Intro/Descriptive Stats for Two Proportions

Last week, we began exploring comparing two groups as opposed to analyzing a single process (such as a dog choosing a cup) or a single population parameter (such as proportion of students with brown hair). This week, we will explore descriptive statistics – both graphical and numerical – that will help us compare two proportions. Today we will investigate diversity in professional sports, based on athleticism/fitness being an important value to our class and diversity an important value in our world.

Diversity in Professional Sports

The data below comes from The Institute for Diversity and Ethics in Sport (TIDES – tidesport.org). From the TIDES website: "TIDES serves as a comprehensive resource for issues related to gender and race in amateur, collegiate, and professional sports and the media. The Institute researches and publishes a variety of studies, including annual studies of student-athlete graduation rates and racial attitudes in sports, as well as the internationally recognized Racial and Gender Report Card, an assessment of racial and gender hiring practices in amateur, collegiate, and professional sports and the media."

The data we will explore today come from the 2021 Major League Baseball (MLB) Racial and Gender Report Card and the 2020 National Football League (NFL) Racial and Gender Report Card, which are the most recent report cards for each professional sports league. The data in the reports was provided to TIDES from the corresponding league offices and, in most cases, is complete data on everyone in the league (thus this is population data, not sample data).

In this data, we will explore the proportions of players, coaches and CEO/Presidents who are persons of color. In this data, *person of color* is a term used to describe anyone who is not considered white, including – but not limited to – Black people, Latinos, Native Americans, and people of Asian descent. The variable "person of color" is self-reported by the participants.

Now, let's talk descriptive statistics…way back in the Week 3 Packet, we introduced *frequency tables* and *bar graphs* for categorical data, including **segmented bar graphs** that allow us show comparisons graphically. We will refresh our memory on segmented bar graphs and introduce **two-way tables**, which are similar to frequency tables, but are used for comparing two categorical variables.

A **two-way table** (also called a **contingency table**) allows us to start to see relationships between two variables. Each cell shows a count of the observational/experimental units that fall into the various categories.

Before we get into the TIDES data, let's return to an old example:

Recall the Organ Donor Study we saw in Week 1…participants were given an organ donor application with one of three different wordings in order to see if the wording impacted whether they chose to become organ donors. There were 161 total participants. Of the 50 participants in the opt-in group, 21 agreed to be organ donors. Of the 56 in the opt-out group, 46 agreed to become organ donors. Of the 55 in the neutral group, 43 agreed to become organ donors. We can organize this data in a two-way table:

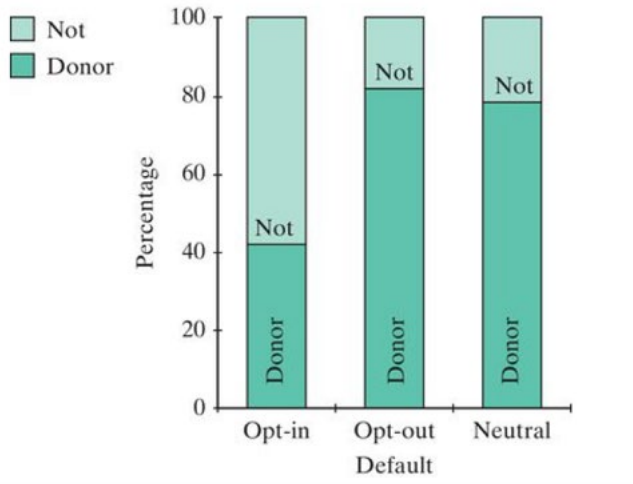
		Donor Choice		
		Donor	Not Donor	Total
Application Wording	Opt-in	21	29	50
	Opt-out	46	10	56
	Neutral	43	12	55
	Total	110	51	161

Notice the "Total" column at the right of the two-way table – this indicates that the number of participants in each application wording group was not exactly the same. When our groups are not the same size, we must be careful about using counts in our two-way table. Note that we would expect the counts to be highest in the "opt-in" row just because we had the most data points in that row (56 vs. 50 or 55). To make our data comparable, we want to convert the counts to **conditional proportions** (proportions computed from different conditions within our data as opposed to on our entire sample data set). Let's create another two-way table for this data, but this time with conditional proportions in the cells instead of counts.

		Donor Choice		
		Donor	Not Donor	Total
Application Wording	Opt-in	21/50 = .42 (42%)	29/50= .58 (58%)	50
	Opt-out	46/56=.82 (82%)	10/56=.18 (18%)	56
	Neutral	43/55=.78 (78%)	12/55=.22 (22%)	55
	Total	110	51	161

Thus, we can see that 42% of participants in the opt-in group, 82% in the opt-out group, and 78% in the neutral group chose to become an organ donor.

Additionally, we can organize this info into a **segmented bar graph**. The entire bar represents 100% of the data. The subgroups are shown as segments of the bar that are in proportion to their percentage of the data. When segmented bar graphs are used to compare subgroups, the bars are lined up adjacent to each with the subgroups in the same order in each bar.



Now that we have some examples, let's get into the data from TIDES!!

See below two-way tables for the 2021 MLB data and 2020 NFL data. Note that POC stands for person of color.

