Day 5 - Statistics

Agenda

- Activity #1: Dice Sampling
  - Sampling Distributions
- Activity #2: Michael and Me
  - Two-variable data
  - Scatterplots
  - Outliers
  - Correlation
- Activity #3: The Case of the Careless ZooKeeper
  - Test of association
  - Hypothesis Testing

The Case of the careless Zookeeper

- The context for this activity is a bit silly/fun
- We are imagining a situation where a careless zookeeper delivers a trainload of animals to another zoo, but uses poor judgment and ships all of the animals (regardless of eating habits) intermixed in the same train cars.

The Case of the careless Zookeeper

- The goal for each group is to answer the following question:
  - Is the type of animal (herbivore or carnivore) related to the injury status (no damage or some damage)?

Test of Association

- The statistics content needed for today is likely to be less familiar than other topics for some of us

- I chose this activity, though, because the Chi-Square test of association is not inaccessible and is very useful

Test of Association

- I will provide an example of the Chi-Square test of association so that we all have a small experience set to draw upon

Test of Association

- Essentially, the test of association is another method to determine if there is a relationship between two (or more) factors

- A common test of association would look like:
  - Null Hypothesis: A and B are independent
  - Alternative Hypothesis: A and B are dependent

Test of Association

- Chi-Square example:

- Guiding question
  - Is there a difference between genders when it comes to whom it is easiest to make friends with?
  - Do females become friends more easily with males, females, neither???

Test of Association

- Given this context we need hypotheses to test. The hypotheses I chose are:

- Null Hypothesis:
  - Gender and finding friends are independent
- Alternative Hypothesis:
  - Gender and finding friends are dependent

Test of Association

- Let's assume that we have the following data from a survey:

	Opposite Gender	Same Gender	No difference	Total
Female	58	16	63	137
Male	15	13	40	68
Total	73	29	103	205

Test of Association

- The major key of a test of association is to test against expectations (null hypothesis)

- If the null hypothesis were true, we would expect numbers to be evenly distributed between genders.

Test of Association

	Opposite Gender	Same Gender	No difference	Total
Female	(Total Females/ Total people) X (column total)	(Total Females/ Total people) X (column total)	(Total Females/Total people) X (column total)	137
Male	(Total Males/ Total people) X (column total)	(Total Males/ Total people) X (column total)	(Total Males/ Total people) X (column total)	68
Total	73	29	103	205

Test of Association

	Opposite Gender	Same Gender	No difference	Total
Female	(137/205) X (73)	(137/205) X (29)	(137/205) X (103)	137
Male	(68/205) X (73)	(68/205) X (29)	(68/205) X (103)	68
Total	73	29	103	205

Test of Association – Expected Values

	Opposite Gender	Same Gender	No difference	Total
Female	48.79	19.38	68.83	137
Male	24.21	9.62	34.17	68
Total	73	29	103	205

Test of Association – Observed Values

	Opposite Gender	Same Gender	No difference	Total
Female	58	16	63	137
Male	15	13	40	68
Total	73	29	103	205

Test of Association

- Formula

$$\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

Test of Association – Sums

	Opposite Gender	Same Gender	No difference
Female	$(58 - 48.79)^2 / 48.79$	$(16 - 19.38)^2/19.38$	$(63 - 68.83)^2/68.83$
Male	$(15 - 24.21)^2/24.21$	$(13 - 9.62)^2/9.62$	$(40 - 34.17)^2/34.17$

$$\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

$$\chi^2 = 8.508$$

Test of Association

- Ok. So $\chi^2 =$ 8.508
- What's the point?

- Well this is where we need to use a table that is based on probabilities and distributions

- For the sake of today's examples and problems – we are dealing with '1 degree of freedom'.
- We do not need to worry about this element of the Chi-Square test today.

Test of Association

- $\chi^2 = 8.508$

- We get to decide our cutoff for how certain we want to be that we are making the 'correct' decision in comparing our hypotheses.

- This is formally called a 'p-value'

Test of Association

- $\chi^2 = 8.508$

- A ‘p-value’ is the probability of obtaining the observed sample results, or even "more extreme" results, when the null hypothesis is actually true (here, "more extreme" is dependent on the way the hypothesis is tested)

Test of Association

- $\chi^2 = 8.508$

- So if we chose a p-value of .05 (5%) it means that if we determine the null hypothesis to be false…then there is a 5% (or less) chance that we could get the observed values, but the null hypothesis still be true.

Test of Association

- $\chi^2 = 8.508$

p-value

df	.25	.2	.15	.10	.05	.025	.01	.005	.0025	.001	.0005
	25%	20%	15%	10%	5%	2.5%	1%	**.5%**	**.25%**	.1%	.05%
1	1.32	1.64	2.07	2.71	3.84	5.02	6.63	**7.88**	**9.14**	10.83	12.12

- Because our Chi-Square value is 8.508, our 'p-value' is somewhere between

  .005 and .0025

  or

  .5% and .25%

Test of Association

- So our 'p-value' is roughly .005
- As a percent it is .5%

- What does this mean?

