Helicopter Activities to Compare Two Sample and Paired Designs

From the
NCSSM Statistics Leadership Institute 2000

Laura Hanna
Parkville Magnet
Baltimore, MD
adam1214@bellatlantic.net

Janet Hassan
Science Academy
Mercedes, TX
SciAcad@aol.com

Beverly Johnson
Ft. Worth Country Day
Ft. Worth, TX
bjohnson@mail.fwcds.pvt.tenet.edu

This packet includes:

- Student Activity Handouts
- Teacher Solution Guide
- Notes for Teachers
- Commentary on Design
- Suggestions for Other Studies
- Activity Templates

A Comparison of Two Sample and Paired Designs

Question: Do long-rotor paper helicopters take a different length of time to fall, on average, than short-rotor paper helicopters?

Materials:

1. helicopters constructed of paper (card stock or copier paper) – both long rotor and short rotor
2. a meter stick with a clothes pin (pinch-type) attached on the end with masking tape, for precision helicopter dropping
3. a stopwatch
4. paper clips
5. a deck of cards (with 30 red cards and 30 black cards) for randomization for Procedure 1 and Procedure 3
6. a coin for randomization for Procedure 2

Procedure One

1. Construct one long-rotor helicopter and one short-rotor helicopter (see diagram and instructions)
2. Shuffle the 60 cards to determine the order of helicopter drops (*e.g.*, red = short-rotor helicopter, black = long-rotor helicopter)
3. Select a height (at least 2 meters) from which to drop the helicopters. (Placing the meter stick dropping device on a step ladder reduces human variability in dropping techniques).
4. Turn over the top card to determine which helicopter to drop; set the card aside (do not reshuffle).
5. Clip the selected helicopter in the clothespin. Have the releaser say "Ready…Set…GO," at which time the releaser releases the helicopter from the clothespin and the timer starts timing. When the helicopter hits the floor the timer stops the stopwatch. Record the time.
6. Repeat steps 4 and 5 until the deck is depleted.

TWO SAMPLE DESIGN DATAQ (Repeated Drops with One Helicopter)

Trial Number	Short Rotor	Long Rotor
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		
21		
22		
23		
24		
25		
26		
27		
28		
29		
30		

I. Graphical Analysis

Using long- and short-rotor helicopter descent times construct the following:

1. a stem plot for each rotor length (or a back-to-back stem plot)
2. a modified box plot for each rotor length using the same scale
3. a histogram for each rotor length using the same scale.

Using the graphical information, compare the shape, center, and spread of the distributions including outliers and any other unusual features.

II. Descriptive Statistics

1. Calculate the mean, variance, and the five-number summary (minimum, first quartile, median, third quartile, maximum) of each of the two sets of data.

2. Using the descriptive statistics, compare the shape, center, and spread of the distributions including outliers and any other unusual features.

III. Inferential Statistics

1. What is the population about which inference can be made?

2. What is the appropriate inference procedure for comparing the mean descent times for the two helicopters? (*e.g.*: paired *t*-test, two independent sample *t*-test, *z*-test, etc.)

3. State and justify the assumptions necessary to apply this inference procedure.

4. Construct a 95% confidence interval for the difference in the mean descent time for each helicopter. Discuss and interpret the meaning of the confidence interval.

5. Use an appropriate significance test to determine whether or not there is a difference in the true mean descent times of the two helicopters.

6. How do the results of the test of significance relate to the observations you made using your confidence intervals?

Procedure Two

1. Construct one long-rotor helicopter and one short-rotor helicopter.
2. Select a height (over two meters) from which to drop the helicopters (Placing the meter stick dropping device on a step ladder reduces human variability in dropping techniques).
3. The helicopters will be dropped in pairs. Flip a coin to determine which of the two (long-rotor or short-rotor) will be dropped first (*e.g.*, heads = short-rotor helicopter, tails = long-rotor helicopter).
4. Clip the selected helicopter in the clothespin. Have the releaser say “Ready…Set…GO,” at which time the releaser releases the helicopter from the clothespin and the timer starts timing. When the helicopter hits the floor the timer stops the stopwatch. Record the time.
5. Repeat step four using the helicopter with the rotor length not used in step four.
6. Repeat steps three, four and five until thirty pairs of helicopter drops have been performed.
7. For each pair, record the difference between the long-rotor helicopter and the short-rotor helicopter descent times (long – short).

RANDOMIZED PAIRS DESIGN DATA

Trial Number	Long Rotor	Short Rotor	Difference
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			
23			
24			
25			
26			
27			
28			
29			
30			

I. Graphical Analysis

A. Using long-rotor descent times, short-rotor descent times, and difference between long- and short- rotor descent times, construct the following:

1. a box-plot for each of the three distributions
2. a histogram for each of the three distributions.
3. a stemplot for the differences of the short- and long-rotor descent times

B. Respond to the following:

1. Using the graphical information describe the shape, center and spread of the distribution of the difference in descent times (include outliers and any other unusual features).

2. Compare the variation in the descent time of the long- and short-rotor helicopters with the variation in the differences.

II. Descriptive Statistics

1. Calculate the mean and variance of each of the three sets of data.

2. What would you expect to be true about the three variances calculated in step one if short- and long-rotor descent times were independent in all pairs? Does it appear to be the case that the descent times are independent?

3. Compare the difference between the means of the long- and short- rotor helicopter descent times with the mean of the differences in the descent times.

III. Inferential Statistics

1. What is the population about which inferences are being made?

2. What is the appropriate inference procedure for comparing the mean difference in descent times for the two helicopters?

3. State and justify the assumptions necessary to apply this inference procedure.

4. Construct a 95% confidence interval for the mean difference in descent time for two helicopters.

5. Does the confidence interval contain 0? What is the significance of this observation?

6. Use an appropriate test to determine whether or not the mean difference in descent time is 0.

7. How do the test results relate to the observation you made about the confidence interval?

Procedure Three

1. Construct thirty long-rotor helicopters and thirty short-rotor helicopters.
2. Shuffle the 60 cards to determine the order of helicopter drops (*e.g.*, red = short-rotor helicopter, black = long-rotor helicopter)
3. Select a height (over two meters) from which to drop the helicopters (Placing the meter stick dropping device on a step ladder reduces human variability in dropping techniques).
4. Turn over the top card to determine which type of helicopter to drop; set the card aside (do not reshuffle).
5. Randomly select a helicopter of the type consistent with your card. Clip the selected helicopter in the clothespin. Have the releaser say "Ready…Set…GO," at which time the releaser releases the helicopter from the clothespin and the timer starts timing. The timer stops the stopwatch when the helicopter hits the floor. Record the time.
6. Repeat steps 4 and 5 until the deck is depleted. Do not reuse a helicopter.

