temperatures.

```
# Fit a model treating temp as numeric
model_numeric <- lm(y ~ type * temp, data = dat)

# Summary of the new model
summary(model_numeric)
```

```
##
## Call:
## lm(formula = y ~ type * temp, data = dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -135.56  -52.50  -35.00   68.83  227.78
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1018.000    230.814  -4.410 0.000243 ***
## typeB          85.000    326.420   0.260 0.797091
## typeC        1072.889    326.420   3.287 0.003516 **
## temp           16.267      1.822   8.926 1.36e-08 ***
## typeB:temp     -1.067      2.577  -0.414 0.683157
## typeC:temp    -10.000      2.577  -3.880 0.000865 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 111.6 on 21 degrees of freedom
## Multiple R-squared:  0.8919, Adjusted R-squared:  0.8661
## F-statistic: 34.64 on 5 and 21 DF,  p-value: 1.865e-09
```

f.

It's not appropriate to treat temperature as a numerical variable. Normally, we used factor for variables that should be treated as categorical variables. From observing the dataframe, the temperature variable belongs to categorical variable instead of numeric(continuous) variable, since it only has 3 levels of 100, 125, 150. Additionally, since the model's RSE are much larger (RSE = 111.6 vs 19.12) and the adjusted R-squared are lower (0.8661 vs 0.9961), we should treat temperature as a factor instead of a numerical variable.