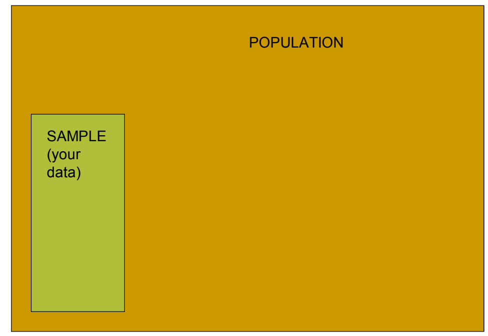
Everything you will ever need to know about statistics…?

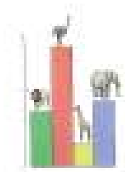
Larkin Powell
Polytechnic of Namibia
November 2009 workshop SNRT
lpowell3@unl.edu



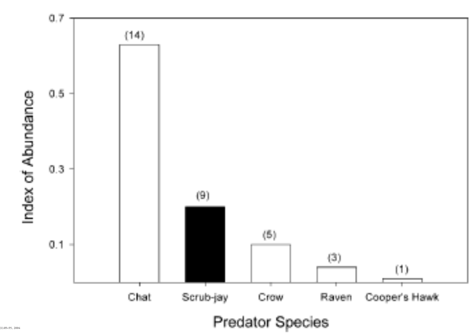
Drawing inferences about a population….using samples

Categorical data

- Summarizing responses to a categorical question, what is your favorite type of animal in Namibia?

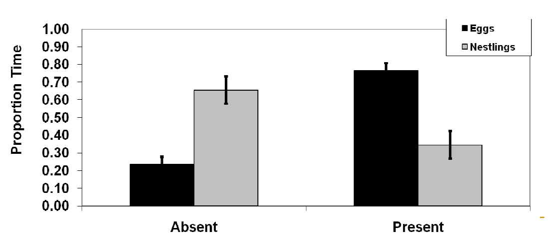


A better example



Categorial and continuous data

- Comparing continuous date between two categories: how often is the female bird absent from the nest?



Relationship between two sets of continuous data

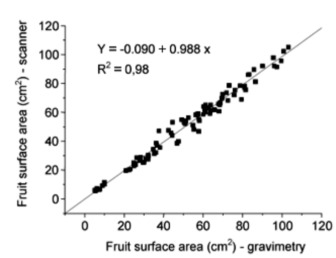


Figure 1 - Relationship between peach palm fruit surface area estimated by the digitalization and the gravimetric methods.

In general

- Categorial data:
  - Use chi-square test to see if frequencies of responses are similar between groups
- Continuous data:
  - Use t-test or ANOVA to see if means of a response differ between groups (groups are categorial data!)
  - Use correlation or regression to evaluate relationship between two continuous sets of data (e.g. rainfall and temperature)

Hypotheses and P-values

- Working: your statement of what you think
  - "I think fertilizer affects growth rates of plants."
- Null: usually, the opposite of what you think
  - Ho: Gf = Gnf
- Alternate: statistics statement of your working hypothesis
  - HA: Gf <> Gnf
- P-value: the probability that your null hypothesis is true
  - P = 0.23: 23% chance that your null is true
  - P = 0.02 2% chance that your null is true

Categorical data: frequency tables

- Goodness of fit: how well do your observations 'fit' your expectations?
- Are the data split among the categories in the way you thought they would be?
- Null hypothesis: the 'expected' proportions of your data.

Examples: calculating expected frequencies

Observed:

Heads	Tails
55	45

Expected:

Heads	Tails
100*?=?	100*?=?

Examples: calculating expected frequencies

A sheep breeder promises the following ratio of offspring from an arranged breeding: 8 brown: 2 white: 3 spotted

A farmer complains that he got fewer brown sheep than expected and wants his money back. From 250 lambs, how many brown, white, and spotted individuals were expected?

Brown: 8/13	= 0.62	0.62*250 = 154
White: 2/13	= ?	?*250 = ____
Spotted 3/13	= ?	?*250 = ____

Chi-square: goodness of fit test

- Compares observed frequency distribution to expected distribution

$$X^2 = \sum_{i=1}^{k} \frac{(O-E)^2}{E}$$

NOTE: the bigger the difference between observed and expected, the bigger the chi-square statistic! So, as chi-square gets bigger, we are showing that our groups may be different than we expected.

