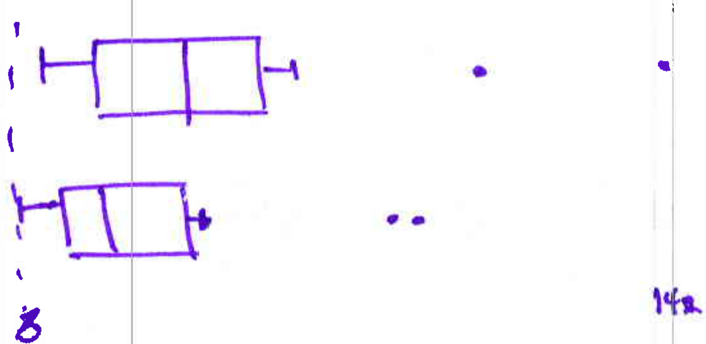
Stat 1350, Problem Solving, Fall 2014 Name KEY

Instructions: This handout runs through a couple different examples of regression/correlation, confidence intervals and hypothesis test problems to help prepare you for the final exam.

1. A study to assess the capability of subsurface flow wetland systems to remove biochemical oxygen demand (BOD) and various other chemical constituents resulted in the accompanying data on x=BOD mass loading (kg/ha/d) and y=BOD mass removal (kg/ha/d).

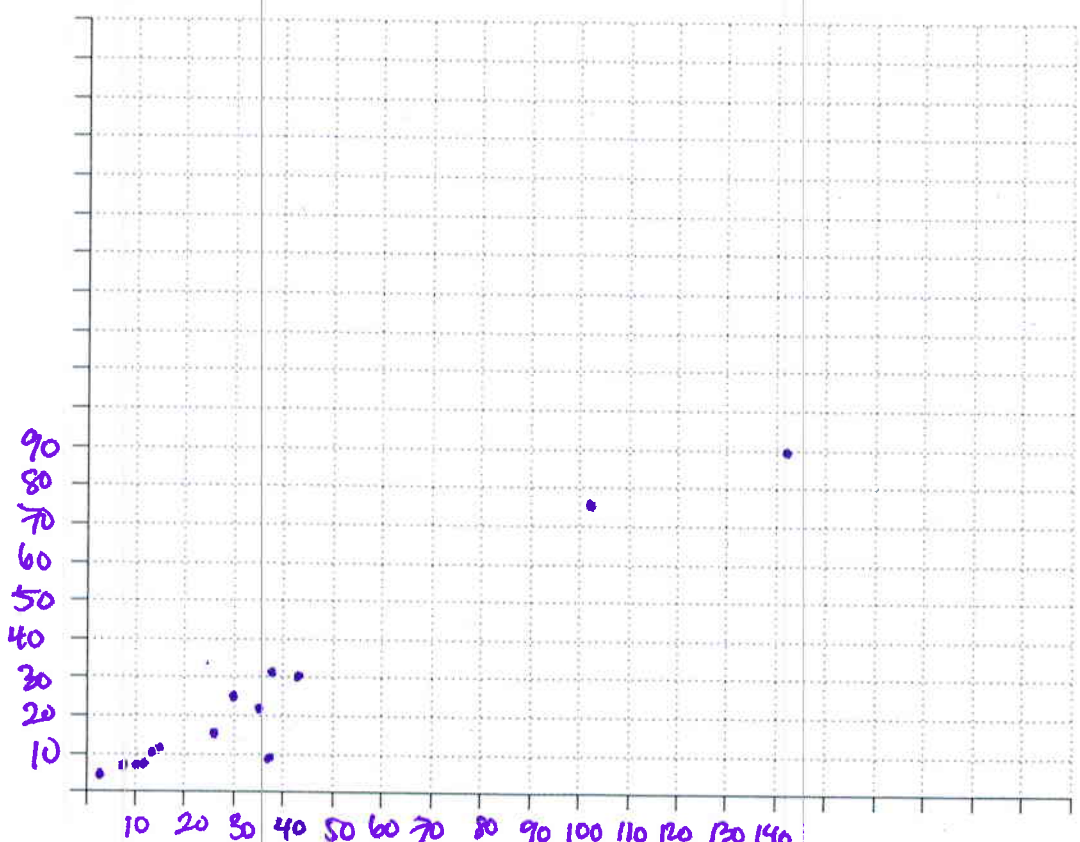
x	3	8	10	11	13	16	27	30	35	37	38	44	103	142
y	4	7	8	8	10	11	16	26	21	9	31	30	75	90

a. Construct boxplots of both mass loading and mass removal. Comment on any interesting features. Draw the graph here. Label all key features.



y-data a bit more skewed, but both have 2 outliers

b. Construct a scatterplot of the data, and comment on any interesting features. Do there appear to be any outliers? Plot each point and label the axes. Indicate your scale.



looks pretty linear but could be affected by 2 large values
possible outlier is (37,9)

c. Which variable is the explanatory variable? Which variable is the response variable?

X= BOD mass loading explanatory

Y= BOD mass removal response

d. Find the linear regression line that best fits the data.

$Y = .6523X + .6261$

e. Does it do a good job modeling the data? What is the value of the correlation coefficient?

yes, pretty good (Strong correlation) $r = .9777$

f. What proportion of data is explained by the relationship between the two variables?

$r^2 = .95597$ or about 95.6%

g. Predict the value of y=BOD mass removal when the BOD mass loading is 75 kg/ha/d.

$y = .6523(75) + .6261 = 49.5486$

h. Could we use this equation to predict the value of BOD mass removal if the BOD mass loading was 200 kg/ha/d?

no. this is beyond the available data range and we do not know that the trend continues

2. Suppose that a random sample of 50 bottles of a particular brand of cough syrup is selected and the alcohol content of each bottle is determined. Let μ denote the average alcohol content for the population of all bottles of the brand under study. Suppose the resulting 95% confidence interval is (7.8, 9.4).

a. What is the margin of error for this confidence interval?

