UCLA Statistics 10
Introduction to Statistical Reasoning
Summer Session C 2019 – Dr. Miles Chen
Lecure Notes
Spenser Talkington

Lecture 1. 5 August 2019

We want to understand how things behave, so we ask questions and seek their answers. Every question has a relevant *population*. However, usually we cannot study the whole population. Often the population is either too large, or it is hard to find the population: ex. folks with lung cancer—they may not know they're members of the population. So, we take a selection of the population. This selection is called a *sample*. From our observations of the sample we try to understand what the answer to the question is for the whole population.

This sample should be *representative* of, or "look like" the whole population. However, its really hard to get a representative sample. A sample which is not representative is said to be *biased*. One method that should, but doesn't necessarily, give a representative sample is *simple random sampling* (SRS). For SRS, first determine the entire population, or the *sampling frame*. Then use a random system to select individuals from the sampling frame. Yet, we may not be able to determine the sampling frame: ex. we don't have a list of the entire population of Los Angeles. Also, each individual should have an equal chance of being chosen.

What we quantitatively know is data. Data is gathered through whatever techniques are sensible. Then it is prudent to organize it. One way to organize it is with a *table* where each separate observation is a row, and each attribute/information/data is a column. Each row may begin with an *identifier* for that datum. Two types of data are:

- *categorical variable*: something that is binned and only distinction matters, ex. gender, area code.
- *numeric variable*: something that is continuous and differences matter, ex. height, age.

Numeric variables may be binned into categorical variables, ex. binning folks by age, but translating categorical variables into numeric variables requires clarification.

Data comes in several forms:

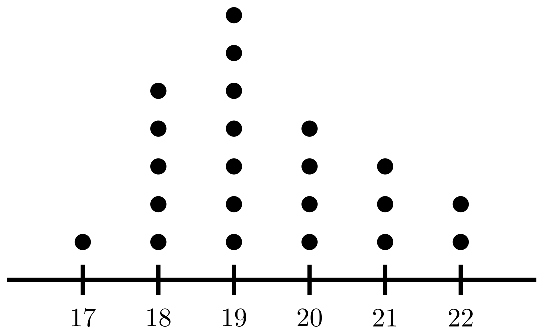
- *anecdotal evidence* is our individual experiences, which are not uniform or repeatable. These experiences may be compiled, as on review sites such as Amazon or Rotten Tomatoes, but it is unwise to draw statistical conclusions from such data sets.
- *observational evidence* comes from the observation of subjects in a sample. Ideally the sample we observe is representative of the population. We try to not interfere with the subjects, other than by taking measurements.
- *experimental evidence* comes from the measurement of characteristics of a sample in a controlled situation. Ideally the researcher can control all aspects of variation in the subjects. The researcher will randomly assign conditions to the subjects. If there is a difference at the end of the experiment, then that difference can be attributed to the different conditions.

With our data, we search for associations between the variables we collected. However, an association, or *correlation* between variables does not mean that there is a cause and effect relationship between variables.

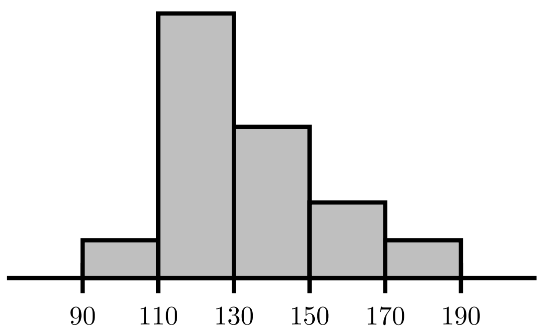
Lecture 2. 7 August 2019

Goal: describe a *distribution* of data, both graphically and numerically. Ex. suppose we have 128 students in the class, and we record each of their ages. The distribution of ages will cluster around say 20, and have a smallest and largest value. This would be "different" from say the ages of folks on an airplane, or at the DMV. Yet this description of ages is vague.

You could, say plot the ages with dots:



However, when you have many data points, such as with weights, it makes more sense bin data and smooth dots into bars for a *histogram*:



Note that absolute numbers, or *frequencies*, are meaningless to compare samples of different sizes, so it makes more sense to compare as fractions, or percentages.