TWO SAMPLE DESIGN DATA (Single Drops with Each Helicopter)

Trial Number	Short Rotor	Long Rotor
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		
21		
22		
23		
24		
25		
26		
27		
28		
29		
30		

I Graphical Analysis

Using long- and short-rotor helicopter descent times construct the following:

1. a modified box plot for each rotor length using the same scale
2. a stem plot for each rotor length (or a back-to-back stem plot)
3. a histogram for each rotor length using the same scale.

Using the graphical information, compare the shape, center, and spread of the distributions including outliers and any other unusual features.

II Descriptive Statistics

1. Calculate the mean, variance, and the five-number summary (minimum, first quartile, median, third quartile, maximum) of each of the sets of data.

2. Using the descriptive statistics, compare the shape, center, and spread of the distributions including outliers and any other unusual features.

III Inferential Statistics

1. What is the population about which inference is being made?

2. What is the appropriate inference procedure for comparing the mean descent times for the two types of helicopters?

3. State and justify the assumptions necessary to apply this inference procedure.

4. Construct a 95% confidence interval for the mean descent time for each helicopter. Compare the two confidence intervals.

5. Use an appropriate test to determine whether or not there is a difference in the mean descent times of the two type helicopters.

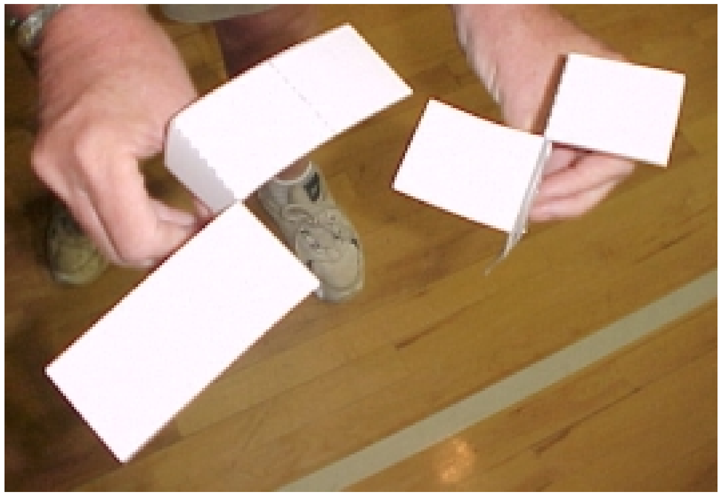
6. How do the results of the test of significance relate to the observations you made using your confidence intervals?

Discussion

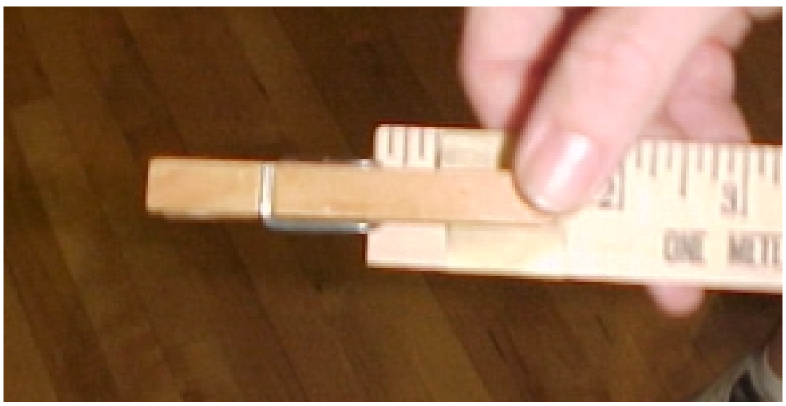
1. Compare and contrast the experimental designs and analyses in Procedure One and Procedure Two.
   What are the advantages and disadvantages of paired versus two-sample design?

2. Compare and contrast the experimental designs and analyses in Procedure One and Procedure Three.
   What are the advantages and disadvantages of using one helicopter of each type and dropping thirty times each as opposed to dropping 30 helicopters of each type one time each?

Pictures



Long-Rotor and Short-Rotor Helicopters



Clothespin Launching Tool (View from Above)

Short-Rotor Helicopter Prepared to be Launched (Side View)

Teacher's Solution Guide to Procedures

The data provided here are results obtained when three AP Statistics teachers attending the Statistics Leadership Institute at the North Carolina School of Science and Mathematics (July, 2000) performed these experiments using these instructions. The graphics and statistical information included here were obtained using JMP IN (pronounced "jump in"), statistical software developed by the SAS Institute.

Procedure One

Procedure One represents a typical two-sample experimental design. The order of the dropping of the helicopter is randomly determined by the deck of cards with the constraint that when the experiment is finished, the short-rotor helicopter and the long-rotor helicopter each had been dropped 30 times.

Note: Graphical and numerical analyses from JMP IN are provided on the following page.

I. Graphical Analysis

The centers of the two distributions are noticeably different. The center of the short-rotor helicopter descent times distribution is between 0.9 and 1.0 seconds while the center of the long-rotor helicopter descent times is between 1.2 and 1.3 seconds.

The shape of the short-rotor helicopter descent times distribution is slightly skewed to the right (towards the higher values). The boxplot indicates that the highest value of 1.41 is an outlier. The shape of the long-rotor helicopter descent times is fairly symmetric and mounded in the middle. The boxplot does not show the value above 1.6 to be an outlier. The stemplots for each of these distributions corroborates these observations.

Most of the short-rotor helicopter descent times were between 0.82 and 1.2 seconds which is a spread of approximately 0.4 seconds. The long-rotor helicopter descent times were between 0.95 and 1.44 seconds which is a spread of approximately .5 seconds. The spreads of these two distributions are approximately the same.

