

Mystery Foot Activity

- What might we be able to tell about the individual who left these footprints
- Possible hypothesis'?



k2992722 www.fotosearch.com

Hypothesis Testing

- to test the viability of the null hypothesis in the light of experimental data. Depending on the data, the null hypothesis either will or will not be rejected as a viable possibility.
- Null hypothesis
 - proposes a general or default position, such as that there is no relationship/difference between two measured phenomena, or that a potential treatment has no effect.
- Alternate hypothesis
 - is a statement of what a statistical hypothesis test is set up to establish. For example, in a clinical trial of a new drug, the alternative hypothesis might be that the new drug has a different effect, on average, compared to that of the current drug.

Mystery Foot

- H_N : There will not be a relationship between ones foot length and height. If there is a small correlation it is due to chance.
- H_A :

IB math requirements for Biology

- **Add, subtract, multiply & divide**
- **Carry out calculations involving means, decimals, fractions, percentages and ratios**
- **Represent and interpret frequency data in the form of bar charts, graphs and histograms, including direct and inverse proportion**
- **Plot graphs involving 2 variables that show linear or non-linear relationships**
- **Plot and interpret scattergraphs to identify a correlation between two variables and appreciate that the existence of a correlation does not establish a causal relationship**
- **Determine mode, median and calculate standard deviation**
- **Select statistical tests appropriate for the analysis of data and interpret the results.**

Statistics

- Mean: Ave
- Median
- Mode
- Range: =Max(B2:B10)-Min(B2:B10) enter

Statistics

- Standard deviation – unbiased estimate of the variability of a population. Reveals the predicted limits w/in which you can make inferences about populations
- Be able to calculate on calculator

$$\sqrt{\frac{\sum(X - \bar{X})^2}{(n - 1)}}$$

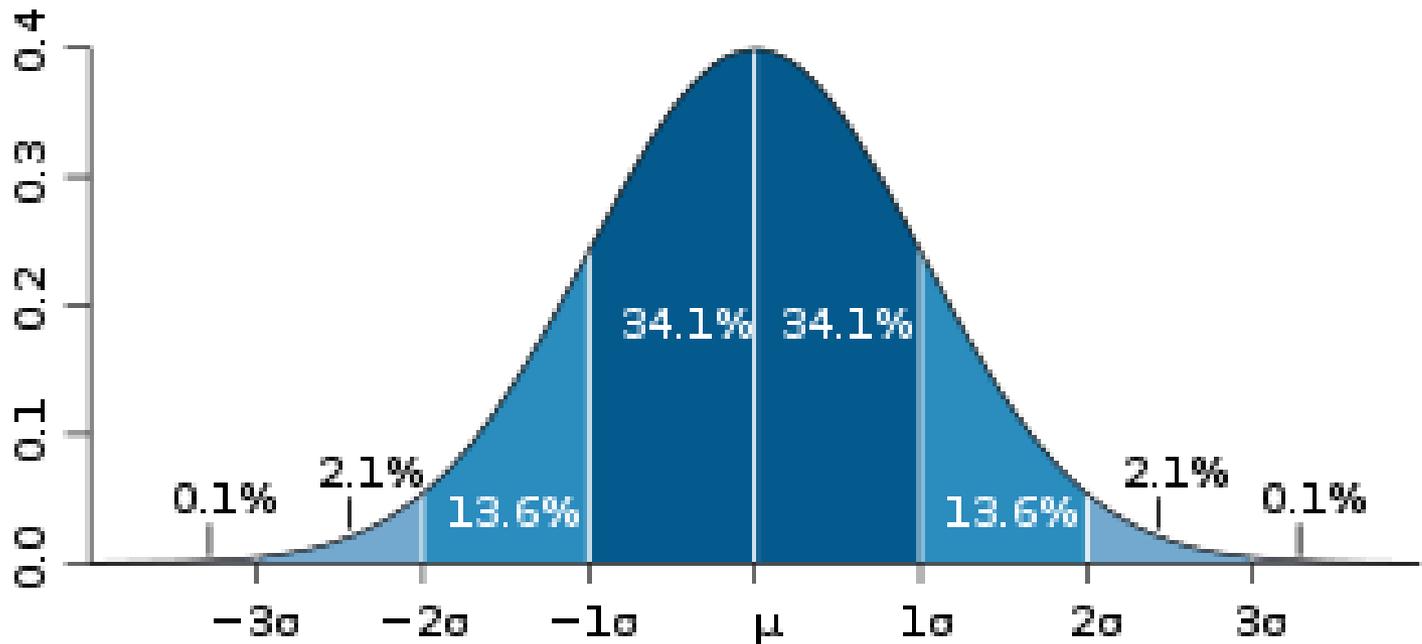
where:

X = each score

\bar{X} = the mean or average

n = the number of values

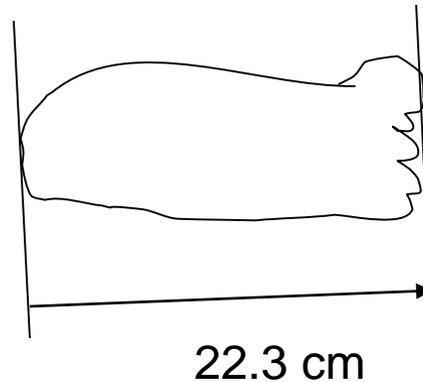
Σ means we sum across the values



68% within 1 standard deviation
95% within 2 standard deviations
99% within 3 standard deviations

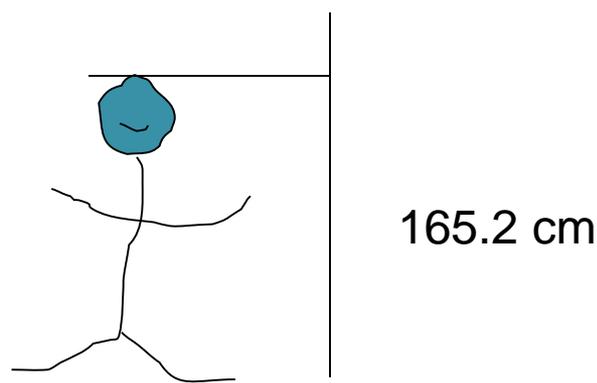
Measuring foot length protocol

- Barefoot
- Stand up straight on top of paper
- Mark back of heel and longest toe
- Measure to nearest 0.1 cm for both feet
- Average



