Case Study 2 (F25)

Due Dates
Lectures 100/200: Monday, October 27th, at 8:00PM Ann Arbor Time via Gradescope
Lectures 300/400: Tuesday, October 28th, at 8:00PM Ann Arbor Time via Gradescope

Late Submission Policy
We offer a 1-hour grace period without any penalty. Submissions uploaded after the 1-hour grace period but within 24 hours of the deadlines listed above will be accepted but are subject to a 10 percent late penalty. No submissions will be accepted thereafter.

Paired Submission Policy
Only *active* lab approach students may submit with a partner. Your partner must be in the same lab section - and you both must be in attendance for the corresponding lab session. To receive the 5 points of extra credit, you must properly tag your partner to the submission via Gradescope. See the submission instructions at the bottom of the lab document (or ask your lab instructor) for additional help.

Purpose
The overall purpose of this case study is to build your skills in conducting an applied statistical analysis in a real-world context. Specifically, this assignment is designed to (1) improve your statistical writing skills; (2) assess your proficiency using testing and interval estimation methods to evaluate associations between variables; and (3) challenge you to make data-based arguments in a setting where decisions based on statistical analyses have substantial real-world implications.

Context - Athletic Evaluations of NFL Scouting[1]
The National Football League (NFL) Scouting Combine takes place annually for the purpose of allowing NFL teams to evaluate prospects. Potential NFL athletes undergo a series of physical and mental tests that coaches and talent scouts reference when making draft decisions. The battery of physical tests are meant to capture an athlete's endurance, strength, and agility; the tests include the 40-yard dash, bench press, vertical jump, 20-yard shuttle run, broad jump, and 3 cone drill.

Recently, some have suggested that the current battery of tests should be modified, given that research has shown an athlete's performance in the Scouting Combine is only weakly predictive of later success in professional sporting leagues and because athletes who tend to perform well in physical tests that emphasize strength (e.g., the vertical jump) may not perform as well on tests that emphasize endurance or agility (e.g., the bench press or the 3 cone drill, respectively).

Task
This case study focuses on three of the physical tests included in the NFL Scouting Combine: the *bench press* endurance test (***bp***), where players bench press a bar bell weighing 225 lbs for as many repetitions as they can before failure; the *vertical jump* strength test (***vj***), where players jump as high as possible from a standing position; and the 3 cone drill *sprint* agility test (***sp***), where players run as fast as possible around three cones placed 5 yds apart that are placed in the shape on an "L." These three tests are designed to capture an athlete's endurance, strength, and agility, respectively. Some of the research linked above suggests that strength and agility are *inversely* related among NFL athletes. Your task is to analyze the available data and decide whether it supports or contradicts this claim.

Primary Research Question: How does an athlete's strength (measured by ***vj***) relate to their agility (measured by ***sp***)?

[1] This case study is inspired by results initially published in Russell, B. & Hogan, P. (2018). Analyzing dependence matrices to investigate relationships between national football league combine event performances. *Journal of Quantitative Analysis in Sports*, *14*(4), 201-212. https://doi.org/10.1515/jqas-2017-0086

Data

The *combine* data set contains variables recorded from a random sample of n = 200 prospective NFL athletes who participated in the NFL Combine between 1999 and 2016. The variables include:

- ***bp.endurance***: An indicator of whether the athlete fell above or below the Combine average number of completed reps before failure (Above Average Endurance or Below Average Endurance)
- ***vj.strength***: An indicator of whether the athlete fell above or below the Combine average vertical jump cleared (Above Average Strength or Below Average Strength)
- ***sp.agility*****:** A variable recording the completion time of the 3 cone drill (measured in seconds)

Assignment

This case study is broken up into two parts: Part 1 (20 points), analyzing the data; and Part 2 (80 points), writing up your results and recommendation in a brief report.

Part 1 - Data Analysis Task: (*Completed during lab - worth 20 points, see below for rubric*)

For *all* research questions below, perform exploratory analyses of the associations between *bp, vj,* and *sprint* by creating graphical representations *and* numerical summaries. For research questions (1) and (4) specifically, conduct inferential analyses as well (i.e., conduct hypothesis tests / construct confidence intervals).

- Note: all data analyses must be performed using functions provided in the Case Study 2 Resources file

Research Questions:

1. What association, if any, exists between athlete strength and agility, as measured by ***vj*** and ***sp*****?**
2. What association, if any, exists between athlete strength and endurance, as measured by ***vj*** and ***bp*****?**
3. What association, if any, exists between athlete endurance and agility, as measured by ***bp*** and ***sp*****?**
4. Does athlete endurance (as measured by ***bp***) influence the association between strength (***vj***) and agility (***sp***)?

