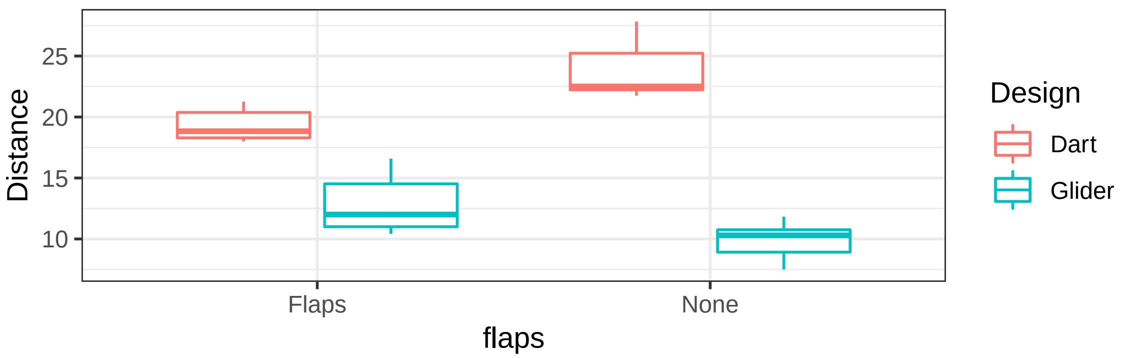
2 Way ANOVA (Highlights from Sleuth3 Chapter 13)

Paper airplanes

A motivated paper airplane thrower measured

- The `Distance` travelled (response)
- The `Design` (dart or glider) and whether or not `flaps` were put on the wings (Flaps or None)

```
ggplot(data = planes, mapping = aes(x = flaps, y = Distance, color = Design)) +
  geom_boxplot() +
  theme_bw()
```



1. Fit a model for the mean distance, using flaps and design as explanatory variables. Allow for interactions between flaps and design in your model fit. Print out a summary of the linear model fit

```
lm_fit <- lm(Distance ~ flaps * Design, data = planes)
summary(lm_fit)
```

```
##
## Call:
## lm(formula = Distance ~ flaps * Design, data = planes)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.4488 -1.5294 -0.8331  1.2741  4.0488
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)             19.3025     0.6787  28.442  < 2e-16 ***
## flapsNone                4.4787     0.9598   4.667 6.90e-05 ***
## DesignGlider            -6.4387     0.9598  -6.709 2.78e-07 ***
## flapsNone:DesignGlider  -7.3937     1.3573  -5.447 8.18e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.92 on 28 degrees of freedom
## Multiple R-squared:  0.9007, Adjusted R-squared:  0.8901
## F-statistic: 84.68 on 3 and 28 DF,  p-value: 3.727e-14
```

2. Write down a single equation for the estimated mean distance function based on the flaps and design variables.

$\hat{\mu}(Distance|flaps, Design) = 19.30 + 4.48 flapsNone - 6.44 DesignGlider - 7.39 flapsNone \times DesignGlider$

3. State how you could use the summary output from part 1 to calculate the estimated mean distances for all four combinations of settings for flaps and design.

$$
\begin{aligned}
\hat{\mu}(Distance|flaps = "Flaps", Design = "Dart") &= 19.30 + 4.48 \times 0 - 6.44 \times 0 - 7.39 \times 0 \times 0 \\
&= 19.30
\end{aligned}
$$

$$
\begin{aligned}
\hat{\mu}(Distance|flaps = "None", Design = "Dart") &= 19.30 + 4.48 \times 1 - 6.44 \times 0 - 7.39 \times 1 \times 0 \\
&= 19.30 + 4.48
\end{aligned}
$$

$$
\begin{aligned}
\hat{\mu}(Distance|flaps = "Flaps", Design = "Glider") &= 19.30 + 4.48 \times 0 - 6.44 \times 1 - 7.39 \times 0 \times 1 \\
&= 19.30 - 6.44
\end{aligned}
$$

$$
\begin{aligned}
\hat{\mu}(Distance|flaps = "None", Design = "Glider") &= 19.30 + 4.48 \times 1 - 6.44 \times 1 - 7.39 \times 1 \times 1 \\
&= 19.30 + 4.48 - 6.44 - 7.39
\end{aligned}
$$

REVISED NUMBER 4: Based on the model fit from part 1, find a 95% confidence interval for the difference in means between DART planes with and without flaps. Interpret your confidence interval in context.

```
confint(lm_fit)
```

```
##                                2.5 %    97.5 %
## (Intercept)              17.912347 20.692653
## flapsNone                 2.512777  6.444723
## DesignGlider             -8.404723 -4.472777
## flapsNone:DesignGlider -10.174055 -4.613445
```

We are 95% confident that in the population of planes folded by this person, the mean distance traveled by dart planes without flaps is between 2.5 nd 6.4 feet farther than the mean distance traveled by dart planes with flaps.