Measuring height protocol

- Barefoot
- Stand upright, heel against wall
- Using straight edge across top of head mark measuring tape
- Measure to nearest 0.1 cm



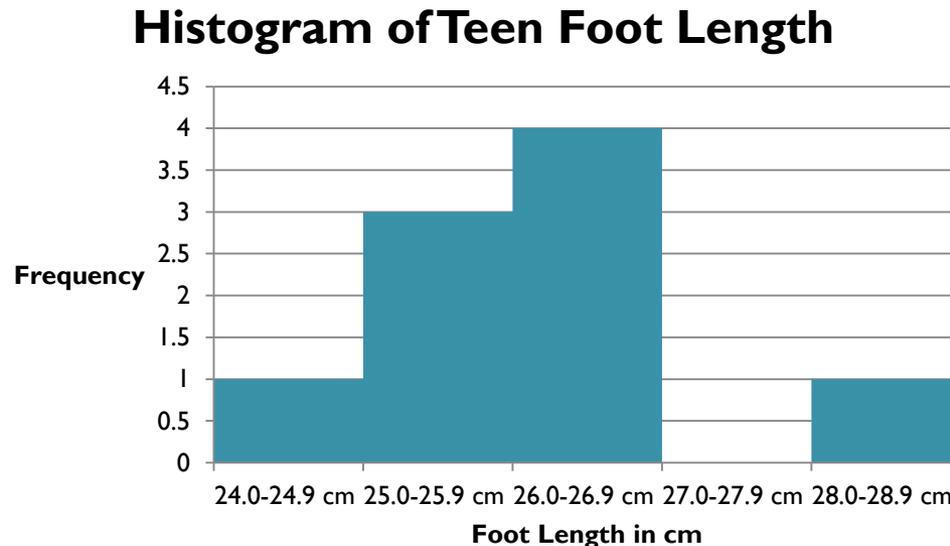
Stem Plot & Frequency Table for Foot Length

24	0.5			
25	0.1	0.5	0.5	
26	0	0	0.5	0.5
27				
28	0.5			

24.0-24.9 cm	1
25.0-25.9 cm	3
26.0-26.9 cm	4
27.0-27.9 cm	0
28.0-28.9 cm	1

Histograms

- Frequency graphs
- Highlight frequency table, column graph, right click on bar format data series, options gap = 0



Homework

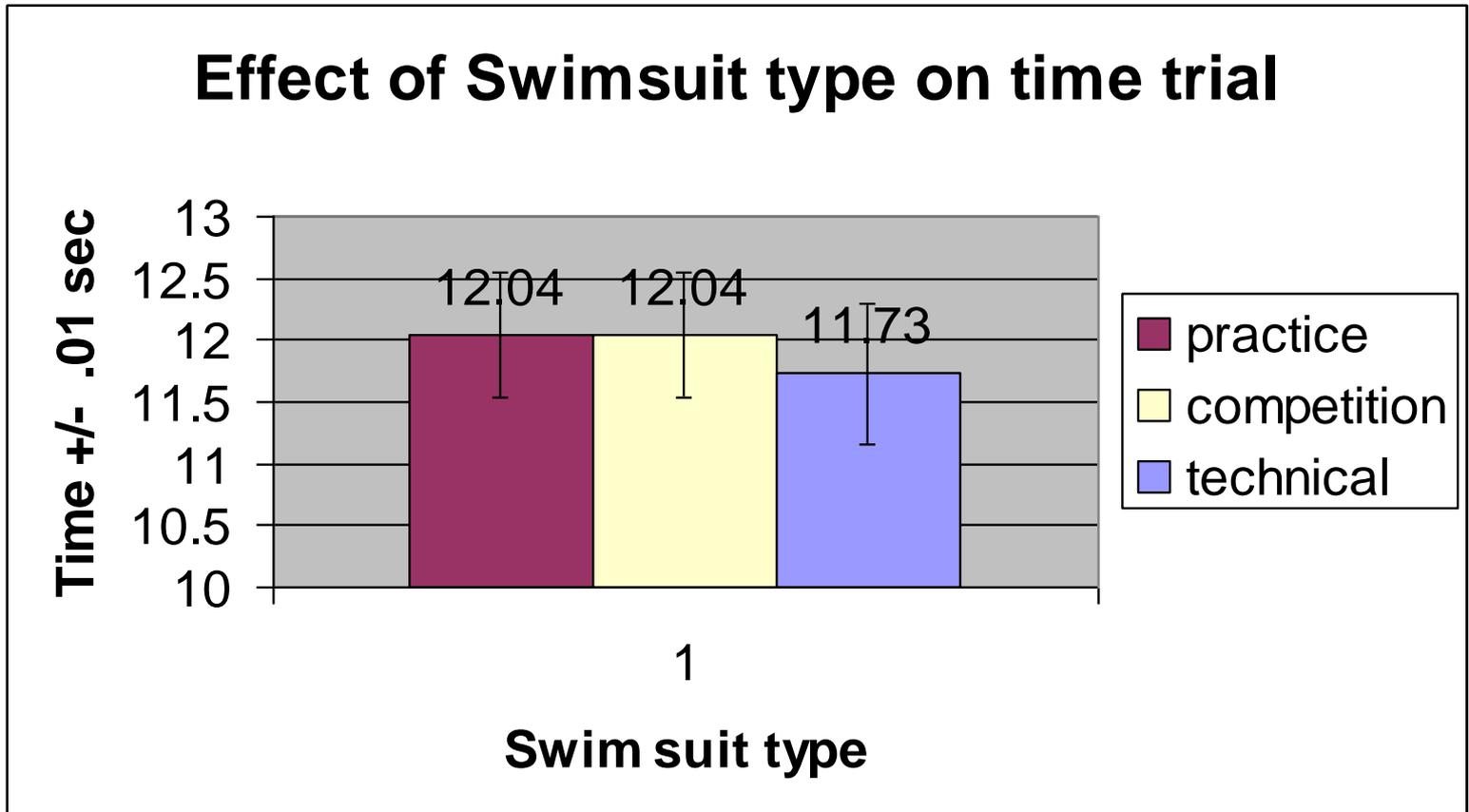
- Determine the mean, median, mode, range, standard deviation for each of the 8 data sets using excel. Choose 2 data sets to make 2 stem plots, 2 frequency tables and 2 histograms (other than teen male foot length)



Can the statistics we did last night (range, mean, median, mode, St Dev, histograms) help us determine if our original hypothesis is supported?