Coin flip example:

55 heads, 45 tails

K = number of categories

$$X^2 = \frac{(55-50)^2}{50} + \frac{(45-50)^2}{50}$$

= 0.5 + 0.5 = **1.0**

Degrees of freedom

- The chi-square statistic can <u>also</u> get larger the more categories you have. So, we have to take the number of categories into account.
- Degrees of freedom: The number of independent pieces of information that go into the estimate of a parameter
  - number of values in the final calculation of a statistic that are free to vary

Degrees of freedom

- For simple (1 dimension) comparisons, df = k -1
  - k = number of categories
- In our coin toss example, df = 2-1 = 1

Determining 'statistical significance'

Our null hypothesis: coin is not weighted (50:50 chance of heads)

Alternative: coin is weighted (not 50% chance of heads)

- In our example:
  - Chi-square = 1.0
  - df = 1
  - P = a value between 0.5 and 0.25 (see table on page 112 of study guide)
  - Do we reject the null (is our observation different than expected)?
    - If $P < 0.05$, then there is less than a 5% chance that your null is correct. If $P < 0.05$, <u>reject null</u>.
    - If $P > 0.05$, there is a >5% chance that the null is correct. If $P > 0.05$, <u>do not reject null</u>.

COMPUTER BREAK!!

Contigency tables: comparing >1 group

	Species 1	Species 2	Species 3
Lived	8	6	13
Died	12	14	7

Null hypothesis: Species 1, 2, and 3 survived at the same rate.

How to get expected values

	Species 1	Species 2	Species 3	Row totals
Lived	8	6	13	27
Died	12	14	7	33
Column totals	20	20	20	**60**

Expected value for each 'cell' = (column total * row total) / grand total

Expected value for "very stressed females" = (20*33)/60 = **11**

Observed	Species 1	Species 2	Species 3
Lived	8	6	13
Died	12	14	7

Expected	Species 1	Species 2	Species 3
Lived	9	9	9
Died	11	11	11

$(8-9)^2/9 = 0.11$ $(6-9)^2/9 = 1.0$ $(13-9)^2/9 = 1.78$

$(12-11)^2/11 = 0.09$ $(14-11)^2/11 = 0.81$ $(7-11)^2/11 = 1.45$

$$X^2 = \sum_{i=1}^{k} \frac{(O-E)^2}{E}$$

Chi-square = 5.24 df = (rows-1)*(cols-1) = 1*2 = 2 $P < 0.05$

COMPUTER BREAK!!

Descriptive stats: central tendency

- How do we summarize continuous data?
  - Does data accumulate around a certain value?
- Mean, median, mode…

Mean

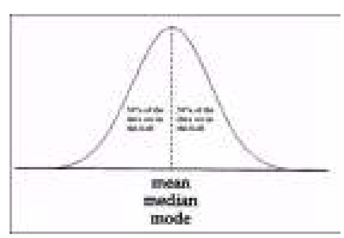
- The mean of a population: μ (mu)
- The mean of a sample: $\bar{x}$
- Calculated as: $\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$
- So, the mean of 9, 15, 22, 22, 24, 29, 30 =
- (9+15+22+22+24+29+30)/7 = 21.6

Mode and median

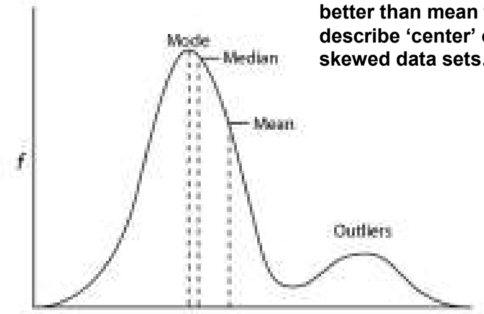
- Median: middle number in ordered list
  - 9, 15, 22, **22**, 24, 29, 30
- Mode: number repeated more than any other
  - 9, 15, **22, 22**, 24, 29, 30