II. Descriptive Statistics

The mean of the short-rotor helicopter descent times distribution is 1.012 seconds and the median is 0.985 seconds. The mean of the long-rotor helicopter descent times is 1.246 seconds and the median is 1.235 seconds. The center of the short-rotor helicopter descent times distribution is less than that of the long-rotor helicopter.

The shape of the long-rotor helicopter descent times is fairly symmetric, which would be expected since the mean and the median are close to the same value (1.246 seconds and 1.235 seconds). The long-rotor helicopter descent times distribution has no outliers since the IQR is 0.19, $1.5 \cdot \text{IQR} = 0.285$, $Q3 + 1.5 \cdot \text{IQR} = 1.635$ and $Q1 - 1.5 \cdot \text{IQR} = 0.875$. All data are between these two values.

The shape of the short-rotor helicopter descent times distribution is slightly skewed to the right, which would be expected since the mean (1.012) is greater than the median (0.985). The short-rotor helicopter descent times distribution has one outlier. Since the IQR is 0.165, any time above 1.34 seconds or below 0.68 seconds can be considered an outlier ($Q3 + 1.5 \cdot \text{IQR} = 1.34$, $Q1 - 1.5 \cdot \text{IQR} = 0.68$). The maximum value, 1.41 seconds, is an outlier on the high side. Since no values were lower than 0.68, there are no outliers in the low range.

The variance of the descent times of the short-rotor helicopters is 0.017 seconds^2 while the variance of the descent times of the long-rotor helicopters is 0.019 seconds^2. The spreads of these two distributions are approximately the same.

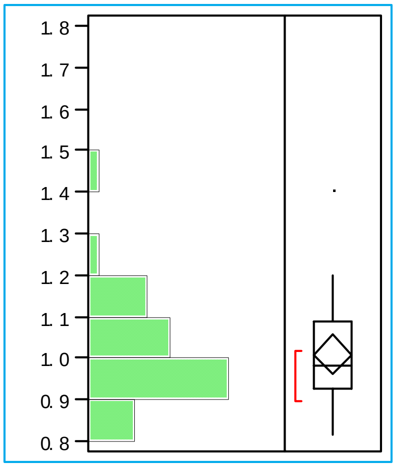
Descent Times of Long-Rotor and Short-Rotor Helicopters
with
1/Short and 1/Long from Procedure 1;
2/Long, 2/Short and 2/Difference from Procedure 2;
and
3/Short and 3/Long from Procedure 3.

3/Short	3/Long	2/Short	2/Long	2/Difference	1/Short	1/Long
1.08	1.76	1.16	1.67	0.51	0.93	1.38
1.18	1.73	1.08	1.73	0.65	0.84	1.27
1.22	1.58	1.06	1.43	0.37	1.14	1.29
1.03	1.33	0.83	1.71	0.88	1.00	1.38
1.27	1.64	1.00	1.32	0.32	0.84	1.02
1.15	1.76	1.17	1.59	0.42	1.08	1.39
1.03	1.58	1.14	1.14	0.00	1.19	1.38
0.99	1.81	1.17	1.79	0.62	1.00	1.44
0.78	1.82	0.94	1.35	0.41	1.05	1.13
1.41	1.86	1.00	1.24	0.24	1.09	1.21
0.84	1.96	0.89	1.68	0.79	0.93	1.40
1.08	1.88	1.20	1.14	-0.06	1.16	0.95
0.99	1.83	1.07	1.44	0.37	1.41	1.24
1.25	1.72	0.92	1.39	0.47	1.20	1.26
1.38	1.50	0.81	1.62	0.81	1.10	1.18
1.29	1.43	1.06	1.85	0.79	1.15	1.13
1.17	1.81	1.25	1.65	0.40	0.82	1.12
1.06	1.79	1.00	1.72	0.72	1.02	1.21
1.26	1.71	1.17	1.36	0.19	0.97	1.23
0.96	1.73	0.80	1.74	0.94	0.96	1.13
1.24	1.21	1.34	1.90	0.56	0.85	1.17
1.10	1.75	1.30	1.77	0.47	0.92	1.20
1.00	1.89	1.16	1.70	0.54	0.92	1.26
1.65	1.67	1.37	1.96	0.59	0.94	1.04
1.13	1.75	1.01	1.45	0.44	0.97	1.34
1.10	1.45	0.90	1.78	0.88	0.90	1.62
1.23	1.60	1.02	1.84	0.82	0.99	1.22
1.4	1.96	1.06	1.80	0.74	1.02	1.19
1.14	2.05	1.06	1.73	0.67	0.98	1.27
1.36	1.70	0.91	1.44	0.53	0.98	1.32

Procedure 1

Histograms, Box Plots, Stem-and-Leaf Plots of Descent Time of a Long-Rotor and Short-Rotor Helicopter (time in seconds for a distance of 2.34 meters)

1/Short Rotor



Quantiles

maximum	100.0%	1.4100
	99.5%	1.4100
	97.5%	1.4100
	90.0%	1.1870
quartile	75.0%	1.0925
median	50.0%	0.9850
quartile	25.0%	0.9275
	10.0%	0.8410
	2.5%	0.8200
	0.5%	0.8200
minimum	0.0%	0.8200

Moments

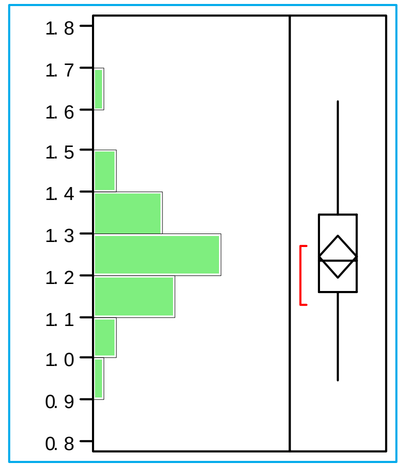
Mean	1.01167
Std Dev	0.12895
Std Error Mean	0.02354
Upper 95% Mean	1.05982
Lower 95% Mean	0.96352
N	30.00000
Sum Weights	30.00000

Stem and Leaf

Stem	Leaf	Count
14	1	1
13		
12	0	1
11	04569	5
10	0022589	7
9	022334677889	12
8	2445	4

