

## STAT 305 D Homework 2 Solutions

Due January 31, 2012 at 12:40 PM in class

1. Vardeman and Jobe Chapter 1 Exercise 2 (page 23): If factor A has levels 1, 2, and 3, factor B has levels 1 and 2, and factor C has levels 1 and 2, list the combinations of A, B, and C that make up a full factorial arrangement.

You can list all the combinations in a table:

A	B	C
1	1	1
1	1	2
1	2	1
1	2	2
2	1	1
2	1	2
2	2	1
2	2	2
3	1	1
3	1	2
3	2	1
3	2	2

2. Vardeman and Jobe Chapter 1 Exercise 10 (page 24): Give an example of a  $2 \times 3$  full factorial data structure that might arise in a student study of the breaking strengths of wooden dowels. (Name the two factors involved, their levels, and write out all six different combinations. You must create the factors yourself. They are not given to you in this problem.) Then make up a data collection form for the study. Plan to record both the breaking strength and whether the break was clean or splintered for each dowel, supposing that three dowels of each type are to be tested.

In a factorial study, the factors are treatments, not blocks. Hence, I must choose factors that describe things that the experimenter does to the dowels. Humidity with levels high, medium, and low is a treatment factor because I can subject dowels to different humidities during the two weeks before testing breaking strength. I can also choose whether or not to soak them in water the day before the test, which gives me another treatment factor with two levels (soak or not soak).

Here is a corresponding data collection form if I assign one dowel per combination of treatment levels:

Humidity	Soak?	Breaking strength	Clean break?
High	Yes		
High	No		
Med	Yes		
Med	No		
Low	Yes		
Low	No		

Since I haven't carried out the experiment yet, I'm leaving the last two columns blank. You could have also filled them with hypothetical data, but just be aware that you don't actually know what the breaking strengths and kinds of break will be until you apply the treatments and do the experiment.

The above data collection form has only one row per treatment combination, so there is only one dowel per treatment group. A better data collection form would thus allow for more than one dowel per treatment group: i.e., multiple *replicates*:

Humidity	Soak?	Breaking strength	Clean break?
High	Yes		
High	Yes		
High	Yes		
High	No		
High	No		
High	No		
Med	Yes		
Med	Yes		
Med	Yes		
Med	No		
Med	No		
Med	No		
Low	Yes		
Low	Yes		
Low	Yes		
Low	No		
Low	No		
Low	No		

3. Vardeman and Jobe Chapter 2 Exercise 3 (page 64):

An experiment is to be performed to compare the effects of two different methods for loading gears in a carburizing furnace on the amount of distortion produced in a heat treating process. Thrust face runout will be measured for gears laid and for gears hung while treating.

- a. 20 gears are to be used in the study. Randomly divide up the gears into a group (of 10) to be laid and a group (of 10) to be hung, using Table B.1 (available at <http://will-landau.com/stat305/tables/random-digits.pdf>). Describe carefully how you do this. If you use the table, begin in the upper left corner.

First, I give each of the 20 gears and 2-digit index from 00 to 19. I move along the random number table, starting at the top left:

12	15	9	66	14	4	05	09	1	13	44	6	45	65	3	13	68	4	66	02	4	91	41	0	51	35	1	22	77	2
30	15	6	90	51	9	95	78	5	47	54	4	66	73	5	35	75	4	11	08	8	67	31	0	19	72	0	08	37	9
59	06	9	01	72	2	53	33	8	41	94	2	65	11	8	71	23	6	01	93	2	70	34	3	25	81	2	62	27	5
54	10	7	58	08	1	82	47	0	59	40	7	13	47	5	95	87	2	16	26	8	78	43	6	39	25	1	64	24	7
99	68	1	81	29	5	06	31	5	28	21	2	45	02	9	57	70	1	96	32	7	85	43	6	33	61	4	29	07	0

I pick the first 10 of these indices I encounter: 12, 15, 05, 09, 11, 02, 14, 10, 19, and 08, which are shown in blue. I assign these gears to be laid. I assign the rest to be hung.

- b. What are some purposes of the randomization used in part (a)?

