The author and publisher of this book have used their best efforts in preparing this book. These efforts include the
development, research, and testing of the theories and programs to determine their effectiveness. The author and
publisher make no warranty of any kind, expressed or implied, with regard to these programs or the documentation
contained in this book. The author and publisher shall not be liable in any event for incidental or consequential
damages in connection with, or arising out of, the furnishing, performance, or use of these programs.

Reproduced by Pearson from electronic files supplied by the author.

Copyright © 2013 Pearson Education, Inc.
Publishing as Pearson, 75 Arlington Street, Boston, MA 02116.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any
form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written
permission of the publisher. Printed in the United States of America.

Video 1: What is an average?

Summary
In this video, we introduce the median and mean by asking, “Do you have more than the average number of feet?” The example given is a sample of pedestrians, one of whom only has one foot due to cancer surgery. As a result, the mean is less than the median. We show the process of calculating a median – by lining up the sample from smallest to tallest – and of a mean – the arithmetic calculations – and emphasize that while someone typically has the median level of some characteristic (e.g., 2 feet), often no-one has the mean level (e.g., 1.89 feet). This demonstrates the difference between a calculated parameter (such as a mean) and one that can be directly observed (such as the median). The video ends by discussing medical risk, arguing that because median risk is generally below mean risk (often, a small number of individuals are at greatly increased risk), many of us get treatments or tests we don’t need.

Use
This video fits in well in the introductory part of any statistics course because a) it covers basic subjects; b) it takes something that seems quite trivial, means and medians, and then connects them to a concrete application, overdiagnosis and overtreatment in the medicine. The video can be accompanied by Chapters 2 and 5 in What is a p-value anyway?

Class discussion questions
1. Does someone always have the median level of some characteristic? This question can provide the opportunity to bring in the concept of interpolation.
2. Are all averages means or medians? You can cite the mode here – though few statisticians use it – or look at the concept of batting average in baseball, which isn’t an average at all but a proportion.
3. What other examples can you think of where most people have more or less than the average? This is true of most things with a non-symmetric distribution (e.g., weight, math scores, marathon times) but it is nice to continue the theme of the video in terms of risk (e.g., most have below average risk of a automobile accident, death by violence, or even, say, getting a date).
4. In the video, the statistician says that scientific statements have to be very precise and we have to think hard about whether they are true. This can lead to discussions about:
   a. How precise is “very” precise? Should we say e.g., that the mean height of US men is 5’10” or would it be more “scientific” to say that it is 5’10.23416”?
   b. How do scientists “think very hard about whether a statement is true?” This is a good opportunity to talk about science in terms of critically evaluating data and methods.
Video 2: When should you use a mean and when should you use a median?

Summary
This video focuses more on when to use a mean and when to use a median. House prices are used to demonstrate that when data are non-symmetric – especially when there are extreme outliers – the median gives a better description of a typical value than the mean. Specifically, the prices of properties on two blocks are compared: in one, all houses are similar and there isn’t much difference between the median and mean; in the other, there is a big expensive block of apartments, so that the mean is nearly twice the median, and far from the cost of any individual property.

But we want to get away from the idea that the data, and only the data, drives the choice of descriptive statistic. The example is given that, if you wanted to buy all the houses in Brooklyn, if you took the median, and multiplied by the number of houses, you wouldn’t have enough cash. So the median is a useful descriptive statistic, but the mean is essential for planning and making decisions.

Use
This video fits well in the early part of a statistics course and is a great introduction to the idea that statistics are used for some purpose, and it is the purpose that drives the choice of statistic. The video can be accompanied by chapters 2, 4 and 5 in What is a p-value anyway?

Class discussion questions
1. Should you use the median or mean to describe a data set if the data are not skewed? This can be a pointer to standard deviation vs. interquartile range.

2. You may read in the newspaper that a study of a new drug for cancer “increased survival by an average of 8 weeks.” It turns out that this is a median, and it is used for complicated statistical reasons. But in a perfect world, would you prefer to know the increase in mean or median survival?