Multiply Stem.Leaf by 10^-1

1/LongRotor



Quantiles

maximum	100.0%	1.6200
	99.5%	1.6200
	97.5%	1.6200
	90.0%	1.3990
quartile	75.0%	1.3500
median	50.0%	1.2350
quartile	25.0%	1.1600
	10.0%	1.0480
	2.5%	0.9500
	0.5%	0.9500
minimum	0.0%	0.9500

Moments

Mean	1.24567
Std Dev	0.13771
Std Error Mean	0.02514
Upper 95% Mean	1.29709
Lower 95% Mean	1.19425
N	30.00000
Sum Weights	30.00000

Stem and Leaf

Stem	Leaf	Count
16	2	1
15		
14	04	2
13	248889	6
12	01123466779	11
11	2333789	7
10	24	2
9	5	1
8		

Multiply Stem.Leaf by 10^-1

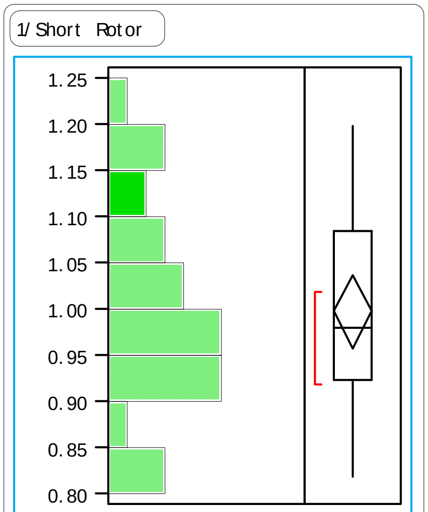
III. Inferential Statistics

1. The population to which inference can be made is all possible descent times for these two particular helicopters.

2. The appropriate inference procedure is a 2-sample *t*-test because the two samples of descent times are independent . The experiment used a completely randomized design.

3. The assumptions necessary are 1) that the underlying populations of the two distributions of descent times are normal and 2) that the observations are independent. The assumption of normality of the short-rotor helicopter descent times is possibly not met. The graphical and descriptive statistics indicate that the distribution is slightly skewed to the right with an outlier. Since the sample size is 30, however, slight skewness may be overlooked but the presence of the outlier may be problematic. Therefore, it is suggested that two analyses be considered: one in which the outlier is excluded and one in which the outlier is included.

Short-rotor helicopter excluding outlier



Quantiles

Moments

Mean	0.99793
Std Dev	0.10658
Std Error Mean	0.01979
Upper 95% Mean	1.03847
Lower 95% Mean	0.95739
N	29.00000
Sum Weights	29.00000

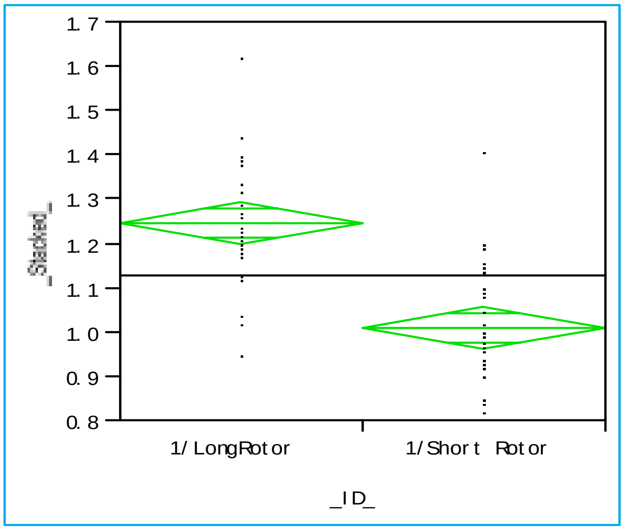
4. From the JMP IN output on the previous page, the 95% confidence interval for the mean of the short-rotor helicopter descent times, including the outlier, is (0.9635, 1.0598). Excluding the outlier, the 95% confidence interval is (0.9574, 1.0385). Thus, we are 95% confident that the mean descent time for the short-rotor helicopter is between 0.9635 seconds and 1.0598 seconds. If we were to repeat this experiment many times, approximately 95% of the intervals constructed from our different data sets will cover the true population mean.

   From the JMP IN output on the previous page, the 95% confidence interval for the mean of the long-rotor helicopter descent times is (1.1942, 1.2971).. Thus, we are 95% confident that the mean descent time for the long-rotor helicopter is between 1.1942 seconds and 1.2971 seconds. If we were to repeat this experiment many times, approximately 95% of the intervals constructed from our different data sets will cover the true population mean.

   All three confidence intervals (short-rotor with outlier, short-rotor without outlier, long-rotor) are approximately the same length (which follows from the fact that the variances of the data for long- and short-rotor helicopters were approximately the same). In addition, the confidence interval for the long-rotor helicopter mean descent time does not overlap either of the confidence intervals for the mean descent time of the short-rotor helicopters. Therefore we can confidently conclude that the two mean descent times are different.

Two-sample Test of Hypothesis of Equal Mean Times

Stacked By _ID_



Oneway Anova

Summary of Fit

t-Test

	Difference	t-Test	DF	Prob>|t|
Estimate	0.234000	6.794	58	<.0001
Std Error	0.034444			
Lower 95%	0.165053			
Upper 95%	0.302947			

Assuming equal variances

Analysis of Variance

Means for Oneway Anova

5. The above analysis is from JMP IN. The diamonds on the plots of the values illustrate the 95% confidence intervals that were calculated previously. The line at the widest part of the diamond is the mean. The vertical endpoints represent the ends of the 95% confidence interval. The smaller lines are called the overlap lines and if the diamonds don't overlap beyond these lines, the two means are significantly different.

Consider these hypotheses:

H_o: $\mu_{short} = \mu_{long}$

H_a: $\mu_{short} \neq \mu_{long}$

The assumptions of normality and independence of the short- and long-rotor helicopter descent times distributions were checked earlier. Because the outlier in the short-rotor helicopter descent times is cause for concern, the analysis can be done with the outlier and without the outlier. In any case, we might proceed with caution in drawing conclusions.