A *symmetric* distribution is one where the left hand side looks like the right hand side: the frequency of values that are higher than *average* is similar to the frequency of values that are below average. A *skewed* distribution may be right skewed with extreme values on the right, or left skewed with extreme values on the left. Ex. baseball salaries are right skewed, and scores on a relatively easy exam are left skewed.

Distributions may have one peak, or be *unimodal*, may have two peaks or be *bimodal*, or have multiple peaks, or be *multimodal.* Ex. number of customers at a restaurant by hour may be bimodal. Google Maps plots business traffic vs. time for restaurants: "looking at this is what I do for fun."

Numeric summaries of data are called *statistics.* In particular some common metrics quantify the "typical values" and the "spread".

1. *mean*: add up the values and divide by the number of them. The mean is a good way to describe the typical value for symmetric distributions. The mean is "pulled" in the direction of the skew, if any.

$$\bar{q} = \frac{1}{n}\sum_{i=1}^{n} q_i$$

2. *median*: find the middle number. If numbers are ordered, the median is $q_{n/2}$. The median is a good way to describe the typical value for skewed distributions.
3. *range*: how far does the data vary?

$$\text{range}(q) = \max(q) - \min(q)$$

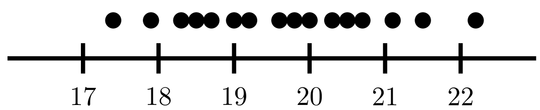
4. *standard deviation*: "where is the middle, and how far is the data from the middle?" This is done be computing the average squared deviation, or *variance* s^2, and taking the square root to find the standard deviation:[1]

$$s(q) = \left[\frac{1}{n-1} \sum_{i=1}^{n} (q_i - \bar{q})^2 \right]^{1/2}$$

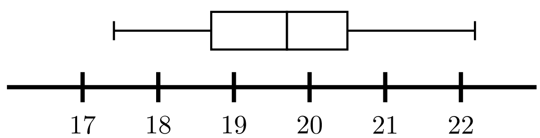
5. *IQR*: how far does half of the data vary? If the data are ordered, the interquartile range (IQR) is $q_{3n/4} - q_{1n/4}$, or the value at the seventy-fifth percentile minus the value at the twenty-fifth percentile.

A way to graphically organize data using *quartiles* is the *box plot*. The box plot draws boxes between $q_{n/4}$ and $q_{2n/4}$, and $q_{2n/4}$ and $q_{3n/4}$, with *whiskers* extending to the minimum and maximum values (or some other point, ex. n standard deviations).

A representation of ages as dots:



The representation of ages as a box plot:



Part of describing a distribution is to point out *outliers*, or observations that are unusual.

The *upper inner fence* and *lower inner fence* are defined by $q_{3n/4} + 1.5 \times IQR$, and $q_{n/4} - 1.5 \times IQR$, and the *upper outer fence* and *lower outer fence* are defined by $q_{3n/4} + 3.0 \times IQR$, and $q_{n/4} - 3.0 \times IQR$. *Mild outliers* lie in between the lower outer fence and the lower inner fence, or the upper inner fence and the upper outer fence. *Extreme outliers* lie below the lower outer fence, or above the upper outer fence.

Another way to describe how unusual an observation is to use a z-score:

$$z(q_i, q) = \frac{q_i - \bar{q}}{s(q)}$$

For unimodal, symmetric data, we have the "68-95-99.7 rule" for percent distributed in $-1 \leq z \leq 1$, $-2 \leq z \leq 2$, and $-3 \leq z \leq 3$ respectively. The percentiles are:

z	percentile
-3	0.13
-2	2.28
-1	15.87
0	50.00
1	84.13
2	97.72
3	99.87

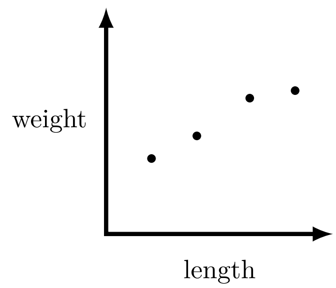
Lecture 3. 12 August 2019

Goal: describe the relationship between two different numeric variables for the same sample.

Ex. is there a relationship between the length of a snake and its weight? Longer snakes tend to weigh more than shorter snakes, but how do we quantify this?