3. If the median house price is $1.9m, does that necessarily mean that half of the houses on the block are worth less than $1.9m and half worth more? This is a good opportunity to bring up the problem of ties.
Video 3: Sampling

Summary
This video discusses sampling in the context of how estimates of population parameters are obtained. It refers to Video 1 (“What is an average?”) where we obtained an “average of 1.89 feet per person.” It points out that applying this statistic depends on thinking through whom the population is meant to be, and that depends on the study question (“If you want to understand your answer, you really have to work out carefully what your question is.”).

As the sample statistic was derived from a bunch of kids heading to the playground, plus a one-legged man who we asked to show up, we conclude that the sample was a bad one. We explain that random sampling is a generally a good way of obtaining a representative sample such that you can be confident that the sample statistic is a good estimate of the population parameter.

Use
The video can be used to introduce the basic concepts of sample statistics and population parameters, and also as a first foray into the issue of study design.

Class discussion questions
1. Are all good samples random? This is an opportunity to bring up opinion polling, which typically tries to obtain views from particular groups (men, women, older, younger, employed, unemployed, Democrat, Republican etc. etc.) and then “weights” the results by the prevalence in the population.

2. Magazines often report surveys giving statistics such as “63% of women expect the man to pay on the first date.” Are these random samples? These surveys are most definitely not random – they are typically click-through from the magazine website – and so can provide an opportunity to discuss the sort of biases that can result from lack of random sampling.

3. Discuss the Literary Digest survey described in Chapter 4 of What is a p-value anyway?
Video 4: Variation 1

Summary
The video begins at the park, with cyclists and joggers going by. We show a very slow old woman going by on a bike, and then a bunch of racing cyclists. We point out that sometimes, what is interesting about a data set is not its average but how much it varies. We then discuss the weather in New York and San Francisco, which have pretty much the same average annual temperature, even though New York has hot summers and cold winters.

Quartiles as a measure of variation are introduced by way of the price of food on take out menus. The video ends with a practical application in medical research, where mean exposure to a toxin is far less interesting than the fact that a small number of individuals are exposed to very high levels.

Use
The video provides a good introduction to the importance of presenting measures of dispersion as well as central estimates. The video can be accompanied by Chapters 3 and 5 in What is a p-value anyway?

Class discussion questions
1. Why is a quartile called a quartile? What is a centile? What is a quantile?

2. What are some examples, other than temperature, where similar averages can be associated with very different distributions? A few thoughts: costs (e.g., cost of buying a song online is the same average cost of driving above the speed limit, assuming that you are only caught speeding occasionally); ERA of pitchers (some are very consistent, some are sometimes brilliant and other times horrible); success rates in surgery (do we want an operation that most surgeons can do pretty well or one in which a few surgeons are nearly perfect and some have very poor results?)

3. Give some practical uses of knowing variation. A few thoughts: what clothes do I need to pack for a trip? Doctors need to know distributions of blood values to know whether a patient is out of range. Industrial engineers need to know distributions, for example the strength of a certain part to see if there is a problem with a manufacturing machine. Clothing manufacturers need to know the distribution of sizes, for example children’s clothes for a certain age.

4. For many years, the New York subway had no air conditioning on the grounds that the average trip was only 15 minutes, and 15 minutes without air conditioning is no hardship, even in the New York summer. Critique this reasoning.
Video 5: Variation 2 (and egg roulette)

Summary
The video starts by comparing means and medians with respect to decision-making. Examples are given, such as putting a seat belt on a child, where decisions based on the median (e.g., waste time putting on the seat belt and don’t have an accident) are inferior to decisions based on the mean.

The standard deviation is introduced as a calculation that allows you to calculate any quantile that you want (e.g., about two-thirds of observations are within one standard deviation of the mean).

Use
This video, along with the Video 4 (Variation 1), can be used when introducing measures of dispersion. The first half is useful when discussing decision analysis, or the application of statistics. The video can be accompanied by Chapters 3 and 21 in What is a p-value anyway?

Class discussion questions
1. Is it true that two-thirds of observations are within one standard deviation of the mean? What about the rule of thumb that 95% of observations are within 2 standard deviations of the mean? This is an opportunity to discuss look up tables for quantiles of the normal distribution.

2. Should you play the lottery? This is a reverse case of the median and mean: the median is to lose; the mean is better, but you still lose. On the other hand, buying a lottery ticket does give you the opportunity to dream about what you’d do with all the money: some economists say that that is why people pay to play the lottery.