The value of the test statistic is t = 6.794 when all data are included. Because the variances were approximately equal in the earlier analysis, the t-statistic was calculated assuming equal variances. The probability of getting a t-value with 58 degrees of freedom at least this far from zero is less than 0.0001 if, in fact, the means of the two distributions of descent times are equal.

Therefore, we conclude that the mean descent times of long-rotor and short-rotor helicopters are not equal.

6. The confidence interval method and the test of significance method led to at the same conclusion.

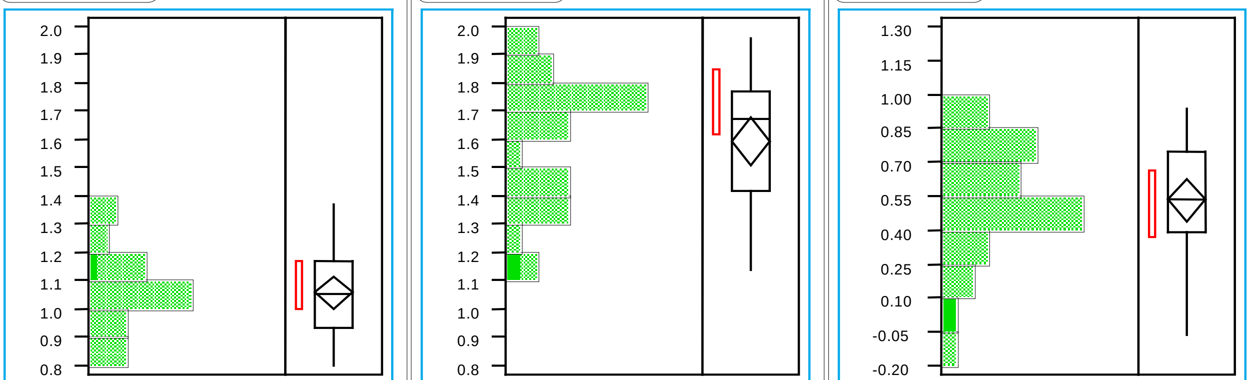
Procedure Two

In Procedure Two, the two helicopters (a short-rotor helicopter and a long-rotor helicopter) were dropped multiple times. What makes this experiment different from Procedure One is that in every two helicopter drops, one of the helicopters was a long- and the other was a short-rotor helicopter. What was randomized was the *order* in which the helicopters were dropped. This experiment has, therefore, a matched pairs design with 30 pairs of helicopter drops.

I. Graphical Analysis

Histograms, Box-plots, Stem-and-Leaf Plots of Descent Times of Paired Long- and Short-Rotor Helicopters

2/Short Rotor | 2/LongRotor | 2/Difference



2/Short Rotor

Quantiles

maximum	100.0%	1.3700
	99.5%	1.3700
	97.5%	1.3700
	90.0%	1.2950
quartile	75.0%	1.1700
median	50.0%	1.0600
quartile	25.0%	0.9350
	10.0%	0.8360
	2.5%	0.8000
	0.5%	0.8000
minimum	0.0%	0.8000

Moments

Mean	1.06167
Std Dev	0.15084
Std Error Mean	0.02754
Upper 95% Mean	1.11799
Lower 95% Mean	1.00534
N	30.00000
Sum Weights	30.00000

2/LongRotor

Quantiles

maximum	100.0%	1.9600
	99.5%	1.9600
	97.5%	1.9600
	90.0%	1.8490
quartile	75.0%	1.7725
median	50.0%	1.6750
quartile	25.0%	1.4200
	10.0%	1.2480
	2.5%	1.1400
	0.5%	1.1400
minimum	0.0%	1.1400

Moments

Mean	1.59767
Std Dev	0.22584
Std Error Mean	0.04123
Upper 95% Mean	1.68200
Lower 95% Mean	1.51334
N	30.00000
Sum Weights	30.00000

2/Difference

Quantiles

maximum	100.0%	0.94000
	99.5%	0.94000
	97.5%	0.94000
	90.0%	0.87400
quartile	75.0%	0.75250
median	50.0%	0.53500
quartile	25.0%	0.39250
	10.0%	0.19500
	2.5%	-0.06
	0.5%	-0.06
minimum	0.0%	-0.06

Moments

Mean	0.53600
Std Dev	0.25050
Std Error Mean	0.04573
Upper 95% Mean	0.62954
Lower 95% Mean	0.44246
N	30.00000
Sum Weights	30.00000

Stem and Leaf

Stem	Leaf	Count
9	4	1
8	1 2 8 8	4
7	2 4 9 9	4
6	2 5 7	3
5	1 3 4 6 9	5
4	0 1 2 4 7 7	6
3	2 7 7	3
2	4	1
1	9	1
0	**0**	1
-1	4	1
-2		

Multiply Stem.Leaf by 10^-1

B. Response

1. The distribution of the differences between the short- and long-rotor helicopter descent times for each trial was slightly skewed to the left (towards lower values). The left whisker (lower values) was longer than the right whisker (higher values) on the box plot. This is confirmed by the shape of the histogram and stem plot. There are no indicated outliers or other unusual features. The distribution of differences is centered at approximately 0.5 seconds.

2. As estimated by the difference between the maximum and minimum values on the box plot, the descent times of short-rotor helicopters showed less variation than did the descent times of the long-rotor helicopters. The box plot of the differences in descent times of each pair of helicopters showed an even greater amount of variation than either of the other two.

II. Descriptive Statistics

1. The variance of the distribution of descent times for short-rotor helicopters was 0.02275 seconds^2. The variance of the distribution of descent times for long-rotor helicopters was 0.05100 seconds^2. The variance of the distribution of difference of descent times of each pair of helicopters was 0.06275 seconds^2.

   If the long- and short-rotor helicopter descent times are independent, the sum of the variances of the individual descent times will equal the variance of the differences of the descent times at each trial. Since this is not the case here (0.02275 + 0.05100 = 0.07375; 0.07375 does not equal 0.06275) we can conclude that descent times of long- and short-rotor helicopters within a pair may not be independent. This is an indication that pairing was effective. If the sum of the variances of the individual descent times of the long- and short-rotor helicopters was about the same or smaller than the variance of the differences in descent time for each pair, pairing was ineffective.

