

Stat 302 Midterm 1

Thursday, February 4, 2016

Student Name _____

Student Number _____

You have exactly 50 minutes to complete this exam.

This test has 6 pages including this one, and tables.

Only non-programmable calculators are allowed for electronics.

That means no graphing calculators and no phones.

This test also includes a t-table. There are no substantial formulae to be applied, so no formulae are provided.

Protips:

- Show your work whenever appropriate. It shows understanding, and that's what's being tested.
- Use the backs of pages if space is an issue.
- If you get stuck on a part, don't abandon the question. Often later parts can be answered without earlier ones.
- Try not to panic, it rarely helps.

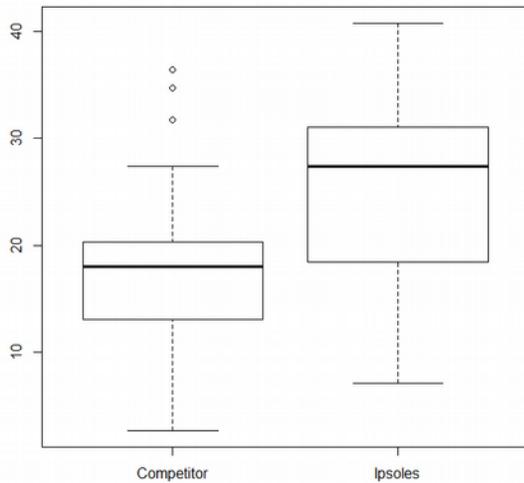
Good luck!

Question	1	2	3	4	5	Total
Out of	6	7	5	5	4	27

Problem 1, Total / 6

Ipsoles has hired you to see if people jump higher while wearing their new line of shoes than in the competitor's shoes.

They found 40 people at the local basketball court wearing Ipsoles, and 60 wearing the shoes of competitors. Below is the results from the two-sample T-test.



```
Welch Two Sample t-test
data: treatment and control
t = 5.3706, df = 73.279, p-value = 4.432e-07
```

(2 pts) Which alternative hypothesis is most appropriate, one-tailed or two-tailed? Why?

(1 pt) Do the boxplots indicate any problems with this data? If so, what?

(3 pts) Can we conclude that Ipsoles shoes cause people to jump higher? If not, how could we change this study improve our prospects of making a causal inference?

Problem 2 Total / 7

Consider the following regression output:

Coefficients:

	Estimate	Std. Error	t value
(Intercept)	10.8034	0.7197	15.010
x	-2.2786	0.7504	-3.037

17 degrees of freedom

(1 pt) Write the regression equation.

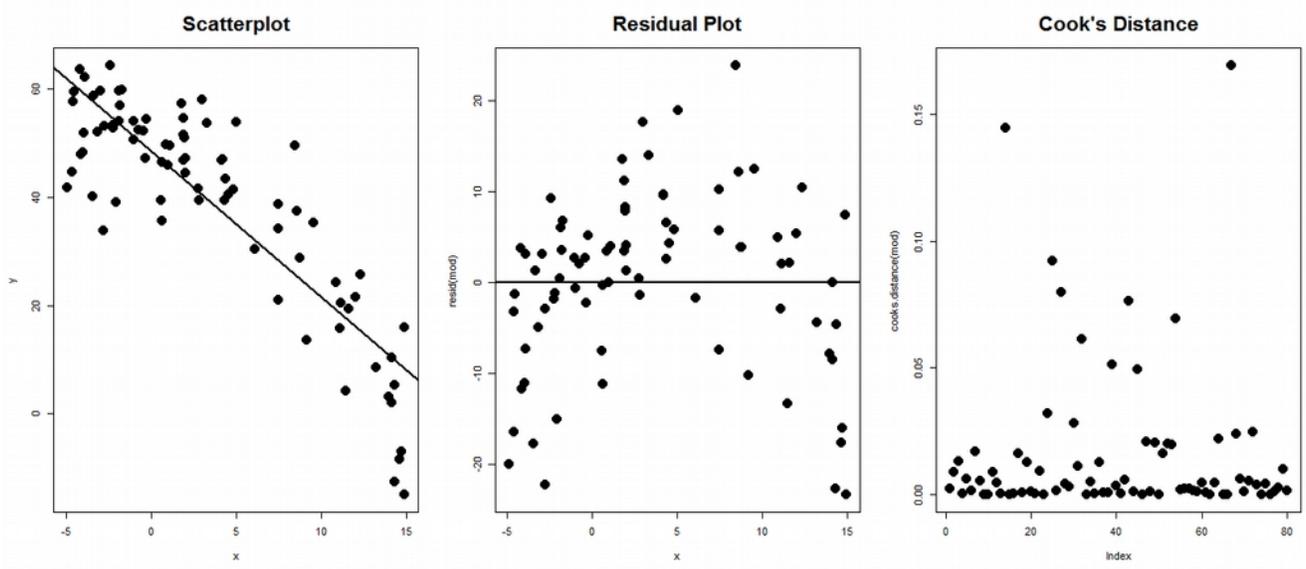
(2 pts) Using the t-table provided. Test the null hypothesis that the slope is zero using the significance level $\alpha = 0.05$. Explain your steps and include a conclusion.

(3 pts) Using the t-table provided. Find the 99% confidence interval of the slope.

(1 pt) 35.17% of the variance in y is explained x. What is the Pearson correlation? Be sure to include the + or – sign.

Problem 3, Total / 5

Here is a scatterplot, residual plot, and Cook's distance plot from a regression.



a) (3 pts) Identify a potential problem, and explain how you found it.

b) (2 pts) What could be done to better model the relationship between X and Y?

Problem 4, Total /5

Consider the following output and table regarding the weight and species of various pets.

Response: y

```
          Df Sum Sq Mean Sq F value    Pr(>F)
animal     3  674655   224885   16.212 2.502e-07 ***
Residuals 46  638097    13872
```

Animal	N	Mean	SD	Cluster
Dog	15	183.7	128.9	A
Cat	13	105.5	73.5	AB
Ferret	11	47.0	63.1	BC
Dragon	11	31.8	39.8	C

(1 pt) From this ANOVA, is there sufficient evidence to suggest that **ANY** of the four means are different from any other at the 0.05 significance level.? How do you know?

(2 pts) From this ANOVA, is there sufficient evidence to suggest that **ALL** four means are different from any other at the 0.05 significance level.? How do you know?

(2 pts) Are there any features of this data that may interfere with the anova results and your conclusions?

Problem 5 Total / 4

A regression equation for a sample of 12 data points $Y = 2.67 + 0.8X$.

(1 pt) A new observation with a value $x = 5$, using this regression equation, what is your best estimate for the prediction of y ?

(3 pts) A hypothesis test of the slope parameter fails to reject the null that slope = 0. Does this mean the slope is zero. If not, why?

Simplified T-Table

(1 tail)	0.025	0.005
(2 tail)	0.05	0.01
df		
1	12.7065	63.6551
2	4.3026	9.9247
3	3.1824	5.8408
4	2.7764	4.6041
5	2.5706	4.0322
6	2.4469	3.7074
7	2.3646	3.4995
8	2.306	3.3554
9	2.2621	3.2498
10	2.2282	3.1693
11	2.201	3.0158
12	2.1788	3.0545
13	2.1604	3.0123
14	2.1448	2.9768
15	2.1314	2.9467
16	2.1199	2.9208
17	2.1098	2.8983
18	2.1009	2.8784
19	2.093	2.8609
20	2.086	2.854