**What is a P-Value?**

**Intro**

* **Chapter 1: I tell a friend that my job is more fun than you’d think: What is statistics?**
  + Statistics involves estimation, inference, and study design
  + Estimation is about trying to work out how large or small something is
  + Inference is about drawing conclusions, usually by conducting a statistical test of a hypothesis
  + A hypothesis is a statement about the world that could be tested to see whether is it true or false
  + Many studies produce numbers; as experts in numbers, statisticians often have a lot to say about how exactly a study should be designed

**Describing Data**

* **Chapter 2: So Bill Gates walks into a diner: On means and medians**
  + What most people call an “average” is what statisticians call a mean. To calculate a mean, think of your data as a list of numbers, add up all the numbers and then divide by the number of items on the list
  + The median is the halfway point of your list of numbers: half of the sample have values higher than the median and half of the values have lower than the median
  + An outlier is when you have an observation that doesn’t follow the pattern of the data
  + When you have outliers, the median often gives a fairer reflection of the data than the mean
  + Generally speaking, means are better than medians for planning and making decisions
* **Chapter 3: Bill Gates goes back to the diner: Standard deviation and interquartile range**
  + Means and medians are useful for describing a data set. Means and medians are types of average, or central tendency
  + You generally want to know not only the average of a data set but how much the data vary around the average: a measure of spread
  + The measure of spread normally reporting with a mean is the standard deviation
  + The measure of spread normally reporting with a median is the interquartile range
  + If data follow something close to a normal distribution, the mean and standard deviation can be used to work out all sorts of things about your data, but you have to do some calculations
  + The median and interquartile range give quick information about the data without the need for an calculation
  + The median and interquartile range are also useful for describing data that are skewed
  + Statistics used to describe a data set, means and median, standard deviations, and interquartile ranges, are known as descriptive statistics
* **Chapter 4: A skewed shot, a biased referee**
  + Skewness describes the distribution of data
  + Data are skewed if there are more observations below the mean than above, or vice versa
  + The greater the proportion of observations above or below the mean, the more skewness you have
  + Bias describes a problem with the design, conduct, or analysis of the study
  + A study is biased if the methods of statistical analyses, cause an estimate to be too high or too low
* **Chapter 5: You can’t have 2.6 children: On different types of data**
  + Statisticians sometimes calculate numbers from a data set, the mean and standard deviation being good examples
  + These numbers often take values that don’t occur on the data set
  + Statisticians sometimes choose particular numbers from the data as being illustrative, such as median and quartiles
  + Means, medians, standard deviations, and interquartile ranges are used to describe variables that can take a large number of different values. These are known as continuous or quantitative variables
  + Variables are sometimes best described in terms of categories: sometimes means and medians aren’t used and the statistician just gives the number and percentage in each category
* **Chapter 6: Why your high school math teacher was right: How to draw a graph**
  + Often in research, we want to understand one thing (such as playing the lottery) in terms of another (such as age)
  + The thing we want to understand can be called ‘y’, and the thing we use to understand ‘y’ is called ‘x’
  + Drawing a graph of y and x, is a good way of understanding the relationship between them
  + You can mark a point and the x and y of each observation - called a scatterplot
  + You can also draw a line or curve closest to each point - best fit

**Data Distributions**

* **Chapter 7: Chutes-and-ladders and serum hemoglobin levels: Thoughts on the normal distributions**
  + The normal distribution is what you get when you add up a large number of random events.
  + It also describes variation of many natural phenomena, such as hemoglobin levels
  + It also describes the variation in results of a study, were we to repeat it many times
* **Chapter 8: If the normal distribution is so normal, how come my data never are?**
  + Despite being called “normal”, you rarely see a normal distribution in some areas of statistics
  + Normal distributions are rare when the data come not from the whole population, but from a special samples, such as medical patients
  + The normal distribution results from the addition of numerous random events
  + Many phenomena result from the multiplication of random events
  + Logarithms change multiplication into addition
  + Non-normal data can sometimes be converted to a normal distribution by using the logarithm of the data
* **Chapter 9: But I like that sweater: What amount of fit is a “good enough” fit?**
  + Many statistical procedures are based on the assumption that the data are normally distributed
  + These procedures include the use of the mean to describe a data set and hypothesis tests such as the t-test, but there are many others
  + There are no clear rules for determining whether a data set is close enough to the normal distribution to make it reasonable to use statistical procedures that assume normality. Ultimately, it is a judgment call on whether the method provides a good enough approximation

