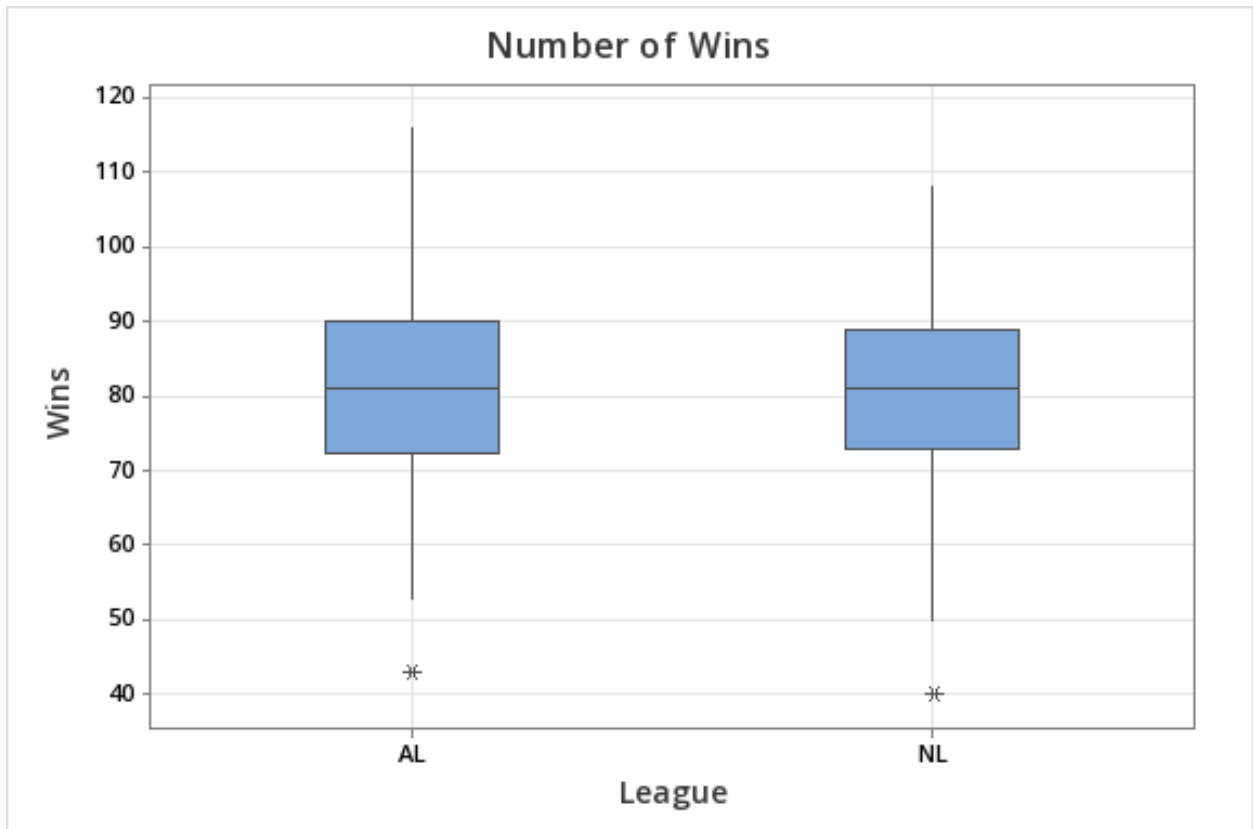


Reflection #3

Part A

1. Select the number of wins (W) and compare the distribution of W for the American League (AL) with that of the National League (NL). Use side-by-side box plots as described in section 4.3.