- If foot length increases then height will also increase because ... (Positive or negative relationship?)
- OR there will be no relationship between foot length and height.

Standard Deviation error bars



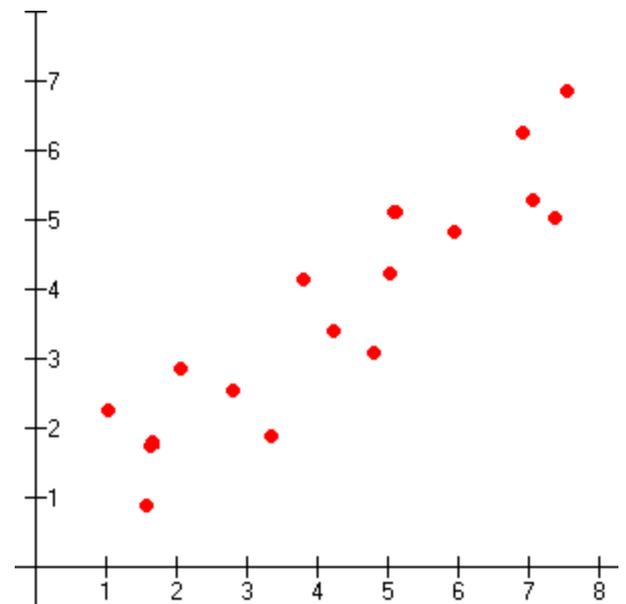
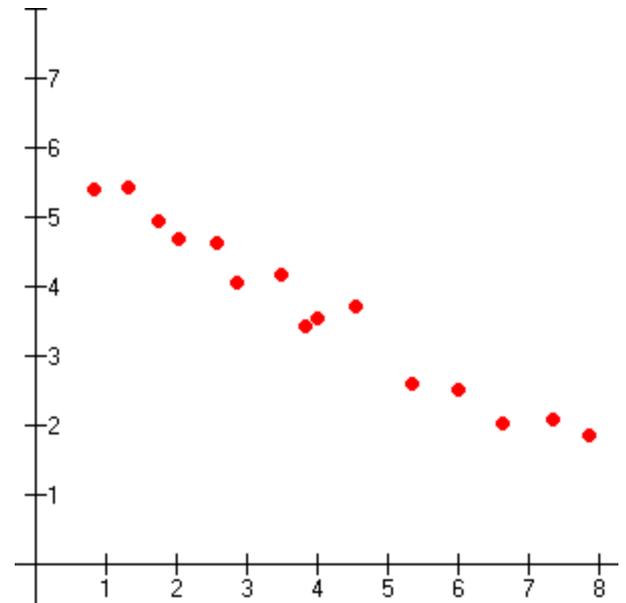
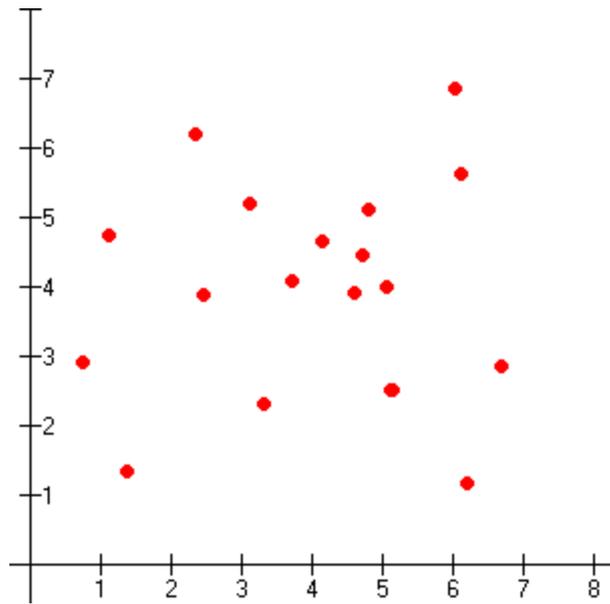
Is there a significant difference in swim suits?

