Name:

Instructions:

- Write your name on this cover page.

- Turn off your cell phone and put it away.

- You **may** use a calculator. However, you **may not** use a calculator on your phone or any other device that connects to the internet.

- You have **50 minutes** to complete the exam.

- You are expected to obey the Honor Code while taking this test. You **may not** discuss the exam with any other students until the exams have been returned.

- You may ask the instructor for clarification during the exam. Students who violate the Honor Code will be referred to the Honor Code Council.

- If you witness others violating the Honor Code, you have a duty to report them to the Honor Code Council.

- Students must pledge to obey the Honor Code by signing below. **Unsigned exams will not be graded.**

I understand and agree to abide by the principles of the Honor Code of Mount Holyoke College.

_____   _____
Signature   *Date*

1. In the 1920s, 50 observations were collected to study the relationship between the speed at which a car travels (miles per hour) and the distance it takes to stop (feet).

   (a) Using the problem description and the relevant R output, check the conditions for conducting an analysis of these data with a linear regression model. You should write an explicit sentence for each condition explaining why it is or is not satisfied, with justification. If you need more information to make a determination, explain what else you need to know. If necessary, find a transformation of the data so that the conditions are as well satisfied as possible.

   (b) Assume the conditions in (a) are reasonably well satisfied. What is the estimated equation of the line describing the relationship between the speed at which a car travels and the distance it takes to stop?

   (c) Assume the conditions in (a) are reasonably well satisfied. What are the interpretations of the estimated intercept and slope? Please interpret the coefficient estimates in context.

(d) Assume the conditions in (a) are reasonably well satisfied. Conduct a hypothesis test of the claim that there is no association between the speed at which a car travels and the distance it takes to stop.

    i. State your hypotheses using equations and a written sentence explaining the meaning of the hypotheses.

    ii. What are the degrees of freedom associated with this test?

    iii. Interpret your results in context.

(e) Are these results generalizable? If so, to what population? If not, explain why not.

2. The data used in this problem come from a series of similar experiments that were conducted at Chatham, Ontario, Canada in 1939 to investigate various control measures for insects in the field. Six treatments were assigned randomly to small plots that were similar in nature, and the number of insects in each small plot was recorded after treatment application.

   (a) Using the problem description and the relevant R output, check the conditions for conducting an analysis of these data with an ANOVA model. You should write an explicit sentence for each condition explaining why it is or is not satisfied, with justification. If you need more information to make a determination, explain what else you need to know. If necessary, find a transformation of the data so that the conditions are as well satisfied as possible.

   (b) For the purposes of this problem, assume that the conditions you checked in part (a) are fairly well satisfied (perhaps after a suitable transformation). Conduct a test to find out whether there are any differences in the mean insect count for the six different insecticide treatments. Use appropariate R output.

       i. Define all parameters involved.

       ii. State your hypotheses in terms of equations involving the parameters and written sentences explaining what the hypotheses mean in context.

iii. Interpret the p-value for your test in terms of strength of evidence against the null hypothesis of the test, stated in context. If you used a transformation, you may intepret your results on the transformed scale.

(c) As in (b), assume that the conditions you checked in part (a) are fairly well satisfied (perhaps after a suitable transformation). Conduct a test to find out whether there are any differences in the mean insect count for sprays A, B and F. You may use the same parameter definitions as you used in (b) - you do not need to define them again.

   i. State your hypotheses in terms of equations involving the parameters and written sentences explaining what the hypotheses mean in context.

   ii. Interpret the p-value for your test in terms of strength of evidence against the null hypothesis of the test, stated in context. If you used a transformation, you may intepret your results on the transformed scale.

(d) Continue to assume the conditions you check in (a) are fairly well satisfied (perhaps after a suitable transformation).

    i. Based on your conclusions from (c), part (ii), is there any reason to further investigate the relationships among these three means? Answer "yes" or "no", and explain briefly.

    ii. **If you answered "yes" for the previous question**, describe briefly how you would construct and interpret confidence intervals with a family-wise confidence level of 95%: one for the difference in mean insect count for spray A and B, one for the difference in mean insect count for spray A and F, and one for the difference in mean insect count for spray B and F. You do not need to construct these intervals.

(e) Would you be comfortable claiming that there is a causal association between insecticide choice and number of insects in the context of this study? Briefly explain why or why not.