3. Is it better to present means and standard deviations or median and interquartile range?
Video 6: The normal distribution

Summary
This video explains the normal distribution via the binomial distribution: the distribution of the number of heads thrown on 20 coins approximates the normal. This is used to explain that the normal distribution is the mathematical consequence of adding up a large number of random events. Some examples are given of normal distributions in the natural world (e.g., mass of ants) and social world (age of marathon runners) and explained in terms of these phenomena resulting from the aggregation of random events.

Use
This video should be assigned whenever the normal distribution is first raised in a course. The video can be accompanied by Chapters 7 and 9 in What is a p-value anyway?

Class discussion questions
1. What is the link between the normal distribution and Video 5 (Variation 2 (and egg roulette)), where we relied on statements such as “about two-thirds of observations are within 1 standard deviation of the mean”? Point out that relationship between quantiles and standard deviations assumes that the distribution of the data does not stray too far from the normal.

2. Do natural phenomena such as hemoglobin levels or the weight of ants really follow a normal distribution? Point out that no data set can ever perfectly match a theoretical distribution.

3. If you add up a large number of random events, you get a normal distribution. How large a number makes a normal distribution? This is an opportunity to bring up asymptotics and approximation.

4. Are the results of studies due to adding up a large number of random events? An opportunity to introduce the idea that the results of studies, if repeated, follow distributions that tend towards the normal and therefore that distributions like the normal distribution are essential for interpreting the results of studies.
Video 7: Not the normal distribution

Summary
The presenter tells a story about when he was first learning statistical analysis of medical research and his lecturer became very excited when he saw a normal distribution (which is how he learned that sometimes, the normal distribution is not very normal at all). Math ability is then used as an example of a non-normal, skewed distribution. This is explained in terms of the data resulting from a process that involves multiplication, with the growth of cancer being an example.

Use
This video can be used at any part of a course discussing data distributions. The video can be accompanied by Chapter 8 in What is a p-value anyway?

Class discussion questions
1. Is the normal distribution used at all to analyze data such as math ability or cancer growth? Log transformation and log normal data can be introduced here.

2. The results of studies are often analyzed using the normal distribution, or distributions close to the normal. Are these applicable where the data are not normal? Central limit theorem can be introduced, explicitly or informally.

3. Think about some characteristics of people that are and are not normally distributed. For example, most things to do with preference or ability are non-normally distributed; biological characteristics such as height, weight (in younger people at least), hemoglobin, albumin etc. tend to have a more normal distribution.
Video 8: Sampling and parameters

Summary
The video demonstrates that repeated samples from a population (M&M’s sampled from a bowl of them) give slightly different estimates. This is used to introduce the concept of confidence intervals and the idea that statistical formulae can be used in place of repeated sampling

Use
The video can be used when standard error is introduced. The video can be accompanied by Chapters 10 and 11 in What is a p-value anyway?

Class discussion questions
1. In the video, the student sampled 50 M&M’s at a time. What effect would it have had if she had only sampled 20? What about 100 or 200? This is an opportunity to introduce the link between sample size and standard error.

2. In real studies, do the data ever get eaten? Well the answer is yes, you often get missing data or other problems with data sets.

3. What would have happened if, instead of counting the number of brown M&M’s, the student had weighed them? Confidence intervals apply to continuous variables as well as binary variables.

4. Why did Van Halen insist on having no brown M&Ms? It turns out that they weren’t being silly, precious rock stars. When they got to the dressing room, they’d check the M&Ms, and if there were any brown ones, they’d know that the venue hadn’t read their contract carefully and that they needed to check the rigging, electricals, lighting etc. etc. The M&Ms were the “canary in the coal mine.”
**Video 9: Why use a p-value anyway?**

**Summary**
The video introduces p-values by imagining that the presenter has to choose whether to cycle down a busy road or along the back streets. The presenter also has a data set where he had recorded the journey times on a few occasions. He analyzes the data set, and then also a larger one after taking more data, to obtain p-values, but in both cases ignores the conclusions, taking the main road when the p-value is high (because he had to rush home from an appointment), but the back streets when the p-value is low (because it is more pleasant to be on the back streets). The point of the video is to emphasize that what p-values do is to test hypotheses, but our decisions are not determined by the results of hypothesis tests.