We can graphically illustrate the relationship between two variables with a scatterplot.

[1]We divide by $n-1$ instead of n because of Bessel's correction. In essence, for any data q, $|\bar{q} - \bar{q}_{\text{true}}| \geq 0 : s^2 \leq s^2_{\text{true}}$.



From a scatterplot, three characteristics to describe the relationship of variables become apparent:

- What is the shape? Is it *linear*, or *non-linear*?
- What is the direction? Is it *positive*, *negative*, or neither?
- What is the strength of the relationship? Is it strong, weak, or moderate?

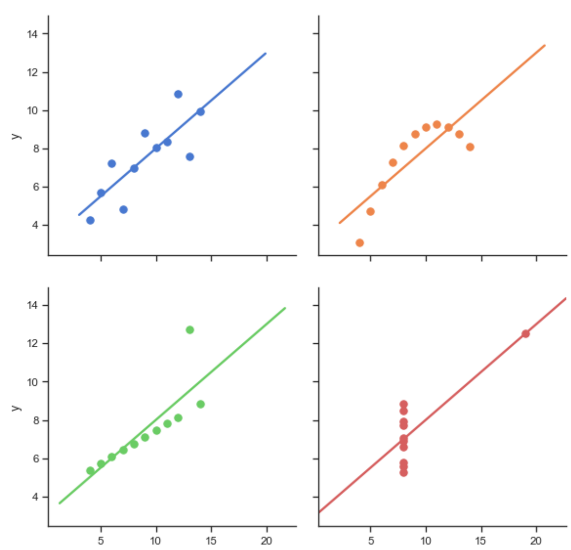
Linear relationships can be approximated with straight lines, and non-linear relationships cannot be approximated with straight lines. A positive relationship means that large (small) x is related to large (small) $f(x)$, while a negative relationship means that large (small) x is related to small (large) $f(x)$. A strong relationship means that knowing the value of one variable enables us to make good predictions of the other variable; alternatively, there is little scatter about the fit. A weak relationship means that knowing the value of one variable does not enable us to make good predictions of the other variable; alternatively, there is a lot of scatter about the fit.

The *correlation coefficient*, R, summarizes the direction and the strength of a linear relationship. The sign of R is the direction of the relationship, and the magnitude of R is the strength of the relationship ($-1 \leq R \leq 1$). Note that:

- Correlation does not tell us anything about the slope.
- Multiplying or adding a constant to all values will not change the correlation: we may use any units.
- The correlation between x and $y(x)$, is the same as the correlation between y and $x(y)$.
- A flat line has $R = 0$.
- Correlation can be greatly changed by outliers.
- Do not *extrapolate*, or predict results for regions where you have no data.
- Correlation is only for linear relationships!
- Correlation is not causation.

Another quantity to describe the strength of a relationship is R^2, or the *coefficient of determination*. For a linear regression, $R^2 = (R)^2$, so R^2 is between 0 and 1. Often R^2 is used to say how "good" a linear regression model is, but it is perhaps better to think about it as "the amount of variation in y that is explained by x."

Anscombe's quartet is illustrative: each data set has a correlation of $R = 0.816$, but only one is linear:



Linear regression is creating a linear equation to describe the relationship between two numeric variables. The variable we use to make the prediction is called the *predictor variable*, the *explanatory variable*, or the *independent variable*. The variable being predicted is called the *predicted variable*, the *response variable*, or the *dependent variable*. Symbolically, our prediction for y, on average, is:

$$\widehat{y} = a + bx$$

The *residual* is defined as the difference between the actual value and the predicted value. The residual may take on positive or negative values:

$$\text{residual} = \text{actual} - \text{predicted}$$

Least squares regression provides a metric to specify what the best fit line is. The method selects the line that minimizes the residual squared. This best fitting line will be specified by:

- the slope is, in terms of the standard deviations of the x and y values:

$$b = R\frac{s(y)}{s(x)}$$

- the intercept is, in terms of the means of the x and y values:

$$a = \bar{y} - b\bar{x}$$

Ex. the width of a book by the number of pages (in millimeters):

$$\widehat{\text{width}} = 6.22 + 0.0366 \times \text{pages}$$

"I don't want you to take this as a cause-and-effect relationship. This is a prediction of the average slope, but it may not hold for individual measurements."

