

Please put your name on the back of your answer book.

Do NOT put it on the front. Thanks.

Do not start until I tell you to.

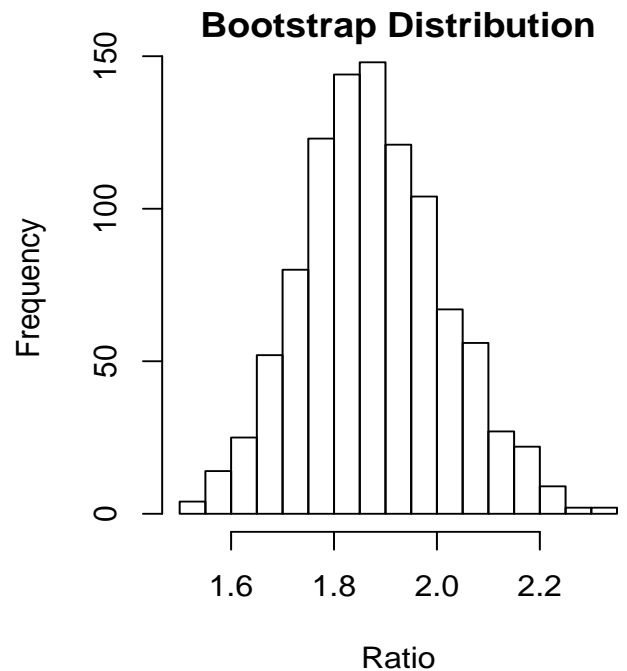
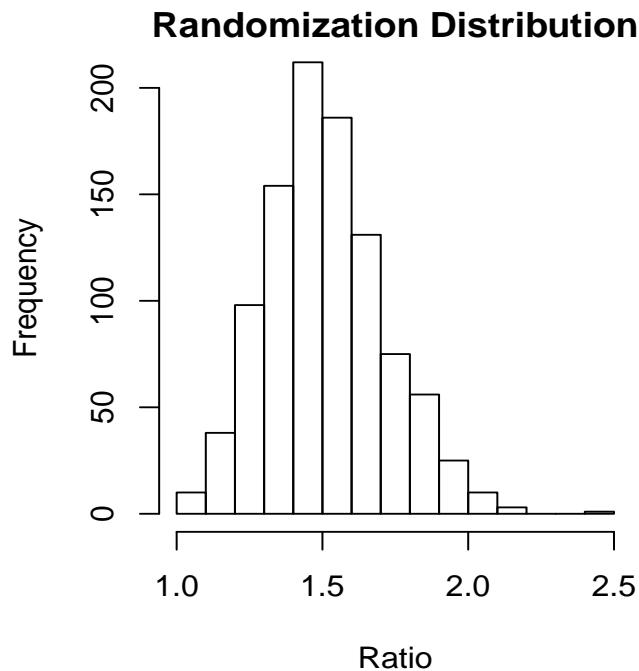
- The exam is closed book, closed notes. Use only the formula sheet and tables I provide today. You may use a calculator.
- Write your answers in your blue book. Ask if you need a second (or third) blue book.
- You have 2 hours (120 minutes) to complete the exam.
Stop working when the end of the exam is announced.
- Points are indicated for each question. There are 130 total points.
- Important reminders:
 - budget your time. Some parts of each question should be easy; others may be hard. Make sure you do all parts you can.
 - notice that some parts do not require any computations.
 - show your work neatly so you can receive partial credit.
- Good luck!

1. 20 pts. The following problem is based on an engineering study comparing two methods (A or B) for one step in making a component. The response is the length of time required for that step. Thirty components were used in the study: 20 were randomly assigned to method A; the remaining 10 used method B. Here are summary data for each group.

Method	n_i	Time required	
		Average	s.d.
A	20	14.65	3.20
B	10	7.34	1.62

In studies like these, the ratio between the two times is the most reasonable way to summarize the treatment effect. The investigators expect that method A will take longer. They want to know if the ratio of means (Method A / Method B) is larger than 1.5.

