Statistics Concepts and Controversies

Chapter 1 Exercises:

1.2 Athlete's salaries

(a) The individuals in the data set would appear to be baseball players.
(b) The variables are team, position, age, and salary. The age and salary have numerical values.
(c) The units of age presumably are expressed in years. The salary is expressed most likely in thousands of dollars, since it says the data is from the 2004 season. If the year was not present, it is conceivable that salaries might have been single dollars if the data was from early 20th century maybe?

1.3 Who recycles?

I think that there might be a number of different ways to try to calculate who recycles rather than just total weight of recycling matter. Maybe comparing the total volume of recycling to volume of garbage, to assess recycled goods as a proportion of total consumption, without the result being skewed by glass weighing more than plastic/aluminum. Or you could take a random sampling of the garbage bins from different neighborhoods, and sort through them to compare the amount of recyclable material that was put in garbage rather than recycling bins.

1.12 Oatmeal and cholesterol

(a) The first method is an observational study because there was no intervention undertaken by the researchers to change any variables. They just measured what differences in variable already existed, potentially because of some other unmeasured variable that would affect the outcome (i.e. commitment to a healthy diet, vegetarianism, Amish upbringing, etc.) The second method is an experiment because two otherwise similar groups were randomly asked to follow a diet regiment that constitutes an intervention, such that the outcome of that intervention can be measured in LDL concentration.
(b) The experiment gives more useful information about whether oatmeal consumption reduces LDL because the experiment has been built around the intervention being given to two different samples within a (presumably) otherwise homogenous group.

1.13 Physical fitness and leadership

(a) No, it isn't an experiment because no treatment was undertaken. The only thing measured was the psychological test among two groups with a different degree of fitness, who had volunteered for the study. It was not designed to measure the variables within a broader population, and no controlled treatment was involved.
(b) It is unclear exactly what population the investigators were interested in. It says only "the relationship between physical fitness and leadership", and then states who was used as subjects. Presumably there is more to be learned that would tell us more if the sample group was not middle-aged executives who had volunteered for an exercise program. They - by definition - are already in positions of leadership; thus anything learned from that group could theoretically be extrapolated (erroneously) to signify that it is correlated with leadership. If they all for example had red hair, you could say that red hair (in your study) indicates leadership skills. Instead, they measured the variables of level of fitness and whatever the psychological evaluation resulted in.

1.14 The cost of paperbacks

The population in this study purports to be all paperback books. The sample is the number of paperback books on the one rack selected by the student, and the variables measured are number of pages and cost of each book.

Chapter 2:

2.1 Letters to Congress
A majority of voters do not necessarily oppose the bill. A majority of the letters received by the senator's office opposed the bill, but they are not necessarily representative of the electorate. It could be that the people most motivated to write the letters are those who oppose any government spending whatsoever, or that people in favor of the proposed bill don't feel the need to write complaining, because it seems that the political tides are in their favor and they are inclined to just go along with that.

2.2 Instant opinion
(a) The sample size is 4315 (Yes) + 8984 (no) + 780 (not sure) + 402 (don't care) = 14,481
(b) The poll cannot be trusted because the participants were not randomly selected. It is possible that they skew toward different biases because participation requires an internet connection, free time, the motivation to fill out an online survey, etc. Any of these might select for a different group or encourage a specific bias that is not being accounted for.
(c) This news clip could encourage survey respondents to feel more inclined to think the advertising targets minors, or in the contrary could encourage them to feel defensive of the beer companies. Either way it clearly could have some impact on the bias of the survey.

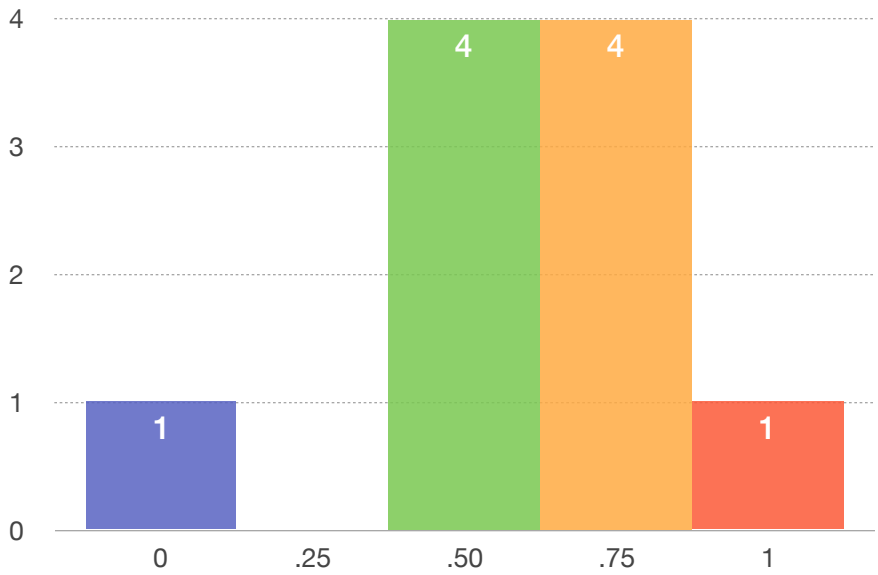
2.3 Ann Landers takes a sample
The sample is biased again because the people who take the time to answer were not randomly selected, and are clearly animated by strong feelings about the issue. I would imagine that there is some skew in terms of who reads that column, and further by those who feel strongly enough to answer in the affirmative, whereas people who have no strong feelings or are content in their relationships might not feel particularly motivated to respond. Furthermore the stigmatization of sex, especially the stigmatization of female desire, could inhibit women from answering in the negative, or could even prevent them from contemplating the issue from a clear perspective without morality or religious bias. I would imagine from a common-sense perspective that many fewer than 72% of women feel that way, but the only way to know conclusively would be to do a proper study.

2.5 Design your own bad sample
(a) The college could run an ad in the student newspaper asking students to respond to an online survey, and use the responses that came in as their sample.
(b) Or instead they could send an email or letter to every person who received a parking ticket from the school in the last year. That would generate some interesting perspectives about student parking!

2.10 Is this an SRS
(a) Every graduate student has an equal chance to be chosen, because the both groups are selected at a 10% rate, i.e. each student has a 1 in 10 chance of being chosen.
(b) It is not an SRS though because the two groups were separated on gender lines. Both the men and the women have a 1 in 10 chance of being selected, but the sample group is guaranteed to be exactly 2 to 1 in gender split. In that was it is not random, even if it is tailored to be more representative.

Chapter 3:

3.3 Voter registration - 18% of all voters in Philadelphia are registered republican is a parameter because it measures a quality of the entire population of subjects in question. The 60% who called in to the show is a statistic, as that is a quality of a selected group from within the population.



3.5 A sampling experiment
(a) 19 22 39 50 = 2 opposed, 2 in favor, 50% in favor
(b) 73 67 64 71 = o f o f, 45 46 77 17 = f f o f, 52 71 13 88 = f f f o, 95 59 29 40 = f f f f, 68 41 73 50 = f f o o, 82 73 95 78 = f o f f, 60 94 07 20 = o o o o, 36 00 91 93 = f f o f 38 44 84 87 = o f f o That's a total of 16 opposed and 24 in favor, out of 40 total. 60% in favor, 40% opposed.
(c) Histogram above.
(d) Of the ten samples, 4 estimated p = .5 correctly. The true value of p = .5 is in the center, but is not the mean or the average of my sample values. More samples would presumable even out the off-center nature of my sample.

3.10 Drugs for seniors
(a) The sample proportion $\hat{p}$ is 497 out of 815, which is .6098, or about 61%. The population parameter in this setting would be 61% of adults who believe that Medicare legislation does not go far enough in covering prescription drugs for seniors.
(b) The confidence statement is 95% confidence that the result is +/- .035 or about 4%.

3.12 Is a larger sample size always better?
No in this case it is not. Despite the large sample size, and resulting (supposed) slim margin of error, the study design precluded accurate results. Because the subjects were not chosen at random, but rather by self-selection, there is bias in the sample group and thus the results do not hold up to margin of error, which is calculated to account for random variation, not selection bias.

3.23 Find the margin of error
The margin of error for a survey of 61,239 people at 95% confidence would be .00404 or roughly +/-0.4%

Chapter 4:

4.1 Not in the margin of error
There are many types of error possible in the poll that are not included in the margin of error. Processing error is one possibility, undercoverage is likely, neither of which would be included in the margin of error. But most notably I think would be the potential for sample error based on response rate. It's likely, especially with such a hot-button issue, that many people who might have felt otherwise would not have participated in the survey. That is nonresponse, a non-sampling error, and would not have been included in the margin of error.

4.2 What kind of error?
(a) Not sampling error, but rather response error
(b) Processing error, not sampling error
(c) Sampling error, because the sample group was self selecting rather than randomly chosen.