Lecture 4. 14 August 2019

Goal: describe how likely, or unlikely something is to happen.

Probability tells us how likely or unlikely certain events are to happen. This can then be extended to tell us how likely or unlikely our data is to be observed. This assumes that certain events are random in the sense that we cannot predict the next outcome, and that there is no pattern in individual outcomes.

Theoretic probabilities are probabilities that are derived using logic from assumptions. Meanwhile *empirical probabilities* are probabilities that are estimates from experimental outcomes or simulations. Theoretical probabilities are preferred, but empirical probabilities are useful when the math is too complicated, or when it doesn't exist.

Call an event A. The probability of A is denoted $P(A)$. The probability of any event is between 0 and 1. If $P(A) = 0$, then A never happens. If $P(A) = 1$, then A is certain to happen.

The *complement* of A is the outcome where A did not happen, and is denoted $P(\texttt{not } A) = 1 - P(A)$.

Simple outcomes are equally likely. If A consists of simple outcomes:

$$P(A) = \frac{\text{number of outcomes } A}{\text{total number of outcomes}}$$

Ex. a coin flip can result in H or T.

$$P(H) = \frac{\text{number of outcomes } H}{\text{total number of outcomes}} = \frac{1}{2}$$

Ex. rolling two six-sided dice. What is the $P(7)$?

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

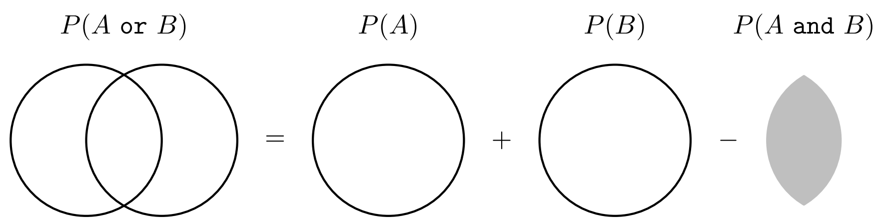
From the tabulation, we find:

$$P(7) = \frac{6}{36} = \frac{1}{6}$$

The probability of A or B is the probability of A alone happening, B alone happening, and A and B both happening, combined. So, in statistics `or` is `inclusive or`. To compensate for double counting:

$$P(A \texttt{ or } B) = P(A) + P(B) - P(A \texttt{ and } B)$$

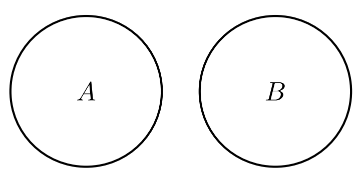
This is graphically represented as:



Suppose we consider 240 people, of whom 80 have a dog, 70 have a cat, and 30 have both a dog and a cat. Work out the total numbers:

	dog	no dog	3
cat	**30**	40	**70**
no cat	50	120	170
	80	160	**240**

Mutually exclusive events are events that cannot happen together, i.e. $P(A \texttt{ and } B) = 0$. Ex. A is when a given number x is even, and B is when x is odd.



Now, we fix the notation for *conditional probabilities*, $P(A|B)$ is the probability of A happening given that B happens. Ex. Conditional probabilities are $P(\text{cat}|\text{dog}) = 30/80 = 3/8$, while $P(\text{dog}|\text{cat}) = 30/70 = 3/7$. Also, note that: $P(\text{cat} \texttt{ and } \text{dog}) = 30/240 = 1/8$. While the total numbers are the same, the probabilities are different. We find that in general:

$$P(A|B) \neq P(B|A)$$

Independent events are events where knowing that one event has happened has no influence on the other event. Mathematically:

$$P(A|B) = P(A|\texttt{not } B) = P(A)$$

Now for a made up example concerning movie preferences:

	Likes Toy Story	Did not like Toy Story	
Likes Frozen	400	100	500
Did not like Frozen	250	250	500
	650	350	1000

From the chart we may work out probabilities:

$$P(T) = \frac{650}{1000} = 0.65$$

$$P(F) = \frac{500}{1000} = 0.5$$

$$P(T \texttt{ and } F) = \frac{400}{1000} = 0.4$$

$$P(T \texttt{ or } F) = P(T) + P(F) - P(T \texttt{ and } F) = 0.65 + 0.5 - 0.4 = 0.75$$

$$P(F|T) = \frac{P(F\&T)}{P(T)} = \frac{400}{650} = 0.615$$

$$P(T|F) = \frac{P(T\&F)}{P(F)} = \frac{400}{500} = 0.8$$

So we find that liking Toy Story is not independent of liking Frozen since $P(T) \neq P(T|F)$, or $0.65 \neq 0.8$.