- (a) 5 pts. Is this a survey, an observational study or an experimental study? Briefly explain.
- (b) 5 pts. The investigators used a randomization test to test the null hypothesis that the ratio of the two means equals 1.5. The test statistic is the ratio of the two sample averages (i.e. Method A average / Method B average). Below are histograms of the randomization distribution given the H_0 : ratio = 1.5 and the bootstrap distribution. Below each figure are the largest 20 and smallest 20 observations in order. There are a total of 999 values in the randomization distribution and 1000 values in the bootstrap distribution.



	Randomization					Bootstrap				
Smallest 20	1.002	1.051	1.051	1.056	1.065	1.533	1.535	1.541	1.543	1.551
	1.080	1.081	1.095	1.098	1.098	1.552	1.558	1.564	1.568	1.576
	1.102	1.103	1.105	1.107	1.111	1.578	1.582	1.583	1.586	1.586
	1.128	1.133	1.136	1.138	1.141	1.589	1.594	1.600	1.605	1.605
Largest 20	2.415	2.129	2.129	2.121	2.079	2.338	2.308	2.265	2.265	2.247
	2.074	2.055	2.054	2.031	2.023	2.246	2.245	2.236	2.227	2.220
	2.019	2.018	2.017	2.012	1.995	2.212	2.203	2.203	2.197	2.192
	1.992	1.985	1.979	1.979	1.966	2.191	2.186	2.185	2.184	2.180

What is the p-value for a one-sided randomization test of ratio = 1.5? Note: the alternative hypothesis is H_A : ratio > 1.5.

- (c) 5 pts. The standard error of the estimated ratio is 0.0946. Calculate the t-statistic to test the hypothesis that the ratio = 1.5.
- (d) 5 pts. Is it reasonable to use t-distributions on the raw data (e.g. what you did in part 1c to approximate design-based inferences about the ratio of Method A to Method B? Explain why or why not.
2. 40 pts. An study evaluated an experimental genetic treatment thought to increase litter size (the number of offspring) of sows (female pigs). 30 sows were chosen from a pig research facility. 10 were randomly assigned to receive the genetic treatment; the remaining 20 served as controls. The size of each sow's first litter was recorded. The data are summarized below.

	Control	Experiment
average	10.3	10.7
s.d.	2.4	1.5
n	20	10

- (a) 5 pts. Which average litter size (control or experimental group) is more precisely known? Make as few assumptions as possible for this part. Explain.
- (b) 10 pts. Use a two-tailed t-test to test the null hypothesis of no effect of the experimental genetic treatment on the mean litter size. Please report your test statistic, an approximate p-value, and a one-sentence conclusion.
- (c) 10 pts. List the assumptions underlying the t-test used in question 2b. For each assumption, list one reasonable method of assessing the validity of that assumption.
- (d) 5 pts. There is evidence that litter size is partly controlled by genetic factors so that sows with the same mother will tend to have similar litter sizes. Describe a modification of the experimental design (leaving the sample sizes the same) that is likely to increase the precision of the estimated difference between the population means. Explain.
- (e) 5 pts. In a related study of the same 30 sows, another researcher measured 500 genetic markers in each sow. For each marker, she did a t-test comparing the mean litter size of sows with the marker and sows without. The smallest 15 p-values, and some additional calculations, are shown on the next page.

Marker #	i	p-value	$p * i/500$	$p * 500/i$	$p*500$
427	1	0.0000036	< 0.000001	0.0018	0.0018
157	2	0.0000059	< 0.000001	0.0015	0.0029
10	3	0.0000065	< 0.000001	0.0011	0.0032
495	4	0.00013	0.0000010	0.016	0.065
86	5	0.00020	0.0000020	0.020	0.10
259	6	0.00062	0.0000074	0.052	0.31
121	7	0.00064	0.0000089	0.046	0.32
208	8	0.0024	0.000038	0.15	1.29
332	9	0.0069	0.00012	0.38	3.42
466	10	0.0095	0.00019	0.47	4.81
463	11	0.050	0.0011	2.29	25.19
99	12	0.058	0.0014	2.43	29.16
110	13	0.073	0.0019	2.83	36.79
122	14	0.083	0.0023	2.98	41.72
254	15	0.086	0.0025	2.87	43.05

Which of these 15 markers would be considered significant at a false discovery rate of 5% = 0.05? If you wish, you are welcome to say “all 15”, or “all except 466” instead of listing most of the markers.