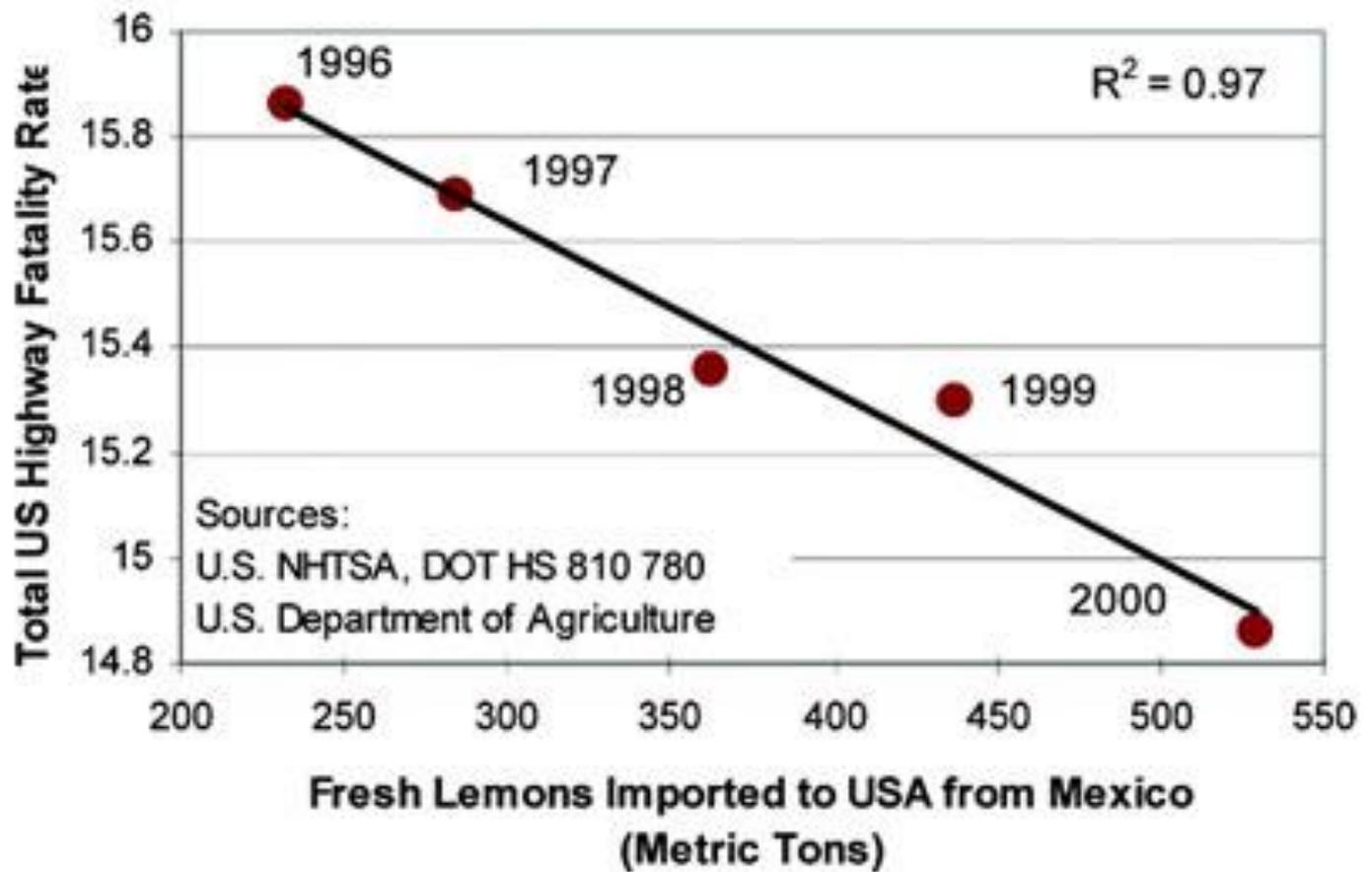
Making custom standard deviation bars

- Make table of ave & St Dev
- Insert bar graph of averages
- use layout option to add error bars
- more error bar options custom,
- specify values and then highlight st dev values for both upper & lower limits

Correlation (Pearson) r

- Measures strength of a linear relationship
- Correlation coefficients – how strong is the relationship
- $+1$ positive correlation
- -1 negative correlation
- 0 no relationship between the 2 measures





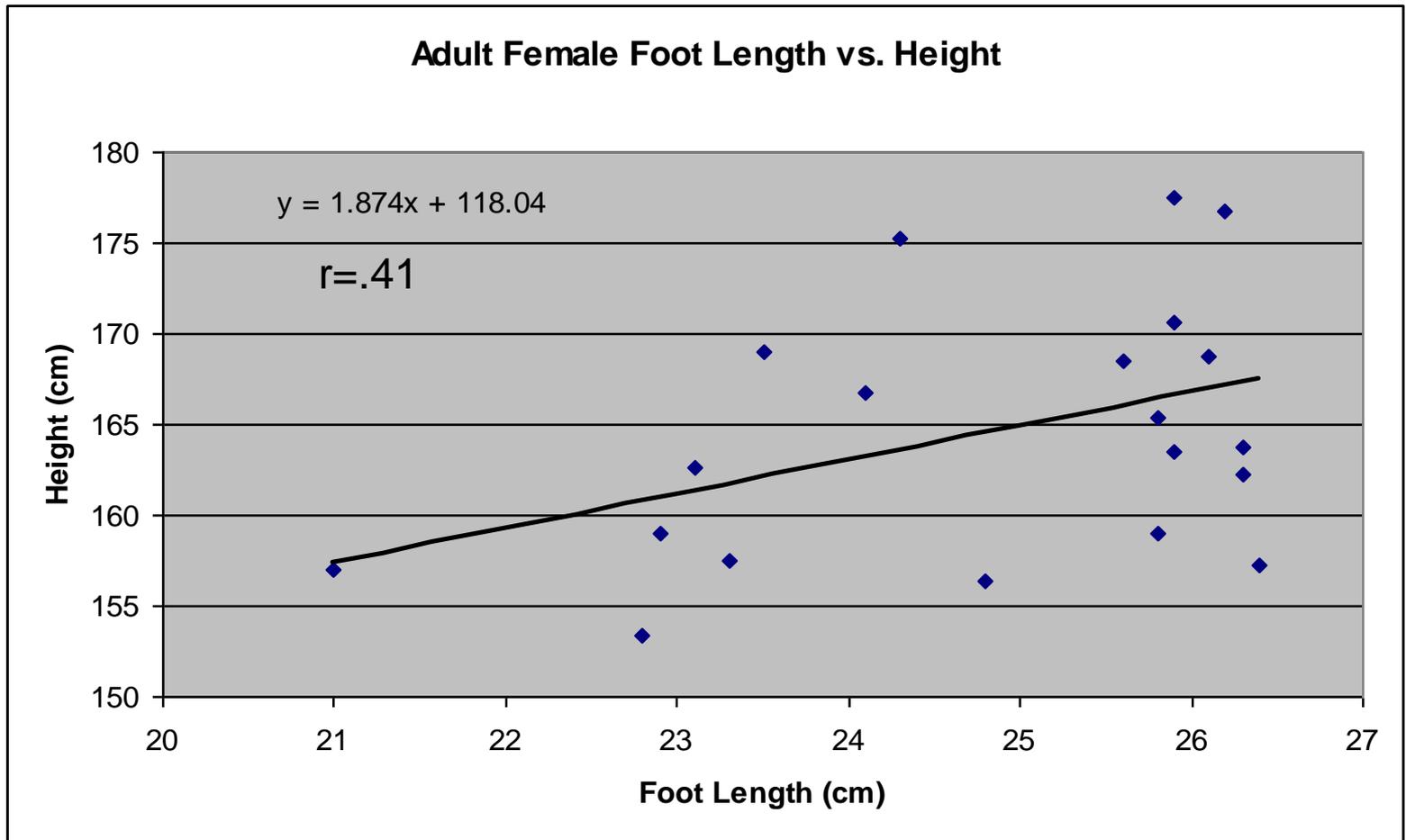
Correlation doesn't necessarily mean causation

