Extra Credit Assignment Due in Lecture W 4 - 7 - 10.

- **This assignment can bump your RAW SCORE on Exam 3 by (up to) 2 points (to a maximum of 22 total raw score).**

  Data is to be collected from the play-by-play broadcast of the NCAA match between MSU and Butler or the final playoff game of the tournament (regardless of who plays).

  Your submission is limited to both sides of a standard sheet of 8.5 by 11 paper.

  Be very clear about your method and your findings.

  Strive for accuracy in data collection and calculations.

  Report any data collection errors or dropouts and what you decided to do with that.

  Cite and check a web reference to this topic.

- **What you will be doing during the game.** Keep pen in hand. Each time MSU (or whichever team you choose) gets the ball, mark + if they score. Otherwise mark -. If the clock is reset but they retain control let's agree to make that a single possession. You can stop with a CONSECUTIVE 40 possessions if you choose. It is vital that they be consecutive. If the team has fewer than 40 possessions we'll just go with whatever they have. Although the method is not generally reliable for $n < 40$, I have a way to give the P-value for $n < 40$. **Be sure to get the data right.**

- **Description of the problem.** Momentum is an important concept in sport. A team or player ***has*** momentum if they are on a run, playing well, especially if they are outplaying their opponent (in which case they have ***the*** momentum). Momentum is often associated with a string (or strings) of successes. Whether such strings constitute real evidence for momentum or are just chance clumping is the question. Thinking about it, if your team appears to have momentum will you rest a key player who could use a break from the action, or instead keep them in the game to preserve momentum? If you believe in momentum, but it is in actual fact only an illusion created by chance clumping, you may over-work your key players to no real advantage.

- **Let's take the example of consecutive possessions by MSU.** We'll use hypothetical data to show what you do. **For this bonus assignment you must use real data from the game.** Each **consecutive** possession can be scored **+** or **-** depending on whether it results in a score or not. Here is the hypothetical data:

  + + + + - - - + + + + - + + + + + - + + + + + + - - + + + + - - + + + + + + + +

  n_1 = (the number of **+**) = 30

  n_2 = (the number of **-**) = 10

  n = total number of possessions = 40

Notice the stretches of unbroken "runs" of **+** or **-** .

Denote by random variable X the number of runs. In the above example $\boxed{X = 11}$:

+ + + + - - - + + + + - + + + + + - + + + + + +
- - + + + + - - + + + + + + + +

If we really have 40 (independent) Bernoulli trials with unknown p (coin flips, with P(+) = p not known, no momentum effect, runs just accidental) then there should be rather more short "runs" than if the scoring comes in longer spurts due to some momentum effect. How to assess whether there are statistically few runs (evidence for momentum) or more runs (evidence that runs are possibly merely random clumps)?

The **conditional distribution** of X, for fixed values of n_1 and n_2, is (for large n) approximately normal with:

$$\textbf{E X (for fixed } \mathbf{n_1, n_2}\textbf{)} = \mu = 1 + \frac{2\, n_1\, n_2}{n} = 1 + \frac{2 \times 30 \times 10}{40} = 16$$

$$\textbf{Variance X (for fixed } \mathbf{n_1, n_2}\textbf{)} = \sigma^2 = \frac{(\mu - 1)(\mu - 2)}{n-1} = \frac{(16-1)(16-2)}{40-1} = 5.38462....$$

$$\textbf{Standard Deviation} = \sigma = \sqrt{5.38462} = 2.32048....$$

$$\textbf{Standard score of X} = z = \frac{x - \mu}{\sigma} = \frac{11-16}{2.32048} = -\ 2.15473$$

The probability that we would see a z-score < - 2.15473 is around 0.0155915.

Conclusion. The event of 11 runs or fewer has a rather small probability of 1.6% in 40 Bernoulli Trials having 30+ and 10-. Were this real data it would offer a cautionary counter-argument against the hypothesis that "Bernoulli trials toss off accidental runs only appearing to us to be momentum."

We have used the "Wald-Wolfowitz Runs Test." Generally, n ⩾ 40 is recommended.

- **Larger n means more power to detect departures from randomness (better chance to detect a smaller momentum effect).** Suppose that the above data held up over n = 160 consecutive possessions. That is, 160 possessions, 120 of which are **+**, with X = 44 runs (everything multiplied by 4 to extend the experience of 40 possessions).

$$\textbf{E X (for fixed } \mathbf{n_1, n_2}\textbf{)} = \mu = 1 + \frac{2\, n_1\, n_2}{n} = 1 + \frac{2 \times 120 \times 40}{160} = 61$$

$$\textbf{Variance X (for fixed } \mathbf{n_1, n_2}\textbf{)} = \sigma^2 = \frac{(\mu - 1)(\mu - 2)}{n-1} = \frac{(61-1)(61-2)}{160-1} = 22.2642....$$

$$\textbf{Standard Deviation} = \sigma = \sqrt{22.2642} = 4.71849....$$

$$\textbf{Standard score of X} = z = \frac{x - \mu}{\sigma} = \frac{44-61}{4.71849} = -\ 3.60285$$

Conclusion. The probability that we would see a z-score < - 3.60285 is around 0.000157375. So the experience of 40 possessions, if replicated over 160 possessions, would be far more unlikely to have been caused by chance and would therefore be more convincing evidence on the side of the momentum point of view.