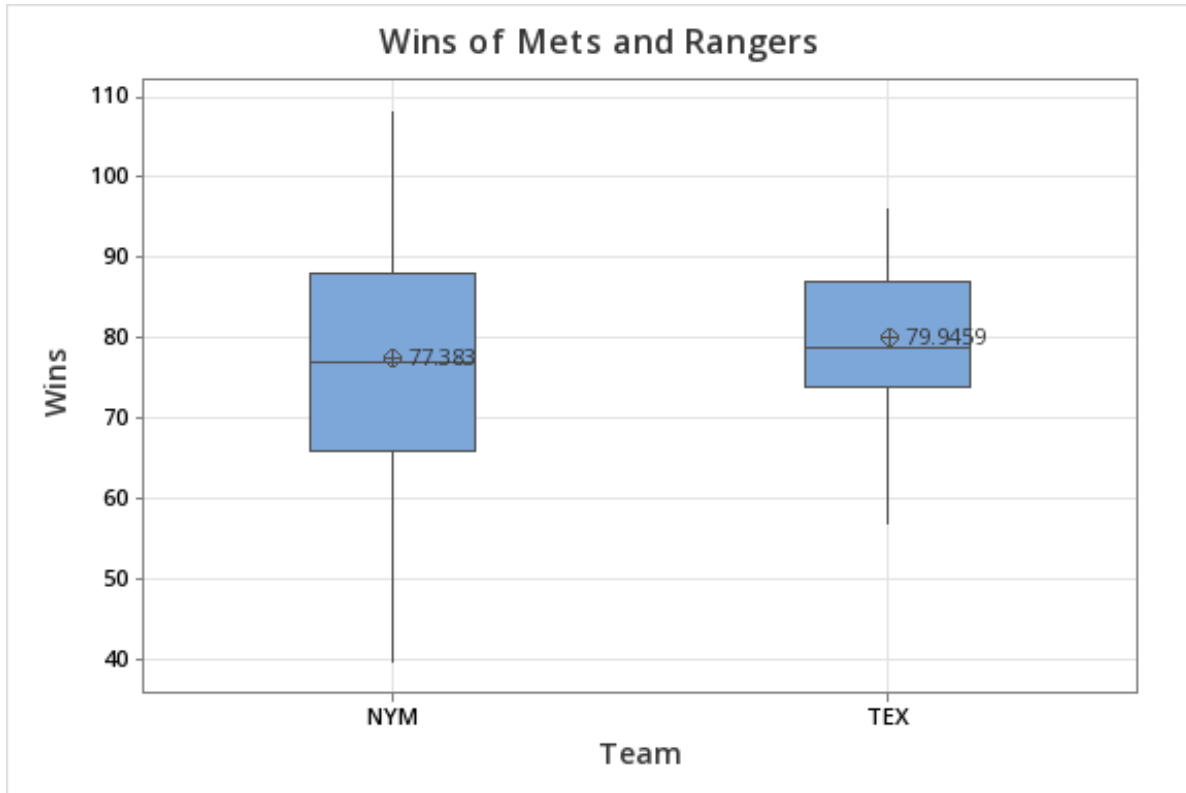


The 5 number summary for both the AL and NL are very similar. They even have outliers that are in the same general area. They share the same median, 81 wins. The AL has lower and upper quartiles at 72.25 and 92, while the NL has 73 and 89. The AL does have the team with the most wins at 162, while the NL has their maximum at 108. Most teams, no matter in the American or National League, tend to win about 72-90 games a season. Since baseball teams play 162 games a season, most teams win about 40-50% of their games.

2. Identify the outliers in both leagues (i.e., the teams that have a total number of wins far from the rest of the teams in their league).

The outlier in the American League was the 2003 Detroit Tigers. They had a total of 43 wins out of 162 games played. The National League outlier was the 1962 New York Mets. They had a total of 40 wins out of 161 games played.

3. Compare the distribution of the Number of wins, W , for NYM and TEX using a side-by-side boxplot and by investigating the numerical summaries of each. (Compare the shapes, means, medians, and the variability).



The distribution of wins between the Mets and the Rangers, unlike the distribution between the AL and NL is very different. The Mets have both extremely good years (108 wins) and extremely bad years (40 wins). Their plot is much more spread out; it contains an upper quartile of 88 and a lower quartile of 66, for an IQR of 22. On the other hand, while the Rangers' best year was just 96 wins, they are much more consistent in their performance. Their minimum was 57 wins, upper and lower quartiles of 87 and 74 respectively for an IQR of 13. The medians and means for the Mets and Rangers were similar with the Mets at 77 and 77.38 and the Rangers at 79 and 79.95, a This shows how different the variability of these two teams were.

**As an aside, something interesting is that the Rangers data doesn't include the 10 years of the team as the Washington Senators from 1962-1971 before they moved to Texas and became the Rangers. According to my partner who is versed in all things baseball statistics (he can tell you any World Series winner, just give him a year!), the Washington Senators 2.0 were not very good. They lost 100+ games in each of their first 4 years, and only actually had one winning season. I know lots more information about them now, but I will spare you all that I learned from him 😊 **

4. Discuss why the discrepancy in variability between the performance of NYM and the performance of TEX didn't cause a similar discrepancy in their respective leagues.

Even though the difference between the Rangers and Mets are obvious, when in the context of almost 500 other data points, those variabilities are not going to have as much of an effect. This is also why the Mets 40 wins is an outlier in the NL graph but not their own. In their own graph, anything under 33 would be an outlier. However, in the NL graph, anything under 49 is an outlier.