Determine r with pearson coefficient using excel

Least Squares Regression Line Trendlines on Scatterplots

- is a method for finding a line that summarizes the relationship between the two variables, at least within the domain of the explanatory variable, x . The least-squares regression line (LSRL) is a mathematical model for the data

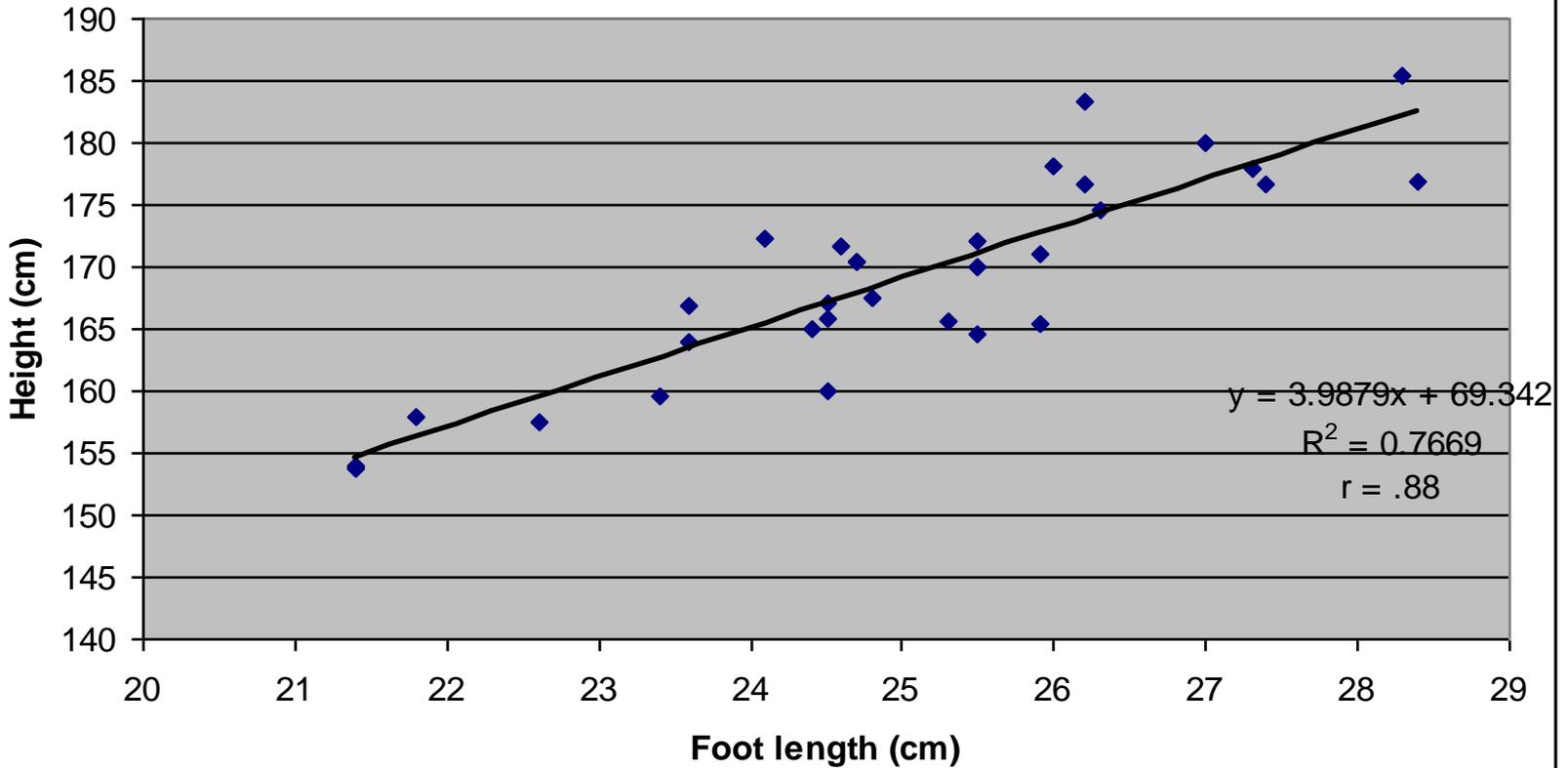
Scatterplot with trend line (LSRL) and correlation coefficient



Correlation Squared r^2

- The fraction of the variance of one variable that is explained by LSRL on the other variable
- If $r^2 = 1$ then all the variation in one variable is accounted for by the linear relationship with other variable
- The higher the r^2 the better the trendline equation is in predicting x or y

Teen Foot Length vs Height



There is a positive relationship between foot length and height. ($r=.88$)