The purpose of randomization is to make sure that there is no systematic difference between the treatment groups before a treatment is applied. That way, we can be reasonably sure that all the systematic differences between the treatment groups that we see after the experiment were caused by the treatment.

4. Now, suppose you want to distinguish between the big gears (10 of them) and the small gears (10 of them). You want to design an improved study to test the effect of gear arrangement on gear distortion. This new study takes the different gear sizes into account. (After all, small gears may distort less than big gears, or vice versa.) You plan to lay 5 small gears, hang 5 small gears, lay 5 big gears, and hang 5 big gears.
- a. Name the design of the study. Also, construct a table with all sample units as rows and variables as columns to show all the combinations of levels of the variables.

This is a randomized complete block design with gear arrangement as a treatment variable and gear size as the blocking variable. The table looks like this:

Gear size	Gear arrangement	Distortion
Big	Laid	
Big	Laid	
Big	Laid	
Big	Laid	
Big	Laid	
Big	Hung	
Big	Hung	
Big	Hung	
Big	Hung	
Big	Hung	
Small	Laid	
Small	Laid	
Small	Laid	
Small	Laid	
Small	Laid	
Small	Hung	
Small	Hung	
Small	Hung	
Small	Hung	
Small	Hung	

I haven't done the experiment, so I leave the gear distortion column blank. I didn't even need to include that column, but I include it to show that, since there is one gear per row, I record one distortion measurement per row during the experiment.

- b. Carry out the appropriate randomization to assign the 20 gears to the appropriate treatment groups for the experiment.

Since this is a randomized complete block design, I randomize gears to treatment groups *within each block*. That means I first randomize the small gears to be laid or hung, and then I randomize the big gears to be laid or hung.

- i. Big gears:

I assign each of the big gears an index from 0 to 9. Starting at the top left corner of the random number table, I move along the random digits, picking the first unique five I see to be laid:

12159	65144	05091	13446	45653	13684	66024	91410	51351	22772
30156	90519	95785	47544	66735	35754	11088	67310	19720	08379
59069	01722	53338	41942	65118	71236	01932	70343	25812	62275
54107	58081	82470	59407	13475	95872	16268	78436	39251	64247
99681	81295	06315	28212	45029	57701	96327	85436	33614	29070

As I show with blue boxes, I pick big gears 1, 2, 5, 9, and 6 to be laid. I pick the rest to be hung.

ii. Small gears:

I assign each of the small gears an index from 0 to 9. Then, picking up where I left off in the random number table, I move along the random digits, picking the first unique five I see to be laid:

12189	68114	05091	13446	45653	13684	66024	91410	51351	22772
30156	90519	95785	47544	66735	35754	11088	67310	19720	08379
59069	01722	53338	41942	65118	71236	01932	70343	25812	62275
54107	58081	82470	59407	13475	95872	16268	78436	39251	64247
99681	81295	06315	28212	45029	57701	96327	85436	33614	29070

As I show with green boxes, I pick small gears 6, 1, 4, 0, and 5 to be laid. I pick the rest to be hung.

Note that it's important to continue where you left off in the random number table from the big gears when you randomize the small gears. If you start from the top left corner for each, you will pick big gears 1, 2, 5, 9, and 6 to be laid and small gears 1, 2, 5, 9, and 6 to be laid. Instead, I want to randomize the big gears independently of the small gears.

