Statistics 101B - Spring 2020
Homework 1 (Due Monday April 13 at 5:00pm)

** Please upload your homework on CCLE. Please write your full name, student ID number, and section number on your homework.
Must be PDF.
NO late work will be accepted!**
You do not need to write the questions in your homework. Please show your work!

1. Problem 1.3
2. Problem 1.8

3. Data collected from a completely randomized design for comparing two treatments A and B are in the following table:

A	30	29	27
B	32	35	34

The experimenter is interested in finding out if treatment B is better than treatment A. Find the p-values using both the randomization test and the t test, and then make the conclusion. (You may use $\alpha = 0.05$)

4. Problem 2.23
5. Problem 2.25
6. Problem 2.29

(**b**) List all of the potential sources of variability that could impact the response.

(**c**) Complete the first three steps of the guidelines for designing experiments in Section 1.4.

1.3. Suppose that you want to compare the growth of garden flowers with different conditions of sunlight, water, fertilizer, and soil conditions. Complete steps 1–3 of the guidelines for designing experiments in Section 1.4.

1.4. Select an experiment of interest to you. Complete steps 1–3 of the guidelines for designing experiments in Section 1.4.

1.5. Search the World Wide Web for information about Sir Ronald A. Fisher and his work on experimental design in agricultural science at the Rothamsted Experimental Station.

1.6. Find a Web Site for a business that you are interested in. Develop a list of factors that you would use in an experiment to improve the effectiveness of this Web Site.

1.7. Almost everyone is concerned about the rising price of gasoline. Construct a cause and effect diagram identifying the factors that potentially influence the gasoline mileage that you get in your car. How would you go about conducting an experiment to determine any of these factors actually affect your gasoline mileage?

1.8. What is replication? Why do we need replication in an experiment? Present an example that illustrates the difference between replication and repeated measurements.

1.9. Why is randomization important in an experiment?

1.10. What are the potential risks of a single large, comprehensive experiment in contrast to a sequential approach?

(a) We would like to demonstrate that the mean shelf life exceeds 120 days. Set up appropriate hypotheses for investigating this claim.

(b) Test these hypotheses using $\alpha = 0.01$. What are your conclusions?

(c) Find the P-value for the test in part (b).

(d) Construct a 99 percent confidence interval on the mean shelf life.

2.20 Consider the shelf life data in Problem 2.19. Can shelf life be described or modeled adequately by a normal distribution? What effect would the violation of this assumption have on the test procedure you used in solving Problem 2.15?

2.21 The time to repair an electronic instrument is a normally distributed random variable measured in hours. The repair times for 16 such instruments chosen at random are as follows:

Hours			
159	280	101	212
224	379	179	264
222	362	168	250
149	260	485	170

(a) You wish to know if the mean repair time exceeds 225 hours. Set up appropriate hypotheses for investigating this issue.

(b) Test the hypotheses you formulated in part (a). What are your conclusions? Use $\alpha = 0.05$.

(c) Find the P-value for the test.

(d) Construct a 95 percent confidence interval on mean repair time.

2.22 Reconsider the repair time data in Problem 2.21. Can repair time, in your opinion, be adequately modeled by a normal distribution?

2.23 Two machines are used for filling plastic bottles with a net volume of 16.0 ounces. The filling processes can be assumed to be normal, with standard deviations of $\sigma_1 = 0.015$ and $\sigma_2 = 0.018$. The quality engineering department suspects that both machines fill to the same net volume, whether or not this volume is 16.0 ounces. An experiment is performed by taking a random sample from the output of each machine.

Machine 1		Machine 2	
16.03	16.01	16.02	16.03
16.04	15.96	15.97	16.04
16.05	15.98	15.96	16.02
16.05	16.02	16.01	16.01
16.02	15.99	15.99	16.00

(a) State the hypotheses that should be tested in this experiment.

(b) Test these hypotheses using $\alpha = 0.05$. What are your conclusions?

(c) Find the P-value for this test.

(d) Find a 95 percent confidence interval on the difference in mean fill volume for the two machines.