4. Based on the model fit from part 1, find a 95% confidence interval for the difference in means between glider planes with and without flaps. Interpret your confidence interval in context.

```
# code for the original question (not the revised question)
planes2 <- planes %>%
  mutate(
    Design = factor(Design, levels = c("Glider", "Dart"))
  )

fit2 <- lm(Distance ~ flaps * Design, data = planes2)
summary(fit2)
```

```
##
## Call:
## lm(formula = Distance ~ flaps * Design, data = planes2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.4488 -1.5294 -0.8331  1.2741  4.0488
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)            12.8637     0.6787  18.955  < 2e-16 ***
## flapsNone              -2.9150     0.9598  -3.037  0.00512 **
## DesignDart              6.4388     0.9598   6.709 2.78e-07 ***
## flapsNone:DesignDart    7.3937     1.3573   5.447 8.18e-06 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.92 on 28 degrees of freedom
## Multiple R-squared:  0.9007, Adjusted R-squared:  0.8901
## F-statistic: 84.68 on 3 and 28 DF,  p-value: 3.727e-14
```

```r
confint(fit2)
```

```
##                           2.5 %     97.5 %
## (Intercept)           11.473597 14.2539026
## flapsNone             -4.880973 -0.9490273
## DesignDart             4.472777  8.4047227
## flapsNone:DesignDart   4.613445 10.1740553
```

We are 95% confident that in the population of planes folded by this person, the mean distance traveled by glider planes without flaps is between 4.9 nd 0.9 feet shorter than the mean distance traveled by glider planes with flaps.

5. Based on the model fit from part 1, find a 95% confidence interval for the mean distance for glider planes that don't have flaps. Interpret your confidence interval in context.

You will have to define a new data frame with variables Design and flaps that have the appropriate variable values, and call predict using that new data frame.

```r
predict_data <- data.frame(
  Design = "Glider",
  flaps = "None"
)

predict(lm_fit, newdata = predict_data, interval = "confidence", level = 0.95)
```

```
##       fit      lwr     upr
## 1 9.94875 8.558597 11.3389
```

We are 95% confident that in the population of planes folded by this person, the mean distance traveled by glider planes without flaps is between 8.6 and 11.3 feet.

6. Based on the model fit from part 1, find a 95% prediction interval for the mean distance for dart planes that don't have flaps.

Again, you'll have to define a suitable data frame with variables Design and flaps.

```r
predict_data <- data.frame(
  Design = "Dart",
  flaps = "None"
)

predict(lm_fit, newdata = predict_data, interval = "prediction", level = 0.95)
```

```
##        fit      lwr      upr
## 1 23.78125 19.61079 27.95171
```

We are 95% confident that a dart plane without flaps randomly selected from the population of planes folded by this person will travel between 19.6 and 28.0 feet.

7. Conduct a test of the claim that adding flaps to a plane will increase the mean distance flown by the same amount, regardless of the airplane design. State your hypotheses in terms of equations involving the model coefficients β_0, β_1, β_2, and β_3, and draw a conclusion in terms of the strength of evidence against the null hypothesis.

$H_0 : \beta_3 = 0$

$H_A : \beta_3 \neq 0$

From the summary output above, the p-value for this test is 8.18e-06. The data provide extremely strong evidence against the null hypothesis that adding flaps to a plane will increase the mean distance flown by the same amount, regardless of the airplane design.

8. Conduct a test of the claim that whether or not you add flaps to a plane has no impact on the mean distance flown. State your hypotheses in terms of equations involving the model coefficients β_0, β_1, β_2, and β_3, and draw a conclusion in terms of the strength of evidence against the null hypothesis.

$H_0 : \beta_1 = \beta_3 = 0$

H_A : At least one of β_1 and β_3 is not equal to 0

We will fit a reduced model that does not include the flaps variable at all, and compare it to our full model using an F test via the `anova` command.

```
fit_reduced <- lm(Distance ~ Design, data = planes)
anova(fit_reduced, lm_fit)
```

```
## Analysis of Variance Table
##
## Model 1: Distance ~ Design
## Model 2: Distance ~ flaps * Design
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     30 217.39
## 2     28 103.17  2    114.23 15.501 2.939e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-value for this test is 2.9×10^{-5}. The data provide extremely strong evidence against the null hypothesis that whether or not a plane has flaps has no association with the mean distance traveled.

9. Conduct a test of the claim that all paper airplanes fly the same distance on average, across both designs and whether or not flaps are used. State your hypotheses in terms of equations involving the model coefficients β_0, β_1, β_2, and β_3, and draw a conclusion in terms of the strength of evidence against the null hypothesis.

$H_0 : \beta_1 = \beta_2 = \beta_3 = 0$

H_A : At least one of β_1, β_2, and β_3 is not equal to 0.

This is an F test; the p-value is reported in the final line of the summary output for the original linear model fit. The p-value is 3.727e-14. The data provide extremely strong evidence against the null hypothesis that the mean distance travelled is the same regardless of which design is used or whether or not flaps are included.