Lecture 5. 19 August 2019

Probability continued.

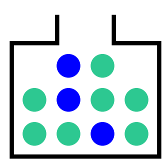
The *multiplication rule* states that:

$$P(A \texttt{ and } B) = P(A) \times P(B|A)$$

If A and B are independent, then $P(B|A) = P(B)$, so, $P(A \texttt{ and } B) = P(A) \times P(B)$. Generalizing:

$$P(q_1 \texttt{ and } q_2 \texttt{ and } \dots \texttt{ and } q_n) = \prod_{i=1}^{n} P(q_i)$$

An example of this is drawing marbles from a jar:



Consider a jar with 10 marbles, of which 3 are blue and 7 are green. Drawing without replacement, what is the probability that the first two draws are both blue?

$$P = P(\text{blue first}) \times P(\text{blue second}|\text{blue first}) = \frac{3}{10} \times \frac{2}{9} = \frac{1}{15}$$

Drawing with replacement, what is the probability that the first two draws are both blue?

$$P = P(\text{blue first}) \times P(\text{blue second}|\text{blue first}) = P(\text{blue first}) \times P(\text{blue first}) = \frac{3}{10} \times \frac{3}{10} = \frac{9}{100}$$

Drawing two marbles with replacement, what is the probability to draw one blue and one green marble in any order?

$$P = P(\text{blue then green}) + P(\text{green then blue}) = \frac{3}{10} \times \frac{7}{10} + \frac{7}{10} \times \frac{3}{10} = \frac{21}{50}$$

These probabilities may be added because they are mutually exclusive.

Drawing four marbles with replacement, what is the probability that at least one marble is blue? Solve this problem by recognizing that the solution is the complement of drawing no blue marbles, or the probability of drawing all green marbles. $P = 1 - P(4 \text{ green}) = 1 - (7/10)^4$.

Drawing four marbles with replacement, what is the probability that two marbles are green and two marbles are blue, in any order? Here we must list all the possibilities:

$$\left\{\begin{matrix} \text{BBGG} \longrightarrow 3 \times 3 \times 7 \times 7/10^4 \\ \text{BGBG} \longrightarrow 3 \times 7 \times 3 \times 7/10^4 \\ \text{GBBG} \longrightarrow 7 \times 3 \times 3 \times 7/10^4 \\ \text{BGGB} \longrightarrow 3 \times 7 \times 7 \times 3/10^4 \\ \text{GBGB} \longrightarrow 7 \times 3 \times 7 \times 3/10^4 \\ \text{GGBB} \longrightarrow 7 \times 7 \times 3 \times 3/10^4 \end{matrix}\right\} = \frac{2646}{10000}$$

Listing all possibilities becomes impractical quickly, so we introduce the formal idea of a *combination* of k "successes" from n possibilities. This is read as n choose k:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Drawing ten marbles with replacement, what is the probability that 4 marbles are blue in any arrangement?

$$P = \binom{10}{4}\left(\frac{3}{10}\right)^4\left(\frac{7}{10}\right)^6 = \frac{210 \times 81 \times 117649}{10^{10}} = \frac{2001209490}{10000000000}$$

Ex. If I buy 10 tickets, each of which have a 1/10 chance of giving any reward, what is the chance that any of them give me a reward? $1 - P(\text{losing all } 10) = 1 - 9^{10}/10^{10} \approx 0.65$.

Lecture 6. 21 August 2019

Goal: describe continuous probability distributions.

