Recap of Single Linear Regression

Dr. Soliman

Make sure you answer each question, and use your lecture notes and any other resources to ensure these answers are complete.

1. Download the data from here, and change the pathname in the chunk above to ensure the data is correctly read in. Run only that "setup" code chuck first.

2. As in lecture, run a linear regression model with `baseball` as the data, `winpercent21` as the dependent variable, and `payroll` as the independent variable. Write the linear regression in the code chunk below, then execute it to see the output.

```
# Write the code for the linear regression below these comments, but within the code chunk.
# It should look something like lm(y ~ x, data)
lm(winpercent21 ~ payroll, baseball)
```

```
##
## Call:
## lm(formula = winpercent21 ~ payroll, data = baseball)
##
## Coefficients:
## (Intercept)      payroll
##    40.12678      0.06569
```

3. Interpret the coefficients from the regression, remembering that `winpercent21` is a percent and that `payroll` is in millions of dollars, and to be as formal as we have been in the lecture notes. Hint: that means that a one unit increase in x here is a one million dollar increase in payroll.

Type your response below this.

The intercept is 40.12678, which captures that if payroll was 0, so the team didn't spend any money, then the average win percentage is about 40.

The slope coefficient, 0.06569, can be interpreted as each additional million dollars spent on payroll, on average, is associated with an increase in the win percentage of 0.06569.

4. As in lecture, use the summary function, `summary()` to obtain the R^2. You will need to use the regression from question 2, and first assign it to an object named `payrollwin_reg` before using the summary function. Do that in the code chunk below.

```
# Write the code below these comments, but within the code chunk.
# It should look something like summary(payrollwin_reg), but you need to assign first...
payrollwin_reg <- lm(winpercent21 ~ payroll, baseball)
summary(payrollwin_reg)
```

```
##
## Call:
## lm(formula = winpercent21 ~ payroll, data = baseball)
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.6739  -5.9535  -0.2274   5.1244  15.2013
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 40.12678    3.77563  10.628 2.47e-11 ***
## payroll      0.06569    0.02316   2.836  0.00839 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.006 on 28 degrees of freedom
## Multiple R-squared:  0.2232, Adjusted R-squared:  0.1954
## F-statistic: 8.044 on 1 and 28 DF,  p-value: 0.008389
```

5. Interpret the R^2 as in the lecture notes below.

Type your response below this. When you are done, click Knit to generate the PDF and bring it to class on Thursday.

The R^2 value is 0.2232, indicating that 22% of the variation in win percentage is explained by variation in payroll.