4.14 Closed versus open questions
It seems to me that closed questions would lead to much greater ease in collating results, as there are a limited number of responses that can more easily be compared to other results, either in other areas or in the past to see changing trends or attitudes. But it constrains the diversity of human thought, and I myself have had difficulty at times with providing such simplistic responses. Perhaps in the example of broccoli, some of the respondents love broccoli but cannot eat it. Or love it cooked but hate it raw.
It also seems to me that closed questions might present more possibility of introducing bias in how the question is worded. A respondent might be more likely to respond in the affirmative to a question asking whether a political candidate seems corrupt if the question is worded to signify that possibility, whereas if they are only asked what their opinion of the candidate is they might say that they like their policies but dislike their personality, or some such more nuanced response.

4.17 A party poll
Every student has a one in ten chance to be chosen for the poll. It is not, however, an SRS because the population was stratified. A truly random sample would have had the possibility that the two groups would be unequally represented, for example it is unlikely but possible that all of the students selected for a truly random sample would come from just the under 21 population. The separation into two different groups and sampling from within those groups, but at the same sample rate, makes them both be equally represented in the results.

4.19 A stratified sample
(a) You would start by taking the alphabetized list and assigning each faculty member a number, from 000 to 999 for the male faculty and from 000 to 499 for the female faculty. Then you would take the first 60 digits (20 subjects times 3 digits) in three digit chunks, and those would be your male selections. The female faculty you will likely need the first 120 digits, as roughly half the three digit chunks will likely be outside your sample window (000 to 499). The first 5 females chosen from line 122 of Table A, discarding numbers over 499, would be 138, 159, 052, 087, and 359. The first 5 males, starting at line 125, would be 976, 461, 214, 937, and 823.
(b) The males have a 1 in 5 or 20% chance of being chosen (200 chosen out of 1000). The females have a 2 in 5 or 40% chance (200 out of 500).

Chapter 5:

5.06 Learning about markets
This is a bad idea because the two groups were self-selecting, not randomly generated, and thus there may be lurking variables. Maybe the students in the 8:30 lecture signed up for that version of the class because they have evening jobs and thus have more experience with markets in real life. Or perhaps the 2:30 cohort is in that lecture because there is another economics class in the mornings and they are taking both simultaneously, and are therefor better prepared.

5.07 The effects of propaganda
The experiment was affected by outside influences that would surely have changed the subjects' perceptions. This confounds the explanatory variables and makes it impossible to

know which variable was the one responsible for the response variable. It was no longer a controlled experiment. It could be visualized this way:

Original attitudes ————> Propaganda————> Later attitudes

Invasion_____↑

5.08 Learning about markets, continued

(a) So I would take the 16 recitation sections, numbered 01 to 16, and randomly assign each by selecting the first eight to be online and the others discussion based. Then as a response variable maybe have a test that quizzes their knowledge of the intricacies of market pricing and the interrelationship of the concepts, and then compare test scores of the two groups.
(b) The online section, chosen from line 119 of Table A, would be group 16, 04, 07, 10, 12, 15, 05, 09.

5.10 Do antioxidants prevent cancer?

(a) The explanatory variable is supplemental antioxidants, either beta-carotene, vitamins C & E, or beta-carotene and vitamins C & E. The response variable is prevalence of colon cancer among the subjects.
(b) The experiment is structured such that the subjects were split into 4 groups as below:

beta-carotene group ——————— \
original group ——> vitamins C & E group ——————— \ ————> compare cancer rates
betacarotene and vitamins C & E — /
placebo group ———————————/

(c) The first five of 864 subjects, chosen from Table A line 118 would be 731, 253, 304, 470, and 296.
(d) No significant difference means that all four groups had similar rates of colon cancer, or that the difference between them was not statistically significant.
(e) There are many potential lurking variables that could explain the outcome. One would be that people who eat more fruits and vegetables (outside of clinical studies) have generally healthier lifestyles, get more exercise, etc. The study design says that the subjects were people who were at risk for colon cancer. Perhaps by being at risk already, it indicates that there are genetic components that exist regardless of antioxidant consumption, or that the explanatory variables are things that happen in early life, long before the study is undertaken, and that that late-stage intervention is not enough to change the risk.

5.13 Sealing Food Packages

(a) The individuals in this experiment are the pairs of packaging material.
(b) The explanatory variable is the temperature at which the packages were sealed. It takes the values (in degrees Fahrenheit) 250, 275, 300, and 325.
(c) The response variable is the amount of force required to open the sealed package.

Chapter 6:

6.1Medical news

Placebo-controlled means that half the patients received a potential remedy, and the other half received something that was indistinguishable from the remedy, for the purposes of eliminating the possibility that patients would improve through the psychological effect of simply being treated, which is common and is known as the placebo effect. In this way the actual effects of the remedy can be isolated from any effects that are due solely to the expectation of effect. Double-blind refers to the fact that both the patient and the administrators of the experiment are blind to what treatment is being administered - whether it is the remedy or a placebo.

6.3 Treating acne
Single-blinded in this case probably means that the person administering the experiment knows which treatment is which. I imagine that is important in this situation because each subject is receiving both the treatment and a control, and thus they cannot be separated by being given a random number or some other method of keeping track of what treatment they get. In addition since the treatment is not a medicine or ointment but rather a physical intervention, it would be very difficult to not have the clinician know which side was which. I suppose it is theoretically possible that it could be double-blind, with both sides of the face receiving some treatment, but one side only UV light or something that could convincingly approximate whatever side-effects there might be of the treatment, both of which performed by an uninvolved clinician who immediately notes down which side was treated in a confidential file not available to the experimenter, and I imagine the same side would be treated every time. It would be very difficult to orchestrate.

6.18 In the corn field
(a) and (b)

Block 1			
Plot 1	Plot 2	Plot 3	Plot 4
C	B	D	A
Block 2			
Plot 5	Plot 6	Plot 7	Plot 8
D	A	C	B
Block 3			
Plot 9	Plot 10	Plot 11	Plot 12
D	A	C	B
Block 4			
Plot 13	Plot 14	Plot 15	Plot 16
C	D	B	A
Block 5			
Plot 17	Plot 18	Plot 19	Plot 20
A	D	C	B

Reading from Table A row 108, I will take the every number from 1 to 4 as the location for each corn variety A, which gives me 4, 2, 2, 4, 1.
Variety B, starting from where I left off on row 108: 2, 4, 4, 3, 1 (already taken by an A), 4.
Variety C: 1, 3, 3, 1, 3.
D fills the remainder.
This strikes me as not entirely the best way to go about this, but that was not the question.

6.19 Speeding the mail?
I would need a little more information to design this exactly, like how the letters are going to be mailed. Will they be dropped off at the post office? In a postal box on a busy corner? Regardless, I would create a large number of identical letters, half of which have the nine digit zip code and the other half with only a five digit zip code. Since the question only asked for two explanatory variables, we will assume that we are only measuring from one set city to another. Then I would take all the hours that one might reasonable expect the mail to get picked up, say from 8AM until 6PM, and 6 days per week that the post office is open, which gives a total of 60 different time slots per week that the mail could theoretically go out. Then for each letter from the two groups I would choose a number from Table A to signify which slot (from 1-60) the

letter would go out, and then record when it arrived at its destination and tabulate all the information from there. I would have a list of number of days it took for each letter to arrive, along with the time (and day of the week) it was mailed, and whether it had a 5 or 9 digit zip code.

6.23 Web-based exercise
The authors of the study found that placebos had no effects in objective or "binary outcomes". I presume this means either something happened or it did not, for example whether the subject had a heart attack or not? They give the binary example of smokers vs non-smokers, which I imagine came from studies of effective strategies for quitting smoking, where the binary choice was whether the smoker quit. But they also found that placebos they did have some effect on pain and on "continuous subjective outcomes". Unfortunately I cannot read the other article suggested because I need a subscription to JAMA to do so.

Chapter 7:

7.7 How common is HIV infection?
This is really tricky. I think that the review board was right to hold the same standards that would apply to anyone in the US. It would set a bad precedent to allow for less human rights just because the subjects live in a country with less robust respect for those rights. However, I think think in this situation the benefit of doing the study outweighs the alternative, which is that the women continue living with a potentially deadly disease and potentially spreading it, without any studies to understand the scope of the problem. I think that more dialogue would be necessary, and if the government really insists that the women cannot be informed because it would overwhelm the health care system, and that Yale is indeed invested in the health and wellbeing of the women there, they could try to negotiate a compromise where they offer assistance to the government in providing health care assistance?

7.10 Not really anonymous
I think if it is construed that there would be the perception of anonymity, it is up to the researchers to be forthright about what exactly that means and what it doesn't. This would be the "informed" part of informed consent. I think it is incumbent on the researchers to go above and beyond in transparency if they expect to have the trust of people who are engaging in helping them do their work.

7.15 AIDS trials in Africa
I think this and the above question about HIV in Africa are both versions of the trolley problem in a way. By getting involved in something that you know is an area involving risk you are taking on some responsibility for the outcome, whereas if you did nothing and people died you could at least deny personal responsibility. Do you not get involved because once you do it is expected that you do everything within your capabilities to prevent transmission and suffering? I think few would agree that this is true; in fact the opposite is widely accepted and viewed as harm reduction in some spheres. It seems like researching ways of stopping transmission is a very noble goal, especially with regard to transmission to utterly innocent newborns, and outweighs the cost of not being as ethical as possible, to a degree that seems unreasonable.