Probability distributions tell us which outcomes are more and less likely. We *define* the probability to be the integral over an interval. This is necessary because while in *discrete distributions* exact values occur, and the probabilities may be summed, the same is not true for *continuous distributions*. The probability is:

$$P(x) = \lim_{\delta \to 0} P(x - \delta, x + \delta)$$

Where we insist on the normalization condition:

$$\int_{x_{\min}}^{x_{\max}} dx\ P(x) = 1$$

Noting that a lot of observations happen to follow a normal distribution, we pay particular attention to this distribution. Recall that the normal distribution is continuous, unimodal, and symmetric. The mean is called μ and the standard deviation is called σ; a notation for distributions is $N(\mu, \sigma)$. The *standard normal distribution*, or z-distribution, has $\mu = 0$ and a $\sigma = 1$. This harks back to the empirical rule of 68-95-99.7.

Two maneuvers are particularly helpful for the z-distribution. Firstly, if we want to know the probability that a value lies between two values on a normal distribution $N(\mu, \sigma)$, it suffices to calculate the z-score of these two values, determine the percentile of each (perhaps with a table), and take the difference.[2] Secondly, from a percentile, or z-score, it is possible to find the values in the distribution $N(\mu, \sigma)$:

$$z = \frac{x - \mu}{\sigma} \implies x = \mu + \sigma z$$

[2]This is because all normal distributions are homomorphic to, and in fact linear scalings of, the z-distribution.

Midterm. 26 August 2019

(Midterm)

Lecture 7. 28 August 2019

An association between variables in data may be real, or it may be a random fluke. How do we quantify the

Ex. 30% of South Campus majors wear glasses and 27% of North Campus majors wear glasses. How can we tell if there really is a dependence, or whether the difference is a fluke based on the sample we selected?

Recall that we study a sample taken from the population since the population is too big to be studied directly. Here we assume that the sample is selected randomly, but take note that this may not always be the case. A population has parameters, such as a mean μ, standard deviation σ, and a proportion p, while a sample has statistics, such as a mean $\bar{q}$, a standard deviation s, and a proportion $\hat{p}$.

The *sampling distribution* is the distribution of a samples statistics' possible values for the range of possible values. The *Central Limit Theorem* says that the distribution of $\hat{p}$ will look very similar to a normal distribution. The center will be equal to the population mean p, and the standard deviation will be:

$$s(\hat{p}(q)) = \sqrt{\frac{p(1-p)}{n}}$$

In other words, $\hat{p}$ is distributed as:

$$\hat{p} = N(p, s)$$

Ex. Suppose that a bucket of candy is 0.30 orange. If a scoop has 60 random pieces, what is the probability that the scoop is 0.35, or more, orange?

Proceed by calculating the z-score:

$$z = \frac{0.35 - 0.30}{\sqrt{\frac{0.30(1-0.30)}{60}}} = 0.845$$

Looking in a table yields a probability of 0.1977.

However, keep these in mind when using the Central Limit Theorem. $\hat{p}$ is discrete, so it follows the *binomial distribution*, which is the discrete distribution. The normal distribution is a good approximation to the binomial distribution for a large sample size n, but a poor approximation for small sample size. The *large sample condition*, or the 10 *yes's condition*, is that:

$$n \cdot p \geq 10 \quad \text{and} \quad n(1-p) \geq 10$$

Also, make sure that the samples are selected randomly and independently. If the sample is selected without replacement, then the *large population condition* is that the population needs to be at least 10 times larger than the sample.

Ex. Suppose that we know that $\hat{p}$ follows the normal distribution, and we find that out of 200 people, 96 have dogs. What are the reasonable values of p? Well, for a normal distribution, 0.95 of values are within 2 standard deviations of the mean, so we can be 0.95 certain that p is within 2 standard deviations of $96/200 = 0.48$.[3] We estimate the standard deviation as:

$$s \approx \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.48 \cdot 0.52}{200}} = 0.0353$$

So the 0.95 confidence interval is 0.48 ± 0.07. This means that we have evidence that the value is between 0.41 and 0.55, even though values outside this interval are not impossible.

[3]Technically 0.95 Confidence Interval is 1.96 standard deviations, and 0.99 Confidence Interval is 2.58 standard deviations.

Labor Day. 2 September 2019

(Labor Day)

Lecture 8. 4 September 2019

Hypothesis testing for proportions seeks to answer: "If this claim were true, could we observe the data that we have?"