**Variation of Study Results: Confidence Intervals**

* **Chapter 10: Long hair: A standard error of the older male**
  + There are two sorts of variation: one is natural variation that can be observed and the other is when a study is repeated
  + Standard deviation is used to describe the natural variation of something you can measure
  + Standard error is used to describe the variation of study results - a statistic such as a mean or proportion calculated from study data - imagining hypothetically you were to repeat a study many times
  + The variation of study results often follows a normal distribution, even if the data from each individual study are non-normal
* **Chapter 11: How to avoid a rainy wedding: Variation and confidence intervals**
  + If we take a sample of individuals, 95% will have values within 2 standard deviations of the mean. This is called the reference range
  + If we repeated a study a large number of times, 95% of the estimates - the mean BMI of students, or the difference in proportions between two types of surgery - would be within 2 standard errors of the true mean; this is called the confidence interval
  + 95% of 95% confidence intervals will include the true value of an estimate. The true value of an estimate is called a parameter
  + Reference ranges are only used for some very specific purposes, such as identifying patients with blood values suggestive of disease. As such, reference ranges are not usually reported in scientific papers
  + Confidence intervals are a useful way of thinking what the results of a study might plausibly be, were you to repeat it
* **Chapter 12: Statistical ties, and why you shouldn’t wear one: More on confidence intervals**
  + You may hear the results of opinion polls described in terms of a margin of error. This is a concept similar to a confidence interval
  + The idea of a margin of error makes it sound as though the true result could be absolutely anywhere within the confidence interval. In fact, is it more like;y to be nearer the middle of the confidence interval than at either extreme
  + A statistical tie means the confidence interval includes no difference - this does not mean ‘might as well flip a coin’
* **Chapter 13: Choosing a route to cycle home: What p-values do for us**
  + P-values test hypotheses
* **Chapter 14: The probability of a dry toothbrush: What is a p-value anyway?**
  + Inference statistics involves testing hypothesis, specifically a null hypothesis
  + A null hypothesis is a statement suggesting that nothing interesting is going on, for example, that there is no difference between the observed data and what was expected, or no difference between the two groups
  + The p-value is the probability that the data would be at least as extreme as those observed, if the null hypothesis were true (no difference)
  + If the data would be unlikely if the null hypothesis were true, we conclude that the null hypothesis is not true
* **Chapter 15: Michael Jordan won’t accept the null hypothesis: How to interpret high p-values**
  + Testing hypothesis consists of the following steps
    - Specify a null hypothesis
    - Apply a statistical test to the data to obtain a p-value
    - If the p-value is less than 0.05 (statistically significant), reject the null hypothesis
    - If the p-value is 0.05 or more, don’t reject the null hypothesis
  + Don’t accept the null hypothesis. If the p-value is high, 0.05 or greater, you can’t say, for example
    - There is no difference between girls and boys handwriting
    - Job training doesn’t improve productivity
    - But rather that we did not find enough evidence to suggest a significant difference that is most likely not just a chance occurrence
* **Chapter 16: The difference between sports and business: Thoughts on the t test and the Wilcoxon test**
  + Statistical tests are applied to data to generate p-values to test hypothesis
  + The t-test and the Wilcoxon tests are two very well known tests