7.18 A right to know?
This is something I haven't ever considered. Since the raw data comes from the public, and the candidate is working (or seeking to work) for the public, it makes sense that the results of the poll should belong to the public. This also means though that there would potentially be a lot of bad information out there, if every poll regardless of quality was released to the public. Yet at the same time allowing pollsters to hoard their information allows them to use what they (or those who are sponsoring or funding their studies) find to manipulate the public. Maybe instead candidates and partisan groups should be prohibited from doing polls altogether.

7.19 Telling the government
I think the problem here is not the asking, though I do think that there should be an option "prefer not to say". Since response is mandatory to my knowledge, it does constitute a bit of an invasion of privacy if they cannot opt out of sensitive questions. The real problem though is that these things are stigmatized. In some places it is illegal to live in a domicile without running water or functional toilet, and admitting to mental impairment or disease could be used as justification to withhold rights or prevent employment. These realities compromise the presumed intention of information gathering by making it less likely for people to answer honestly. What is needed is to de-stigmatize these things, which would be facilitated by a more comprehensive understanding of how prevalent they are, with which the census presumably could assist. Does it invade your privacy if it is anonymous? How can we be assured that it truly is anonymous?

Chapter 8:

8.3 School bus safety
The counts of deaths aren't at all convincing because we don't know the proportion of each group that died in their means of transportation. To be useful at all we would need to know the rates or ratios of deaths for each category, and even that doesn't go far enough. The numbers that are given are for children who die each year in school bus accidents vs children who die during school hours in auto accidents. Were all those children on their way to school at the time of death? That number would have to be sifted from the data before we can begin to make useful comparisons. Even if all children did ride the bus, there will still be automobile deaths of school-age children during school hours, for those who aren't in school, or who are skipping school, or who leave school during school hours for doctor's appointments, etc.

8.6 Measuring intelligence
Answering questions like this is not a test of problem solving ability or cognition, it is a test of trivia, based solely on one's experience with certain things (relatively unimportant, at least with regard to intelligence) and would be heavily biased toward some segments of the population, whether it be those who are in more urban settings and thus have more exposure to popular media, or whether it is subjects of populations who might be more likely to watch the World Cup.

8.8 Measuring pain
I know that pain has historically been difficult to objectively measure. You could ask the patient for their own estimation, but I have read that women have a different experience of or relation to pain, so that right away makes the subjective reporting more difficult. Since we learned that in measurements with low reliability taking multiple samples increases accuracy, we could ask each patient to write down their average pain every hour for a week leading up to the appointment, or maybe once a day for a month, and take an aggregate of those measurements? But that would average out one-time occurrences like migraines or other temporary pain episodes. We could have a standardized pain scale and try to trust self-reporting by asking each patient to compare their momentary and chronic pain scales to a set list of common pain events, for example a pain scale from pinprick to broken limb. Or we could take fMRI scans and compare the activity in regions of the brain known to process pain. That would be an expensive diagnostic test though.

8.10 Testing job applicants
I suppose to determine this you would take a group of employees who had taken the GATB and then worked at various different jobs, and have their employers fill out surveys of their degree of satisfaction with the employees, and then compare that with the GATB scores across different groups of people to see if the GATB actually correlates with employer satisfaction for different groups. It is possible that even though some groups such as white people score

higher, that their actual on-the-job performance is equal to that of lower-scoring groups, or even less than, while it is equally possible that black subjects taking the test underscore but overdeliver. Evaluations of on-the-job performance could also be tainted by racism though, so if possible you could get direct information of job performance in some fields, like say student test scores for employees who are teachers, or number of orders fulfilled or some other such objective measure of performance, so as to sidestep the possibility that workers of color are rated more poorly than their peers even when they are equally productive.

8.19 Measuring pulse rate
Because a greater sample size will even out the variability in a smaller sample size. There may be aberrations in only a 6 second sample size that would average out over a longer time, and the small number of beats in only a 6 second sample would translate poorly when scaled up. What if your heart beat a split second before the sample started and a split-second after it ended? You would undercount by practically a whole beat. Taking the longer sample is essentially taking ten separate samples and averaging them together to get a more reliable measurement.

Chapter 9:

9.6 Trash at sea?
The coffee cups alone are highly implausible. If the crew was the same size as the passengers, 222,000 coffee cups would equal 111 per person, which is almost 16 per day for a week cruise. Add on to that 5 soda cans per person per day, almost 3 beer cans or bottles per person per day, and about .75 wine bottles per person per day. That's a very well-hydrated, jittery, and drunk cruise (and crew)!

9.11 Stocks go down
The Dow that day dropped 685 points, from its opening of 9605. So the math is:
685/9605=.0713, x100 = 7.13 percent

9.15 Greeks on campus
They added the two numbers, rather than averaging them or combining them in some other statistically weighted way (i.e. if women were only a quarter of the total population, giving a pure average of the two numbers would overstate the proportion of "Greeks" since the men have a higher percentage by gender). But since we can assume that the genders have relatively equal representation, it would make most sense to combine the two numbers *and then divide by half,* rather than just adding the two numbers. Very critical mistake, and good evidence that journalists should have to take statistics.

9.13 We don't lose your baggage.
I would not believe this claim, unless they could somehow prove that they lost zero bags during that period. Maybe they ceased operations six months prior, and thus had no baggage to lose?

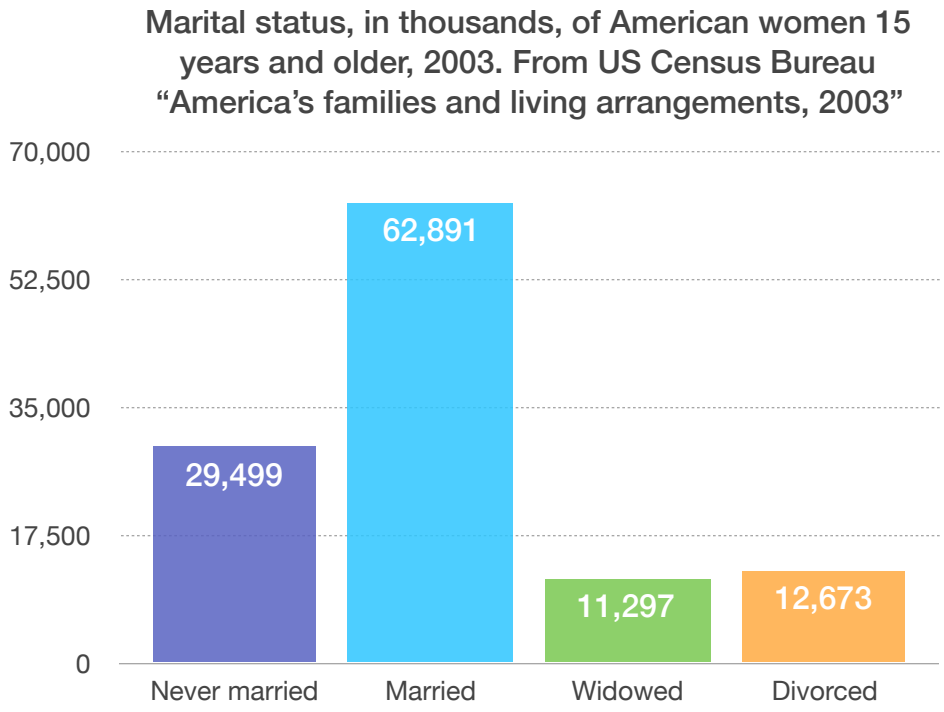
9.24 Time off
If 40% of sick days taken are Mondays and Fridays, and the office has a standard five day work week, that is exactly 40% of the week, which seems somehow unsurprising.

Chapter 10:

10.3 Marital status

(a) According to the graph, there were 29,499 (in thousands) never married women, plus 11,297 widows and 12,673 divorced women, for a total of 52,469 thousand married women, or 53,469,000 unmarried women.

(b)



(c) It would not be appropriate to use a pie chart, unless we were representing percents of total American women. You cannot compare separate values a pie chart.