**Use**
This video should be used when p-values and hypothesis testing are first introduced. The video can be accompanied by Chapters 1 and 13 in *What is a p-value anyway?*

**Class discussion questions**
1. In the video, the scientist focuses pretty much exclusively on the p-value. Is this really what scientists do? Should they? *This can form the basis for discussing the fact that many scientists and other non-statisticians who work with statisticians often seem to think that statistics is exclusively about p-values and that statisticians are simply machines for producing p-values.*

2. The statistician ended up ignoring the p-value and focusing on how long it would take to get home on average following each route. This reflects two different sides of statistics. What are they called? *An opportunity to contrast inference with estimation.*

3. Why bother with inference at all? Why not just focus on estimation? *Discussion point 1 of Chapter 12 of What is a p-value anyway? can also be introduced here.*
Video 10: What does a p-value mean?

Summary
The video gives the definition of a p-value by telling a story about a toothbrush: a mother comes home and asks her son whether he has brushed his teeth; the son says “yes”; the mother goes to the bathroom, finds that the toothbrush is dry and concludes that the son has not brushed his teeth. So the p-value is the probability of a dry toothbrush. The hypothesis is that the son has brushed his teeth, and the data are the bristles of the dry toothbrush; the data would be unlikely if the hypothesis were true, so we reject the hypothesis. Hence, the p-value is the probability of the data, or something more extreme, if the null hypothesis were true.

Use
This video should be used when p-values and hypothesis testing are first introduced. The video can be accompanied by chapter 14 in What is a p-value anyway?

Class discussion questions
1. Scientists often misinterpret the p-value as the probability of the hypothesis. For example, they’ll say that “the p-value is the probability that the null hypothesis is true.” Why do you think that they make this mistake? An opportunity to discuss the fact that the probability of the hypothesis is what you really want to know about. Statistics (unless you go Bayesian) can’t tell you the probability of the hypothesis.

2. Could the child in the video have brushed his teeth? Of course, he could have dried the toothbrush with a hair dryer, or it could be a particularly hot and dry day. But this is very unlikely. This is an opportunity to discuss the way that statistics gives probabilities and then rules of thumb based on those probabilities (e.g., \( p < 0.05 \), reject the null). It doesn’t offer “proof.” Chapter 13 discussion question 1 from What is a p-value anyway? is a good reference on this point.

3. If the toothbrush was wet, does that mean the child did brush his teeth? No. If all Spaniards are Europeans, this does not mean that a European you meet must be Spanish. Similarly, if the probability of the data would be high if the hypothesis were true, this doesn’t mean that the hypothesis is true given the data. It is for this reason that we don’t accept hypotheses, we just reject or fail to reject.
Video 11: The p-value is the probability of the data, not the hypothesis

Summary
The video emphasizes that the p in “p-value” refers to the data, not the hypothesis, in two ways. First, it is pointed out that whilst you can calculate the probability of rolling a double-six on two fair dice, there is no easy way to calculate whether the dice were fair, given the fact that you just rolled two sixes. Second, an example is given of implausible cancer treatment. It would remain implausible even if a study was published and the p-value was low.

Use
This video should be used when p-values and hypothesis testing are first introduced. The video can be accompanied by Chapter 14 in What is a p-value anyway?

Class discussion questions
1. Can you ever use statistics to work out the probability of a hypothesis? Yes, if you specify a prior probability and use a Bayesian approach.

2. Has there been a study showing that a cancer zapper works? This is a nice opportunity to point out that many crazy cancer therapies have actually been tested, and they have generally been found NOT to work. A reference that is freely available online is Vickers AJ. Alternative cancer cures: "unproven" or "disproven"? CA: A Cancer Journal for Clinicians. 2004;54(2):110-8. It is a good example of how statistics can be used to help people.

3. Does that mean you can continue to believe the null hypothesis even if you have rejected it? Here’s an opportunity to compare the results of a study with how your beliefs change.
Video 12: What is statistical significance?

Summary
The video first points out that what statistical significance means, and all that it means, is that the p-value from a study is less than alpha (which has to be pre-specified). It goes on to point out that statistical significance does not mean significant in other ways. The example is given for typing speed: even if the p-value comparing the typing speed of men and women was statistically significant, you may not change an opinion you’d have or decision you’d make, if the differences between genders were small.