2. The estimated mean descent time for short-rotor helicopters was 1.0617 and the estimated mean descent time for long-rotor helicopters was 1.5977. The difference between these is 0.536. The estimated mean of the difference of the descent times for each pair of helicopters was 0.536. This is not surprising since the difference of the means of two distributions is always equal to the mean of the differences.

III. Inferential Statistics

1. The population to which we can infer is the population of differences in time for the two particular helicopters which were used in this experiment.

2. Because the experimental design paired the two types of helicopter in each trial, a paired *t*-test is appropriate. Each trial consisted of dropping either a long- or short- rotor helicopter followed by the remaining helicopter.

3. The assumptions are 1) that the distribution of the differences in the descent times of the pair of helicopters at each trial is normal, 2) that we have a simple random sample of differences in descent time (the descent times of the long- and short-rotor helicopters need not be independent). The distribution of the descent times of the differences was slightly skewed. However, since the sample size is sufficiently large (n =30) and no outliers are present, it is safe to proceed with the paired *t*-test.

4. The 95% confidence interval for the mean of the distribution of differences is (0.44246, 0.62954). This means that we are 95% confident that the population mean of the difference in the descent times for each pair of helicopters is between 0.44246 seconds and 0.62954 seconds, meaning that long-rotor helicopters take between 0.44 and 0.63 seconds longer, on average, to fall from this height.

5. This interval does not contain 0. We can, therefore, conclude with confidence that the mean of the differences is not equal to 0.

6.

H_o: $\mu_{differences} = 0$

H_a: $\mu_{differences} \neq 0$

```
2/Difference
Quantiles
Moments
Test Mean=value
Hypothesized Value              0
Actual Estimate             0.536
                  t Test
Test Statistic  11.7199
Prob > |t|       <.0001
Prob > t         <.0001
Prob < t         1.0000
```

The *t*-statistic value with 29 degrees of freedom is 11.7199. The probability of getting a value this extreme or greater when the mean of the differences is really 0 is less than 0.0001.

We conclude that the mean of the differences of the descent times for pairs of helicopters is not 0 seconds. That is, the mean descent times of long- and short-rotor helicopters are truly different.

7. The results of the confidence interval construction led us to conclude that since the interval did not include 0, the mean of the differences in the descent time for each pair of helicopters was not 0. The results of the test of significance that the mean of the differences was equal to 0 also led us to conclude that the mean difference was not equal to 0. Since the hypothesis test was a two-sided test (we did not specify whether or not one was greater than the other), the results of the hypothesis test and the confidence interval were the same.

Procedure Three

Procedure Three and Procedure One are essentially the same in that the order of the dropping of the helicopters is randomly determined and whether or not a long-rotor helicopter is dropped has no bearing on whether the next helicopter dropped is a long- or short-rotor helicopter. The difference between Procedure One and Procedure Three is that in One, one helicopter of each rotor length is dropped thirty times and in Three, 60 different helicopters (30 of each rotor length) are dropped once each. The difference in the experiments then, will be in the scope of inference - not in the analysis procedures.

Note: Graphical and numerical analyses are on the following page.

I. Graphical Analysis

The centers of the two distributions are noticeably different. The center of the short-rotor helicopter descent times distribution is about 1.15 seconds while the center of the long-rotor helicopter descent times is about 1.75 seconds.

The shape of the distribution of the short-rotor helicopter descent times is approximately symmetric with gaps between 0.85 and between 0.95 seconds and between 1.45 and 1.65 seconds. The graph of the distribution is mounded in the middle. The boxplot indicates that there is an outlier on the high end. The shape of the distribution of long-rotor helicopter descent times is fairly symmetric, mounded in the middle. The boxplot reveals one outlier on the low end. The stemplots for each of these distributions corroborate these observations.

Most of the short-rotor helicopter descent times (between the gaps) were between 0.95 and 1.45 seconds which is a spread of approximately 0.5 second. The long-rotor helicopter descent times were between 1.15 and 2.15 seconds which is a spread of approximately 1 second.

II. Descriptive Statistics

The mean of the short-rotor helicopter descent times distribution is 1.159 seconds and the median is 1.145 seconds. The mean of the long-rotor helicopter descent times is 1.709 seconds and the median is 1.740 seconds. The center of the short-rotor helicopter descent times distribution is less than that of the long-rotor helicopter.

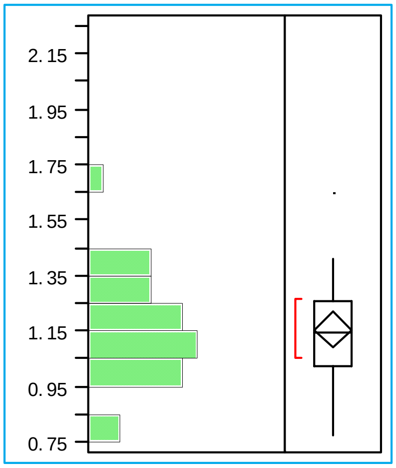
The shape of the long-rotor helicopter descent times is fairly symmetric, which would be expected since the mean (1.709 seconds) and the median (1.740 seconds) are close to the same value. The long-rotor helicopter descent times distribution has one outlier since the IQR is 0.228, $1.5 \cdot \text{IQR} = 0.342$, $\text{Q3} + 1.5 \cdot \text{IQR} = 2.165$ and $\text{Q1} - 1.5 \cdot \text{IQR} = 1.253$. Therefore, the minimum time, 1.210, is an outlier.

The shape of the short-rotor helicopter descent times distribution is approximately symmetric, which would be expected since the mean (1.159 seconds) is approximately equal to the median (1.145 seconds). The short-rotor helicopter and so descent times distribution has at least one outlier since the IQR is 0.233, $1.5 \cdot \text{IQR} = 0.3495$, $\text{Q3} + 1.5 \cdot \text{IQR} = 1.6125$, and so the maximum value of 1.650 is an outlier. Also, $\text{Q1} - 1.5 \cdot \text{IQR} = 0.6805$. Since no times were lower than 0.6805, there are no outliers in the low range.