$\frac{7.8+9.4}{2} = 8.6$

$8.6 \pm .8$

$7.8 - 8.6 = \pm .8$

Margin of error is ± 0.8

b. Would a 90% confidence interval calculated from this same sample have been narrower or wider than the given interval? Explain your reasoning.

narrower.
less confidence means we can give a more accurate prediction

c. Calculate the 90% and 99% confidence intervals. Be sure to write your answers as intervals.

$.8 = 1.96 \frac{s}{\sqrt{n}}$ $\frac{s}{\sqrt{n}} = .40816$

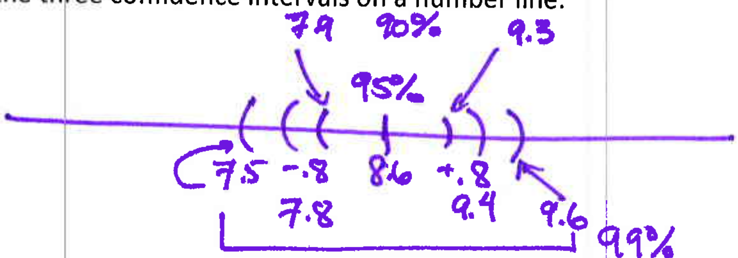
90% margin of error = $1.64 * .40816 = .669$
$(8.6 \pm .669) = (7.931, 9.269)$

99% margin of error = $2.57 * .40816 = 1.0489$ $(7.551, 9.649)$

d. Consider the following statement: There is a 95% chance that μ is between 7.8 and 9.4. Is this statement correct? Why or why not?

yes. Confidence intervals can be interpreted as the true value of the mean being 95% likely to be in this range.

e. Draw the three confidence intervals on a number line.



3. An article reports that for a sample of 50 kitchens with gas cooking appliances monitored during a one-week period, the sample mean CO_2 level (ppm) was 654.16, and the sample standard deviation was 164.33.

a. Calculate and interpret a 95% confidence interval for true average CO_2 level in the population of all home from which the sample was selected.

Z Interval (Stats) $(608.61, 699.71)$
$\sigma = 164.33$
$\bar{x} = 654.16$
$n = 50$ C level = .95

b. Suppose that other features of the data remained the same but the sample size was 100 instead of 50. What is the new 95% confidence interval? Is it wider or narrower than before? Why?

Change $n = 100$

$(621.95, 686.37)$

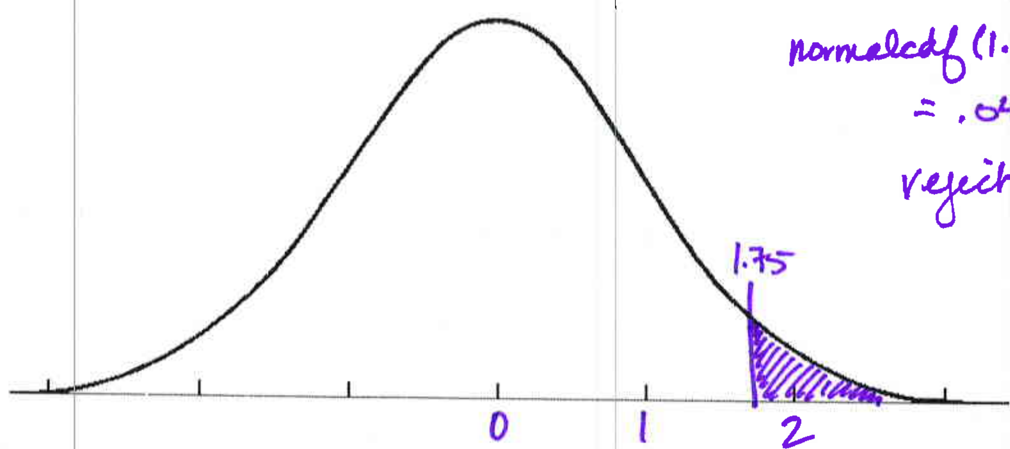
narrower less variability in sampling distribution owing to larger sample size

4. For the following pairs of assertions, indicate which do not comply with our rules for setting up hypotheses and why.
   a. $H_0: \mu = 100, H_a: \mu > 100$
   b. $H_0: p \neq 0.25, H_a: p = 0.25$ equal sign goes on Ho not Ha
   c. $H_0: \mu = 120, H_a: \mu = 150$ can't have 2 equal signs
   Ha: μ > 120

5. A manufacturer of nickel-hydrogen batteries randomly selects 100 nickel plates for test cells, cycles them a specified number of times, and determined that 14 of the plates have blisters. Does this provide compelling evidence for concluding that more than 10% of all plates blister under such circumstances?
   a. State and test the appropriate hypotheses.

   Ho: p = .10
   Ha: p > .10
   1 PropZTest
   Po = .10
   X = 14
   n = 100
   z = 1.33
   p = .0912 > α
   fail to reject Ho

   b. Using a significance level of 0.05, test the appropriate hypothesis. What is your z-score for the test? Mark your score on the normal curve below along with the critical value $z = 1.75$.



   normalcdf(1.75, E99)
   = .040059 < .05
   reject Ho

   c. What is the P-value for the test? How does it compare to the significance level?

   p = .0912 > .05 it is more (a)
   p = .040059 < .05 it is less (b)

   d. Do you reject the null hypothesis or fail to reject it?

   fail to reject Ho (a)
   reject Ho (b)

   e. What does that mean in the context of the problem? Do you have good evidence that more than 10% of plates blister?

   There is not sufficient evidence (a) to think the blister rate is higher than 10%