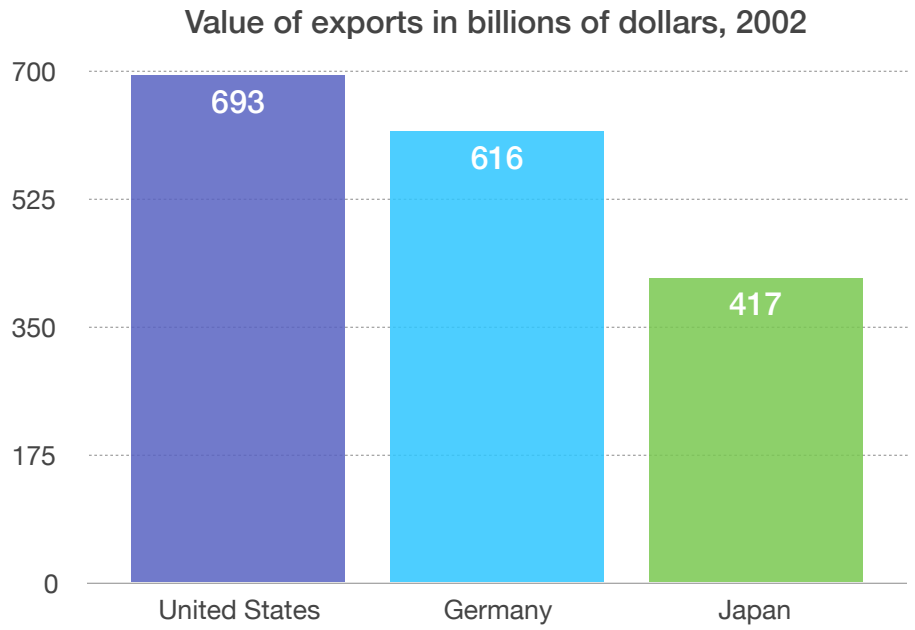
10.5 We sell CDs

CDNow obviously had the lions share of online CD sales, at least according to the graph. The percentages in the graph only add up to 49%, so there is 51% of the market unaccounted for. Presumably, since the graph is labeled "percent market share for the three leading online sellers" that the rest of the total is made up of smaller retailers, but we don't really know. The graph does not fairly represent the data, because the pictogram shows a giant CD for CDNow with their 33% of the market, and a much smaller CD pictogram for N2K with 12%. I will assume that the width of the images are proportional to the company's size of market share, but in terms of overall area that the image takes up, it makes it seem like CDNow has maybe ten times the sales of the nearest competitor, when in reality it is less than 3 times.

10.9 Exports

(a) It isn't an accurate graph because it does not represent the values in a proportional or easily comparable way, and misrepresents the size of difference between the exporters.

(b) Chart below



10.13 Trends

(a) I would expect the number of students entering university bringing a typewriter with them to trend downward every year (though I would note that I was one of the few who did!)
(b) The number of students entering with a personal computer I would expect to trend upward.
(c) The percent of adult women who do not work outside the home I would expect to trend downward.

10.18 Trucks versus cars

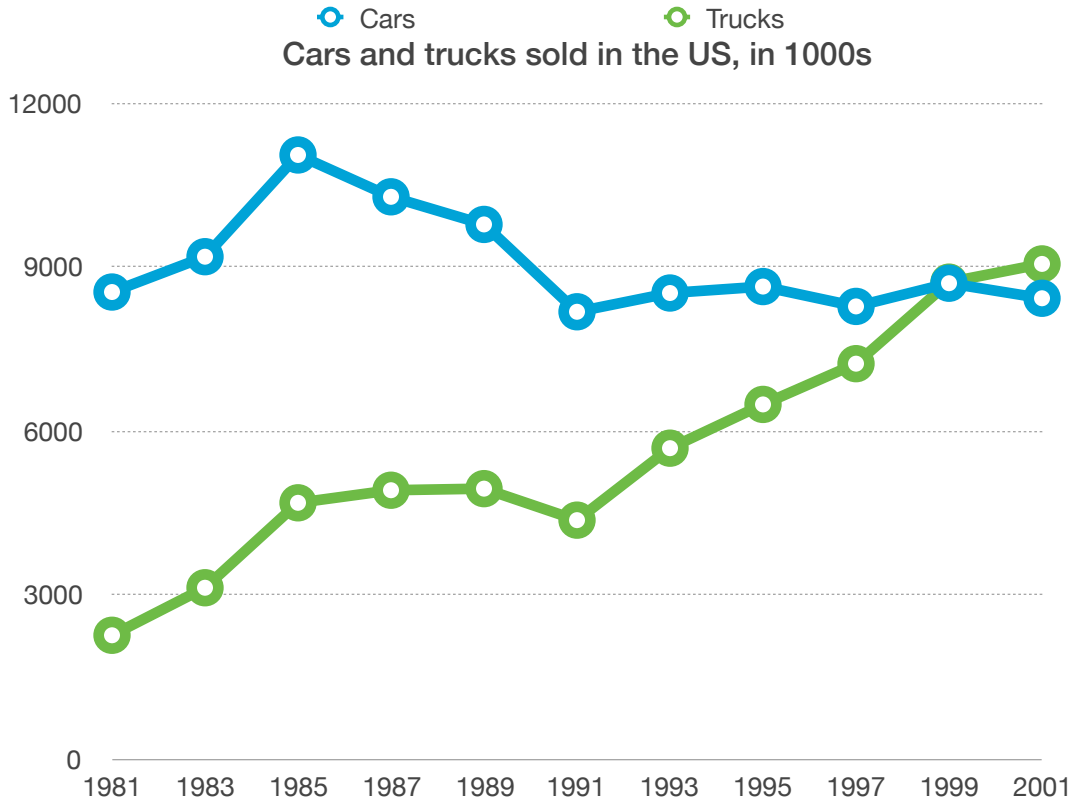


Chart is above. It seems clear from the line graph that the total number of cars over the time period rose but then dipped back down to stay at around the same level throughout the course of the time period, while the number of trucks had a pretty steady increase across the same time period, just beginning to outdo car sales by the end of the data in 2001.

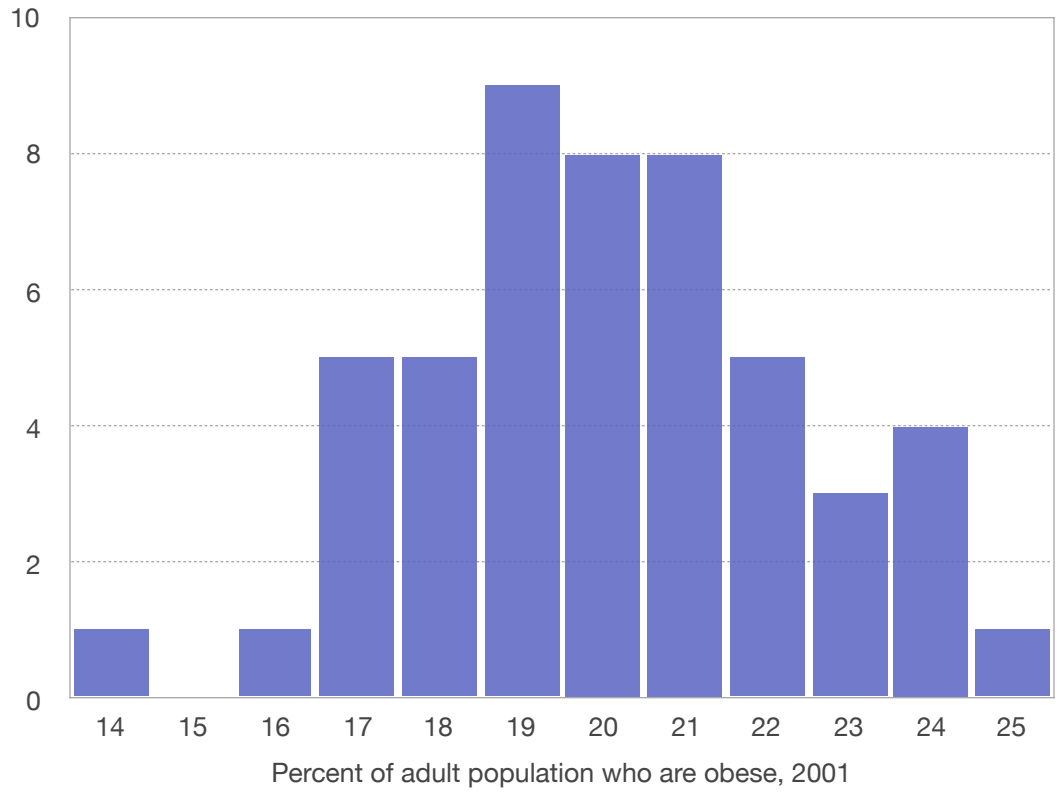
Chapter 11:

11.2 Where do the young live?
(a) Utah and North Dakota have 12.0 and 13.3 percent of their residents age 18 to 24. It is not clear from the graph which is which.
(b) Ignoring those two, the rest of the stemplot has a symmetrical shape, with a center around 9.5%.
(c) The distribution is much less spread out. This stemplot, discarding the two extremes, has only 4 values, whereas the other was spread across 8 values.

11.4 Returns on common stocks
(a) The shape is symmetrical with a broad spread and a single peak.
(b) The approximate center of the distribution is from 10-20% return. The smallest and largest total returns were from -70 to -60%, and 100-110%.
(c) About 23% of the stocks lost money.

11.7 The obesity epidemic

Here is a chart showing the distribution of the percentage of adults who are obese in each of



the 50 states, organized in number of states with each percent represented, rounded to whole

numbers. Colorado is an outlier with 14%, the rest of the states for a slightly right-skewed shape, with a center at around 20%.

11.13 How many calories does a hot dog have?

10 | 7
11 |
12 |
13 | 5689
14 | 067
15 | 3
16 |
17 | 2359
18 | 2
19 | 015
Calories in different brands of hot dogs

My guess is that the Eat Slim Veal Hot Dogs are the outlier, at 107 calories. The rest of the distribution is dispersed, with two peaks around 130 calories and another around 170 calories. The center of the distribution (discounting the outlier) is 160 calories.