Use
This video should be used when p-values and hypothesis testing are first introduced. The video can be accompanied by Chapter 13, 14 and 15 in What is a p-value anyway?

Class discussion questions
1. In Chapter 27 of What is a p-value anyway?, discussion question 3, introduces the “Prosecutor’s fallacy.” As the video introduces court cases, this might be an interesting opportunity to discuss this statistical error.

2. What should alpha be?

3. Let’s assume you do a study with an alpha of 0.05. Is there a big difference if the p-value is 0.051 vs. 0.049?
Video 13: Basketball players won’t accept the null hypothesis

Summary
In the video, the presenter, who has never played basketball before, gets creamed by a kid called Ernie who is out there on the courts every weekend. However, the difference in score is not statistically significant. Other players still would prefer Ernie on their team. They are correctly refusing to accept the null hypothesis.

Use
This video should be used when p-values and hypothesis testing are first introduced. The video can be accompanied by Chapter 15 in What is a p-value anyway?

Class discussion questions
1. The basketball player was much better than the presenter was. How come there was no statistically significance difference between them?

2. Demonstrate that the p-value is actually 0.07. This can be an opportunity to do calculations by hand i.e. $2 \times (2^9 + 8 \times 2^8)$ as an introduction to binomial distribution.

3. Imagine that Ernie and the statistician were going to play a long game, best out of a 100 or so. What analysis could you to predict what the final score might be? This is a chance to introduce the confidence interval.
Video 14: The fish and chip guy won’t accept the null hypothesis

Summary
The problem of accepting the null hypothesis is illustrated by a famous study finding that women eating a low-fat diet do have lower rates of breast cancer, but with a p-value of 0.07. This was generally interpreted as “low-fat diets don’t work,” a case of accepting the null hypothesis.

The presenter then discusses accepting the null more broadly, pointing out that the typical claim that there is “no evidence” for something, is often made when no studies have been done.

Use
This video should be used when p-values and hypothesis testing are first introduced. The video can be accompanied by Chapter 15 in What is a p-value anyway?

Class discussion questions
1. Give some examples of arguments that are a type of reflecting the null hypothesis. Some possible examples: “I have no reason to doubt his integrity” = “I have never examined his work closely enough to know”; “The safe level of alcohol to drink during pregnancy is not known, so the only safe approach is not to drink alcohol at all”; “There is no evidence that Governor Smith’s tax policies will have any effect on job creation” = “We haven’t passed the new tax laws and waited a couple of years to find out their effects.”

2. What should we conclude from the breast cancer study? This is explored in discussion question 3 in chapter 15 of What is a p-value anyway?

3. Why do you think people do commonly accept the null hypothesis? Discuss that “failing to reject” is somewhat unsatisfying as a conclusion. People like certainty, even when there isn’t any.
Video 15: Two types of variation

Summary
The video introduces the difference between natural and statistical variation. The presenter conducts a study on hair length, presenting the results in the form of a histogram. This shows variation in hair length between individuals (i.e. standard deviation, although this term is not used in the video). The study is then repeated and it is shown that the result of the study, mean hair length, also varies (i.e. standard error, although again the term is not presented). The point is made that there is variation you can see (men’s hair length really varies) and other variation that is theoretical (the variation of research results were a study to be repeated).

Use
This video should be used in classes introducing estimation. The video can be accompanied by Chapter 10 in What is a p-value anyway?

Class discussion questions
1. The study question, as described by the presenter, is to find out the average length of man’s hair in Brooklyn. To do this, they measure hair length at the farmers’ market. Is this a good or a bad idea? *Men shopping at the farmers’ market may not be representative of Brooklyn men.*

2. The presenter shows the variation of mean hair length. Would you see something similar if you plotted medians? *Yes, in theory, but there is no easy formula to work out how medians would vary.*

3. How did the presenter choose which strand of hair to measure? *Discuss the fact that scientific research follows a specific protocol that specifies these sorts of points very precisely.*
Video 16: Standard error and standard deviation

Summary
This is a continuation of Video 15, but now introduces the terms “standard error” and “standard deviation” explicitly, showing the relationship between the two in terms of the square root of the sample size. A survey question “Do the New York Yankees stink?” with a binary response is then used to show that you can have a standard error even when there is no standard deviation. Specifically, because the results of the survey study would follow a distribution close to the normal if repeated a large number of times, the distribution can be used to assess how much faith one should put in the results of any one particular study.

