Case Study 1

Jonathan Sneh, Ishani Tarafdar, Georges Durand, Raul Higareda

2023-03-28

```
grades <- read.csv("grades.csv", header=TRUE)
grades.mlr <- lm(exam2 ~ . ,data=grades)
summary(grades.mlr)
```

```
##
## Call:
## lm(formula = exam2 ~ ., data = grades)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -57.414  -6.793   0.850   7.831  27.124
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   35.26655    5.46283   6.456 5.03e-10 ***
## exam1          0.34756    0.05853   5.939 8.88e-09 ***
## project        0.01576    0.03971   0.397    0.692
## cs             0.02337    0.05811   0.402    0.688
## hw             0.40359    0.06267   6.440 5.49e-10 ***
## participation -0.02741    0.05210  -0.526    0.599
## semester      -7.51155    0.66993 -11.212  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.06 on 268 degrees of freedom
## Multiple R-squared:  0.5405, Adjusted R-squared:  0.5302
## F-statistic: 52.54 on 6 and 268 DF,  p-value: < 2.2e-16
```

We want to make a model with 95% confidence (i.e. α)

Based on the summary we may want to look into dropping project, cs, and participation from the dataset since the p-values in the summary output for project, cs, and participation are greater than 0.05.

Our null and alternative hypothesis are as follows: Hi

$$\begin{cases} H_0, & \beta_{project} = \beta_{cs} = \beta_{participation} \\ H_A, & \text{Either } \beta_{project}, \beta_{cs}, \text{ or } \beta_{participation} \text{ is not equal to zero} \end{cases}$$

```
grades.reducedmlr = lm(exam2~exam1 + semester + hw, data=grades)
summary(grades.reducedmlr)
```

```
##
## Call:
## lm(formula = exam2 ~ exam1 + semester + hw, data = grades)
```

```
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -57.029  -7.070   0.881   7.921  28.145
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 36.02239    4.73371   7.610 4.56e-13 ***
## exam1        0.34759    0.05717   6.080 4.08e-09 ***
## semester    -7.54042    0.66197 -11.391  < 2e-16 ***
## hw           0.40989    0.04775   8.584 7.28e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.01 on 271 degrees of freedom
## Multiple R-squared:  0.5393, Adjusted R-squared:  0.5342
## F-statistic: 105.7 on 3 and 271 DF,  p-value: < 2.2e-16
```

```
anova(grades.mlr, grades.reducedmlr)
```

```
## Analysis of Variance Table
##
## Model 1: exam2 ~ exam1 + project + cs + hw + participation + semester
## Model 2: exam2 ~ exam1 + semester + hw
##   Res.Df   RSS Df Sum of Sq      F Pr(>F)
## 1    268 38963
## 2    271 39065 -3   -102.64 0.2353 0.8717
```

Since the p-value is greater than an alpha level of 0.05 in the anova ouput, we fail to reject the null hypothesis with 95% level of confidence.