The *Null Hypothesis* that the population proportion is equal to p_0, $H_\emptyset : p = p_\emptyset$. *Alternate hypotheses* may be that the proportion is not equal to, is greater than, or is less than. For example: $H_A : p \neq p_\emptyset$, $H_A : p < p_\emptyset$, or $H_A : p > p_\emptyset$.

Ex. a company says that "over ninety percent of customers are satisfied." How would we test this? Set the null hypothesis $H_\emptyset : p = 0.90$, and the alternative hypothesis $H_A : p < 0.90$. We are only concerned if the proportion is less than 0.90.

To conduct a hypothesis test:

1. Write your hypothesis, $H_\emptyset$ will always have an equals sign.
2. Organize: choose a significance level, check to make sure you can use the Central Limit Theorem.
3. Compute: standard error, test statistic, p-value
4. Interpret: Make a conclusion about your hypothesis.

Ex. 0.47 of California's registered voters voted. Is the rate different for Poli Sci majors?

1. $H_\emptyset = 0.47$, $H_A \neq 0.47$
2. Organize: choose a significance level, $\alpha = 0.05$. We assume that the selection of individuals was random and independent, there was a large sample and a large population. Say that our sample was 54 students from one class, of whom 40 voted, and the department has 800 students in total.
3. Compute: If the null hypothesis were true, what is the probability of getting the data we have?

$$\hat{p} = \frac{40}{54} = 0.74$$

Is it possible from random sampling alone that we got 0.74? Yes, but if $p = 0.47$, then the probability of getting 0.74 is, $s = \sqrt{0.47 \cdot 0.53/54} = 0.068$, so $z = (0.74 - 0.47)/0.068 = 3.97$, which only occurs 0.000036. Therefore our *p-value* is 0.000072, so our data happens less than 0.000072 of the time.
4. Interpret: Our probability calculations show us that our data is unlikely to occur if the null hypothesis were true. So we conclude that Poli Sci majors vote more than the whole population.

This naturally extends to hypothesis testing between two samples. Ex. a survey:

	2012	2019	
support	645	870	1515
oppose	855	630	1485
	1500	1500	

Did the proportion of supporters change from 2012 to 2019? Select the null hypothesis $H_\emptyset : p_{2012} = p_{2019}$, and the alternate hypothesis $H_A : p_{2012} \neq p_{2019}$. Then select the significance level $\alpha = 0.05$. If $H_\emptyset$ were true, $\hat{p}_{2012} - \hat{p}_{2019}$ would be expected to be close to 0. In fact, we expect:

$$\hat{p}_{2012} - \hat{p}_{2019} = N\left(0, \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_{2012}} + \frac{1}{n_{2019}}\right)}\right) = N(0, 0.0183)$$

While the real difference is 0.15, which is a z-score of 8.22, which strongly indicates that there is a difference.

Lecture 9. 9 September 2019

Hypothesis test with one sample, ex. $H_\emptyset : p = 0.42$, and the alternative may be $H_A : p \neq 0.42$. Not-equals means that we have a *two-sided test*, while $>$ and $<$ are *one-sided tests*. This comes down to the p-value, which is the probability of observing our data, or something more extreme than our data if we assume the null hypothesis were true. When $H_\emptyset$ is true, only randomness is responsible for the variation in our data. A large p-value means that there is a high probability of getting our data from randomness alone, while a small p-value means that there is a low probability of getting our data from randomness alone. To make this rigorous, we select a *significane level*, $\alpha = 0.05$. When $p < \alpha$, we have evidence to reject the null hypothesis, and when $\alpha < p$, we do not have evidence to reject the null hypothesis. "Anytime you draw an arbitrary line, weird things happen."

Two types of errors, known as *false positives* and *false negatives* are:

1. Type I error: In truth, $H_\emptyset$ is true, but our data by random chance led us to reject $H_\emptyset$
2. Type II error: In truth, $H_\emptyset$ is false, but our data by random chance led us to not reject $H_\emptyset$

α is the probability of generating a false positive, while β is the probability of generating a false negative. Selection of α depends on the relative consequences of making errors. Also, not rejecting $H_\emptyset$ is not the same as saying $H_\emptyset$ is true.

Final. 11 September 2019

(Final)