I can be 77% confident that I can use the line equation to predict x or y

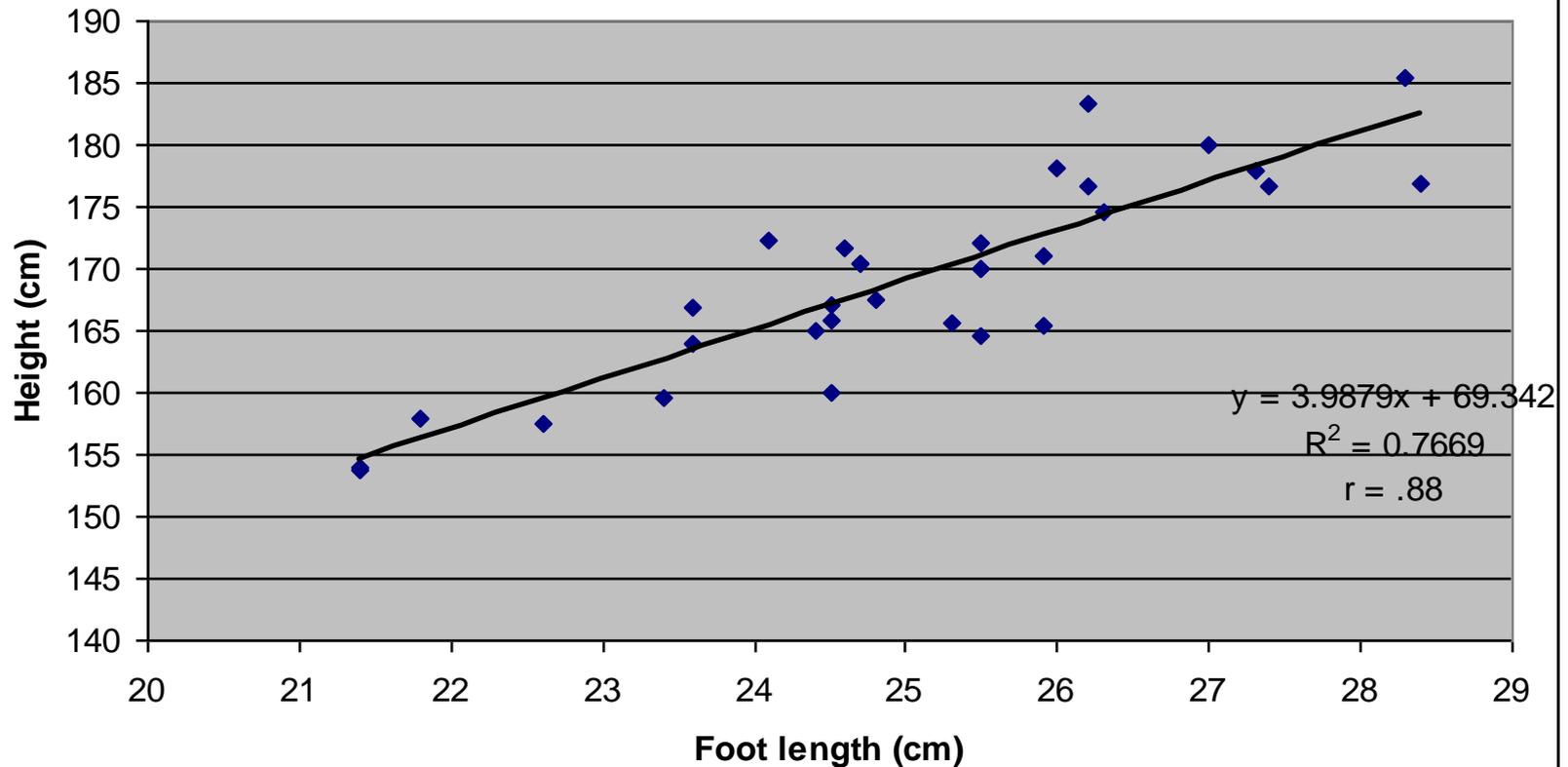
Homework

- Make a bar graph of foot length ave & st. dev error bars for all 4 groups
- Make a scatterplot with trendline (equation), r & r^2 values for all data, all teens, all adults and 2 other data sets
- Write some conclusions about the original 2 hypothesis, commenting on which groups have a stronger correlation & why

Critical Values for r

- Level of significance depends on number of trials in data set.
- Use Critical r value table and 95% level of significance
- Male teen $r=.8$, 24 pairs of data
- 95% significance needs an $r = .41$
- With $r = .8$ we are over 99% confident that there is a positive relationship
- Smaller sample size needs larger r values

Teen Foot Length vs Height



There is a positive relationship between teen male foot length and height. ($r=.88$)

I am 99% confident in the positive correlation according to the critical r value table.

I can be 77% confident that I can use the line equation to predict x or y

Discussion of correlations between teens & adults, males vs females.

Comparing Means – T - test

- Compares 2 means
- Miracle Grow ave = 15 cm vs control ave = 13 cm Is there a significant difference? How about 20 cm vs 15 cm?
- To obtain p value Best to use Vassarstats
- 2 tailed allows you to test null hypothesis
- Biology requires 95% certainty $p < .05$ if less than .05 results are significantly different and can reject null. If greater than .05 can't reject null, results are not significantly different

T-test continued

- Teen female height 165.3
- Adult female height 164.5
- $P = .70$
- Can't reject null, only 30% confident that the averages are significantly different, 70% chance of error
- Practice t-test