  + Both tests are used when two groups are compared with respect to a continuous variable
  + The t-test involves calculating an estimate addressing the hypothesis of the study as well as its standard error. The p-value is calculated by comparing the estimate to the standard error
  + To run a Wilcoxon test, the data must first be converted to ranks. The p-value is calculated by comparing differences in ranks to an expected distribution of differences in ranks
  + The t test can be unreliable if the data are very skewed
* **Chapter 17: Meeting up with friends: On sample size, precision and statistical power**
  + When you are planning a study, you have to work out the sample size you need
  + If the main purpose of your study is estimation, the sample size you’ll need depends on how precise you’d like your estimate to be. Precision can be thought of in terms of the width of your confidence interval
  + The precision follows the inverse square law: if you want to halve the width of your confidence interval, you have to quadruple your sample size
  + If the main purpose of your study is hypothesis testing, then the sample size you’ll need depends on power. Power is the probability that, if there is truly an effect of the given sample size, you will reject the null hypothesis
  + Power is typically set at 80% or 90%
  + There are formulas to work out sample size exactly
* **Chapter 18: When to visit Chicago: About linear and logistic regression**
  + Regression is the process of producing an equation - at its simplest form y =mx + b
  + y is what we want to understand, x is what we use to predict y, m is the coefficient and is the constant or intercept
  + y is sometimes called the dependent variable
  + x is sometimes called the independent variable, predictor or covariate
  + one type of x is called univariate, multiple x is called multivariable or or multiple regression
  + Linear regression is used when y is a continuous variable such as time to run a marathon
  + Logistic regression is used when y is binary
* **Chapter 19: My assistant turns up for work with shorter: About regression and confounding**
  + We often think that one thing is associated with another, when in fact both are caused by something else entirely. This is known as confounding
  + If the world doesn’t turn out how you predict, and you think your prediction is a good one, it often means that something else is going on
  + Regression makes predictions
  + If you think that an x is associate with a y, but are worried about confounders, you can add these confounders as covariates in a multiple regression
  + You can calculate the statistical significance of a predictor x in a regression by comparing its coefficients to its standard error
  + Multiple regression is not magic, and doesn’t make the problem of confounding go away
* **Chapter 20: I ignore my child’s cough, my wife panics: About specificity and sensitivity** 
  + Sensitivity is the probability of a positive diagnostic test given that you have the disease
  + Specificity is the probability of a negative diagnosis test given that you don’t have the disease
  + If you have two diagnostic tests and you want to know which one is better, sensitivity and specificity are often not very helpful
  + A different way to compare which test is better is to instead ask, what are the consequences? Of diagnosing a healthy person sick, or a sick person healthy?
  + If you have a patient’s test come back positive/negative, the probability that this result is true is called the positive/negative predictive value
* **Chapter 21: Avoid the sales: Statistics to help decisions**
  + The first stage of decision analysis involves four steps
    - write down each possible decision
    - workout all possible outcomes of each decision
    - work out the probability of each outcome
    - work out the value of each outcome
  + The expected result of each decision is calculated by multiplying the probability of each outcome by its value and adding together