(f) 5 pts. Which of these 15 markers would be considered significant at an experiment-wise error rate of 0.05 if you used a Bonferroni correction to adjust for testing 500 hypotheses?

3. 70 pts. The U.S. Government has very strict rules for food safety. These require sterilization of most processed foods to reduce or eliminate unwanted bacteria. One common sterilization method is to heat food to a specified temperature and hold it at that temperature for a specified length of time. This kills any bacteria that are present in the food. Unfortunately, heating food also affects its quality. The following problem is based on a study of canned peas. This study compared five different combinations of time and temperature. Three (A, B, and D in the table below) meet the requirements for sterilization and two (C and E), labeled ‘super-sterile’, exceed the requirements. This study examined the quality of the peas after sterilization and storage for 6 months.

Because of the time required to change temperatures and inherent variability in raw material and plant conditions, the experiment was conducted over 6 days. Each day, five batches of peas were randomly assigned to each of the five temperature and time combinations (one batch per treatment per day). Treatments are labeled by the temperature / time, so 55/30 was heated to 55 degrees C for 30 minutes. Batches were then stored in a single warehouse for six months. The response is an average quality score measured for each batch of canned peas. This score ranges from 0 (very crisp = high quality) to 10 (very mushy = low quality). The data and treatment means are shown below:

Type	Treatment	Temp/Time	Day						Mean
			1	2	3	4	5	6	
Sterile	A	50/60	3.2	3.8	1.2	0.6	0.6	1.5	1.82
Sterile	B	55/30	1.1	2.4	2.8	1.6	1.8	1.3	1.83
Super-sterile	C	55/45	4.9	4.5	2.2	3.0	2.0	1.1	2.95
Sterile	D	60/15	2.3	1.6	2.2	1.9	1.3	3.6	2.15
Super-sterile	E	60/20	4.6	6.5	5.6	3.8	4.1	4.3	4.82

Additional SAS output that may be useful is included at the end of the exam. The treatments are listed in “SAS order”.

- (a) 5 pts. What is the experimental unit in this study? Briefly explain.
What is the observational unit? Briefly explain.
- (b) 10 pts. Complete the entries in the ANOVA table :
- | Source | df | SS | MS | F |
|------------|----|-------|-------|----|
| Days | ?? | 11.68 | | |
| Treatments | ?? | ?? | ?? | ?? |
| Error | ?? | ?? | 0.968 | |
| Total | ?? | 69.29 | | |
- (c) 5 pts. Write an appropriate statistical model (i.e. the equation relating observations to parameters) that corresponds to the above ANOVA table.
- (d) 5 pts. Compute the efficiency of blocking.
- (e) 10 pts. The F statistic for Treatments in the part 3b tests a specific null hypothesis (H_0):
- identify that null hypothesis
 - approximate the p-value for that test
 - write an appropriate one sentence conclusion.
- (f) 5 pts. Which of the following is the best approach to answer the question ‘which pairs are treatments are significantly different?’. Briefly explain your choice.
T-tests (LSD), Protected LSD, Tukey-Kramer HSD, Bonferroni, Scheffe, Something else.
- (g) 10 pts. The investigators are specifically interested in two questions:
How large is the mean difference between the two super-sterile treatments (C and E)?
How large is the difference between the mean quality of the three sterile treatments (A, B, and D) and the mean quality of the two super-sterile treatments (C and E).
- Write out the coefficients to estimate each of these contrasts.
 - Are these two contrasts orthogonal? Explain why or why not.
- (h) 5 pts. Construct a 95% confidence interval for the difference between the two ‘super-sterile’ treatments, (C and E)
- (i) 5 pts. The investigators are also interested in whether there are any differences in quality among the three ‘sterile’ treatments, (A, B, and D). That is, they want to test $H_0: \mu_A = \mu_B = \mu_D$. If this is possible **without computing Sums-of-Squares from the raw data**, please test H_0 : no differences between treatments A, B, and D; report the F statistic and approximate the p-value. If this test is not possible without additional information, say so and tell me what additional information you need.
I am not expecting you to compute SS from the raw data.
- (j) 5 pts. The investigators will repeat this study. It turns out to be very difficult to change the temperature many times during the day. It would be easier to use only one treatment (time/temperature combination) each day. That design requires 30 days to collect 6 replicates of these 5 treatments. The investigators ask for your opinion. Should they continue using all five treatments in a day, or should they use just one treatment per day? Provide a short justification for your recommendation. Note: **one more part of this question on the next page**