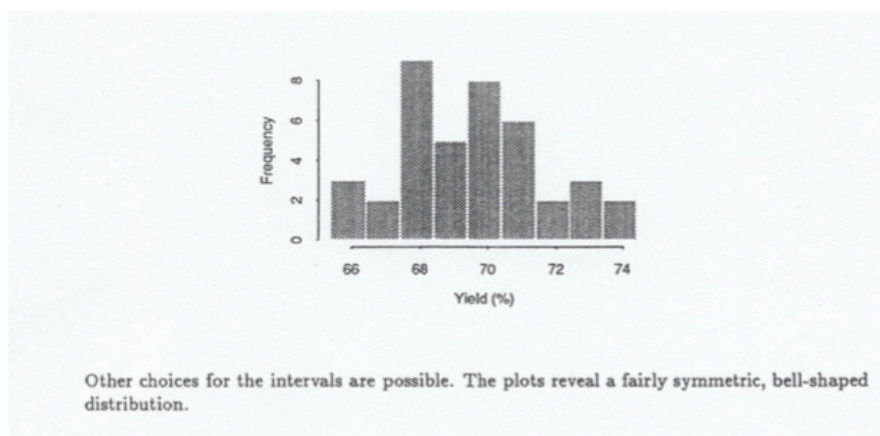
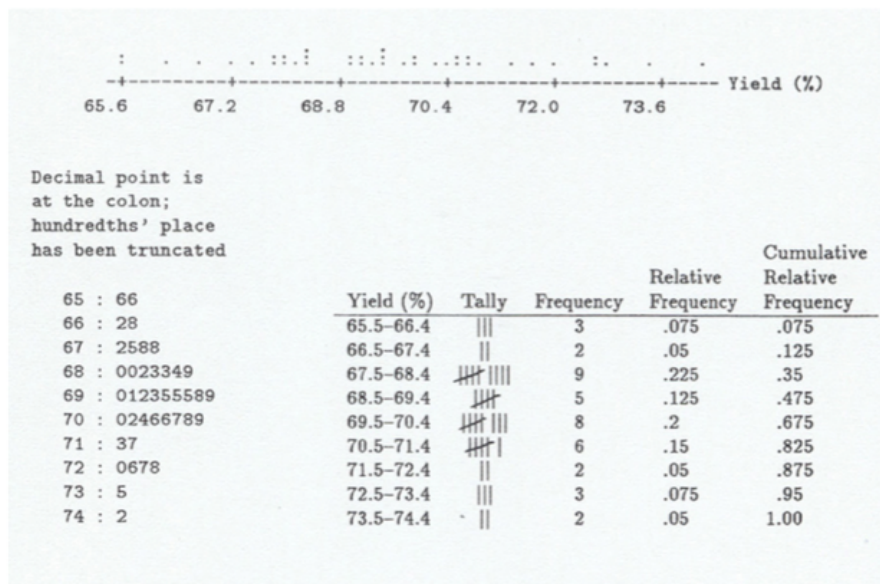
5. Vardeman and Jobe Chapter 3 Section 1 problem 1 (page 77):

The following are percent yields from 40 runs of a chemical process, taken from J. S. Hunter's article "The Technology of Quality" (*RCA Engineer*, May/June 1985):

65.6, 65.6, 66.2, 66.8, 67.2, 67.5, 67.8, 67.8, 68.0, 68.0, 68.2, 68.3, 68.3, 68.4, 68.9, 69.0, 69.1, 69.2, 69.3, 69.5, 69.5, 69.5, 69.8, 69.9, 70.0, 70.2, 70.4, 70.6, 70.6, 70.7, 70.8, 70.9, 71.3, 71.7, 72.0, 72.6, 72.7, 72.8, 73.5, 74.2

Make a dot diagram, a stem-and-leaf plot, a frequency table, and a histogram of these data.

**Roughly, what is the distributional shape of the data?**



6. Vardeman and Jobe Chapter 3 Section 2, part (a) of problem 1 (page 92):

The following are data (from *Introduction to Contemporary Statistical Methods* by L. H. Koopmans) on the impact strength of sheets of insulating material cut in two different ways. (The values are in ft lb.)

Lengthwise Cuts	Crosswise Cuts
1.15	.89
.84	.69
.88	.46
.91	.85
.86	.73
.88	.67
.92	.78
.87	.77
.93	.80
.95	.79

Find the medians, quartiles, and the .37 quantiles of the two datasets.

