ENS 495 Fall 2017
Problem Set: Types of Models & ANOVA (Ch 9, 12, 15, 17)

Name: _____

Discussion among students is encouraged but all answers must be written in your own words. Points will be deducted if you answers are identical to other students.

Chapter 9: Contingency analysis: associations between categorical variables

Termites

Ch 9	NEW	How is the data for this question organized: contingency table (its categorical data, but the key idea is that is categorical data in a contingency table)	
21	NEW	What is the response variable? Immobilized (columns)	
pg	NEW	What is the predictor variable? Color source of liquid (blue vs. white)	
264	NEW	What type of data is the response variable? Categorical (binary)	
	NEW	What type of data is the predictor variables? Categorical (technically qualitative)	
	NEW	How many groups are there in the predictor variable? 2 (blue, white)	

Genotypes at codon 129

Ch 9	new	How is the data for this question organized: contingency table
33	new	The predictor variable is "age". How many groups are there? two
pg	new	The predictor variable is "genotype". How many groups are there? three
267	new	If "age" was given as a number (eg 28 years old) what type of variable would it be? numeric
	new	What if the values for the MM and MV genotypes were combined so that the response variable had 2 groups, M_ and VV (where _ = M or V). If we treat age as a real number (not just old vs young) and use this new response variable (M_ vs VV), what type of regression would we use? (this was first mentioned on Wed 10/15 and followed up on 10/27; also discussed in Ch. 17.9) logistic regression
	new	Explain when this type of regression is used: Logistic regression is used when you have a binary categorical response variable (here, genotype) and a numeric predictor variable (here, age). The continuous variable is used to predict the value of the response.

From "Review Problems" after chapter 9

MathWorld Web Page (page 272)

Rev Prob 14		Explain why this statement is false: This statement is false. Hypothesis testing does not tell you whether the null hypothesis is true or false, or whether Ha is true or false. It tells you how likely the data you collected are to have occurred IF the null hypothesis is false. This definition is goofy, and it what makes p-values easy to mis-interpret.

Problem Set: Chapter 12: Comparing 2 means (t-tests)

		Spinocerebellar ataxia
33	**new**	**What type of error bars are these? 95% CIs**
pg 365	**b**	Read section 12.6 (pg 346-347) to answer this question Explain why it is possible to predict the outcome of a statistical test based on this graph, and what is that outcome? State in terms of an approximate p value (p>0.05, p = 0.05, p<0.05). The confidence intervals only overlap by a small amount (less than half their length). THe p-value will therefore be less than 0.05. ("p<0.05"). This is called "inference by eye."
		What term do I use to describe the difference between the 2 means? **Effect size**
	new	**What type of test would you use: 2-sample t-test**
	new	Redraw the graph in this box graph so that the opposite conclusion would be reached **graph should be redrawn so that either a)the means are closer togther or b) make the confidence intervals longer**

In the scatter plot to the right, what type of variables is on…

x axis: Numeric / Numeric count / Factor / Other

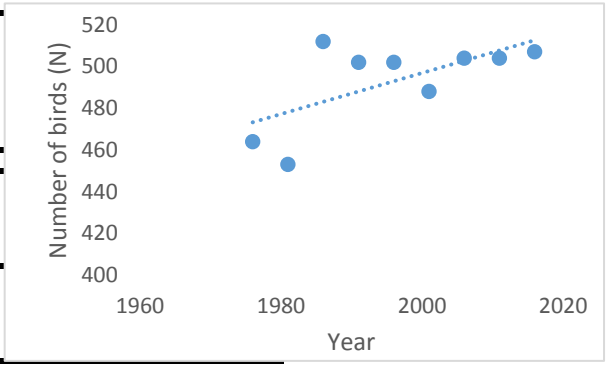
y axis: Numeric / Numeric count / Factor / Other

What type of plot is this?

scatterplot

What type of method would you use to analyze these data?

regression / linear regression



The following statement has a minor mistake, but it drives me nuts. What is it?