- (k) 5 pts. The investigators are intrigued by the apparent slight decrease in quality in treatment D (60/15), compared to treatments A and B. They want to conduct another study using just these three treatments (A, B and D). Again, they will use a block design and want you to suggest a sample size (number of days). They are concerned primarily about the contrast between treatment D and the average of treatments A and B. You agree to use a standard deviation of 1. They want an $\alpha = 5\%$ test to have 80% power to detect the relatively small difference of 0.35. How many days should they use?

SAS code and part of the SAS output for problem 3 (quality of canned peas)

```
data food;
  infile 'foodqual.txt';
  input day trt quality;

proc glm;
  class day trt;
  model quality = day trt;

  lsmeans trt /stderr;

  estimate 'ave of A,B,D - ave of C,E' trt 2 2 -3 2 -3 /divisor = 6;
  contrast 'ave of A,B,D - ave of C,E' trt 2 2 -3 2 -3;

  estimate 'ave of A,B,C - ave of D,E' trt 2 2 2 -3 -3 /divisor = 6;
  contrast 'ave of A,B,C - ave of D,E' trt 2 2 2 -3 -3;

  estimate 'difference C - E' trt 0 0 1 0 -1;
  contrast 'difference C - E' trt 0 0 1 0 -1;

  estimate 'difference D - E' trt 0 0 0 1 -1;
  contrast 'difference D - E' trt 0 0 0 1 -1;

run;
```

Class Level Information

Class	Levels	Values
day	6	1 2 3 4 5 6
trt	5	50/60 55/30 55/45 60/15 60/20
	Number of Observations Read	30
	Number of Observations Used	30

Note: Some output deleted.

Continuation of the SAS output for problem 3 (quality of canned peas)

Least Squares Means

trt	quality LSMEAN	Standard Error	Pr > t
50/60	1.81666667	0.40163555	0.0002
55/30	1.83333333	0.40163555	0.0002
55/45	2.95000000	0.40163555	<.0001
60/15	2.15000000	0.40163555	<.0001
60/20	4.81666667	0.40163555	<.0001

Dependent Variable: quality

Contrast	DF Contrast	SS	Mean Sq.	F Value	Pr > F
ave of A,B,D - ave of C,E	1	27.378	27.378	28.29	<.0001
ave of A,B,C - ave of D,E	1	11.858	11.858	12.25	0.0023
difference C - E	1	10.453	10.453	10.80	0.0037
difference D - E	1	21.333	21.333	22.04	0.0001

Parameter	Estimate	Standard Error	t Value	Pr > t
ave of A,B,D - ave of C,E	-1.95000000	0.36664141	-5.32	<.0001
ave of A,B,C - ave of D,E	-1.28333333	0.36664141	-3.50	0.0023
difference C - E	-1.86666667	0.56799844	-3.29	0.0037
difference D - E	-2.66666667	0.56799844	-4.69	0.0001