5. Based on historical data, the probability that in a given year the NYM will make the playoffs is $p = 7/47 = 0.149$. Let X be the discrete random variable that gives the total number of Playoffs made by NYM in the last 20 years, i.e., from 1993 to 2012.
 - a. Assume that the outcomes for the NYM in these years are unknown for us. Also assume that the outcome in any of the years is independent of the outcome in any other year. Under these assumptions, what would be the distribution of X ? Why?

The distribution of X under these conditions would be a Binomial Random Variable. First, there are only 2 outcomes. Either the Mets make the playoffs, *success*, or they do not, *failure*. The probabilities of winning from year to year are independent of each other as stated above. There are 20 trials, and the probability of success is the same for each trial.

- b. What is the probability that the total number of playoffs made by NYM during this 20-year period is exactly three?

For part b-part d, I used the binompdf function on my TI-84 calculator to perform the actual calculations, though the formula used is $P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$.

For $x = 3$, the probability that the Mets make the playoff in these 20 years is about **0.2428**.

- c. What is the probability that the total number of playoffs made by NYM is at most 3?

For this, I calculated the probabilities when $x = 0,1,2,3$ and added these together, getting a probability of **0.6526**

- d. What is the probability that the total number of playoffs made by NYM is at most 18?

To find this, I calculated the probabilities of $x = 19,20$ added these together, then subtracted from 1, getting a probability of $1 - (3.3518 \times 10^{-15}) \approx$ **1.000** when rounding.

- e. What is the probability that the total number of playoffs made by NYM is at least 15?

For part e, I calculated for $x = 15, 16, 17, 18, 19, 20$ and added these together, getting a probability of $2.897 \times 10^{-9} \approx 0.0000$ when rounded.

- f. What is the expected number of playoffs that NYM will make in this 20-year period?

The expected number of playoffs that the Mets will make can be found by finding the Expected Value, $\mu = E(X) = np = 20 * 0.149 = 2.98$ playoff games.

- g. Find the variance of the number of playoffs that NYM is expected to make in this 20-year period?

The variance is found by calculating $\sigma^2 = V(X) = np(1 - p) = 20(0.149)(1 - 0.149) = 2.53598$

- h. Can we use the Poisson distribution with $\lambda=2.98$ to model the number of playoffs that NYM will make? Why?

No, we could not use the Poisson distribution. While successes occur one at a time (either playoffs or not) and only happen one at a time (cannot make the playoffs twice in once year), and the events are independent, the variance and expected value are not the same, $\mu \neq \sigma^2$. In Poisson, $\mu = \lambda = \sigma^2$.

Part B

2. Given that 10% of us are left-handed and some desks are made for left-handed students. To determine the number of such desks to order and place in our classrooms, we want to compute the probability of getting exactly 3 left-handed students in a class of 15 students.

- a. This is a binomial experiment. Explain why.

This is a binomial experiment because it meets all the criteria. First, there are only 2 outcomes. A student is either left-handed, success, or not left-handed, failure. There are 10% of left-handed people, so the probability is fixed (0.1). There are 15 students in the class, or 15 trials. These trials are independent, so it must be a binomial experiment.

- b. Use the binomial probability formula to find the probability of getting 3 left-handed students in a class of 15 students, given that 10% of us are left-handed.

Again, I used the binompdf function on my TI-84 calculator to perform the actual calculations.

For $x = 3$, the probability that there will be 3 left-handed students in the class is about 0.1285.

- c. Use Minitab or Excel to generate probability distribution for each random variable for this experiment. Copy and paste your table.

Binomial with $n = 15$ and $p = 0.1$

x	$P(X = x)$
0	0.205891
1	0.343152
2	0.266896
3	0.128505
4	0.042835
5	0.010471
6	0.001939
7	0.000277
8	0.000031
9	0.000003
10	0.000000
11	0.000000
12	0.000000
13	0.000000
14	0.000000
15	0.000000

- d. Use the table you generated to find the probability of (i) exactly 3 successes and (ii) at least 3 successes.

Using the table, the probability of exactly 3 successes is **0.128505**, the same value I got for part b.

The probability of at least 3 successes is easier to find by adding up $x = 0, 1, 2$ and subtracting the sum from one. The sum of those 3 events was 0.815939. Subtracting that from 1 shows that the probability of at least 3 successes is **0.184061**.

- e. Find the mean and standard deviation for this event.