The variance of the descent times of the short-rotor helicopters is 0.03264 seconds^2 while the variance of the descent times of the long-rotor helicopters is 0.03527 seconds^2. Therefore the spreads of these two distributions are approximately the same.

Histogram, Boxplots, Stem-and-Leaf Plots of Descent Times of 30 Long-Rotor and 30 Short-Rotor Helicopters (time in seconds for a distance of 2.34 m)

3/Short Rotor



Quantiles

maximum	100.0%	1.6500
	99.5%	1.6500
	97.5%	1.6500
	90.0%	1.3980
quartile	75.0%	1.2625
median	50.0%	1.1450
quartile	25.0%	1.0300
	10.0%	0.9630
	2.5%	0.7800
	0.5%	0.7800
minimum	0.0%	0.7800

Moments

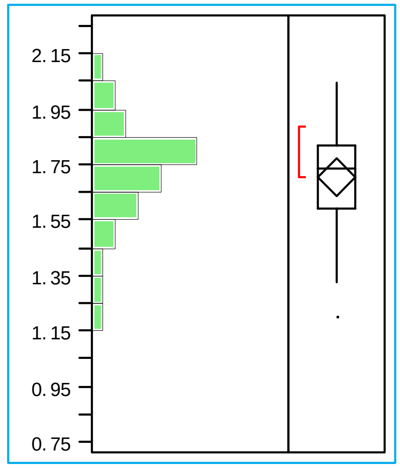
Mean	1.15900
Std Dev	0.18066
Std Error Mean	0.03298
Upper 95% Mean	1.22646
Lower 95% Mean	1.09154
N	30.00000
Sum Weights	30.00000

Stem and Leaf

Stem	Leaf	Count
16	5	1
15		
14	01	2
13	68	2
12	2345679	7
11	0034578	7
10	033688	6
9	699	3
8	4	1
7	8	1
6		

Multiply Stem Leaf by 10^-1

3/Long Rotor



Quantiles

maximum	100.0%	2.0500
	99.5%	2.0500
	97.5%	2.0500
	90.0%	1.9530
quartile	75.0%	1.8225
median	50.0%	1.7400
quartile	25.0%	1.5950
	10.0%	1.4320
	2.5%	1.2100
	0.5%	1.2100
minimum	0.0%	1.2100

Moments

Mean	1.70867
Std Dev	0.18785
Std Error Mean	0.03430
Upper 95% Mean	1.77881
Lower 95% Mean	1.63852
N	30.00000
Sum Weights	30.00000

Stem and Leaf

Stem	Leaf	Count
20	5	1
19	66	2
18	1123689	7
17	0123355669	10
16	047	3
15	088	3
14	35	2
13	3	1
12	1	1

Multiply Stem Leaf by 10^-1

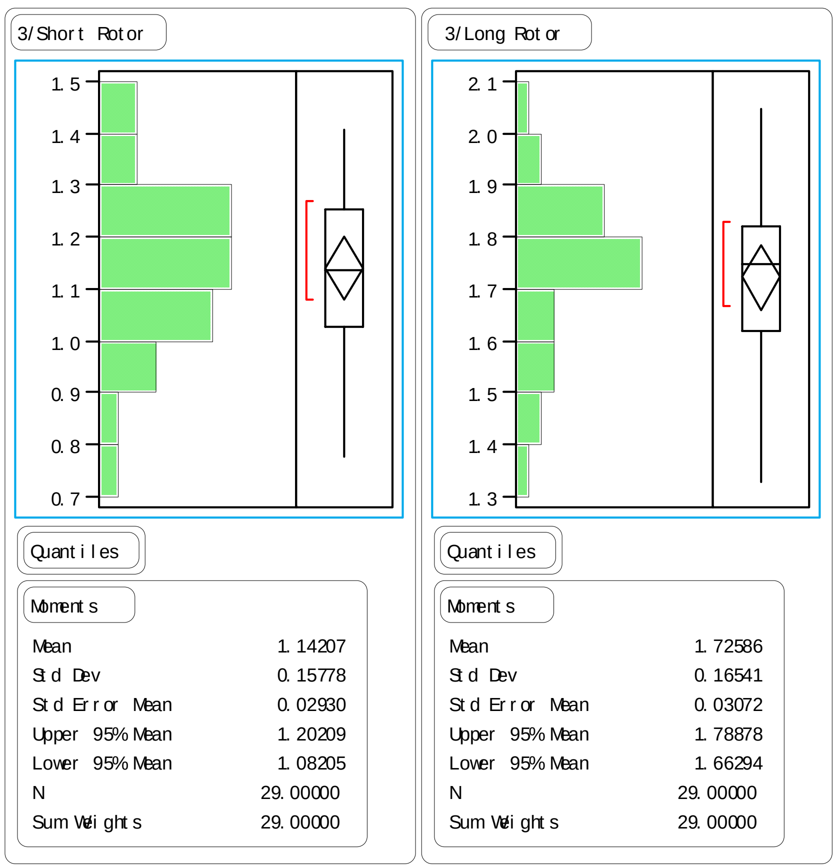
III. Inferential Statistics

1. The population to which inference can be made is the descent times of long- and short-rotor helicopters of these two types, *in general*. (See Commentary for further discussion of this.)

2. The appropriate inference procedure is a 2-sample t-test because the randomization procedure did not pair the drops. Rather, we have two independent samples.

3. The assumptions necessary are that the underlying populations of the two distributions of descent times are normal and that the observations are independent. The assumptions of normality of the helicopter descent times are possibly not met. The graphical and descriptive statistics indicate that the distributions are roughly symmetric but each has an outlier. Since the sample size is 30, slight skewness may be overlooked but the presence of the outliers may be problematic. Therefore, it is suggested that two analyses be considered: one in which the outliers are excluded and one in which the outliers are included.

