Biostatistics

Biostatistics: analyze biological data using statistical methods.

Challenges:

1. Collect data in a way that it can be analyzed properly.
2. Reduce and organize data.
3. Choose appropriate statistical test.
4. Correctly interpret output of the test

"A taste for the Beautiful" (pg 306, 9th ed text)

Crests (plume, feather) were affixed to the heads of male Finches.

5 stimulus sets.

Each stimulus set =

1. Male with white crest
2. Male with green crest
3. Male with red crest
4. Male with no crest

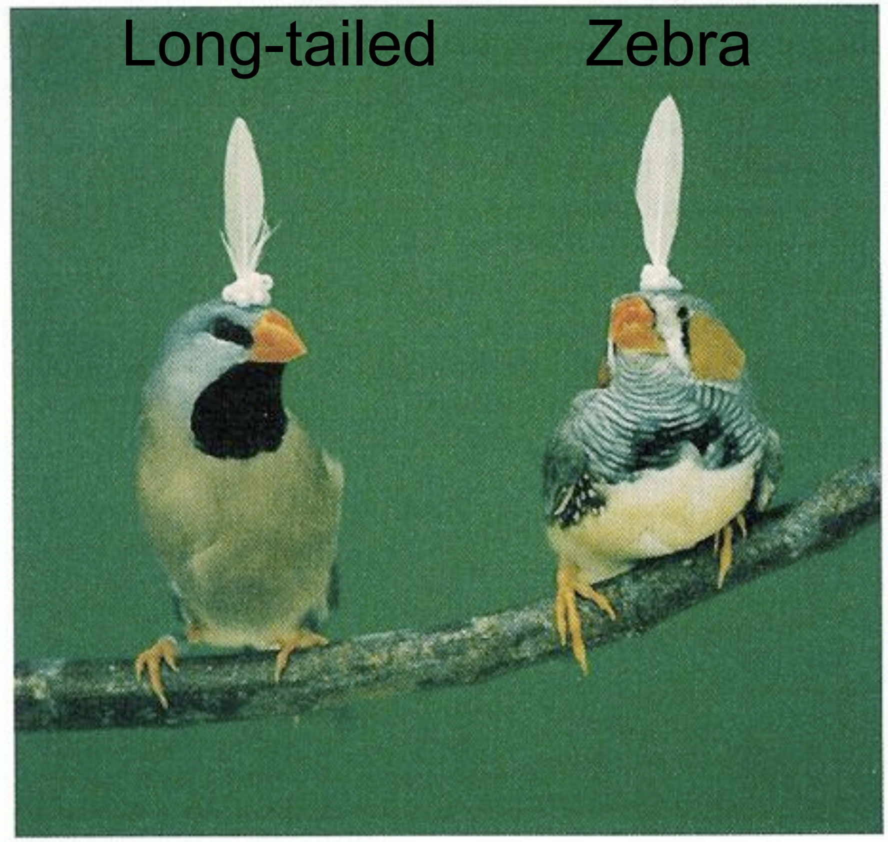


Figure 9.21 Mate preferences for a novel ornament. (A) A male long-tailed finch (left) and a male zebra finch (right) have been outfitted with bizarre white plumes. (B) The addition of white plumes made male zebra finches more attractive to females than control males without plumes or those given head-dresses of red or green feathers. A, photograph by Kerry Clayman, courtesy of Nancy Burley; B, after Burley and Symanski [174].

“A taste for the Beautiful” (pg 306, 9th ed text)

15 females were tested.

Each female was tested with two stimulus sets and her response was averaged.

Females were introduced to each stimulus set for two hours.

Results were measured in percent social time spent with males of each crest type.

It appears that female Zebra Finches preferred males with white crests (left graph), but is there really a significant difference?