11.16 Back-to-back stemplot

```
                69 | 1 |
              455 | 2 | 25
         334477 | 3 | 45
       0255669 | 4 | 1166679
               | 5 | 449
               | 6 | 0
             3 | 7 |
Barry Bonds    |   |  Babe Ruth
```
Number of home runs per season

From the graph you can see that Bonds had the greatest number of home runs in one season, though that season was far and above his season average. Meanwhile Babe Ruth was more consistent in hitting a higher number of home runs per season.

Chapter 12:

12.5 Where are the young more likely to live?
(a) The shape of the distribution is roughly symmetrical, slightly skewed to the right but only slightly, and the vast majority of the points are concentrated firmly around the peak, which will mitigate the degree to which the one outlier can affect any difference between the median and the mean.
(b) There are 50 data points, so the median is between the 25th and 26th, which is 10.0%. The mean, meanwhile (pun intended) is 10.128.

12.9 How many calories does a hot dog have?
Since I already did the stemplot above, I just need the five-number summary. There are 17 data points, so the median is the 9th data point (easy in this case without the formula, but (n+1)/2 is 9), which is 153. The first quartile is in between the 4th and the 5th data points, so an average of the two or 138.5. The third quartile is between points 13 and 14, or 180.5 calories. The minimum value is 107, and the maximum is 195. So the five-number summary for this is:
107 138.5 153 180.5 195

12.11 Minority students in engineering

(a) The positions of the five-number summary for 152 observations would be:
1 38.5 76.5 114.5 152

(b) To read the data off the histogram, which doesn't give precise numbers but rather ranges (1-5, 6-10, etc), is a little tricky. I will assume the lowest value is 1, and both the Q1 and median still fall within the first bar of the graph which could be any value from 1-5, so I will approximate the values for a five-number summary something like 1 3 5 10 47. With this much of a skew to the spread, you only need to graduate more than 10 students to be in the top quarter.

12.12 The statistics of writing style

Tricky, this one. I think that the five-number summary here would be
1 2.5 3.5 6.5 15
Such that the lowest letter count in a word is 1, roughly 25% of the words have 2.5 letters or less, roughly half the words have 3.5 letters or less (or more), and 75% of the words have 6.5 letters or less, with the longest word having 15 letters.

12.14 Immigrants in the Eastern states

There are 26 states listed, the median of them all is between plots 13 and 14, or 10.3 thousand immigrants. Without New York and Florida the median is the average of data points 12 and 13, which is 7.4 thousand immigrants. The mean with Florida and NY is 21.15, the mean without them is 13.8. The median without the two outliers is 71.8% of the median with them included, whereas the mean without them is 64.2% of the mean with them. The mean in this case changes more dramatically when the two outliers are excluded.

Chapter 13:

13.5 IQ test scores

Since the mean score is 111, and the standard deviation is 11, 100 points is one standard deviation below the mean. Using the 68-95-99.7 rule we know that 68% of scores are between the -1 and 1 standard deviations. Since this is located right at the border of the lower standard deviation, we know that the amount between score 100 and score 111 is 34%, and what is below score 100 is 16%. So roughly 16% of the seventh graders have an IQ score of less than 100.

13.7 Length of pregnancies

(a) If the mean is 266 days and the standard deviation is 16 days, that means that the middle 95% of pregnancies fall within two lengths of the standard deviation, or from 266-32 to 266+32 - from 234 days to 298 days.

(b) By extension, since 95% of the pregnancies are within that range, among the remainder 5% half are less than this middle amount and half are more than it. So the longes 2.5% of pregnancies are longer than 298 days.

13.18 NCAA rules for athletes

If the mean is 1026 and the standard deviation is 210, then a score of 820 would be about one standard deviation below the mean. Thus again the students with less than 820 would be the bottom 16% (the lower half of the 32% that lie outside the middle standard deviations) of scorers.

13.19 More NCAA rules

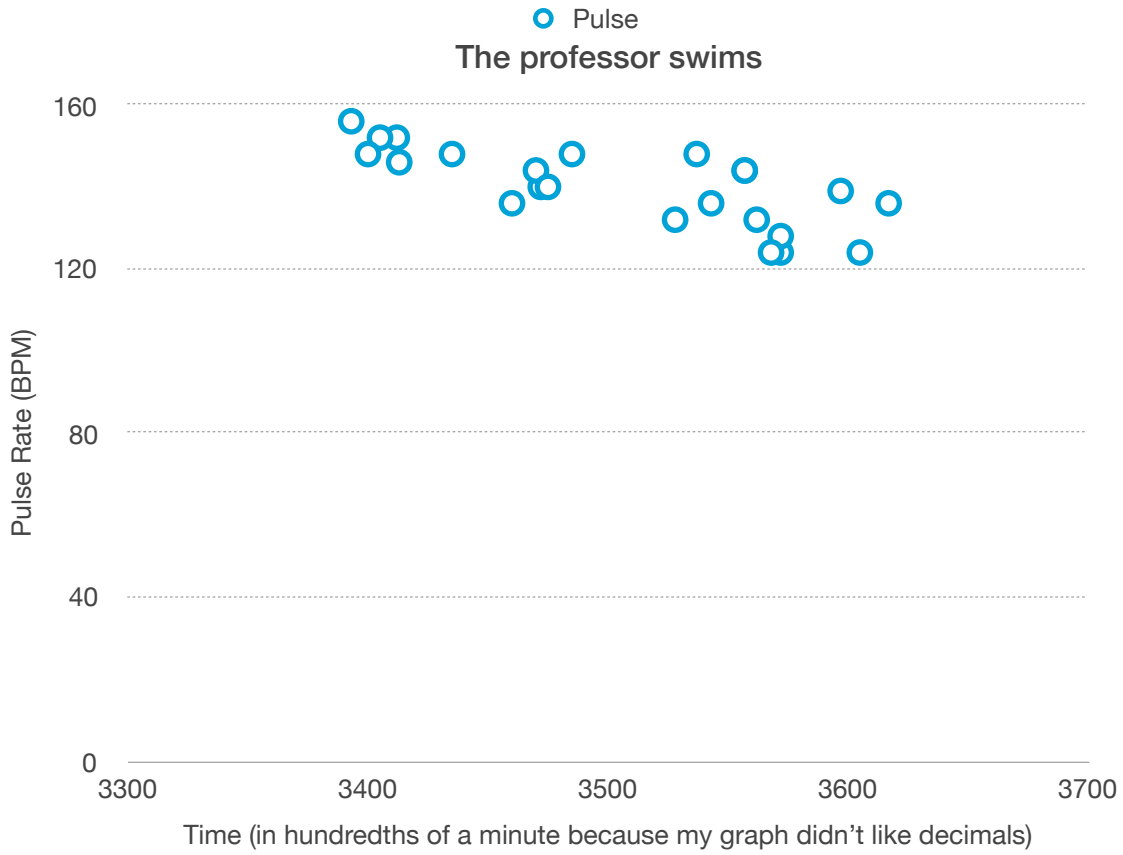
The mean of 1026 - the low score of 720 is 306, or about 1.5 standard deviations. Table B tells us that -1.5 standard deviations puts those students in the bottom 6.68 percentile, so only 6.68 percent of students have less than a score of 720.

13.23 Japanese IQ scores
So the mean for American and European scores is 100, with a standard deviation of 15. Thus the Japanese mean of 111 is a little less than the standard deviation, more exactly it is about 0.7 of the the standard deviation, or about the 75.8th percentile.

Chapter 14:

14.9 The professor swims
The mean of the times is 35.06 and the standard deviation is 0.73. The mean and standard deviation for the pulse rates are 140 and 9, respectively.

(a) Below is the scatterplot. I mapped time as the explanatory variable because I figured speed (or lack thereof) would be correlated with pulse rate.



(b) The association is negative. I would explain this by saying that the time it takes per session is inversely related to the speed of swimming, which I would expect to be positively associated with heart rate. In other words the longer the time, the slower the speed, and thus the lower the heart rate.

(c) Without calculating *r* I would say that the relationship looks quite strong, with the points being somewhat but clearly dispersed around a straight line angled slightly downward.

14.13 The professor swims (some more)

(a) My graphing program just returned a value of -1, which seems roughly correct but seems too tidy a number to entirely trust. Is it rounding to the nearest integer, or tenth or something? Or do I not know how to use it properly? The main gist of the answer seems

right though, signifying that there is a very strong negative correlation between the amount of time it takes and the heartrate.

(b) The unit of measurement would make no difference, because the thing being calculated here is the relationship between the two variables as they change, and whether they change in tandem or not.

14.16 Body mass and metabolic rate

As above, changing the unit of measurement would have no effect on the relationship between the two variables. It would move the entire plot of scatter points around on some grid, and maybe they would be more scattered out over a larger area, but the overall effect would be the same and the *r* value would not change.

14.18 Teaching and research

The newspaper report is wrong, because what the psychologist was actually saying is that there is no relationship between how highly a professor is rated by their students and the productivity of their research. Rather than saying that "good researchers tend to be poor teachers" what they psychologist was actually saying is that there isn't a relationship between the two - some great teachers may also be doing very productive research while some poor teachers may also be doing great research - the data show that the two factors have no connection.

14.20 Guess the correlation

(a) Substantial negative correlation. The greater the age, the less the price.