Descriptions of continuous data: what is 'Normal' data

- Large samples
- 'even' distribution around the mean
- The mean is a good descriptor of the central tendency of 'normally distributed' data

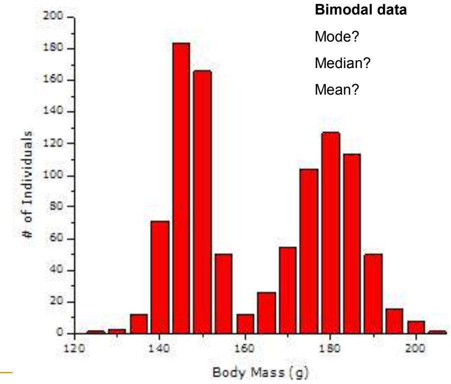


http://www.duncanwil.co.uk/norm_files/image007.jpg



Mode or median are better than mean to describe 'center' of skewed data sets…

http://www.biomedcentral.com/content/figures/1471-2105-4-31-1.jpg



Bimodal data

Mode?

Median?

Mean?

Standard deviation: 'average distance from the mean'

- "Variance" gives us squared-units—normally does not make sense (kg^2???)

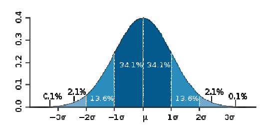
$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2$$

- So, to change it back to 'regular' units, we take the square root. This is the SD:

$$s = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})^2},$$

So, if s^2 = 20.3, **s = 4.5 (take square root of 20.3)**

Using SD to make confidence intervals

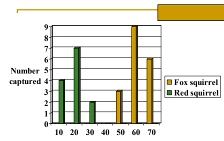


68% of measurements are within 1 SD of the mean.

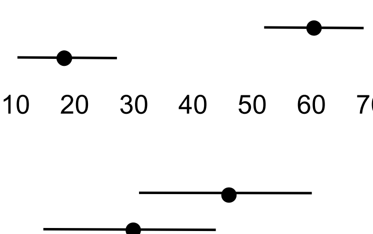
95% of measurements are within 2 SD of the mean.

99.7% of measurements are within 3 SD of the mean.

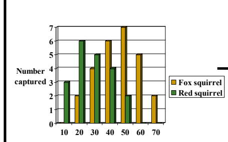
We often use a 95% confidence interval. We can say "if we do the experiment again, we are 95% certain the mean of our new sample would fall within this confidence interval (2 SD from your mean)."



Using confidence intervals



10 20 30 40 50 60 70



COMPUTER BREAK!!

Tests of normality—useless? Mostly.

- Usually done to determine:
  - Are data normally distributed?
  - Can I use a parametric test (t-test, ANOVA)?
- Many statisticians suggest “normality tests are less useful than you’d guess.”
  - Small samples almost always pass test for normality
  - Large samples (more power) will fail test because of small deviations which will not affect t-test/ANOVA
  - “The best way to evaluate your data is to look at a graph and see if the distribution deviates grossly from a bell-shaped normal distribution.”

www.graphpad.com/library/BiostatsSpecial/article_197.htm

More reasons to not worry about normality tests

- Data do NOT need to be normally distributed to use a t-test or ANOVA
  - The means of the sample need to be normally distributed
  - This will be assured for all but the most “perverse” data
- Non-parametric tests do not test for differences in means
  - They test whether the data distributions are identical between two sets
  - A ‘significant result’ only tells you that they differ in some way (mean, variance, shape, etc.)
  - We normally want to infer something about the parameters (e.g., means) estimated by parametric tests. It is illogical to use a non-parametric test and then report a mean…!

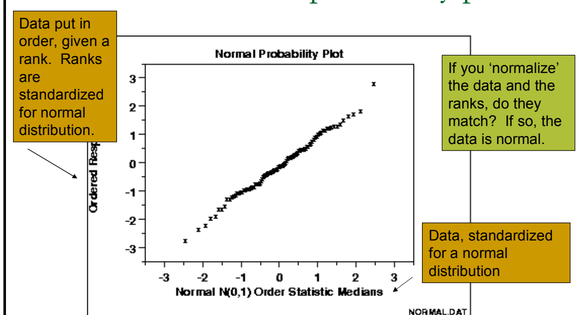
Johnson, D. H. 1995. Statistical sirens: the allure of nonparametrics Ecology 76: 1998-2000.