Part 2 - Writing Task (*Completed outside of lab - worth 80 points, see below for rubric*)

Write a report summarizing your findings from Part 1. Your report should provide a brief summary of the goals of your analysis, demonstrate a clear understanding of its results, and provide a statement that summarizes whether the data supports or contradicts the claim. Successful statistical reports should follow the structure recommended in the outline below and range between approximately 700 and 1000 words. **Submissions that are between 1100 and 1500 words will receive a 5-point penalty, and those that exceed 1500 words will receive a 10-point penalty.**

Introduction: Describe the purpose of the analysis and justify its importance. Summarize the data that have been collected. Successful summaries will be one paragraph in length and will…

a. state the observational units of the study, note the sample size, and describe all variables analyzed

Summary of Findings

Summarize the methods and findings of your graphical and numerical summaries, along with the results of any tests and confidence intervals. Highlight and interpret the key results. A successful analysis will be separated into at least three paragraphs and will…

a. Summarize your analysis of the association between remote athlete strength and agility, *without* accounting for athlete endurance (Research Question 1). Clearly define the parameter(s) you tested / estimated and provide a written conclusion in the context of the research question they address. In this section, provide written *definitions* of your test statistic, *p*-value, estimated effect size, and confidence interval.
b. Then, summarize your analyses corresponding to Research Questions 2 and 3.
c. Finally, summarize your analyses corresponding to Research Question 4 about the association between athlete strength and agility *after* accounting for differences in athlete endurance. Include a passage that compares these results to those presented in (a). You do *not* need to provide written definitions of your test statistic, *p*-value, estimated effect size, or confidence interval here.

Discussion and Recommendation

Provide a discussion of whether the analyses you conducted suggest athlete strength and agility are inversely related. A successful discussion will be one to two paragraphs in length and will…

a. Include a statement that clearly illustrates a full understanding of the role confounding factors play in studies of variable associations.
b. A clear explanation of *whether* and *why* the variable that approximates an athlete's endurance confounds the association between their strength and agility.
c. A final statement regarding how athlete strength appears to relate to agility *after* the athlete's endurance has been accounted for.

Criteria for Success:

Rubric for Data Analysis Task (20 points)

		Research question (i) **(5 points)**	Research question (ii) **(5 points)**	Research question (iii) **(4 points)**	Research question (iv) **(6 points)**
Statistical Analysis	**Exemplary** Correctly addresses all elements of the prompt, demonstrating expert understanding and interpretations of required concepts / skills.	5	5	4	6
	Proficient Correctly addresses all elements of the prompt, demonstrating understanding and interpretations of required concepts / skills with just one minor error.	4	4	3	5
	Emerging Correctly addresses most elements of the prompt, demonstrating solid understanding and interpretations of required concepts / skills. May include a few minor computational/conceptual mistakes.	3	3	2	4
	Needs Improvements Demonstrates limited understanding of statistical concepts / skills. Large errors or conceptual mistakes are present.	2	2	1	2
	Missing Completely missing any attempted analysis.	0	0	0	0

Criteria for Success:

Rubric for the Written Report (80 points)

		Overall Report **(5 points)**		
Presentation	**Proficient** Report flows in a logical order, containing minimal grammatical or spelling errors. Writing is clear, creative, and informative.	5		
	Emerging Report includes passages that are vague, unclear, or include substantial grammatical or spelling errors.	3		
	Needs Improvements Report displays little logical order, suffering from many grammatical or spelling errors.	1		
		Introduction **(5 points)**	Analysis Overview **(40 points)**	Discussion **(30 points)**
Statistical Reasoning	**Exemplary** Correctly addresses all elements of the prompt, demonstrating expert understanding and interpretations of required concepts / skills.	5	40	30
	Proficient Correctly addresses all elements of the prompt, demonstrating solid understanding and interpretations of required concepts / skills. May include one or two minor conceptual mistakes.	4	36	27
	Fair Correctly addresses most elements of the prompt, demonstrating an understanding and interpretations of required concepts / skills. Includes several mistakes, one of which may be major.	3	32	22
	Emerging Correctly addresses some elements of the prompt, demonstrating an understanding and interpretations of required concepts / skills. May include more than two minor conceptual mistakes OR demonstrates major misunderstandings.	2	28	18
	Needs Improvements Demonstrates very limited understanding of statistical concepts / skills. Includes more than one major misunderstanding in core concepts.	1	21	12