Use
This video should be used in classes introducing estimation. The video can be accompanied by Chapters 11 and 12 in What is a p-value anyway?

Class discussion questions
1. If you repeated the survey about the New York Yankees a large number of times, would the results really follow a normal distribution? This is a chance to state that the results would follow the binomial, but that this approximates to the normal when samples sizes are large.

2. The presenter says that studies looking for small differences, such as in cancer survival, need large sample sizes. How large is “large enough”? A chance to introduce the general concept of a sample size calculation and the role that statisticians play in research design.

3. In the video, some of the interviewees responded “don’t know” to the question about the Yankees. Yet the results of the study were presented in terms of “63% say the Yankees stink and 37% answered no.” How were respondents answering “don’t know” handled? Point out that there are various options. They could be assumed as responding “no” (i.e. they didn’t say that the Yankees do stink), they could be excluded from the analysis (i.e. sample until you have 100 people who give you an answer one way or another) or the results could be presented as e.g., 50% said “yes”, 34% “no” and 16% “don’t know.” The key issue is that there is no correct approach other than the fact that this issue has to be thought about in advance, a careful decision made and justified and then the method to be used pre-specified in a study protocol.
**Video 17: How does statistical testing work?**

**Summary**
The video introduces hypothesis testing via a permutation test. The presenter sets out to compare apples and oranges (he ends up comparing apples and peaches) by weighing each piece of fruit and writing the weight on a post-it note that is then attached to the fruit. He conducts a permutation test by removing the post-it notes, shuffling them and re-attaching them to randomly selected pieces of fruit. This demonstrates that the difference in weights follows a distribution close to the normal, with a mean of zero. The experiment is then repeated but instead of the difference in weights being recorded each time, the $t$ statistic is calculated. The presenter points out that there is no need to calculate the $t$ distribution by hand, repeating an experiment many times, because the $t$ distribution has been worked out mathematically. The video concludes by generalizing the experiment to many forms of statistical testing: the statistician calculates a test statistic, for which the distribution under the null is known.

**Use**
This video should be used in classes introducing inference. The video can be accompanied by Chapter 16 in *What is a p-value anyway?*

**Class discussion questions**
1. Does the difference in mean weights follow a normal distribution? *An opportunity to say that this is only approximately true, that the $t$ distribution is close to, but not exactly normal.*

2. Why do we do $t$ tests instead of getting a computer to do the random sampling experiment? *The $t$ test was invented long before the availability of computers. Moreover, even with modern computers, permutation tests can take a long time.*

3. Do statisticians ever do the experiment mixing up the results of a study? *Yes, in complicated situations when there is no obvious test statistic.*
Video 18: Regression

Summary
The video introduces regression in terms of an experiment as to how far the presenter can walk down the block before being hit by a rubber chicken. The presenter introduces the idea that regression is a prediction and that regression can be represented as a simple equation. Multivariable regression is introduced with a marathon-running example: the more a runner trains, the faster he or she will complete the marathon, but there are other factors, such as age, that also influence running times.

Use
This video should be used in classes introducing regression. The video can be accompanied by Chapter 18 in What is a p-value anyway?

Class discussion questions
1. The presenter describes regression as a particularly powerful and useful technique. Is this true, or was it just the sort of general chit-chat used to introduce videos? An opportunity to say that for many statisticians, regression consists a large part of their everyday work.

2. Why do you think there is such an emphasis on prediction? Some statisticians say that statistics is primarily about prediction. As a simple example, a t-test to compare pain scores in a clinical trial of a drug can be thought of in terms of a prediction as to whether a patient seen by a doctor in the future would do better on the drug than on control.

3. In reality, no one tries to predict marathon times in terms of age, gender and training miles. What would you use instead? See discussion question 1 in Chapter 18 of What is a p-value anyway?
Video 19: Confounding

Summary
The video discusses the use of regression to adjust for confounding via two simple examples. In the first, height is a confounder in the question of whether men or women have better jumping ability. It is explained that regression is a mathematical alternative to trying to find men and women who are the same height and comparing how high they can jump. In the second example, smoking is a confounder of the association between coffee drinking and cancer.