Final quote

- *“If ecologists are careful about randomly sampling the populations about which they want to draw inferences, standard parametric methods will ordinarily be adequate; if they are not nonparametric methods will not protect them from sailing off course.”*

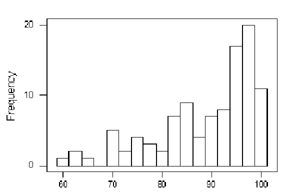
Johnson, D. H. 1995. Statistical sirens: the allure of nonparametrics Ecology 76: 1998-2000.

A “normal” normal probability plot

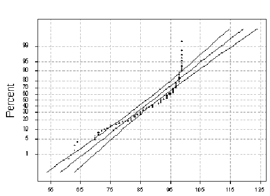


Fat pencil test: can points be covered by a ‘fat pencil’? If so, the data is approximately normally distributed.

Non-normal data…



The year printed on pennies (1970, 1980, 1990, 2000, etc.)



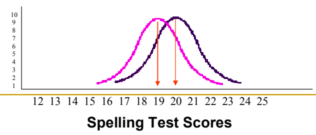
COMPUTER BREAK!!

Suppose we conducted a study to compare two strategies for teaching spelling.

Group A had a mean score of 19. The range of scores was 16 to 22, and the standard deviation was 1.5.

Group B had a mean score of 20. The range of scores was 17 to 23, and the standard deviation was 1.5.

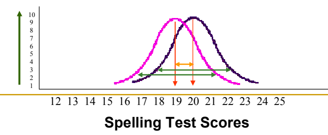
How confident can we be that the difference we found between the means of Group A and Group B occurred because of differences in our reading strategies, rather than by chance?



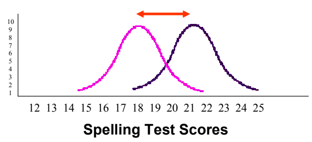
A t-test allows us to compare the means of two groups and determine how likely the difference between the two means occurred by chance.

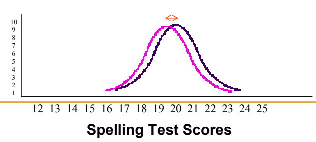
The calculations for a t-test requires three pieces of information:

- the difference between the means (mean difference)
- the standard deviation for each group
- and the number of subjects in each group.



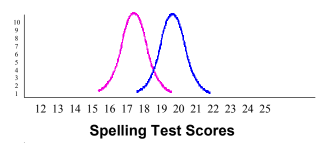
All other factors being equal, large differences between means are less likely to occur by chance than small differences.

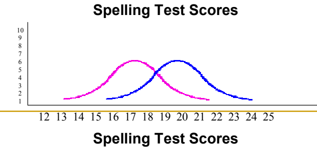




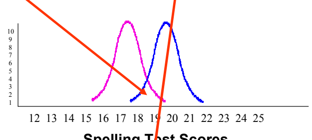
The size of the standard deviation also influences the outcome of a t-test.

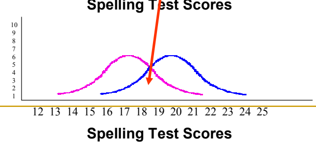
Given the same difference in means, groups with smaller standard deviations are more likely to report a significant difference than groups with larger standard deviations.



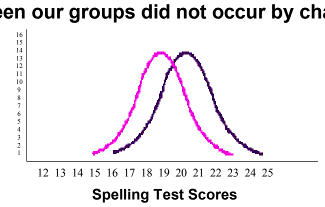


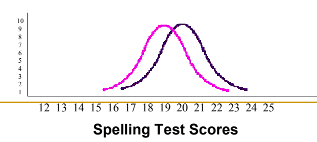
From a practical standpoint, we can see that smaller standard deviations produce less overlap between the groups than larger standard deviations. Less overlap would indicate that the groups are more different from each other.





