

**Directions:** Type or clearly handwrite your solutions to each of the following exercises. Partial credit cannot be given unless all work is shown. You may work in groups provided that each person takes responsibility for understanding and writing out the solutions. Additionally, you must give proper credit to your collaborators by providing their names on the line below (if you worked alone, write “No Collaborators”):

---

1. **[+10]:** For a study about a new treatment for pancreatic cancer, 100 pancreatic cancer patients are recruited. For ethical reasons every patient in the study is given the highest current standard of care which consists of surgery to remove cancerous tumors from the pancreas followed by a period of radiation therapy. After receiving the standard treatment, each patient is given the opportunity to have a second experimental treatment that consists of coating some of the cancerous cells removed from the patient during surgery with an antigen that the human body will reject and injecting the coated cells back into the pancreas of the same patient. The basic idea is that this second treatment will help prevent recurrence of cancer tumors because the patient’s immune system will attack the coated cells and also attack other cancerous cells in the pancreas that the surgery and radiation treatments may have missed. The 57 patients who volunteer for the second treatment and the 43 patients who refuse the second treatment are followed for three years to determine how long each patient survives.

(a) Is this study an experiment or observational study? Justify your answer.

(b) Identify the experimental or sampling units.

(c) Identify the response variable.

(d) Identify the treatments or study/population conditions.

(e) Describe whether or not control of extraneous variation was used in this study.

(f) Describe whether or not randomization was used in this study.

(g) Describe whether or not replication was used in this study.



3. [+5]: Perform a randomization test in SAS of the research question given in problem 2 by computing the difference in the sample means for each of 25,000 data sets created from new random assignments of 30 patients to one of the two drugs.

(a) Obtain an approximate  $p$ -value for the randomization test.

(b) Interpret the result of the randomization test in the context of this study. Is it consistent with your response to problem 2(d)?

(c) Which sampling distribution is the randomization histogram approximating?

**Total:** 25 points    **# correct:** \_\_\_\_\_    **%:** \_\_\_\_\_