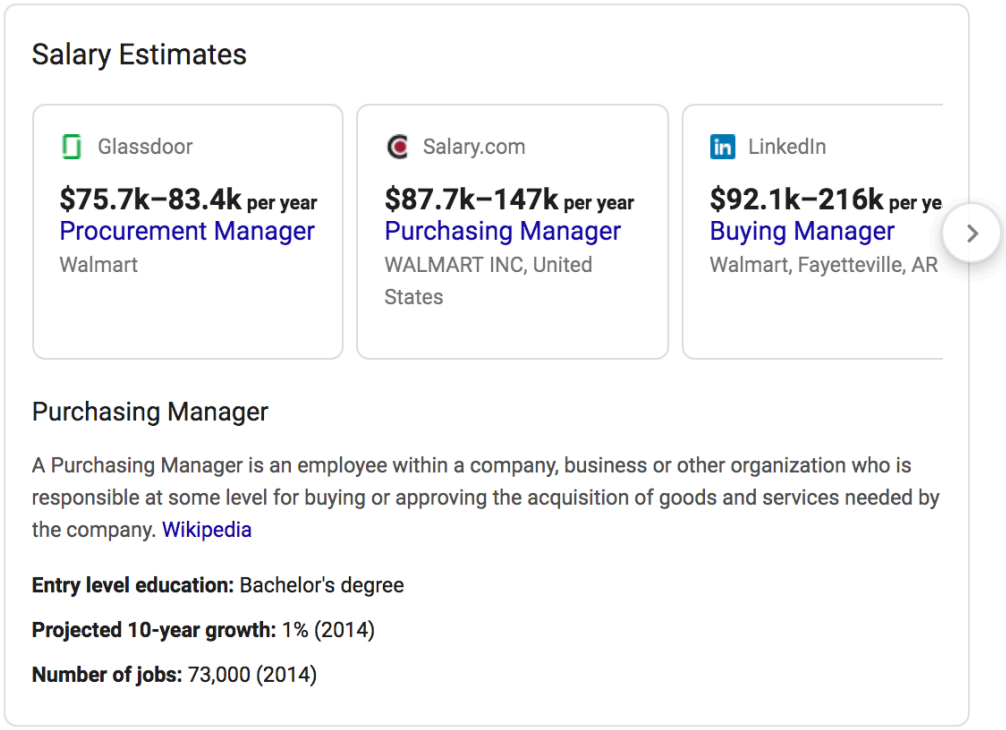
DECIDING WHICH STATISTICAL MODEL IS MOST APPROPRAITE TO ANSWER YOUR RESEARCH QUESTION

Options

Frequencies
Descriptives
Bivariate correlations
Chi-squared test
Independent-samples t-test
Paired-samples t-test
Linear regression
Logistic regression
Poisson or negative binomial regression
Sensitivity/specificity

1. You are the Purchasing Manager of a Walmart branch in the Bahamas.



You don't like to disclose your salary, but your annual "merit raise" is practically its own salary. You make a lot. But they expect a lot in return. They expect you to make the company *way* more money than they pay you (that's how *all* employment works). To accomplish that, you need to stock your shelves with products that will actually be purchased. To do *that*, it is a fundamental requirement that you know statistics… very, very well.

Cyclone season is coming up. And you have a database of past consumer activity. So you look up the last time a natural disaster was announced and you assess purchasing patterns from that moment until the moment the disaster arrived.

A) What are the first statistical tests you would run? Rudimentary tests, i.e., what products are consumers (very generally) purchasing?

B) You want to compare those purchasing patterns to the patterns of a previous year when there was no disaster en route. What test?

C) You want to know what products consumers purchase *together* (so that you can display them together in the same aisle). What test?

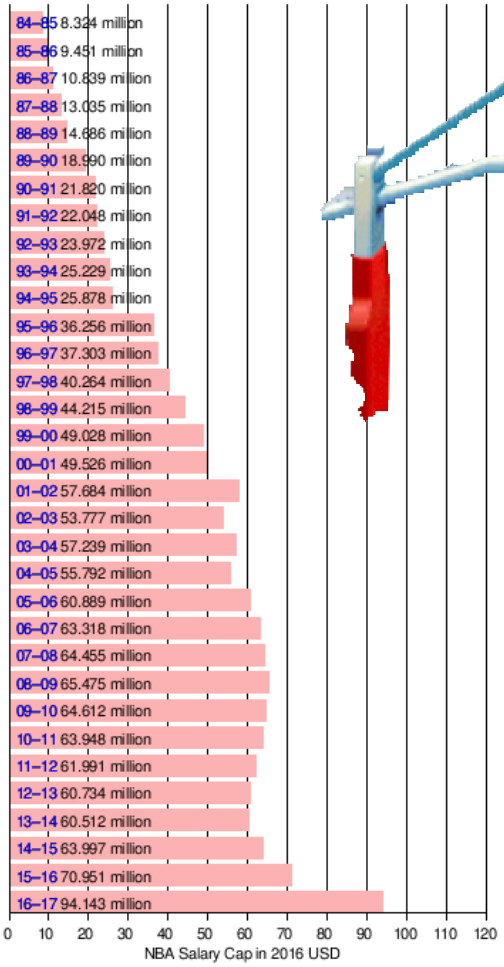
2. You work in the Human Resources office at University of the Pacific.