Use
This video should be used in classes introducing regression. The video can be accompanied by Chapter 19 in What is a p-value anyway?

Class discussion questions
1. Imagine that in the coffee example, the researchers had found a statistically significant regression coefficient for coffee, even after adjusting for smoking. Could you conclude that coffee causes cancer? An opportunity to discuss correlation vs. causation.

2. Given the choice, would you rather find men and women who are the same height and have them jump, or would you be better off using regression? Experiment is always better than observation! Regression can help when the experiment would be too expensive or take too long.

3. What else might you want to include in the regression for coffee, smoking and cancer? A chance to discuss that, often, many variables are entered into a regression. Also an opportunity to discuss whether including a variable such as diet in a regression really does control for diet as a confounder. See chapter 19 of What is a p-value anyway?, particularly the section entitled "Roll up for the magic show!"
Video 20: Extrapolation

Summary
The video gives various examples of inappropriate extrapolation of regression lines, such as assuming that a toddler could run a 5k in 35 minutes based on a regression of times from high school runners, or that a salsa made of mashed garlic and jalapenos would score 20 out of 10 on a taste test.

Use
This video should be used in classes on regression. The video can be accompanied by Chapter 25 in *What is a p-value anyway?*

Class discussion questions
1. Give some examples of relationships between two variables that follow curves. *Answer: pretty much everything!* Indeed, the difficult thing is finding relationships that are linear. We generally only see linear relationships because we are looking at only a small range of possible values.

2. In the video, the presenter says “you can’t score 20 out of 10” and the student answers, “the regression line says you can.” Who is right? Well obviously the presenter. In the case of statistics, you have to be very careful to make sure that math follows reality. But note that this isn’t necessarily the case in other fields. For example, the Higgs boson was predicted many decades ago because it makes some math come out right.

3. Think of some examples of inappropriate extrapolations. *Some possibilities: predicting future trends, particularly with respect to human performance, such as world record times OR exercise studies done on a small number of athletes said to be the “new way for everyone to exercise.” The latter examples can help students think more generally about applying the results of studies and the problem of extrapolation.*
Video 21: Absolute vs. relative risk

Summary
The video uses examples from shopping to explain the difference between absolute and relative risk and why it is that absolute risk should be used for decision making. Much of the video concerns whether to buy an expensive bike lock that halves the risk that your bike will be stolen. The presenter shows that what you really need to know is not the relative risk – i.e. that risk will be halved – but the absolute risk of theft with the expensive bike lock compared to a standard model.

Use
This video can be introduced towards the end of an introductory class on statistics, when discussing how to use statistics in everyday life or to make practical decisions. The video can be accompanied by Chapter 21 in What is a p-value anyway?

Class discussion questions
1. If absolute risk is so much more useful than relative risk, why do you think scientists and newspaper reporters focus so much on relative risk? Perhaps because it sounds more exciting and dramatic to say something “doubles your risk of death” than “increases risk of death by 0.1%.”

2. Do a decision analysis to see whether you are better off betting on a number in roulette (1 in 38 chance of winning, $1 bet gets you $36) compared to betting on red vs. black (18 in 38 chance of winning, $1 bet gets you $2).

3. Statisticians involved in medical research often do an analysis like that conducted for the bike lock to help make decisions such as whether a woman should have chemotherapy for breast cancer. In the bike lock example both the decision (which bike lock to buy) and the consequence (having a bike stolen) are financial and can be represented in dollars. For breast cancer, neither the decision (going through months of chemotherapy) nor the consequence (avoiding recurrence) are really about money. How do you think these analyses are done? The statistician has to use the same scale for both, such as quality adjusted life years or simply assuming that “a doctor would treat 100 patients to prevent one recurrence, so a recurrence is worth 100 treatments.”
Video 22: A result is not a conclusion

Summary
The video introduces the general form of a scientific experiment: Introduction, Hypothesis, Methods, Results, Conclusion. The presenter emphasizes that: a) each part of the experiment flows logically from the previous one (e.g., the hypothesis determines the methods); b) a conclusion has to be something practical, not just a restatement of the results.

