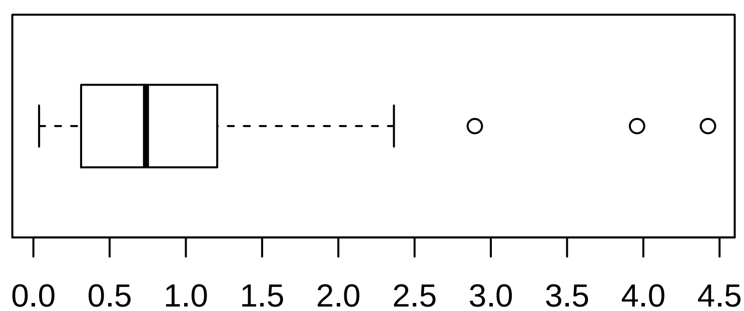
Stat 185 - Sample Final Exam

Most, but not necessarily all, of our final exam this Tuesday, July 18 will consist of problem that look like one of the following.

1. Suppose that a study is conducted by asking students what score they think they are going to get on a test and then measuring their blood pressure. The goal is to see if expected score has an effect on blood pressure.
   (a) What is the predictor variable and what is the response variable? For each variable determine if it is categorical or numeric.
   (b) Is this study an observational study or a designed experiment?

2. Answer the following questions about this box and whisker plot (you only have to get reasonably close):



   (a) Estimate the mean
   (b) Estimate the IQR
   (c) Estimate the highest observation
   (d) Estimate the 4th highest observation
   (e) Estimate the lowest observation
   (f) Discuss the shape; do you think a normal distribution would be useful to model this data?

3. Draw a box and whisker plot from the following summary of a list of numeric data:

Min	1st Qu	Median	Mean	3rd Qu	Max.
-0.8356	-0.5462	0.2566	0.1322	0.5537	1.5953

4. Suppose that 20% of all people have green eyes. Use a normal distribution to estimate the probability that more than 20 of the sixty lucky students taking Stat 185 with me next semester will have green eyes.

5. An organic farmer sells bags of Brussels sprouts labeled as one pound each. Let's use a t-test to investigate the farmer's claim that bags each contain one pound. I purchased one bag each of 4 weeks during the Summer farmer's market, weighed them at home and recorded the following weights:

1.05	1.15	1.2	1.1

   (a) Write down the hypothesis statement for the problem.

   (b) Write down a formula showing that the mean is $\mu = 1.125$

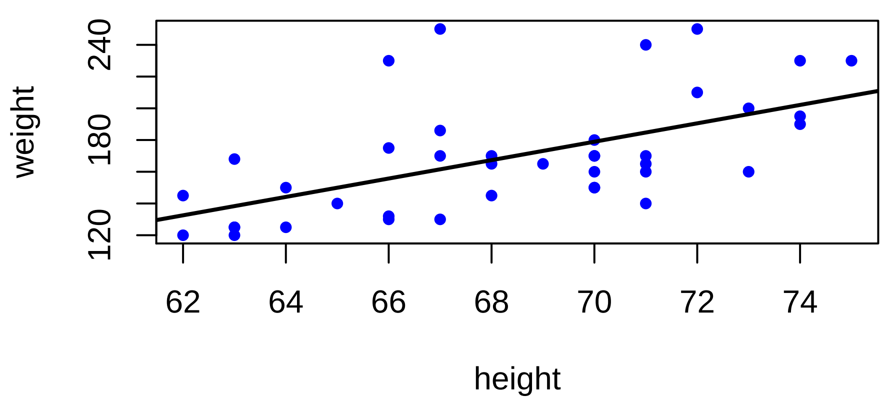
   (c) Write down a formula showing that the standard deviation is $\sigma = 0.06454972$.

   (d) Assuming the mean $\mu_0 = 1$ is correct, compute the t-score for the observed mean.

   (e) Compute a 95% confidence interval for the average weight of a bag of these Brussels sprouts.
   Be sure to use a t-distribution with the correct degrees of freedom.

   (f) Use a t-table to find the critical t-value for a 95% level of confidence for this problem.

6. The figure below shows a scatter plot of weight (in pounds) vs height (in inches) for a random sample 40 American men and the table shows linear regression table for that same data. We are curious if there is a relationship between height and weight.



	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-226.9576	89.4932	-2.54	0.0154
height	5.7984	1.3065	4.44	0.0001

(a) Write down the hypothesis statement that we would check with this linear model.

(b) Write out the formula for linear model.

(c) What does the model predict for the weight of a man who is 70 inches tall?

(d) State the conclusion of the hypothesis test from part (c).

7. A professor who teaches a large introductory statistics class (197 students) with eight discussion sections would like to test if student performance differs by discussion section, where each discussion section has a different teaching assistant. The summary table below shows the average final exam score for each discussion section as well as the standard deviation of scores and the number of students in each section.

	Sec 1	Sec 2	Sec 3	Sec 4	Sec 5	Sec 6	Sec 7	Sec 8
n_i	33	19	10	29	33	10	32	31
$\bar{x}_i$	92.94	91.11	91.80	92.45	89.30	88.30	90.12	93.35
s_i	4.21	5.58	3.43	5.92	9.32	7.27	6.93	4.57

The ANOVA output below can be used to test for differences between the average scores from the different discussion sections.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
section	7	525.01	75.00	1.87	0.0767
Residuals	189	7584.11	40.13		

Conduct a hypothesis test to determine if these data provide convincing evidence that the average score varies across some (or all) groups. Check conditions and describe any assumptions you must make to proceed with the test.

8. A college statistics professor expects to assign grades in the following proportions

Grade	A	B	C	D	E
Expected Percentage	10%	30 %	40%	20%	0%
Actual count	3	6	8	1	0

Perform a χ^2 test to examine whether these grade counts agree with the expected grade counts.

9. *I expect there will be one problem, like this one, where you need to use R.*
Suppose we wish to compare the climates of Asheville, NC and Columbus, OH. We randomly pick 8 days and find the daily high temperature for that day for each of the two cities giving the results below. Find a 95% confidence interval for the mean difference in the daily high (city A - city B). Interpret your confidence interval.

Day	1	2	3	4	5	6	7	8
Columbus high	54	72	70	86	82	73	48	52
Asheville high	52	70	66	88	78	70	50	51