Recently, there have been complaints about compensation. Specifically, administration is "exhibiting sexism in salary assignment" within the social sciences. That's the claim. It's your job to investigate.

A) What is the first statistical test you run to see *whether* there's a difference between sexes in salaries?

B) If you do find a gross difference in mean salaries (and even if you don't find a difference) what is the next test you would run to measure the precise effect of sex on salary while holding other explanatory variables constant?

3. You are a galaxy-renowned basketball player.



Although you have the top salary in the league, your annual earnings remain stifled by salary cap restrictions. There's nothing more you can do on the court to prove (without earning) your worth. So you turn to statistics. The table to the left shows inflation-adjusted salary caps for every season between 1984-1985 and 2016-2017.

You also have data on the salaries of each individual team during those years, the salary of the highest paid player on each team, average ticket sales per game for each of those teams, average per-game television viewership, and gross NBA income.

A) What *initial* statistical test would you run to see how all of those variables relate to each other?

B) What test would you run next? In other words: what statistical model would be most appropriate to justify an increase in the salary cap?

4. You recently completed your undergrad in some health-related field.

You want to go into physical therapy, but you don't want to spend all of that money applying ($1,500 or so) if you're not *certain* you'll be accepted somewhere. So you get data on all applicants to the program where you hope to go. 300 applicants, 40 of them admitted, 260 rejected. Among those applicants, you have demographic information (age, sex, etc.), overall GPA, science-specific GPA, total GRE score, GRE subscale scores, whether they had research, how many publications they had, how many clinical hours they completed, and what undergraduate institution they graduated from.

A) You want to see what the average applicant looks like (age, GPA, what percentage did research, etc.). What tests do you run?

B) You want to know the difference in GPA, science-specific GPA, and GRE scores between admitted and rejected applicants. What test?

C) You want to know how important research is to the likelihood of acceptance. First, not controlling for any other variables, is there a difference between admitted and rejected applicants in whether they had published. What test?

D) You find a significant difference in whether accepted and rejected applicants had published research. So now you want to isolate its exact value. Holding constant variables like GPA and GRE, what effect does research have on the odds of acceptance? What test?

5. You are the principal of Shankstabbing High School in Stockton.

"Statistically," your school has the worst crime in the district, or so you've been told by the district folks. First, you have some qualms about the misuse of the word "statistically" there. (What stats are they talking about, exactly? Or are they just employing the word colloquially in a careless attempt to sound objective? It's usually the latter.) Second, you're wondering if you should implement a school uniform policy to remedy your school's crime syndicate reputation.

There are twenty schools in your district. One of them is yours. Ten of them have uniforms; the other ten (yours included) are being pressured to adopt them, but haven't done so yet. You have an enormous amount of data at your disposal. Instead of disposing of it, you analyze it.

A) Let's start with some means. What are we looking at (district wide)? How much crime per school on average?

B) Ten schools enforce uniforms; the other ten don't. All of them have data on the number of crimes, number of suspensions, number of expulsions, and average GPA. What test would you run to see if there's a difference between uniformed and non-uniformed schools?

C) All schools also have data on the percentage of students who graduate and the percentage of graduates who are admitted to college. What test would you run to see if there's a difference between uniform schools and non-uniform schools?

D) All ten schools who adopted uniform policies have before-and-after data. Same variables as above. What test would you run to see if there's a pre-post difference in those variables?

E) How would you estimate the effect of uniforms on mean GPA while holding constant the size of the school and teacher-to-student ratio?

6. You are a PE teacher in one of the newly-uniformed Stockton schools.

Budget cuts occur every year. During budget-cutting season, if you fail to justify the importance of your job, you will lose it. Administrators are not moved by subjective arguments ("I'm improving the confidence and self-worth of my students!" is a justification that will result in your unemployment unless you manage to quantify that claim). So you set out to quantify your value. Objectively. Numerically. Statistically.

You have data on all twenty schools in your district. Some of them have mandatory PE in the curriculum every semester; others have elective PE.

The dependent (outcome) variables of interest are student GPA, graduation rate, SAT score, and percentage who go to college. Your independent (explanatory or predictor) variables are size of school, student-to-teacher ratio, whether the school is uniformed or not, demographic information, and whether students have PE.

A) You want to know what variables are associated with PE. Just a basic "what's related to what?" test. What do you run?

B) You want to look at the difference in mean GPA between seniors who had PE every semester and seniors who had PE for no more than two semesters total. What test?

