Quiz 6 — 3/7/2024

Instructions

This take-home quiz is due on **Thursday, March 7 at 23:59**.

You may use your own course materials, as well as any materials directly linked from the course website. **No collaboration allowed.**

Type your answers **directly in this Jupyter notebook**, and submit this notebook (just the `ipynb` file) using the submission form on the course website.

Feedback. Overall, most of you had the right idea with this quiz.

When interpreting coefficients and their confidence intervals, be very careful about specifying under which circumstances your interpretation holds. See Lessons 15 and 16 for details.

Also, be careful with the units!

Problem 1

The data frame `MLB2007Standings` from the `Stat2Data` library contains data on many variables for Major League Baseball teams from the 2007 regular season. The winning percentages are in the variable $WinPct$ and scoring variables include $Runs$ (scored by a team for the season) an ERA (essentially the average runs against a team per game).

Run the cell below to load and preview this data.

In [1]:
```
library(Stat2Data)
data(MLB2007Standings)
head(MLB2007Standings)
```

A data.frame: 6 × 21

	Team	League	Wins	Losses	WinPct	BattingAvg	Runs	Hits	HR	Doubles	⋯	RBI	SB	OBP	SLG
	<fct>	<fct>	<int>	<int>	<dbl>	<dbl>	<int>	<int>	<int>	<int>	⋯	<int>	<int>	<dbl>	<dbl>
1	Arizona Diamondbacks	NL	90	72	0.556	0.250	712	1350	171	286	⋯	687	109	0.321	0.413
2	Atlanta Braves	NL	84	78	0.519	0.275	810	1562	176	328	⋯	781	64	0.339	0.435
3	Baltimore Orioles	AL	69	93	0.426	0.272	756	1529	142	306	⋯	718	144	0.333	0.412
4	Boston Red Sox	AL	96	66	0.593	0.279	867	1561	166	352	⋯	829	96	0.362	0.444
5	Chicago Cubs	NL	85	77	0.525	0.271	752	1530	151	340	⋯	711	86	0.333	0.422
6	Chicago White Sox	AL	72	90	0.444	0.246	693	1341	190	249	⋯	667	78	0.318	0.404

a.

Fit a multiple linear regression model to predict $WinPct$ based on $Runs$ and ERA.

Provide **only** the summary output for this part.

In [ ]:

b.

In the 2007 season, the Boston Red Sox scored 867 runs and had an ERA of 3.87. Use this information and your fitted model from part a to predict the Red Sox's winning percentage. Use the code cell below as a calculator to show your work.

In [ ]:

c.

It turns out that the Boston Red Sox actually had a winning percentage of 0.593 for the 2007 season. Find the residual. Use the code cell below as a calculator to show your work.

In [ ]:

d.

Interpret the estimated coefficient of ERA.

Write your answer here. Double-click to edit.

e.

What is the test statistic and associated p-value for the t-test on the coefficient of ERA?

Write your answer here. Double-click to edit.

f.

Based on your answer to part e, what do you conclude about the relationship between $WinPct$ and ERA? Assume a significance level of 0.05.

Write your answer here. Double-click to edit.

g.

Use the `confint()` function in R to find a 90% confidence interval for $Runs$ and ERA.

In [ ]:

h.

Interpret the 90% confidence interval for $Runs$ that you found in part g. Your answer should discuss the relationship between $Runs$ and $WinPct$.

Write your answer here. Double-click to edit.

i.

What is the test statistic and associated p-value for the ANOVA test for your model?

Write your answer here. Double-click to edit.

j.

Based on your answer to part i, what do you conclude about the effectiveness of your model as a whole? Assume a significance level of 0.05.

Write your answer here. Double-click to edit.

k.

What is R^2_{adj} for your model?

Write your answer here. Double-click to edit.

Grading rubric

Problem	Weight
1a	1.0
1b	0.5
1c	0.5
1d	0.5
1e	0.5
1f	0.5
1g	0.5
1h	0.5
1i	0.5
1j	0.5
1k	0.5
Max Score	**60**