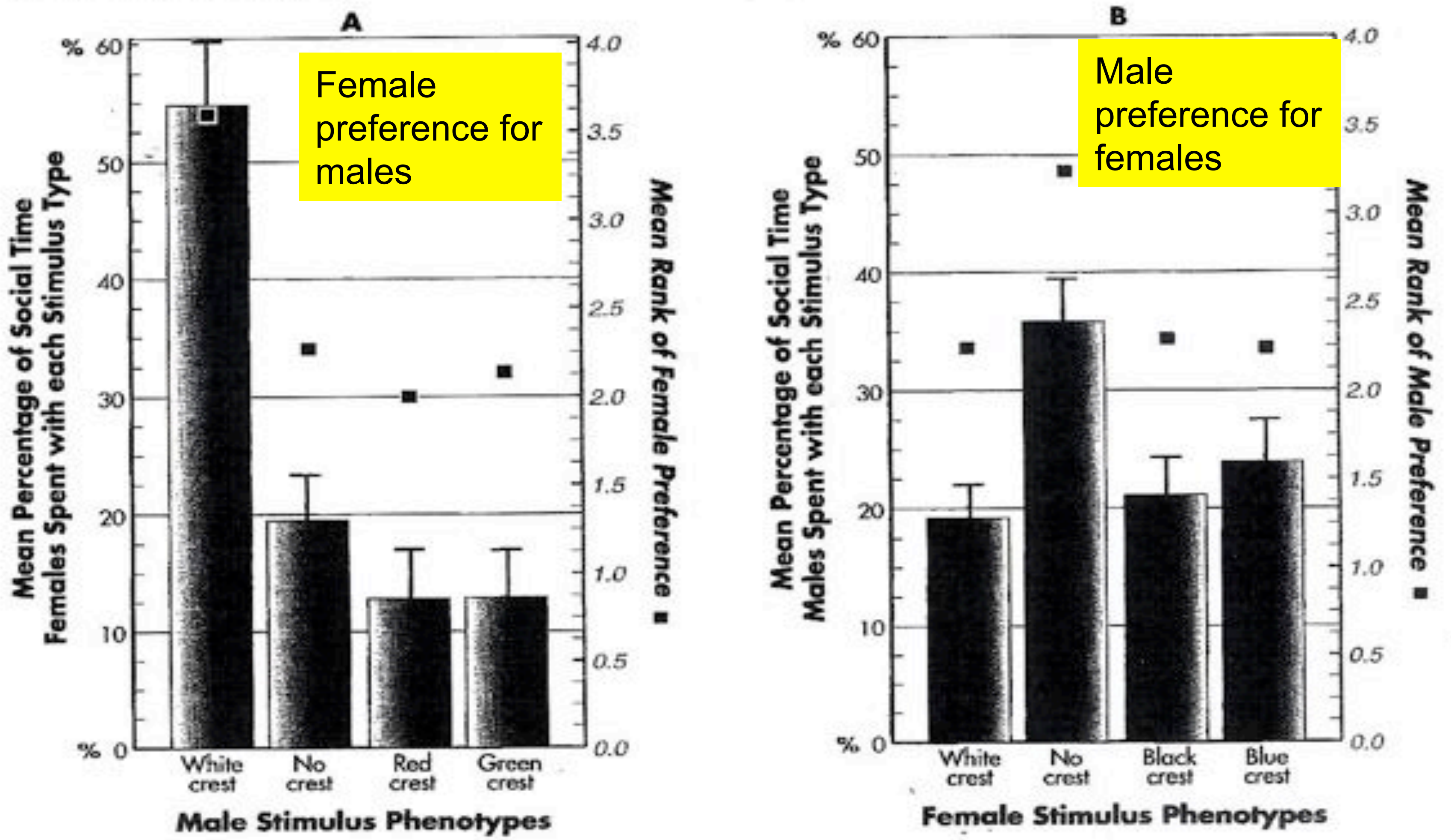


Figure 3: Mate preferences of zebra finches for nonornamented ("No crest") versus artificially crested conspecifics. Bars represent the mean proportion (+SE) of time spent in social contact with each stimulus type; dots represent mean within-trial rank of social preferences (1–4, with 4 being most preferred). *A*, Female preference for males. *B*, Male preference for females.

Results from an experiment may be due to chance.

It is *possible* to flip a coin five times and have it come up heads every time!

Yet the conclusion that flipped coins are more likely to come up heads than tails is incorrect.

It is *unlikely* that you flip a coin five times and get heads every time but it is not impossible. So how confident must we be to accept the conclusion that the coin is loaded?

The probability of getting five heads in a row is $(0.5)^5 = 0.03$.

We use a statistical test to determine the probability that the results of an experiment are due to chance.

P-values

- In general, if the probability that your result could have occurred by chance alone is less than or equal to 5% ($p = 0.05$) then results are accepted as statistically significant.
- There is no good mathematical reason for setting significance at 0.05, it is just a widely accepted value. The lower the p-value the better!
- P-values are calculated using a number of statistical tests.
- P-values are found in an appropriate table depending on the statistical model used.

What about the zebra finches?

It is possible that the researchers had some bad luck and the female Zebra Finches in the study just happened to spend more time with white-crested males even though they do not have a true preference for them. That is, the result they got was a chance outcome.