The size of our sample is also important. The more subjects that are involved in a study, the more confident we can be that the differences we find between our groups did not occur by chance.





Comparisons involving means

- **Comparing a sample's mean to some hypothesized value**
  - Seed packets are supposed to hold 200g. We took a sample of 30 packets. Is the mean statistically different than 200g?
  - Two-tailed t-test

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

df = n-1

Null hypothesis: no difference between sample and hypothesized value (standard)

If P<0.05, reject the null. There is a difference.

Comparisons involving means

- **Comparing the means of two independent samples**
  - The weight of kudu bulls are compared in two samples from farms in northwestern Namibia and farms near Windhoek.
  - Sampling one herd does not affect the sampling of the other herd.
  - Standard t-test (2-tailed)

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

df = $(n_1\text{-}1)\text{+ }(n_2\text{-}1)$

Null hypothesis: no difference between means

If P<0.05, reject the null. There is a difference.

Comparisons involving means

- **Comparing the means of two <u>dependent</u> samples**
  - The weight of kudu bulls are compared in two samples from before winter and after winter.
  - Each bull is weighed twice. The samples are not independent (same animals).
  - t-test of differences (2-tailed)

$$t = \frac{\bar{d} - \mu_d}{s_d / \sqrt{n_d}}$$

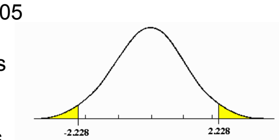
df = n-1

Null hypothesis: difference is 0

If P<0.05, reject the null. The animals do not weigh the same.

Two-tailed t-tests

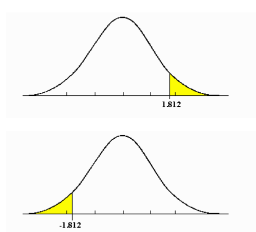
- Suppose we use the 0.05 level of significance.
- A two-tailed test divides the 0.05 into the two tails: 0.025 in each.
- Performed if the results will be interesting if difference occurs in either direction (Pop 1 is either greater or less than Pop 2).



Critical t-value corresponding to 0.05 level of significance.

One-tailed t-tests

- Suppose we use the 0.05 level of significance.
- A one tailed test places all 0.05 in one tail (you decide which one).
- Performed if the results are only interesting if they occur in a certain direction (Pop 1 is greater than Pop 2).



Critical t-value corresponding to 0.05 level of significance.

HOW TO: t-test between means is conducted the same way, except that you use 2*alpha to look up the critical t-statistic. For example, instead of 0.05, you use 0.10.

Comparing 1-tailed and 2-tailed

- Are your workers putting more than 200g in the seed packets?
  - Mean content = 202 g
  - SD = 4.431 g
- Two-tailed question: is mean more or less than 200g?
  - Use 0.05 to find critical t value.
- One-tailed: is mean more than 200g?
  - Use 0.10 to find critical t value.

df = n-1

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

T-test: car mileage comparison

t-Test: Two-Sample Assuming Equal Variances

	Variable 1	*Variable 2*
Mean	**16.02531646**	**30.48101266**
Variance	17.61473548	37.30412204
Observations	79	79
Pooled Variance	27.45942876	
Hypothesized Mean Difference	0	
df	156	
t Stat	**-17.33771307**	
P(T<=t) one-tail	1.63382E-38	
t Critical one-tail	1.654679996	
P(T<=t) two-tail	**3.26764E-38**	
t Critical two-tail	**1.975287473**	

COMPUTER BREAK!!

Wilcoxen Rank Sum Test

- To compare two continuous samples
- Non-parametric (does not assume normality)
- Uses 'ranks' of the data

Weights of kudu's harvested by two Professional Hunters. Are the weights of PH A heavier than PH B?