(b) One is tempted to think it would be negative - but we need more information. Hybrid cars weigh more than similar non-hybrid models but get better gas mileage. If we exclude hybrids, I imagine there would be somewhat of a negative correlation - greater weight, lower MPG.

(c) I would guess there would be a mild correlation. For men of roughly equal fitness the greater the height the more a man would weigh. Fitness would play a huge role here though.

(d) No correlation.

Chapter 15:

15.4 State SAT scores

(a) That the correlation is negative says that an increase in the explanatory variable leads to a decrease in the response variable, and vice versa. Thus in this case the more students take the SAT in a given state, the lower the state's average score.

(b) If r is -0.85, then r^2 would be 0.7225, so 72% of the average score is predicted by the number of students taking the test.

15.8 2000 and 2004 presidential elections

(a) There is a very strong positive association between the two variables. It is an upward slanted line at an almost 45° angle, and almost all the dots are very close to the imagined line. There is one data point that seems to be on the line, but is very far removed from the other cluster. It is an outlier in how low a percentage voted for Bush, but not in the degree of correlation between the two election year outcomes.

(b) I don't have a plot to draw this line on because I don't have the data and I'm not writing in the book.

(c) If the correlation is r=0.973 then r^2 would be .95, so 95% of the variation in 2004 percentages is explained by the 2000 percentages.

15.10 Wine and heart disease

If the regression line is y = 260.6 - 22.97x, then death rate = 260.6 - 22.97(wine quantity)
A country with 1 liter per person would have a death rate of 237.63 per 100,000 people.
A country with 8 liters per person would have a death rate of 76.84 per 100,000.

15.21 Lots of wine
Using the above formula, a country with yearly per person wine consumption of 150 liters would have -3,184.9 deaths from heart disease per 100,000 people. This is obviously absurd, as you cannot have negative deaths. This goes back to what they said in the lesson, that you can make predictions within the data, and potentially extend a bit beyond the data, but this far extended from the observations the regression breaks down.

15.26 Is math the key to success in college?
I would venture to guess that the students taking math in college are not benefiting from some "magic" property of the math classes, but are the students who would succeed anyway, be it from having supportive parents, being smarter than their classmates, or being more motivated and capable from the beginning. It may be correlative, but not causative.

Chapter 16:

16.7 The Food Faddist Price Index
If we calculate the index at the same quantities of goods that he bought in 1995, but at 2005 prices, then the FFPI (1995=100) would be 1940.5/1547 X 100 = 125

16.11 Living too long?
2003 $ = 12,000 x (2003 CPI)/(1975 CPI) = $41,041

16.13 Good golfers
Vijay Singh (2004) = $10,905,166.00
Ben Hogan (1940) = $10,655 = (2004) $143,766.39
Jack Nicklaus (1976) = $266,499 = (2004) $884,739.21

16.15 Calling London
The $12 phone call in 1976 was worth $40.28 in 2004 dollars. The actual cost in 2004 of only $15 represents a price decrease of about 63%.

16.20 Rising incomes?
If median income rose from $39,547 to $43,318 in real dollars, that represents an increase of about 9.5% (43,318/39,547 = 1.09535). The threshold for the top 5% of households went from $115,000 to $154,120 (154,120/115,000 = 1.34017) which means about 34%.

Chapter 17:

17.8 Winning a baseball game
No, I would not. While over time that average has emerged, it is an average of disparate factors over a long period of time, not a result of random chance. Furthermore, the best team from each league playing each other promises to be a more competitive game than the best from each league playing other teams within their league who are (by dint of the fact that this team is the champion) lesser teams.

17.9 Will you have an accident?
(a) I would guess .1
(b) I don't drive that much, and my car has many safety features that other cars do not. I also don't text and drive, drive drunk, or otherwise drive distracted. I think that these things would contribute to having a lower probability, though the probability still exists.

Nevertheless, I think there are probably people who drive particularly unsafely and thus get into more than their fair share of accidents, which must average against people who get into fewer. I would hope to be in the "fewer" category. At the same time I understand to some extent it is truly random.

(c) Most people probably think that they are exceptional and that statistics only happen to other people.

17.12 Personal probability?

The design and construction of space shuttles is an insanely complex and dynamic process. It seems surprising to me that there aren't more failures. While the 1 in 100,000 number was obviously misjudged, and seems both too small a chance and too easy a number, I imagine that you could calculate some rough probability, by looking at the number of complex systems and calculations and communication successes that would be necessary for a successful launch, and tally them all up along with rough probabilities for some sort of error within each part of each system, and get a more accurate idea of probability.

17.14 Playing "Pick Four"

(a) Each four digit number has the same probability, because each number is essentially being put in a hat, mixed around, and drawn out. 2873 has the same chance as 1234 or 6666.

(b) The point of the question being asked here is that most people would assume a number that "looks random" is more likely to be picked than one that doesn't "look random", like four of the same number or whatever the four digit year is. But I think the authors are also missing the fact that many people probably don't choose numbers that "seem random", but rather pick numbers that are personally significant to them that specifically *don't* seem like random numbers, for example birth days, special years, 6666 and 7777, auspicious numbers or combinations of auspicious numbers, like 2323, etc. However if you did pick a number that was numerically unlikely to be one of those selected by other people (e.g. one that couldn't be a combination of month/day, or was too high to be a year, or didn't have any special meaning attached), they are right that you would win a bigger pot.

17.20 An unenlightened gambler

(a) The ball has no idea what is "due", there is a 50/50 chance *every single time* no matter what the previous nth results were.

(b) Here again the gambler is wrong. A card game is zero sum, such that there are a limited number of cards in the deck. If you already have five of the 26 red cards, the chances that the next is red is significantly less than 50%. It is still random, but the odds are no longer even.

Chapter 18:

18.8 Birth order

(a) GGG, GGB, GBG, GBB, BBB, BBG, BGB, BGG. Since there are 8 possibilities, the chance of any one is about 12.5%.

(b) Out of the above possibilities, 3 are some arrangement of two girls and one boy, so the probability of that outcome is 3 x 0.125 or 37.5%.

18.11 Colors of M&M's

(a) 0.13 + 0.13 + 0.14 + 0.16 + 0.20 = 0.76, so the proportion of blue M&M's is 0.24.

(b) 0.12 + 0.12 + 0.15 + 0.15 + 0.23 = 0.77, so the proportion of blue peanut M&M's is 0.23, and therefore the probability of getting a blue one at random is 23%.

(c) A milk chocolate M&M chosen at random has a 47% chance of being red, yellow, or orange. A peanut M&M has a 50% chance of being one of those three colors.

18.12 Legitimate probabilities?

(a) The given probabilities satisfy the logic of probability, in that there are only two outcomes, and together they add up to 1. I would say that the probabilities given are not likely, but they satisfy the rules.
(b) No, this does not satisfy the rules, as the total sum of all possibilities is more than 1.
(c) No this doesn't add up to 1 either.

18.14 What will we die of?
(a) If the mean probability is 0.75 and the standard deviation is 0.014 then for the poll to get a sample with more than 77.8% of respondents saying yes would be twice the standard deviation, and half of those outside the standard deviation are in the left of the curve and the other half are to the right of the curve. So with the 68 - 95 - 99.7 rule, we know that the portion above the second deviation line is 2.5% of the the total density. Thus there is 2.5% chance that they will get a sample in which more than 77.8% think that.
(b) The probability of getting an answer that misses the truth by 2.8% or more is 5%, because it follows the same logic as above but could be over or under the truth, thus coming to the total of 5% that is outside two standard deviations.

18.16 What will we die of (optional)?
So 71.5% would be 3.5% less than the mean, and since the standard deviation is 1.4%, that is minus 2.5 standard deviations, which puts it in the .62th percentile, so about .62% likelihood.

Chapter 19:

19.4 Simulating an opinion poll
So since exactly 70% of women feel this way, we would construct a simulation where each digit represents one woman, and the numbers 1-7 mean that she feels her husband does his fair share, whereas 8-0 mean she does not. To run one repetition of the poll using line 112, we would need 100 total numbers, and starting from line 112 takes us to line 114. The numbers are:
59636 88804 04634 71197 19352 73089 84898 45785
62568 70206 40325 03699 71080 22553 11486 11776
45149 32992 75730 66280
That is, in response to whether the husband does his fair share:
YNYYY NNNNY NYYYY YYYNY YNYYY YYNNN NYNNN YYYNY
YYYYN YNYNY YNYYY NYYNN YYNNN YYYYY YYYNY YYYYY
YYYYN YYNNY YYYYN YYYNN
Which is 33 Ns and 67 Ys.
To run a simulation, you would do this many times with different number sets and tabulate the results of each run, then make a distribution graph, charting the mean and the standard deviation and thus figuring out what the likelihood is of getting 80 Y.