Data from that study gave a p-value of 0.002, or a 0.2% probability that the results of this experiment were due to chance. (1 in 500!)

Given the low p value, everyone would accept that the results are really due to the experimental stimulus and not the result of chance.

Different statistical tests for different kinds of data sets…

Statistical tests are chosen based on:

1) Type of data (parametric/ nonparametric/ binomial)

2) Number of groups being compared

3) Whether or not values are shared between groups (referred to as ‘paired’ for two groups and ‘matched’ for three or more)

Common Statistical Tests
(rows=purpose of test, columns= type of data)

	Parametric	**Non-parametric**	**Binomial (two possible outcomes)**
Describe one group	Mean, SD	Median	Proportion
Compare a group to a hypothetical value	One-sample t test	Wilcoxon test	Chi-square test
Compare two unpaired groups	Unpaired t test	Mann-Whitney test	Fisher's test
Compare two paired groups	Paired t test	Wilcoxon test	McNemar's test
Compare three or more unmatched groups	One-way ANOVA (analysis of variance)	Kruskal-Wallis test	Chi-square test
Compare three or more matched groups	Repeated-measures ANOVA	Friedman test	Cochrane Q

Type of Data

Parametric tests are preferable when you can trust that your data were sampled from a population that follows a normal distribution (bell curve).

When it is unclear whether or not data follows a normal distribution, a non-parametric test is used.

Binomial data is data in which there are only two possible outcomes to each trial.

http://stat.stanford.edu/~naras/jsm/NormalDensity/NormalDensity.html

-or-

http://www.math.csusb.edu/faculty/stanton/m262/normal_distribution/normal_distribution.html

Parametric vs Non-parametric

- Both tests should get similar results, although a parametric is more accurate if applicable.
- Use a parametric test if there are many data points and you can look at the distribution to get good idea of whether or not it follows a bell curve.
- Use a parametric test if previous studies show that data fits a normal distribution.
- Best to use non-parametric if there are few data points and the distribution is unknown.
- When the exact distribution is unknown, all commonly used nonparametric tests rank the outcome variable from low to high and then analyze the ranks.

Number of Groups

- Number of data sets being compared
- In the Zebra Finch example four data sets are compared (white crest, red crest, green crests, and no crest)

Paired/matched *vs.* Unpaired/unmatched

Paired data typically involves measurements made on the same individuals before and after a treatment or event.

Example: testing the reaction time of the same group of individuals before and after consuming alcohol.

	RT before	RT after
Subj 1:	100 ms	125 ms
Subj 2:	115 ms	138 ms
…	…	…

Note that this is not the same as "Repeated Measurements" statistical analysis, which is more complicated.

Paired/matched *vs.* **Unpaired/unmatched**

Used to compare between different individuals, species, etc.

Example: Measure reaction time in one group of individuals before alcohol:

Subj 1 = 110 ms, subj 2 = 118 ms, …

Next measure reaction time in a second group of individuals after alcohol:

Subj 22 = 135 ms, subj 23 = 121 ms, …

Use a test to determine if means of the two populations (alcohol vs control) are same (H_0) or different.

Sample Size (N)

Small sample size: it is more likely that chance will have a hand in the results!

Suppose you are looking for the average height of students in the classroom, but only sample the heights of three students who are picked at random.

It is possible that you end up with three of the tallest or shortest students in the room. The more students you measure the more likely it is that your results are accurate.

In the zebra finch experiment the researchers assert “in the fifth set, the un-crested male received the most social time (42.1%), with the white-crested male in second place (37.5%).”