2.24 Two types of plastic are suitable for use by an electronic calculator manufacturer. The breaking strength of this plastic is important. It is known that $\sigma_1 = \sigma_2 = 1.0$ psi. From random samples of $n_1 = 10$ and $n_2 = 12$, we obtain $\bar{y}_1 = 162.5$ and $\bar{y}_2 = 155.0$. The company will not adopt plastic 1 unless its breaking strength exceeds that of plastic 2 by at least 10 psi. Based on the sample information, should they use plastic 1? In answering this question, set up and test appropriate hypotheses using $\alpha = 0.01$. Construct a 99 percent confidence interval on the true mean difference in breaking strength.

2.25 The following are the burning times (in minutes) of chemical flares of two different formulations. The design engineers are interested in both the mean and variance of the burning times.

Type 1		Type 2	
65	82	64	56
81	67	71	69
57	59	83	74
66	75	59	82
82	70	65	79

(a) Test the hypothesis that the two variances are equal. Use $\alpha = 0.05$.

(b) Using the results of (a), test the hypothesis that the mean burning times are equal. Use $\alpha = 0.05$. What is the P-value for this test?

(c) Discuss the role of the normality assumption in this problem. Check the assumption of normality for both types of flares.

2.29 The diameter of a ball bearing was measured by 12 inspectors, each using two different kinds of calipers. The results are as follows:

Inspector	Caliper 1	Caliper 2
1	0.265	0.264
2	0.265	0.265
3	0.266	0.264
4	0.267	0.266
5	0.267	0.267
6	0.265	0.268
7	0.267	0.264
8	0.267	0.265
9	0.265	0.265
10	0.268	0.267
11	0.268	0.268
12	0.265	0.269

(a) Is there a significant difference between the means of the population of measurements from which the two samples were selected? Use $\alpha = 0.05$.

(b) Find the P-value for the test in part (a).

(c) Construct a 95 percent confidence interval on the difference in mean diameter measurements for the two types of calipers.

2.30 An article in the journal *Neurology* (1998, Vol. 50, pp. 1246–1252) observed that monozygotic twins share numerous physical, psychological, and pathological traits. The investigators measured an intelligence score of 10 pairs of twins. The data obtained are as follows:

Pair	Birth Order: 1	Birth Order: 2
1	6.08	5.73
2	6.22	5.80
3	7.99	8.42
4	7.44	6.84
5	6.48	6.43
6	7.99	8.76
7	6.32	6.32
8	7.60	7.62
9	6.03	6.59
10	7.52	7.67

(a) Is the assumption that the difference in score is normally distributed reasonable?

(b) Find a 95 percent confidence interval on the difference in mean score. Is there any evidence that mean score depends on birth order?

(c) Test an appropriate set of hypotheses indicating that the mean score does not depend on birth order.

2.31 An article in the *Journal of Strain Analysis* (Vol. 18, no. 2, 1983) compares several procedures for predicting the shear strength for steel plate girders. Data for nine girders in the form of the ratio of predicted to observed load for two of these procedures, the Karlsruhe and Lehigh methods, are as follows: SS

Girder	Karlsruhe Method	Lehigh Method
S1/1	1.186	1.061
S2/1	1.151	0.992
S3/1	1.322	1.063
S4/1	1.339	1.062
S5/1	1.200	1.065
S2/1	1.402	1.178
S2/2	1.365	1.037
S2/3	1.537	1.086
S2/4	1.559	1.052

(a) Is there any evidence to support a claim that there is a difference in mean performance between the two methods? Use $\alpha = 0.05$.

(b) What is the P-value for the test in part (a)?

(c) Construct a 95 percent confidence interval for the difference in mean predicted to observed load.

(d) Investigate the normality assumption for both samples.

(e) Investigate the normality assumption for the difference in ratios for the two methods.

(f) Discuss the role of the normality assumption in the paired t-test.

2.32 The deflection temperature under load for two different formulations of ABS plastic pipe is being studied. Two samples of 12 observations each are prepared using each formulation and the deflection temperatures (in °F) are reported below:

Formulation 1			Formulation 2		
206	193	192	177	176	198
188	207	210	197	185	188
205	185	194	206	200	189
187	189	178	201	197	203