The mean can be found by finding the Expected Value. $\mu = E(X) = np = 15 * 0.1 = 1.5$

The standard deviation is found by calculating the variance and taking the square root. $\sqrt{\sigma^2} = \sqrt{V(X)} = \sqrt{np(1-p)} = \sqrt{15(0.1)(1-0.1)} = \sqrt{1.35} \approx 1.1619$

3. Create/formulate one real life problem for a discrete uniform probability distribution.

Yesterday I went out for ice cream with my partners little sister. At the ice cream shop, there were 20 different flavors (Pavarotti, Raspberry Chocolate Chip, Vanilla, Chocolate, Mint, etc.). Natalie was allowed to get one flavor. What is the probability that she picked Raspberry Chocolate Chip as her ice cream flavor?

- a. Write the general discrete uniform probability distribution function.

The probability distribution function for a discrete uniform distribution with n possible outcomes is $P(X = x) = \frac{1}{n}$

- b. Solve the problem you created and provide its solution.

Since there are 20 options for ice cream, and Natalie has an equally likely chance of picking any of the flavors, the $P(\text{Raspberry Chocolate Chip}) = \frac{1}{20} = 0.05$

- c. Find the mean and standard deviation for the problem you created.

Mean is the same as Expected Value, $\mu = \sum[x \cdot p(x)]$, where $p(x) = P(X = x)$. Since there are 20 options, $\mu = \frac{1}{20} + \frac{2}{20} + \dots + \frac{20}{20} = \frac{210}{20} = 10.5$

To solve for Standard Deviation, we use the formula, $\sigma = \sqrt{V(X)} = \sqrt{\sum[(x - \mu)^2 p(x)]}$.

$$\sigma = \sqrt{\sum_{x=1}^{20} (x - 10.5)^2 (0.05)} = \sqrt{33.25} \approx 5.766$$

4. Create/formulate one real life problem for a Poisson probability distribution.

In 2023, MNUFC scored a total of 43 goals over 34 games, or an average of about 1.265 goals per match. What is the probability that in their next match, they will score at least 4 goals?

- a. Write the general Poisson probability distribution function.

The general Poisson probability distribution function is: $P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}$

- b. Solve the problem you created and provide its solution.

First, since the rate is given as 1.265 goals per game, we know that $\lambda = 1.265$.

To find the probability of $P(X \geq 4)$, we can take $1 - P(X \leq 3)$. First, I calculated the probability for $X = 0, 1, 2, 3$ and added those together, getting $P(X \leq 3) = 0.9603$. Thus, $P(X \geq 3) = 1 - 0.9603 = .0397$

- c. Find the mean and standard deviation for the problem you created.

In Poisson distributions, $\mu = \sigma^2 = \lambda$. Since $\lambda = 1.265$, this means that the mean $\mu = 1.265$. To find the standard deviation, we need to take $\sigma = \sqrt{\lambda} = \sqrt{1.265} \approx 1.1247$.

5. Create/formulate one real life problem for a hypergeometric probability distribution.

My partner and I try to do movie night at least 4 times a month. We currently have 80 movies on our to-be-watched list, with 20 of them being comedies. During our next movie night, we randomly select 3 movies from the list. What is the probability that at least 2 of them are comedies?

- a. Write the general hypergeometric probability distribution function.

The general hypergeometric probability distribution function is: $P(X = x) = \frac{\binom{k}{x} \binom{N-k}{n-x}}{\binom{N}{n}}$,

where k =total successes, N =total size of population, n =size of sample, and x =number of successes in the sample

- b. Solve the problem you created and provide its solution.

For this problem, $N = 80$ (movies on our watch list), $k = 20$ (total comedies), $n = 3$ (number of movies selected). Because I want to find $P(X \geq 2)$, we need to find

$P(X = 2) + P(X = 3)$, so $P(X \geq 2) = P(X = 2) + P(X = 3) = \frac{\binom{20}{2} \binom{60}{1}}{\binom{80}{3}} + \frac{\binom{20}{3} \binom{60}{0}}{\binom{80}{3}} \approx .1388 + .0139 \approx .1526$

- c. Find the mean and standard deviation for the problem you created.

To find the mean, we need to take $\mu = n \left(\frac{k}{N} \right) = 3 \left(\frac{20}{80} \right) = 0.75$. To find the standard

deviation, $\sigma = \sqrt{n \left(\frac{k}{N} \right) \left(1 - \frac{k}{N} \right) \left(\frac{N-n}{N-1} \right)} = \sqrt{3 \left(\frac{20}{80} \right) \left(1 - \frac{20}{80} \right) \left(\frac{80-3}{80-1} \right)} = \sqrt{3 \left(\frac{1}{4} \right) \left(\frac{3}{4} \right) \left(\frac{77}{79} \right)} = \sqrt{\frac{693}{1264}} \approx 0.7404$