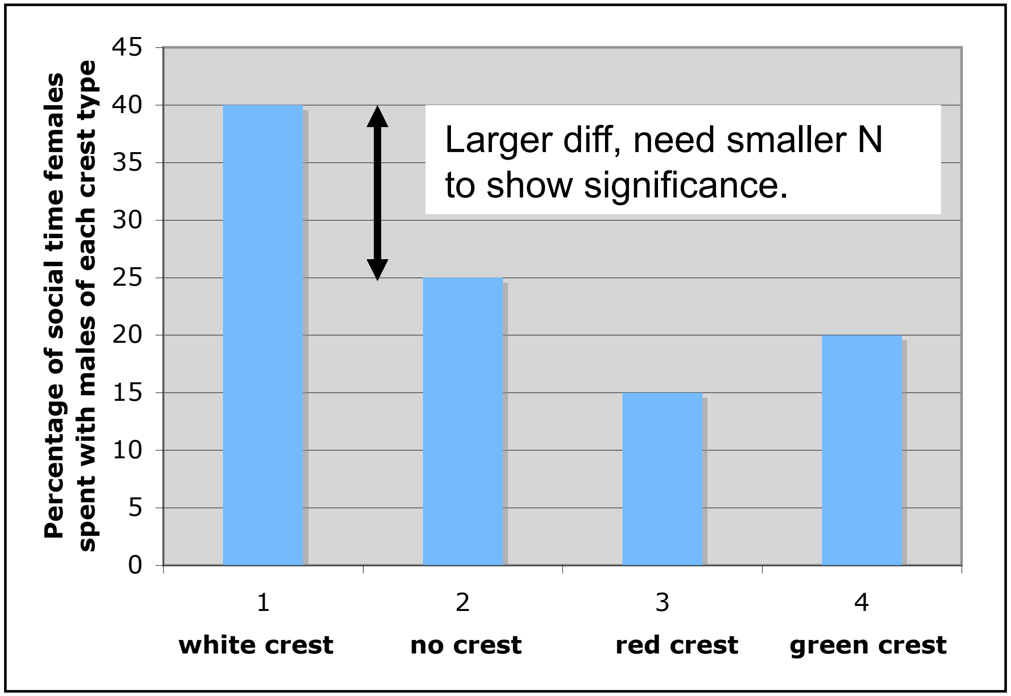
However, with a large enough sample size (N = number of female finches tested) it becomes clear that females actually spend significantly more social time with white-crested males than with other crest types.

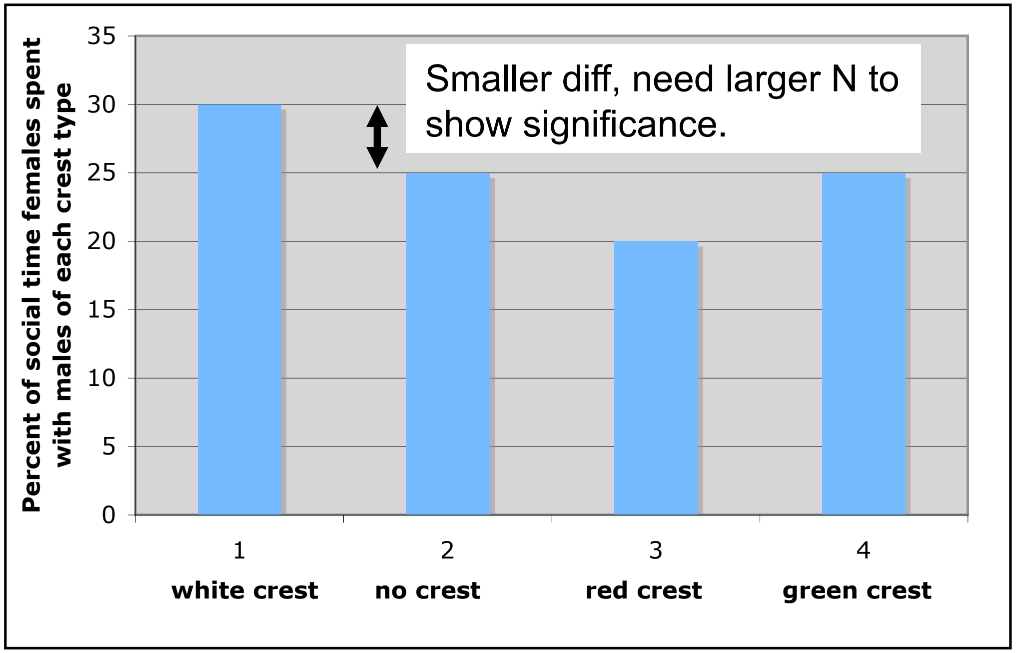
Do female zebra finches show a significant preference for males with white crests?

They do if the sample size (number of female finches tested) is large enough.

The bars represent the mean across multiple females.

As the number of females increases it becomes less likely that the difference in time spent with each crest type is due to chance alone.





So, what test was used for the Zebra finch study?

Statistical Tests
(rows=purpose of test, columns= type of data)

	Parametric	**Non-parametric**	**Binomial (two possible outcomes)**
Describe one group	Mean, SD	Median	Proportion
Compare a group to a hypothetical value	One-sample t test	Wilcoxon test	Chi-square test
Compare two unpaired groups	Unpaired t test	Mann-Whitney test	Fisher's test
Compare two paired groups	Paired t test	Wilcoxon test	McNemar's test
Compare three or more unmatched groups	One-way ANOVA (analysis of variance)	Kruskal-Wallis test	Chi-square test
Compare three or more matched groups	Repeated-measures ANOVA	Friedman test	Cochrane Q

The Friedman test produced a p-value of 0.02, or a 2% chance that the results were due to chance.

