Cheat Sheet for Statistics 101

Why is having some background in statistics important?

"In many cases, judges review the statistical analysis produced by consultants, as presented and critiqued by competing expert witnesses on behalf of the parties at trial. The legal system in general, and judges in particular, become the audience for local government data gathering and analysis."

"Documenting Disparity in Minority Contracting: Legal Requirements and Recommendations for Policy-Makers" Public Administration Review, 2007

What are statistics?

Simply information, or data, of any kind, but usually in the form of numbers.

How are they used?

In three ways:

Descriptive statistics:	To describe someone or something. Context.
Inferential statistics:	To find patterns, usually in an effort to find (infer) a relationship between two things

How are statistics abused?

To only present part of a picture – they part the presenter wants you to see. Or to imply a relationship between two things when the pattern is weak, or, even if the pattern is strong, to imply that a relationship actually matters in the context of all the information – in the big picture.

What are some of the most common statistical terms I'll see?

What does a typical person/place/thing look like?

Mean	average
Median	middle observation

Is the people/places/things pretty similar or is there a lot of variation?

Range	lowest to highest value
Standard deviation	how spread out the data are

Does X track with Y? Is X related to Y? Does X influence Y? Does X cause Y?

Correlation	where two sets of data move together – e.g. as one goes up, the other goes up
Causation	where data indicate that one thing causes or has an influence on another thing

Significance

Statistical significance	when a pattern in the data is so clear it is not likely to be random. It is very likely that something is going on, that something is influencing the data to follow that pattern.
Material significance	whether or not the pattern, even if crystal clear, matters in the big picture.

High Quality Statistical Analysis **Uses this** 	Low Quality Statistical Analysis **Not this** 
Simple, Basic In interviews, 38 out of 50 clients said they were unhappy with the service	Unnecessarily complicated – using fancy methods when simple ones suffice. An index of client satisfaction was created using 10 different kinds of data gathered from 6 sources and merged through hierarchical linear modeling to create a predictive model of behavior showing most would report being unhappy if asked.
Can be explained in common sense terms We asked if the client was satisfied with the service.	Can't be explained unless the audience has a technical background. Trust me
Not changed much by extreme values The median number of complaints was only two per client. Two clients made 365 complaints each, skewing the average.	Open to manipulation by an extreme case There was an average of 7 complaints per client last year.
Can be summarized in a basic chart or graph without extra explanation	Need a lot of explanation to understand the data presented
Precision The police received 1048 complaints about this property in FY 2011.	Vague descriptions or terms A bunch of people complained.
Uses a lot of data, gathered systematically The analysis looked at all complaints (N=1048) for FY 2011 for this property.	Uses only a few pieces of data, gathered in a sloppy way We looked at the first dozen or so complaints that came in that day

High Quality Statistical Analysis **Uses this** 	Low Quality Statistical Analysis **Not this** 
Uses data appropriately Out of 50 clients, we spoke to four in person. Three said the complaints were minor.	Uses numbers to make the study seem bigger or more important than it really is. 75% of people interviewed said complaints were minor.
Context Those 1048 complaints for this property represented only 2 percent of all property complaints in FY 2011. Of the 1048 complaints, the bulk came from just two people – a married couple, each making a complaint each day (365 days) for the entire FY.	No context It was a big bunch of people.
Clear, precise definitions To be recorded as a complaint, the client had to provide a written description, in hard copy or electronically, of a specific problem and the context of the situation to the central office. General expressions of dissatisfaction were not recorded as complaints.	Overly complicated or vague definitions A complaint was when someone said they were unhappy with something.
Comparable data We compared FY 2011 data to FY 2012 data	Combining different kinds of data We compared FY 2011 data to School Year data for 2011-2012

High Quality Statistical Analysis	**Low Quality Statistical Analysis**
Uses this 	**Not this** 
Data measured in the same way In these studies, minority owned businesses were consistently measured as being at least 50% owned by a member of a federally recognized and defined minority.	Data measured in different ways In one of these studies, minority owned businesses was measured as being at least 50% owned by a member of federally recognized and defined of minority group. In another one, only businesses with 51% or greater ownership were classified as minority-owned.
Clearly documented sources Source: American Community Survey, U.S. Census Bureau, 2010.	No source A bunch of people told us… It is common knowledge that… Everybody agrees that…
Recognizable, credible and transparent sources Source: American Community Survey, U.S. Census Bureau, 2010.	If you have to ask "what is this group/who are these people?" The Society of Important NC Citizens told us… The Association of Knowledgeable people reported is common knowledge that… Lawyers representing people who were unhappy reported…
Caveats This analysis is limited to situations… This indicates… This supports the idea that… This challenges the idea that…	Strong claims 'this proves…' 'it is clear that…'
Correlation is not causation Popsicle sales and murder rates are correlated, but both happen in the summer. Popsicles do not cause murder. Sunlight causes murder. (Joking)	Correlation is sufficient to argue causality. Popsicle sales and murder rates are correlated. As one goes up, the other goes up, and as one goes down the other goes down. Therefore popsicles spur people to murder.