19.10 Tonya's free throws
(a) To simulate a shot if the probability is 0.8, take a random digit and 1-8 the possibility of making the shot, while 9 and 0 are misses. To simulate 5 independent shots, just take any five numbers at random from the table, translate them into Scores (S) or Misses (M), and those are your 5 shots.
(b) Using table A starting at line 125, we have:
MSSSS SSSSM SSSSS SSSSS SSSSS SSSSM SSSMS SMSSS MSMSS SMMSS SSSMM
SSSMS SSSSS SSSSS SSSMM SSMMS SSMMM MMSSS SSSSM SSSSM SMMSS SSMSS
SSSSS SSMSS SSSSM SSSSS MSSSS SSSSS SSSSS SMSSS SSMSS SSSMM SSSSM
SSMSS SSSSS SSMSS SSSSS SSSMS MMSMS MSSSS SMMSS SSSSS SSSSS MMSSS
SSSSS MSSSM SSSMS MSSSS MSMMS SSSSS

That’s 50 simulations, and 3 of them had three misses, with a rough probability of 6% of missing three (or more, though in this simulation three misses was the max) out of five free throws.

19.13 Gambling in ancient Rome

(a) To simulate the tossing of a single astragalus, you would take random numbers and assign each side a number range, so for example 1=narrow flat side (which we will call side A), 2-5=broad concave side (B), 6-9=broad convex side (C), and 0=narrow hollow side (D). Each digit you select represents one of the four sides being “up”. To throw four independently, you just take four random digits and assign each digit to one of four simulated astragali.
(b) Using the numbers starting at line 115, and using only the first four of each 5 digit cluster for ease of navigation, we have:

CADB, CCCC, CBBB, BBCD, CBCC, ABBB, BACC, BBBC, ABBB, BCDA, BABB, CDBC, CBAD, CBBB, BDBC, CAAC, BCAC, CCBB, CBAC, CDCB, BBBA, BCAC, BABB, CBBB, CBAC

That’s 25 simulations of a four astragali roll, and only 3 of them were “Venus” rolls, which are odds of about 12%.

19.15 Two warning systems

(a) You could simulate System A by choosing numbers at random and assigning 1-9 as functioning in time, with only 0 as a failure, and then choosing a series of numbers to represent successive runs of the simulation.
(b) You can do the same with system B, except using 9 and 0 as the fail digits.
(c) Run simulation

if 1-9 system A warns the pilot
System A (choose digit) <
if 0 go to System B (choose digit) <
if 9 or 0 fail
if 1-8 system succeeds

Basically you take two digits, if the first digit is 1-9 then System A has done its job and alerted the pilot. But if the first digit is a 0 then you look at the second digit, and if it is 1-8 then System B has done its job and alerted the pilot, but if not then both systems have failed. Starting from line 118 of Table A:

73 19 03 25 33 04 47 02 96 69 84 40 70 07 85 65 95 68 63 82 95 85 70 71 18 87 66 49 20 99 58 80 66 69 79 98 62 48 48 26 35 47 65 59 72 39 42 16 58 50 04 26 63 54 35 43 74 21 19 37 71 48 70 99 84 29 07 71 48 63 61 68 34 70 52 62 22 45 10 25 13 87 38 15 98 95 05 29 09 08 73 59 27 51 86 87 13 69 57 61

In terms of Pass/Fail it is:

P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P P F P P P P P P P P P P P

That was 99 out of 100 simulations succeeded in warning the pilot. Basically only a 00 or 09 cause both systems to fail, giving a 98% chance of success.

19.17 The airport van

I take 9 digits for each van, each digit represents one passenger, with 1-6 representing that they show up, 7 or 8 that they are a no-show, ignoring 9s and 0s and choosing the next number instead. Then run a simulation starting at line 101 of Table A:

122353457	265652711	752114748	
562871364	388830746	782735782	

125314254	227411858	874751181	
482853736	484876752	676553438	
764715412	573552476	314836472	
727754426	711481677	241786824	
488242536	537117275	436176568	
245467717	335684173	764188833	
775585328	513155272	613651541	
632485822	765858576	236388545	

That's 30 simulations, during which 5 of the 30 had more than 7 people show up to ride in the van. Not great odds if you want your customers to trust you.

19.22

Chapter 20:

20.4 More Pick 4
Here are the probabilities:

Outcome:	$0	2604	$104
Probability	0.9976	0.0001	0.0023

You win $2604 if you pick 1234 and the selected numbers are 1234, so there is only a 1 in 10,000 chance of that happening. You win $104 if you select those same numbers in any different order. Those possibilities are 1243, 1324, 1342, 1423, 1432, 2134, 2143, 2314, 2341, 2413, 2431, 3124, 3142, 3214, 3241, 3412, 3421, 4123, 4132, 4213, 4231, 4312, 4321. That is a total of 23 other possibilities, out of the total possibilities for a four digit number, which are 10,000. The remainder of the 9976 possibilities (10,000 - 24) all net $0. So if we add up the products of these probabilities, you have:
(0.9976 x $0) + (0.0001 x $2604) + (0.0023 x $104) = $0 + $0.26 + $0.2392 = $0.4992 as the expected amount.

20.12 An expected rip-off?

Sex of child	**Probability**	**Psychic profit**	**Probability x Payout**
Boy	0.51	$10	$5.10
Girl	0.49	$0	$0

Every time the mother has a boy, the psychic keeps $10. Every time it's a girl, the psychic nets $0. The expected amount is $5.10 per customer. Definitely a profitable racket.

20.14 Life insurance
(a) The expected amount in this case is 0.0015 x $100,000 + 0.9985 x $0 = $150, which in theory nets you $50 in profit. However in reality since you are only selling one policy and you're not a life insurance company, either you have $200 that you grifted from your friend or you have a dead friend and you're $99,800 in the hole.

(b) I don't really think that's a great deal, and you are ill-equipped to handle the payout if your friend does die. The odds only work based on the law of large numbers, and in this situation you do not have the large numbers.
(c) A large company would be profiting off the odds that they are raking in many orders for the life insurance, and have those funds available to pay out to the rare cases when their customers do die. They can actually benefit from the odds, whereas an individual can only really lose.

20.17 We really want a girl
I would simulate this situation by taking two digits at random and setting 1-49 as resulting in a girl, and 50-00 as resulting in a boy. We would then read off a chart like Table A and take whatever number of two digit numbers necessary until a girl results, at which point we record the number of total children resulted from each simulation.
Starting at row 126 of Table A, we have:
96 B, 92 B, 71 B, 99 B, 31 G (5) - 36 G (1) - 80 B, 97 B, 41 G (3) - 92 B, 77 B, 56 B, 78 B, 87 B, 41 G (6) - 48 G (1) - 40 G (1) - 94 B, 19 G (2) - 03 G (1) - 43 G (1) - 90 B, 99 B, 94 B, 77 B, 25 G (5) - 33 G (1) - 06 G (1) - 43 G (1) - 59 B, 40 G (2) - 08 G (1) - 51 B, 69 B, 25 G (3) - 85 B, 11 G (2) - 73 B, 60 B, 71 B, 15 G (4) - 68 B, 91 B, 42 G (3) - 27 G (1) - 06 G (1) - 56 B, 51 B, 43 G (3) - 74 B, 13 G (2) - 35 G (1) - 24 G (1).
That totals 25 runs of the simulation, with a total of 53 children and an average number of 2.12 children.

20.19 A multiple-choice exam
I would set up a simulation where each random digit represents one question, with 1 and 2 representing the correct answer (C), and the rest of the numbers representing the wrong answer (W). Then reading off Table A starting at line 130 we have WWWWC, WWWCW, WWCWW, WWWCW, WWWWW, WWWWW, WWCWC, WWCWW, WWWWW, CWWWC. Oops, that was 50 iterations, not 20. The results for 20 repetitions were exactly 20% correct, the results for all 50 was a little less, 0.18 or 18%.

Chapter 21:

21.4 Gun violence
(a) The population proportion *p* is the actual proportion of all adults in the US who believe that gun violence will become more common, as opposed to whatever percent the survey found. The survey is just an attempt to approximate *p* without being able to actually ask every single person.
(b) The sample parameter or $\hat{p}$ is 706/1009 = 0.6997, or effectively just 0.7. So the 95% confidence interval for *p* is $\hat{p}$ +/- $z^* \sqrt{(\hat{p}(1-\hat{p})/n)}$ = 0.7 +/- $1.96\sqrt{(0.7 \times 0.3)/1009}$ = 0.7 +/- 1.96 x $\sqrt{0.21/1009}$ = 0.7 +/- 1.96 x .0144 = **0.7 +/- 0.028, or in other words *p* is from 67.2% to 72.8% with 95% certainty.**
(c) My work agrees with this margin of error, though they obviously rounded up, as my margin was 2.8%.

21.6 Gun violence
Since the margin of error changes with the square root of the sample size, I believe you would need to have four times the sample size, which would make it 4,036 people to decrease the margin of error by half.

21.12 Teens and their TV sets
(a) Using the formula from page 435 and a z^* of 1.96 for 95% confidence, we have 66% +/- 2.9% have a TV in their room, and 18% +/- 2.3% would choose Fox as their favorite network.

(b) The “19 cases out of 20” simply refers to the confidence interval of 95%, since that is 19 out of 20. And in both cases the margin is no more than 3 percentage points; one is 2.9 and the other 2.3 percentage points.