"There was a significant differene between the control and the group that received the anti-malarial treatment (t = 2.43432, p = 0.0034243, df = 16, difference = 13.34123, CI = 10.012 - 15.3434)"

What is the miskate? _____--

The values should be rounded, eg t = 2.4, p = 0.003, diff = 13.3, CI = 10.0-15.34

In the plot with error bars to the right . . .

x axis: Numeric / Numeric count / Factor / Other

y axis: Numeric / Numeric count / Factor / Other

What type of method would you use to analyze these data?

1-way ANOVA

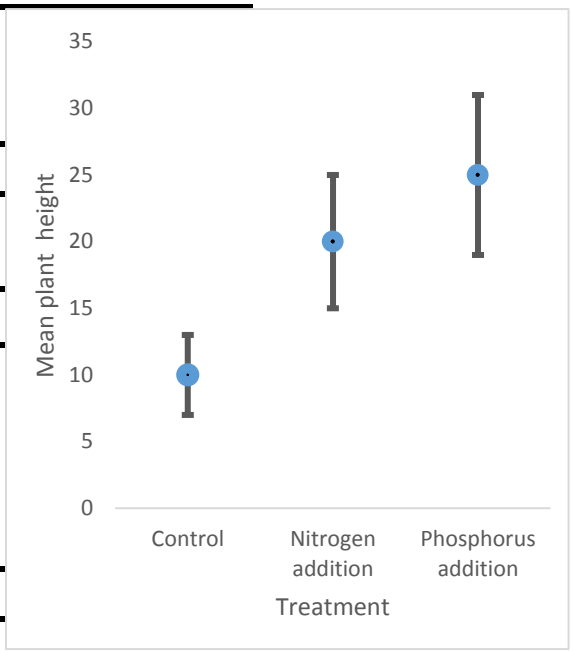
Using the confidence intervals to estimate whether the there is a significant difference between….

Control vs. Nitrogen? Not significant / Significant

Control vs. Phosphorus? Not significant / Significant

Nitrogren vs. Phosphorus? Not significant / Significant

What is the typical threshold for "significance" **alpha = 0.05**



ENS 49
Problem
Name

Chapte

Ch 9

21

pg

264

Ch 9

33

pg

267

From "I

Rev Prob 14

Proble

33

pg 365

In the s

x axis

y axis

What ty

scatt

What ty

regre

The foll

What

The val

In the **x axis** **y axis**
What t **1-wa**
Using t signific Contr Contr Nitrog
What is

5 Fall 2017
n Set: Types of Models & ANOVA (Ch 9, 12, 15, 17)
: _____

er 9: Contingency analysis: associations between categorical variables

Some people answered "Freuency table"; this is not the typical term used |
specifically it's a 2 x 2 contingency table

Both variables are categorical; it's a bit confusing though, because the data get tabulated as counts of the number of individuals in each category. This is a common mistake. The key idea is what is the fate of the organism - alive or dead.

The key here is that the response variable is categorical AND the numeric variable is numeric.

Review Problems" after chapter 9

m Set: Chapter 12: Comparing 2 means (t-tests)

Some people wrote "inferential" this is true, but not very precise). You can do inference w/ them.

catter plot to the right, what type of variables is on…

: Numeric / Numeric count / Factor / Other

: Numeric / Numeric count / Factor / Other

ype of plot is this?

erplot

ype of method would you use to analyze these data?

ession / linear regression

lowing statement has a minor mistake, but it drives me nuts. What is it?

is the miskate? _____--

ues should be rounded, eg t = 2.4, p = 0.003, diff = 13.3, CI = 10.0-15.34

plot with error bars to the right . . .

: Numeric / Numeric count / Factor / Other

: Numeric / Numeric count / Factor / Other

ype of method would you use to analyze these data?

y ANOVA

he confidence intervals to estimate whether the there is a
ant difference between….

ol vs. Nitrogen? Not significant / Significant

ol vs. Phosphorus? Not significant / Significant

gren vs. Phosphorus? Not significant / Significant

s the typical threshold for "significance" alpha = 0.05