Use
This video can be introduced towards the end of an introductory class on statistics, when discussing how statistics is used in the scientific process. The video can be accompanied by Chapters 1 and 33 in What is a p-value anyway?

Class discussion questions
1. The presenter gives the hypothesis of the study as “toiletries are no more expensive at the local pharmacy than at the national chain.” Normally, we think about statistical hypotheses in terms of the null hypothesis, and this is often described in terms of “no difference between the local pharmacy and the national chain.” Is the presenter wrong to talk about “no more expensive”? An opportunity to discuss one- vs. two-sided testing. This is a good example of when to use one-sided testing because your action (shopping locally) would be the same whether there is no difference in price or the local store is cheaper.

2. Imagine you do a study and find pretty good evidence in favor of a hypothesis, but not enough to be sure. How might you phrase your conclusion instead of saying something such as that the hypothesis “may” be true?

3. Does a scientific study always have a hypothesis? An opportunity to make a distinction between inference and estimation. Some studies aim only to make better estimates. Still there is a scientific question there, the introduction has to make the case as to why the question is important, the methods have to flow logically from the question and so on.
Video 23: Multiple testing

Summary
The video gives various examples of multiple testing, including whether some passersby are going to give money to a street musician, astrology, and genetic testing. The key point is that “if you keep on asking questions over and over again, you’ll eventually get a yes answer to at least some of them.”

Use
This video can be shown for classes on inference, although it is suggested that students should have a basic grounding in hypothesis testing. The video can be accompanied by Chapter 28 in *What is a p-value anyway?*

Class discussion questions
1. The presenter states, for a large number of unrelated null hypotheses that are true, the null will be rejected one in 20 times. What is the probability that if you test 20 true null hypotheses, you will have at least one p-value less than 5%? *This is a good way of showing that probabilities for independent events multiply rather than add* (it isn’t true that you’d definitely reject one null hypothesis if you tested 20 that were true). You can also show that the answer approximates $1 – 1/e$ on the grounds that the probability of getting zero successes where in $n$ attempts when the probability of success is $n$ is $(1-1/n)^n$, i.e. the derivation of $e$. See also discussion question 2 in Chapter 8 of *What is a p-value anyway?*

2. The presenter gives an example of some scientists doing a study in which a protein is not found to be associated with cancer recurrence. The scientists then start asking about the protein in different subgroups of patients. Assuming that the protein has absolutely no role in cancer, what is the probability that the scientists will come up with at least one statistically significant result if they look at subgroups defined by age (old, young), gender (male, female) and time to diagnosis (short, long)? *First, you have to note that there are actually 26 subgroups. There are 8 subgroups defined by all three characteristics (e.g., old, males, recent diagnosis; young, female, recent diagnosis); 12 by two characteristics (e.g., old males; females with short time to diagnosis); and 6 by one characteristic (i.e. old; young; male; female; short; long). However, the probability of rejecting at least one null is not $1 – 0.95^{26}$ because the null hypotheses are not independent. If you know that there is no difference in older people, there is probably no difference in males, because some men are older. This is a good example of non-independence in hypothesis testing.*

3. Do geneticists really testing thousands of hypotheses? *Yes, but they do it in a special way to reduce the problems of multiple testing, such as by applying a correction factor to the p-value.*
Video 24: Statistics is about people, even if you can’t see the tears

Summary
The video makes the point that though nobody wants to think that they are a statistic, acting as if you are a statistic can help you make better decisions. That said, statisticians should never forget that the numbers they analyze correspond to real people, who have friends, relatives and stories to tell.

Use
This video can be shown at the end of introductory classes on statistics. The video can be accompanied by Chapter 34 in What is a p-value anyway?

Class discussion questions
1. Why do you think that people often feel that “the statistics don’t apply to me”?
2. The reason why so many of us now live long, healthy lives is due to statistical analysis of health data. What other statistical analyses have had a large impact on how we live our lives?
3. Why do you think statistics often has a bad name? An opportunity to contrast “statistics”, the name of an academic field of study, with “statistics” as the plural of “statistic.” Many people misuse use statistics, in the second sense, by quoting numbers in a misleading way. Statistics in the first sense, the academic discipline, can actually help identify and correct this sort of thing. See discussion point 4 in Chapter 1 of What is a p-value anyway?