**Some Common Statistical Errors, and what they teach us**

* **Chapter 22: One better than Tommy John: Four statistical errors, some of which are totally trivial, but all of which matter a great deal**
  + Don’t accept the null hypothesis. Instead say something like “we could not show/prove a difference”
  + Don’t report p-values for baseline difference between groups created at random. Simply report estimates for each group separately
  + Don’t give a p-value 0, instead say p < 0.001
* **Chapter 23: Weed control for p-values: A single scientific question should be addressed by a single statistical test**
  + Each and every p-value you report addresses a specific null hypothesis and therefore a specific scientific question
  + Questions about how things change over time, or vary with dose, can be addressed by a single regression equation, instead of multiple comparisons between selected pairs of doses or times
* **Chapter 24: How to shoot a TV episode: Statistical analyses that don’t provide meaningful numbers**
  + The value of any statistical analysis depends on whether the numbers it provides are meaningful
  + Many statistical methods only provide p-values. As such, they can help test hypothesis (inference) but do not provide estimates
  + Many statistical tests should be accompanied by other tests, such as combine an inference with an estimate test such chi-squared with regression
* **Chapter 25: Sam, 93 years old, 700 pound Florida super-granddad: Two common errors in regression**
  + Law of diminishing returns - additional effort does not equate to proportionate rewards received
  + Regression cannot predict too far in the future and predictions should not be taken at face value. Critical thinking should be applied to confirm the numbers make sense given the context
  + Turning x into x^2 is one way to turn straight line regression into a curve
* **Chapter 26: Regression to the Mike: A statistical explanation of why an eligible friend of mine is still single**
  + If you see something that is far from the mean, it is likely to be closer to the mean the next time you check. This is common and called regression to the mean
* **Chapter 27: OJ Simpson, Sally Clark, George and me: About conditional probability** 
  + One common error of calculating probability is calculating the probability of something that has already occurred and then come to conclusions about that caused it based on whether that probability is high or low
  + The probability of something given information that something else is true is called conditional probability
  + Conditional probability depends on both the probability before information was obtained and the value of the information
* **Chapter 28: Boy meets girl, girl rejects boy, boy starts multiple testing**
  + The more statistical tests you conduct, the greater the chance that one will come up statistically significant, even if the null hypothesis is true
  + A small study has a good chance of failing to reject the null hypothesis, even if it is false
  + Subgroup analysis involves increases both the risk of falsely rejected the null hypothesis when it is true and falsely failing to reject the null hypothesis when it is false
  + If your p-value is greater than 0.05, you don’t accept the null hypothesis, you fail to reject the null hypothesis
* **Chapter 29: Some things that have never happened to me: Why you shouldn’t compare p-values**
  + P-values measure strength of evidence, not the size of the effect
  + To test a hypothesis you need a single p-value, not two separate p-values addressing entirely different hypotheses
  + DON’T COMPARE P-VALUES
* **Chapter 30: How to win the marathon: Avoiding errors when measuring things that happen over time**
  + A statistical analysis of something that changes over time involves decisions of when to ‘start the clock’
  + Many over time statistical errors occur because of starting the clock at the wrong time
  + One common error is when assessing the the association between the predictor and the time to something happening is to start the clock before you have the information you need
  + Lead time bias - when finding the problem earlier does not yield better results. Such as catching prostate cancer early leading to longer time being miserable when it is removed before inevitably dying either way
  + Intention to treat analysis - To work out whether something is worth it, ei an action leading to desired result - one must look at all trials, not just ones that were successful
* **Chapter 31: The difference between bad statistics and a bacon sandwich: Are there rules in statistics?**
  + Statistics is used to help scientists analyze data, but is itself a science
  + Statistics is not subject to a set of unchanging rules - they are constantly being developed, tested, and updated
* **Chapter 32: Look at your garbage bin: It may be the only thing you need to know about statistics** 
  + Two key characteristics of science are reproducibility and the systematic attempt to avoid error
  + For statistical analysis to be scientific, it needs to be reproducible. Writing programing code for an analysis helps reproducibility
  + Many errors in scientific reports result not from flawed study design, poor experimental, technique or inappropriate stats, but from simple errors in data collection or typing
* **Chapter 33: Numbers that mean something: Linking math and science**
  + What statistics should be about is linking math to science
  + Think through the science and develop statistical hypotheses in the light of specific scientific questions
  + Interpret the results of analyses in terms of their implications for those questions
  + Computers now make complex statistical analysis extremely easy to conduct. As a result, these methods might be misused and misinterpreted
* **Chapter 34: Statistics is about people, even if you can’t see the tears**
  + Remember that the data being worked are associated with real people, events, personal factors and it is important to not distance oneself from that
  + Data on cancer is collected from real people, with real diseases, and real families and lives that are impacted