ANOVA

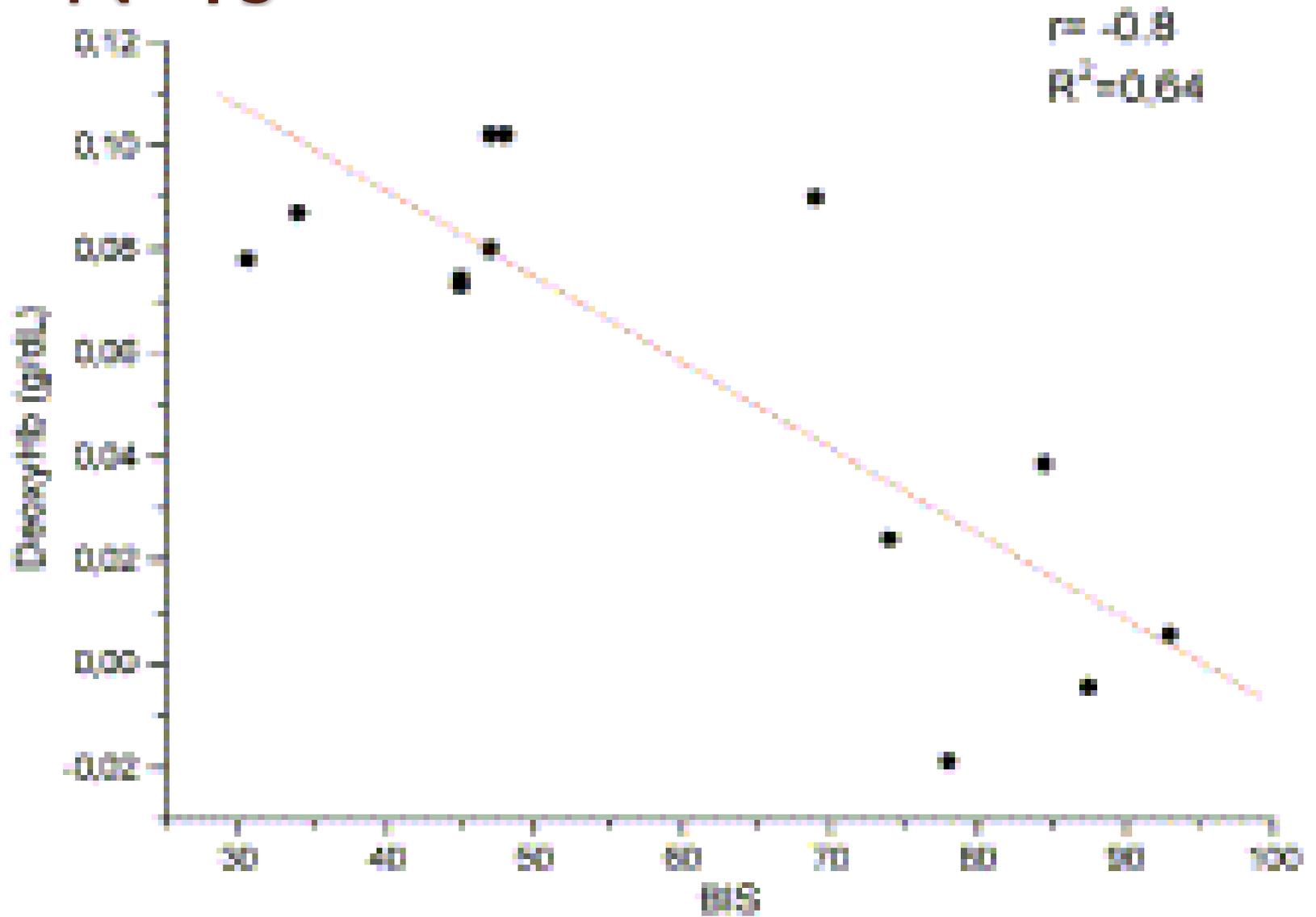
Analysis of Variance

- Compares 3-5 sets of means
- Gives p value
- Use Vassarstats
- Comparing 4 sets of foot data
- 26.4 24.2 27.1 24.8
- $P = <.0001$
- Over 99% confident that there is a significant difference in the 4 sets of foot data means, can reject null

Homework

- Write 5 H_A and 5 H_N for
- **means comparisons**
 - Eg H_A Adult males are taller on average than adult females
 - H_N There is no difference in height between adult males and females
 - Evaluate each with a T-test, write a concluding statement.
 - Use ANOVA to evaluate the ANOVA practice set, write a H_A & H_N first