(a)

$i$	$\frac{i-.5}{10}$	$Q_L(\frac{i-.5}{10})$	$Q_C(\frac{i-.5}{10})$
1	.05	.84	.46
2	.15	.86	.67
3	.25	.87	.69
4	.35	.88	.73
5	.45	.88	.77
6	.55	.91	.78
7	.65	.92	.79
8	.75	.93	.80
9	.85	.95	.85
10	.95	1.15	.89

For the lengthwise sample:

$$\text{Median} = Q(.5) = \frac{.88 + .91}{2} = .895$$

$$1st \text{ Quartile} = Q_1 = Q(.25) = .870$$

$$3rd \text{ Quartile} = Q_3 = Q(.75) = .930$$

$$Q(.37) = .880$$

For the crosswise sample:

$$\text{Median} = \frac{.77 + .78}{2} = .775$$

$$Q_1 = .690$$

$$Q_3 = .800$$

$$Q(.37) = (.8)(.73) + (.2)(.77) = .738$$

You can read  $Q(0.25)$  and  $Q(0.75)$  directly from the table since 0.25 and 0.75 are each  $\frac{i-.5}{10}$  for  $i = 3$  and  $i = 8$ , respectively. And as I said in class, since there is an even number of observations for both the crosswise data and the lengthwise data, the median in each case is just the average of the middle two values. Hence, since  $n = 10$  for both lists of numbers, we average  $x_5$  and  $x_6$  to get the median in each case.

For the lengthwise sample, here is how I calculate  $Q(0.37)$ . First, I pick  $i' = np + 0.5 = 10 \cdot 0.37 + 0.5 = 4.2$ . Then, I calculate  $Q(0.37)$  as follows:



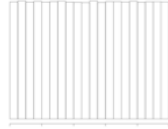
$$\begin{aligned}
Q(0.37) &= (\lceil i' \rceil - i')x_{\lfloor i' \rfloor} + (i' - \lfloor i' \rfloor)x_{\lceil i' \rceil} \\
&= (\lceil 4.2 \rceil - 4.2)x_{\lfloor 4.2 \rfloor} + (4.2 - \lfloor 4.2 \rfloor)x_{\lceil 4.2 \rceil} \\
&= (5 - 4.2)x_4 + (4.2 - 4)x_5 \\
&= (0.8)x_4 + (0.2)x_5 \\
&= (0.8)(0.88) + (0.2)(0.88) \\
&= 0.88
\end{aligned}$$

For the crosswise sample, here is how I calculate  $Q(0.37)$ . As before, I pick  $i' = np + 0.5 = 10 \cdot 0.37 + 0.5 = 4.2$ . Then, I calculate  $Q(0.37)$  as follows:

$$\begin{aligned}
Q(0.37) &= (\lceil i' \rceil - i')x_{\lfloor i' \rfloor} + (i' - \lfloor i' \rfloor)x_{\lceil i' \rceil} \\
&= (\lceil 4.2 \rceil - 4.2)x_{\lfloor 4.2 \rfloor} + (4.2 - \lfloor 4.2 \rfloor)x_{\lceil 4.2 \rceil} \\
&= (5 - 4.2)x_4 + (4.2 - 4)x_5 \\
&= (0.8)x_4 + (0.2)x_5 \\
&= (0.8)(0.73) + (0.2)(0.77) \\
&= 0.738
\end{aligned}$$

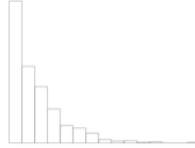
7. Identify the following distributional shapes:

a. (4 points)



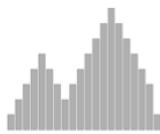
Uniform, symmetric,  
multi-modal

d. (4 points)



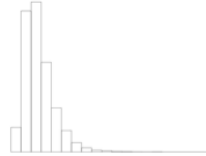
Truncated, ~~unimodal~~, right-skewed,  
asymmetric

b. (4 points)



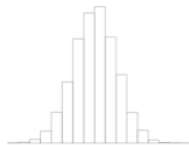
Asymmetric, bimodal

e. (4 points)



Mound-shaped, right-skewed,  
asymmetric

c. (4 points)



Symmetric, ~~unimodal~~, bell-shaped

8. Weekly feedback. You get full credit as long as you write something.
1. Is there any aspect of the subject matter that you currently struggle with? If so, what specifically do you find difficult or confusing? The more detailed you are, the better I can help you. **You got full credit as long as you wrote something.**
  2. Do you have any questions or concerns about the material, class logistics, or anything else? If so, fire away. **You got full credit as long as you wrote something.**