2021 MLB Data

		POC Status		
		POC	Not POC	Total
Role in Organization	Player	342	564	906
	Head Coach	6	24	30
	Assistant Coach	526	641	1167
	CEO/President	2	28	30
	Total	876	1257	2133

2020 NFL Data

		POC Status		
		POC	Not POC	Total
Role in Organization	Player	1264	418	1682
	Head Coach	4	28	32
	Assistant Coach	295	299	594
	CEO/President	4	28	32
	Total	1567	773	2340

As you can see in the far-right column, our groups do not have the same number of people, so comparing counts is not useful. What we really need are the conditional proportions in each category (for example, the proportion of all players who are POC and Not POC).

1. Fill in the following two-way tables with the conditional proportions (include percent as well) instead of the counts (use the Organ Donor table as an exemplar):

2021 MLB Data

		POC Status		
		POC	Not POC	Total
Role in Organization	Player			906
	Head Coach			30
	Assistant Coach			1167
	CEO/President			30
	Total	876	1257	2133

2020 NFL Data

		POC Status		
		POC	Not POC	Total
Role in Organization	Player			1682
	Head Coach			32
	Assistant Coach			594
	CEO/President			32
	Total	1567	773	2340

2. Based on your tables, discuss what associations you see between a person's role in the organization and whether they are a person of color. Explain what evidence you have to support each association.

3. Now construct segmented bar graphs for each of the MLB and NFL data:

MLB Data Segmented Bar Graph

NFL Data Segmented Bar Graph

4. In what ways do the segmented bar graphs make the comparisons and/or associations easier/harder to see versus the two-way tables?

5. What decisions did you make in terms of constructing your segmented bar graphs?

We've already computed numerous conditional proportions to help describe our data. You've probably already looked at the **differences** in at least some of these conditional proportions to help explain the associations you see in the data. For example, back to our organ donor study, 42% of people in the opt-in versus 82% of people in the opt-out group chose to become organ donors. This is a difference of 40 percentage points, which seems to indicate an association between these wording types and the choice to become a donor. Another way to compare the conditional proportions is to take their ratio, with the larger of the two values in the numerator – this is called the **relative risk**. For example, .82/.42 = 1.95. Thus, people in the opt-out group are 1.95 times more likely to choose to be an organ donor than people in the opt-in group.

6. Choose two comparisons to make from the data that you find interesting – one from the MLB data and one from the NFL data. For each comparison, compute the difference and relative risk (show computations for each) and interpret them.

   MLB data comparison (state what you are comparing)

   Difference

   Interpretation

   Relative Risk

   Interpretation

NFL data comparison (state what you are comparing)

Difference

Interpretation

Relative Risk

Interpretation

Now let's think about what we've explored tells us about diversity in the MLB and NFL.

7. Why do we care about diversity, in professional sports or in any other profession?

8. In what ways do you see diversity/lack of diversity in the MLB and NFL? Explain how the data above supports your answers.

9. In what ways does a lack of diversity impact an organization and potential future members of an organization? What messages does it send to potential future members if all the power is held by a non-diversified group of people?

In 1999, MLB was the first professional sports league to institute a hiring policy to help address the lack of diversity in the coaching staffs and top executives. Called the "Selig Rule," it required every team to consider minority candidates for all general manager, assistant general manager, field manager (head coach), director of player development and director of scouting positions. In 2003, the NFL adopted the "Rooney Rule," which required teams to interview at least one non-white candidate for head coaching positions. Let's look at some data to see if these initiatives helped the leagues diversify their coaching staffs and top executives.

1998 MLB Data (pre Selig rule)

		POC Status		
		POC	Not POC	Total
Role in Organization	Head Coach	4/30 (13%)	26/30 (87%)	30
	Assistant Coach	51/202 (25%)	151/202 (75%)	202
	CEO/President	Data unavailable		
	General Manager	0 (0%)	30 (100%)	30
	Total	55	207	262

2021 MLB Data

		POC Status		
		POC	Not POC	Total
Role in Organization	Head Coach	6/30 (20%)	24/30 (80%)	30
	Assistant Coach	526/1167 (45%)	641/1167 (55%)	1167
	CEO/President	2/30 (7%)	28/30 (93%)	30
	General Manager	4/30 (13%)	26 (87%)	30
	Total	538	719	1257

2003* NFL Data (first year of Rooney Rule - 2002 data unavailable)

		POC Status		
		POC	Not POC	Total
Role in Organization	Head Coach	3/32 (9%)	29/32 (91%)	32
	Assistant Coach	169/510 (33%)	341/510 (67%)	510
	CEO/President	0 (0%)	32 (100%)	32
	General Manager	2/32 (6%)	30/32 (94%)	32
	Total	174	432	606

2020 NFL Data

		POC Status		
		POC	Not POC	Total
Role in Organization	Head Coach	4/32 (12.5%)	28/32 (87.5%)	32
	Assistant Coach	295/594 (50%)	299/594 (50%)	594
	CEO/President	4/32 (12.5%)	28/32 (87.5%)	32
	General Manager	2/31 (6%)	29/31 (94%)	31
	Total	305	384	689

10. In what ways does the data suggest these diversity initiatives have/have not been successful in diversifying the coaching staffs and top executives in the MLB and NFL?

11. What ideas does your group have for why these initiatives have not been successful enough in all hiring categories?

12. What suggestions does your group have to better help diversify professional sports (or other organizations), especially with regards to positions of power within the organizations?