- It means:
  - There is a .5% chance that the Null hypothesis is in fact true, but we still got the sample results
  - Stated another way: It is very unlikely that if the null hypothesis is true that we would get a sample like this
  - Typically a p-value less than .05 is considered '**significant**'

The Case of the careless Zookeeper

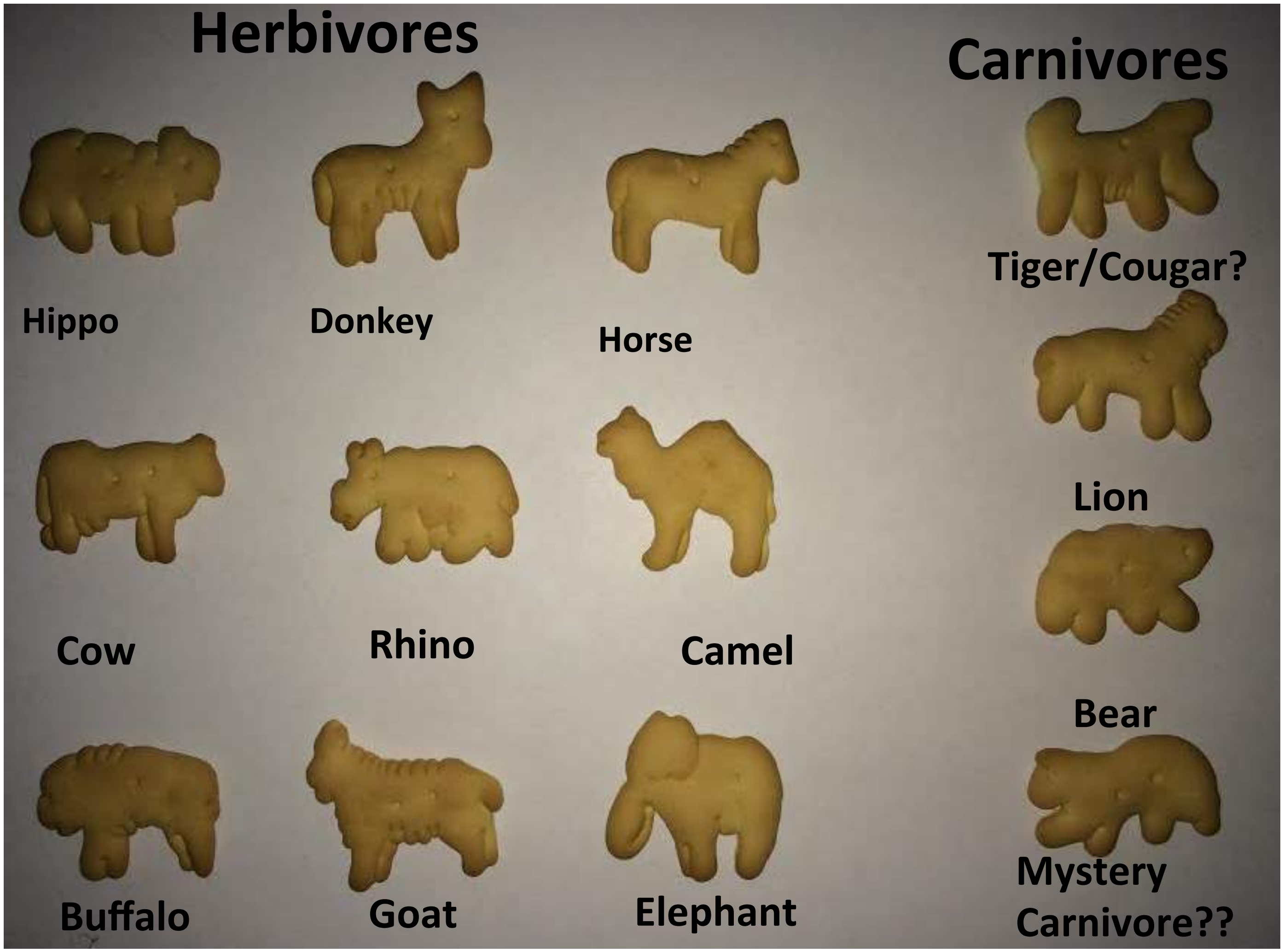
- Back to the beginning.
- We are imagining a situation where a careless zookeeper delivers a trainload of animals to another zoo, but uses poor judgment and ships all of the animals (regardless of eating habits) intermixed in the same train cars.

The Case of the careless Zookeeper

- The goal for each group is to answer the following question:
  - Is the type of animal (herbivore or carnivore) related to the injury status (no damage or some damage)?

The Case of the careless Zookeeper

- Here is my best attempt to categorize the animals I believe you will encounter:



Discussion