21.16 68% confidence

(a) Since the “68” part of the 68-95-99.7 rule requires only one standard deviation from the mean, the formula is even easier; you just tabulate the $\hat{p}$ +/- $\sqrt{(\hat{p} \times (1-\hat{p})) / n}$

(b) In this situation 68% confidence means that if you took many sample statistics from the same population, 68% of them would have the true value (that is, the value you are looking for that is true within the entire population, not just within your sample group) inside the range of the quoted numbers.

(c) The math is the same, you just nix the 2 that is multiplying the numbers inside the square root. So your range is 0.0636 +/- 0.0047 or between 5.89% and 6.83%. This is a much tighter spread, and thus seems more definite, but is less reliable because it only describes 68% of the results we would get if we did this survey many times.

21.19 Polling women

The z* for a 99% confidence interval is 2.58, so you follow the formula with 2.58 as the multiplier for the calculation within the square root. The proportion of women who feel they don’t get enough time for themselves is 0.47 +/- 0.04, or between 43% and 51% with a 99% confidence. The interval is larger but we have greater confidence that this spread contains the truth within it.

Chapter 22:

22.2 Students’ earnings

It means that there was a difference between the earnings of men and women that would only be found by chance in 4% of samples, or that there is only a 4% chance that the difference is attributable to anything other than an actual difference between the earnings of men and women. Whereas the difference between the earnings of black and white students is minimal, and the distribution of the data shows that the sample reaches conclusions that we could expect to see in about 48% of samples if indeed there is no difference between the earnings of the two groups.

22.8 Attending church

(a) This is one of the issues, along with race (as discussed much earlier in the book), about which we might think people feel uncomfortable answering honestly, or about which they might feel guilty or judged, and thus answer in a way that is aspirational rather than completely honest.

(b) The null hypothesis would be that 40% go to church, whereas the alternative hypothesis is that less than 40% attend. Or H_0: $p = 0.40$, H_a: $p < 0.40$.

22.17 Significance means what?

I think that the statement is meant to read about what is considered statistical significance, and sets the bar such that the P-value must be 0.05 or less before results call into question the null hypothesis. In other words, the student was close to the truth, but neglected to mention that results at that level might occur but only 5% of the time.

22.21 Teens and their TV sets

The question can be summed up as: If only 50% of the population of teens had a TV, the result *from this* sample of 66% having a TV would be how much of an anomaly?

I did the math below to show I know how, but I believe that I need the standard deviation from the sample, which data they did not give. The calculations call for using *s* from the sample data

to calculate a revised *s* that can be used to determine the significance. Without that I don't know how to get the correct answer.
So H_0: $p = 0.50$, H_0: $p > 0.50$
$\hat{p} = 692/1048 = 0.66$
Standard deviation: $\sqrt{(.5 \times .5)/1048} = 0.015$
Standard score for $\hat{p} = 0.66$: $(0.66-0.50)/0.015 = 10.67$
This doesn't entirely make sense to me, that standard score seems absurdly high. Is it because I should have used the $\hat{p}$ in the standard deviation calculation instead of *p*? If so, it would instead be:
Standard deviation: $\sqrt{(0.66 \times 0.34)/1048} = 0.0146$, which isn't very different. I think I need that *s* figure.

22.26 Mice in a maze
The null hypothesis is that the mice will complete the maze in the mean time of 19 seconds, and the alternative hypothesis is that they will complete it faster. Or in other words H_0: $\mu = 19$, H_0: $\mu < 19$

Chapter 23:

23.5 Searching for ESP
(a) Four out of 500 is less than 1%, yet the statistical significance is 1%. I think that within a group of 500, you would expect to find around five (more or less) that outperform random guessing. Finding several that beat the odds is just proof of the randomness that exists.
(b) Once the researcher has identified the people they think have ESP, they should put them through a controlled experiment. Then you can see whether the results really have significance.

23.6 Comparing package designs
The difference is statistically significant, but what is the size of the difference? If it is only slightly more, the degree of significance may be meaningless. One design is bound to be bought more than the other, but the degree to which that is due to preference can only be known when we have both the P-value and the hard numbers.

23.8 Blood types in Southeast Asia
To really make sense of the findings, we would need to know more information about the results and especially the sample size. If your sample size is only ten people, you would expect to find some diversity of blood type. Without knowing the sample size we can't really know how meaningful it is.

23.9 Why we seek significance
I think the student is close, but one thing they are missing is that there is not one "significance", but rather that the P-value tells us how likely it is that the results could be due to chance, or some other factor, rather than solely the explanation given for whatever result.

23.10 What is significance good for?
(a) No, it cannot tell us that.
(b) Yes, that is the function of significance tests.
(c) No, it doesn't really tell us that either. Combined with the results, it can tell us how confident we can be in a small effect, but not really the importance of the effect itself.

Chapter 24:

24.1 Extracurricular activities and grades

Of the 119 students surveyed who spent less than 2 hours a week on extracurricular activities, 55% got a C or better, while 45% got a D or F. For the students who spent 2 to 12 hours per week, those numbers were 74.7% C or better, 25% D or F. And among those students who spent more than 12 hours, 37.5% got a C or better, while 62.5% got a D or F. It would seem that those students who spent a moderate amount of time on extracurricular activities got better grades than those who did few or too many extracurricular activities. But this is correlative, not causative. It could be that the students who are getting the worst grades are less motivated in general, and thus fare poorly in classes and also do not have the energy for activities outside class. Or they have full-time jobs that impact the amount of time they have to devote to studying or activities.

24.2 Smoking by students and their parents
For the students whose parents both smoke, 22.5% smoke and 77.5% do not.
For the students with one smoking parent, 18.6% smoke and 81.4% do not.
For the students with nonsmoking parents, 13.9% smoke and 86.1% do not.
You can see that as the parental smoking declines, student smoking also declines.

24.8 Airline flight delays
(a) About 13.3% of all Alaska flights were delayed, whereas only about 10.9% of America West's flights were delayed.
(b) Alaska's delayed flights by airport: LAX 11.1%, PHX 5.2%, SD 8.6%, SFO 16.9%, SEA 14.2%. America West's delayed flights by airport: LAX 14.4%, PHX 7.9%, SD 14.5%, SFO 28.7%, SEA 23.3%.
(c) America West had a higher percentage of delayed flights at all the airports, and dramatically more at SFO and SEA. But they flew relatively few flights to those destinations, whereas they flew a lot of flights to Phoenix, which I presume is their hub, with a relatively low delayed flight percentage. The sheer number of flights that were not delayed there (even though the percentage of delays was greater than Alaska's, who had few flights there) overpowers the few number (but high percentage) of flights that were flown from Seattle and San Francisco.

24.9 Race and the death penalty
(a) Here is the chart:

	White Defendant	Black Defendant
Death Penalty	19	17
No death penalty	141	149

(b) 11.88 percent of total white defendants were sentenced to death, where only 10.24% of Black defendants were. But for Black defendants/white victim, that number is 17.46%, and Black defendant/black victim it is 5.83%. Meanwhile for white defendant/white victim the numbers are 12.58%, and white defendant/Black victim it is 0%.
(c) You can see that while the overall death sentence rate for white defendants is higher, they never got the death sentence for killing Black victims. Their higher death sentence rate is the result of them killing more white people, who it seems have more value in the justice system, or just have better lawyers advocating for their families. Where Black defendants were sentenced for killing Black victims, they were sentenced to death much less often than when they were convicted of killing white victims. The fact that white defendants killed more white people brings the percentage of their death sentence up. In other words, because the proportion of people killed by Black people was mostly Black people, their rates are lower, but if the Black/white proportion of people killed the same for the Black defendants, their rates would be much higher.

24.10 Majors for men and women in business

(a) Yes, two of the cells are small, but the criteria was that no more than 20% be less than 5, and all counts are 1 or greater. There are no cells less than 5 here, and even if one was it would be less than 20%. If 2 cells were less than 5 it would not be eligible for the chi-square test.

(b) It would seem that women were more likely than men by a significant margin to choose administration, and slightly more likely than men to choose accounting. Accounting: 54% female, administration: 69% female, economics: 45% female, finance: 51% female. Another way to look at it: only 2.2% of all females went into economics, whereas 3.7% of males did. But to really discover the relationship between gender and major, we would have to run the numbers. First we need to tabulate the expected counts, assuming that the genders are equally represented, which would be our null hypothesis. For accounting: there would be 72.28 females, administration 76.36 females, economics 6.4 females, finance 69.95 females. Comparing these numbers to what is in the chart, we see that aside from administration, the numbers aren't all that far from what would be expected. If we then calculate the chi-square statistic, we get $x^2 = 10.712$, which for a table with 4 rows and 2 columns you would use 3 degrees of freedom, and looking at the chart on page 494 it seems that for that degree of freedom, our results are significant at 0.05, but not quite 0.01. That shows a high degree of significance.

(c) Only 386 out of 722 students responded, so just over half. This certainly would weaken any conclusions we can draw from the data.

	Females	Males	Totals
accounting	68	56	124
administration	91	40	131
economics	5	6	11
finance	61	59	120
Total	225	161	386

	Observed		**Expected**	
	Female	Male	Female	Male
Accounting	68	56	72.3	51.7
Administration	91	40	76.4	54.6
Economics	5	6	6.4	4.6
Finance	61	59	70	50