C) You want to know the graduation rate difference between students who had PE at least six times during high school and students who had PE no more than twice. What test?

D) Ignore *overall* GPA. Now you want to compare the *single-semester* GPAs of students while they had PE and while they didn't. Same group being compared in two different scenarios. What test?

7. You lose your PE job. You didn't want it anyway. Now you're a coach. College basketball. But you have the same problem: annual budget cuts. You still have to justify your job every year.

The school now cares about a lot more than they did when you were teaching high school PE. Athletically, they care about ticket sales, winning percentage, and whether your team makes it to the playoffs. Academically, they care about your athletes' mean GPA, retention (quantified as the number of semesters that the student remains enrolled), and graduation rate.

A) All of your athletes undergo preseason cardiovascular screening. Do they have hypertrophic cardiomyopathy? Sickle cell anemia? Some other potentially risky condition? Two of your athletes test positive for sickle cell and one tests positive for hypertrophic cardiomyopathy. How sure are you that those aren't false positives (what statistic?)? What statistic assures you that the negatives aren't false negatives?

B) You want compare in-season GPA (that semester) to off-season GPA (*that* semester) in all of your athletes. What test do you run?

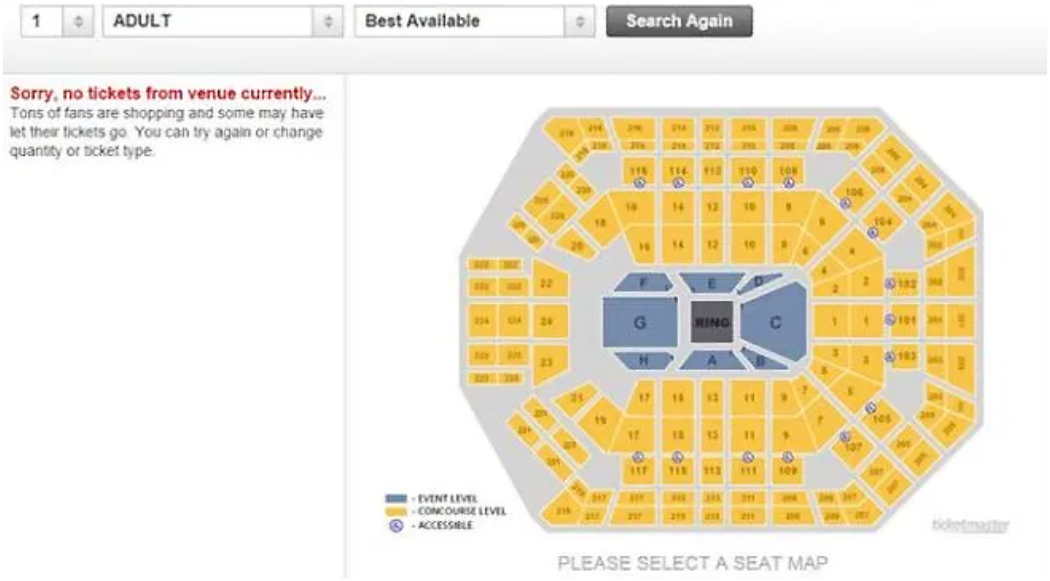
C) You want to compare the graduation rates of basketball players to the graduation rates of other-sport athletes *and* non-athletes. What test?

D) You want to compare retention (number of semesters enrolled) of basketball players to non-athletes. What test?

E) Now you want to make that same comparison, but you want to create a prediction model, assessing the *isolated* effect of basketball-player-status on number of semesters enrolled holding constant academic major and SAT score. What test?

F) You want to prove that your appointment as coach increased the team's winning percentage. But you don't want to just compare crude percentage numbers… because that's not real statistics. That's how idiots appraise things. You want to hold constant your program's inflation-adjusted budget, school ranking, and number of applicants. *Then* what's the difference on win percentage? What test does that?

8. You outsource the marketing portion of your stats to someone who claims to be talented in matters of sports marketing.



You need to demonstrate an increase in ticket sales. Your job depends on it. How should that sports marketing person figure out the important variables?

A) *Tons* of data exist. In every program. Guaranteed. You get those data. And you pass it along to that self-proclaimed "sports marketing guru" for analysis. What is a crude test he/she should run to see what kind of values exist? (In this sample, what kind of sales data are we working with here? What's the average ticket cost? How many tickets are being sold? What percentage of the seats are filling up?)

B) Now Mrs. Guru (I've decided it's a she, and she's married, so she's a misses) wants to know which variables are related to which other variables. And in what direction. And how strongly. What basic test?

C) Now that MG has a basic understanding of the variables she's working with (including their values and relationships with other variables), it's time to construct the statistical evidence that you (coach) are filling the bleachers better than any previous coach. Since you arrived, more tickets are being sold. What independent (i.e., predictor) variables are of interest (i.e., what should be controlled for?)? And what statistical test should MG conduct?