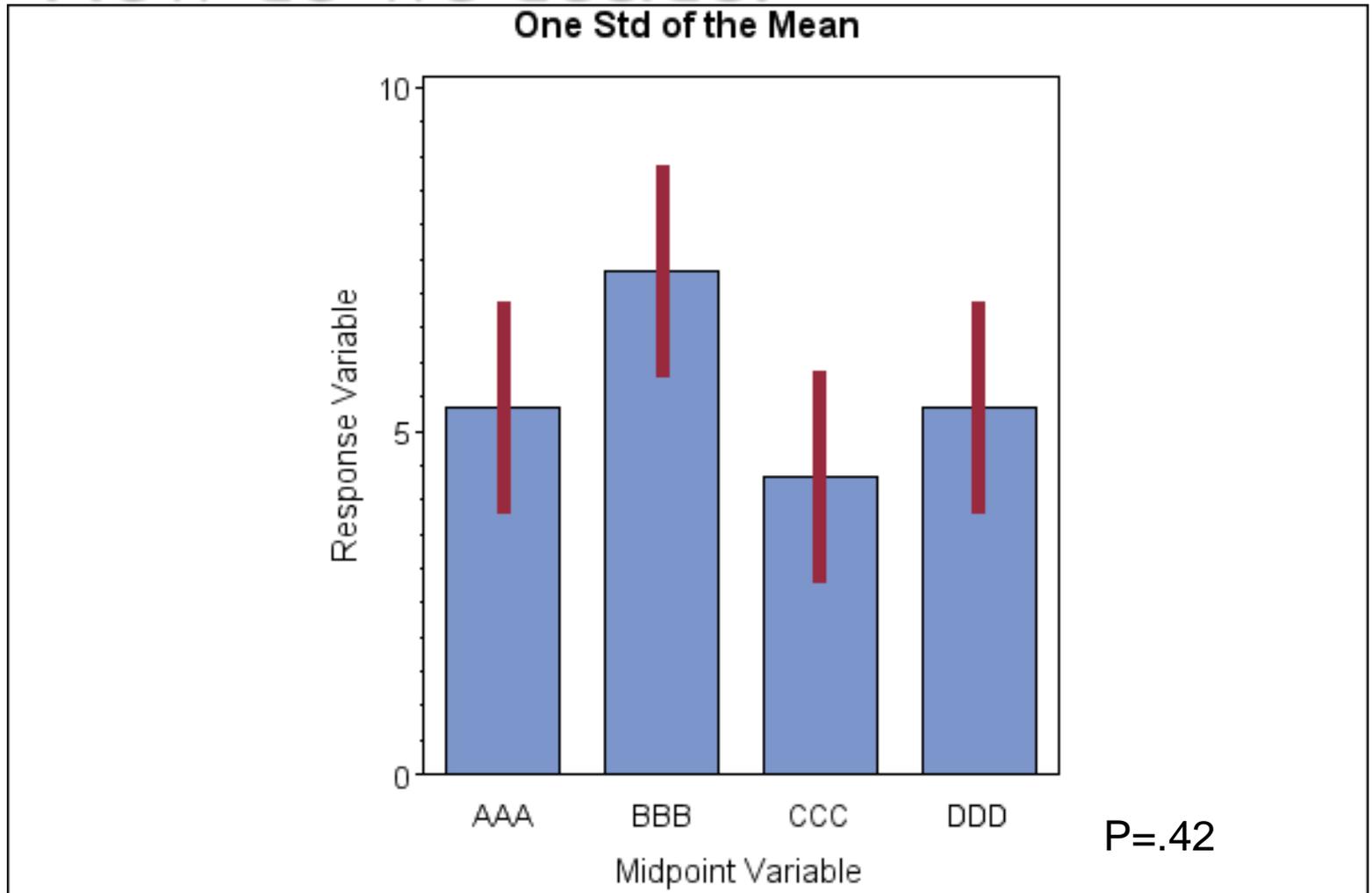
N=13



- Over 99% confident that there is a strong negative correlation $r = -.8$ (critical r value for 99% with 14 samples only needs to be .74)
- We can be 64% confident in using the line equation (LSRL) for predicting x or y ($r^2 = .64$)

Are the differences significant?

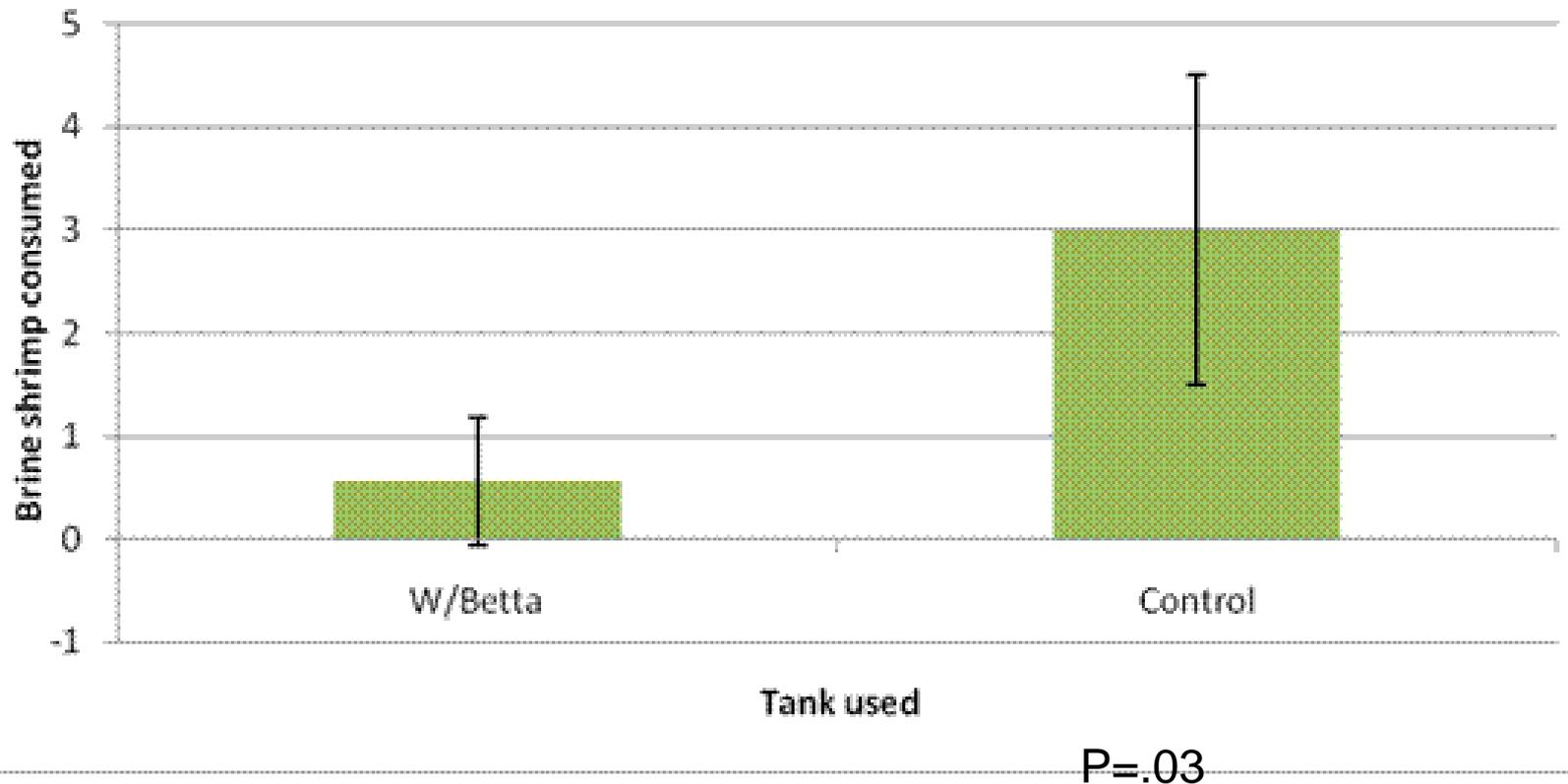
How do we decide?



Anova

- $P=.42$
- Can't reject the null; we only have 58% confidence that the averages of the 4 groups are significantly different ($.42 > .05$)
- (42% chance of error)
- Standard deviation error bars all overlap a little bit with each other.

Is there a difference between the 2 experimental groups? How do we find out?



Paired T-Test

- $P=.03$
- We can reject the null hypothesis, we have 97% confidence that the 2 groups have significantly different means. We can accept the alternate hypothesis.
- Standard deviation error bars do not overlap.

Stats Review

- To show how common a data set is?
- To show if there is a difference between the average height of teen females and adult females?
- To show if there is a correlation between teen foot length and height?
- To know the confidence in the correlation?
- To be able to make a prediction about one's height from their foot length you need?
- To know the confidence you have in making the above prediction?
- To know if there is a difference between adult male, adult female and teen male heights?