Since 0.02 is less than 0.05, this is statistically significant.

How about another example…

The eastern chipmunk trills when pursued by a predator, possibly to warn other chipmunks.

Burke da Silva et al. (2002) released chipmunks either 10 or 100 meters from their home burrow, then chased them (to simulate predator pursuit).

Out of 24 female chipmunks released 10 m from their burrow, 16 trilled and 8 did not trill.

When released 100 m from their burrow, only 3 female chipmunks trilled, while 18 did not trill.

Applying the correct test, the proportion of chipmunks trilling is significantly higher (P=0.0007) when they are closer to their burrow

Choosing the correct test..

- How many outcomes are there?
- How many groups are being compared?
- Are the groups matched/paired?

- There are two possible outcomes (the chipmunks either trill or do not trill) therefore these data are binomial.

- Two groups are being compared, the group of chipmunks released 10m from their home burrow and the group released 100m from their home burrow.

- The groups consist of different individuals and therefore are unpaired.

- So, according to the table …

Statistical Tests
(rows=purpose of test, columns= type of data)

	Parametric	**Non-parametric**	**Binomial (two possible outcomes)**
Describe one group	Mean, SD	Median	Proportion
Compare a group to a hypothetical value	One-sample t test	Wilcoxon test	Chi-square test
Compare two unpaired groups	Unpaired t test	Mann-Whitney test	Fisher's test
Compare two paired groups	Paired t test	Wilcoxon test	McNemar's test
Compare three or more unmatched groups	One-way ANOVA (analysis of variance)	Kruskal-Wallis test	Chi-square test
Compare three or more matched groups	Repeated-measures ANOVA	Friedman test	Cochrane Q

Fisher’s Exact Test

Data from Fisher’s Exact test fits into a 2X2 table, and if you copy this into the following handy website that accommodates a 2x2 table and hit ‘compute’ you should get a 2-tail p-value of 0.0007.

http://www.langsrud.com/fisher.htm

	10m	100m
trilled	16	3
Did not trill	8	18

… and finally, one of the most simple tests: Comparing two means.

Suppose you observe a bird at the zoo and measure the time spent pecking (1) three hours before feeding and (2) 20 minutes before feeding. You do this on six different occasions.

Our hypothesis is that birds peck more in anticipation of food. The null hypothesis is that the time spent is equal for the two conditions.

Here are some results (data are in minutes):

3 hours: 5, 7, 12, 3, 8, 10

20 minutes: 8, 15, 22, 18, 12, 14

Do we accept or reject the hypothesis?

What test?

3-hour group
20-minute group

So compare two groups. The birds are not paired up.

Is pecking rate normally distributed? We don’t know, so we will be conservative and pick a non-parametric test:

Statistical Tests
(rows=purpose of test, columns= type of data)

	Parametric	**Non-parametric**	**Binomial (two possible outcomes)**
Describe one group	Mean, SD	Median	Proportion
Compare a group to a hypothetical value	One-sample t test	Wilcoxon test	Chi-square test
Compare two unpaired groups	Unpaired t test	Mann-Whitney test	Fisher's test
Compare two paired groups	Paired t test	Wilcoxon test	McNemar's test
Compare three or more unmatched groups	One-way ANOVA (analysis of variance)	Kruskal-Wallis test	Chi-square test
Compare three or more matched groups	Repeated-measures ANOVA	Friedman test	Cochrane Q

We can plug these values into a web calculator:

http://elegans.swmed.edu/~leon/stats/utest.html

Copied from the web site:

U Test Results

n_1	n_2	U	P (two-tailed)	P (one-tailed)
6	6	33.0	0.01515152	0.00757576
normal approx z = 2.40192			0.01630918*	0.00815459*

*These values are approximate.

So with p value less than 0.02 we would reject H_0 and conclude it is unlikely that the results obtained were by chance, thus the means of the two experimental groups are truly different, and we accept our hypothesis.

Where to get statistical programs…

1. Web (as seen)
2. Purchase
3. Download open source freeware

Some for-purchase products:

PSU bookstore has SPSS

Other good products are SAS, JMP, Statistica, Minitab. SAS is free on the PSU mainframe (odin).

Most are very expensive, but with reasonable student versions

Best open source: R