PH A	PH B
600	620
650	640
633	650
700	500
710	705

PH A	PH B
600	620
650	640
633	650
700	500
710	705

	500	600	620	633	640	650	650	700	705	710
PH	B	A	B	A	B	A	B	A	B	A
Rank	1	2	3	4	5	6	7	8	9	10

	500	600	620	633	640	650	650	700	705	710
PH	B	A	B	A	B	A	B	A	B	A
Rank	1	2	3	4	5	6.5	6.5	8	9	10

Sum of ranks:

PH A: 2+4+6.5+8+10 = 30.5

Ph B: 1+3+5+6.5+9 = 24.5 ←----- smaller of the two rank sums

Calculate U, where W = smaller rank sum, n1 is size of smallest sample, and n2 is size of other sample.

$$U = W - \frac{n_1(n_2+1)}{2} \qquad U = 24.5 - \frac{5(5+1)}{2} = 24.5 - 15 = 9.5$$

Now, use data to calculate:

$$\mu_U = \frac{n_1 n_2}{2} = \frac{5*5}{2} = 12.5$$ ←mean of U

$$s_U = \frac{\sqrt{n_1 n_2 (n_1 + n_2 + 2)}}{12}$$ ←SD of U

$$s_U = \frac{\sqrt{5*5(5+5+2)}}{12} = \frac{\sqrt{25*12}}{12} = \frac{17.3}{12} = 1.44$$

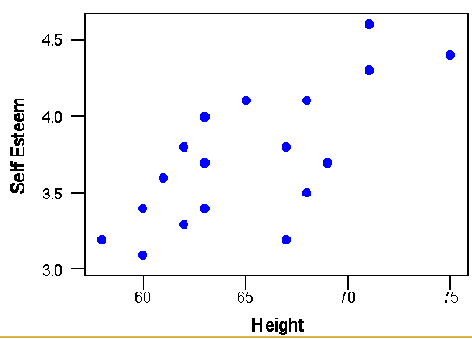
Now, calculate the Z statistic:

$$z = \frac{U - \mu_U}{s_U} = \frac{9.5 - 12.5}{1.4} = -2.14$$

Critical value of Z for P=0.05 is 1.96.
--Drop the negative sign and compare. Is 2.14 greater than 1.96? If so, reject the null hypothesis.

COMPUTER BREAK!!

Correlation: are two things related?



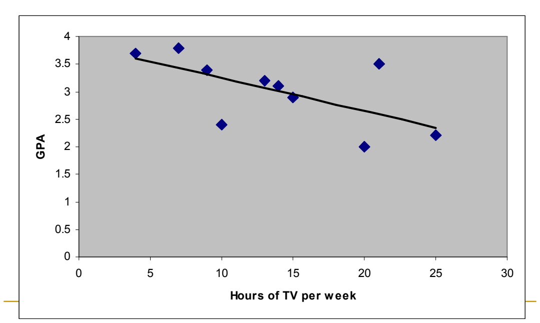
NOTE: correlation does not imply causation!

Correlation analysis: r-statistic

R-statistic measures how far (on average) each data point is from the line drawn through the 'cloud'. The farther, on average, data is from the line, the LOWER the r-statistic.

R ranges from -1 to 1

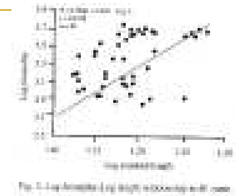
Correlation analysis: TV hours correlated with low grades?



COMPUTER BREAK!!

Linear regression

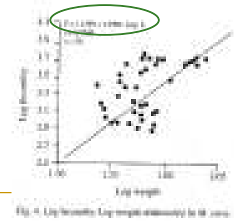
- Describes the exact relationship (estimates a slope)
- Use of linear regression IMPLIES causation, but does not guarantee causation
- Allows prediction from your sample within the range of data you collected



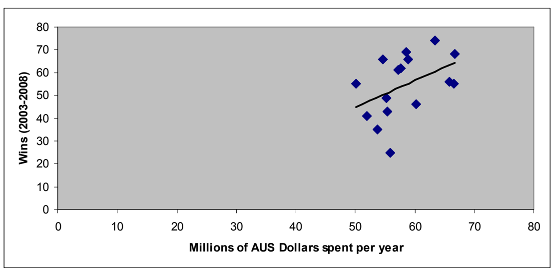
y = mx + b

m = slope

b = intercept



Regression analysis: does more money spent on football result in more wins?



COMPUTER BREAK AND END!!