Short-Rotor and Long-Rotor Descent Times With Outlier Excluded From Each Set

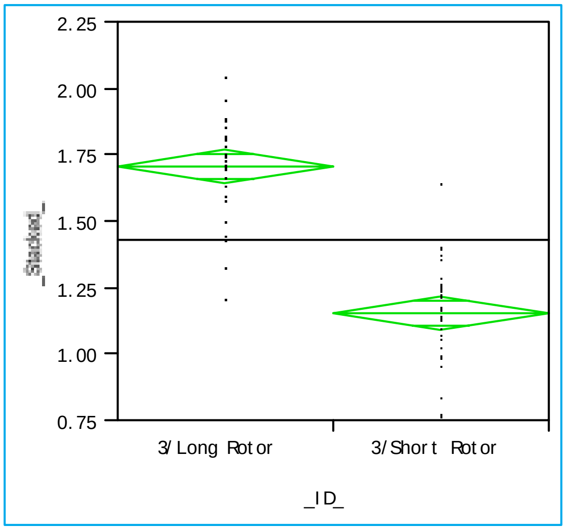


4. The 95% confidence interval for the mean descent time for short-rotor helicopters is (1.09154, 1.22646) while the confidence interval excluding the outlier is (1.08205, 1.20209). Notice that the interval based on data that excludes the outlier is shifted to the left and is somewhat more narrow. The 95% confidence interval for the mean descent time for long-rotor helicopters is (1.63852, 1.77881) while the confidence interval excluding the outlier is (1.66294, 1.78878). Notice that the interval based on the data that excludes the outlier is shifted to the right and is somewhat narrower. However, in either case the confidence intervals for the mean of short-rotor descent times do not overlap the confidence intervals for the mean of long-rotor descent times. So in either case, the confidence intervals indicate that the mean descent times of long-rotor and short-rotor helicopters are not the same.

30 Long- and 30 Short- Rotor Helicopter Descent Times

Including Outliers

Stacked By _ID_



Oneway Anova

Summary of Fit

t-Test

	Difference	t-Test	DF	Prob>|t|
Estimate	0.549667	11.552	58	<.0001
Std Error	0.047583			
Lower 95%	0.454419			
Upper 95%	0.644915			

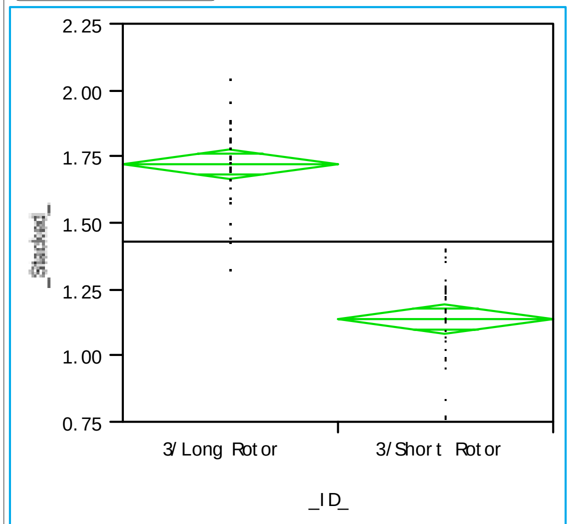
Assuming equal variances

Analysis of Variance

Means for Oneway Anova

Excluding Outliers

Stacked By _ID_



Oneway Anova

Summary of Fit

t-Test

	Difference	t-Test	DF	Prob>|t|
Estimate	0.583793	13.753	56	<.0001
Std Error	0.042449			
Lower 95%	0.498757			
Upper 95%	0.668829			

Assuming equal variances

Analysis of Variance

Means for Oneway Anova

4. The above analysis is from JMP IN. The diamonds on the plots of the values illustrate the confidence intervals that were calculated previously.

   Consider the following hypotheses:

   H_o: $\mu_{short} = \mu_{long}$
   H_a: $\mu_{short} \neq \mu_{long}$

   The assumptions of normality and independence of the short- and long-rotor helicopter descent times distributions were checked earlier. Because the outliers in the short-rotor helicopter and long-rotor helicopter descent times are cause for concern, the analysis can be done with the outliers or without the outliers. In any case, we might proceed with caution in drawing conclusions.

   The value of the test statistic when outliers are included is t = 11.552. Because the variances were approximately equal in the earlier analysis, the t-statistic was calculated assuming equal variances. The probability of getting a t -

value with 58 degrees of freedom at least this extreme is less than 0.0001 if, in fact, the means of the two distributions of descent times are equal.

The value of the test statistic when outliers are excluded is $t = 13.752$. Because the variances were approximately equal in the earlier analysis, the *t*-statistic was calculated assuming equal variances. The probability of getting a *t*-value with 56 degrees of freedom at least this large is less than 0.0001 if, in fact, the means of the two distributions of descent times were equal.

Therefore, we conclude that the mean descent times of long-rotor and short-rotor helicopters are not equal using either method.

Typically when performing a two-sample *t*-test, we elect *not* to pool the variances and calculate a single population variance, since two populations may have equal means (as the null hypothesis usually assumes) yet still have unequal variances. When we do not pool, the degrees of freedom are calculated using a formula that can be found on page 633 in the Yates, Moore and McCabe text, *The Practice of Statistics*. However, JMP IN calculates the value of the test statistic for a two-sample *t*-test using a "modified *t*-test for unequal variances." When the data gathered for Procedure Three was analyzed using a TI-83 and the two-sample *t*-test with unpooled variance, the value for the *t*-statistic, and the number of degrees of freedom were identical to the values reported when JMP IN was used with unequal variances. It is worth noting that, because the sample variances were close in value, the difference between the pooled value and the unpooled values was very small.

Teacher Notes

1. Groups of three students work very well. One student is the timer, one the dropper and one the recorder. A group of two would work, but the third person serves as a quality control (for example, the dropper may not have noticed that some of the helicopters were not in the launcher the same as others) and makes the data collection a bit more efficient. Groups of four or five would probably be too big.

2. Groups may decide to have more than one timer and then average the times or use the median time. This would require larger groups.

3. There are two different ways to construct the short-rotor helicopters. For this experiment we decided to use the short-rotor helicopter that matched the body shape of the long-rotor helicopter (see the diagrams). You may have reasons to use the other design where the actual area of the two helicopters is the same.

4. Remind the students that the calculator calculates *standard deviation* – you must square to get variance.

5. Dropping the helicopters from a height of at least 2 meters is recommended (2.34 meters was used in our experiment). At lower heights the ratio of variance to mean or "noise to signal" is much greater—the timing error may be large relative to the time taken to drop.

6. Using a ladder (custodial ladder) is a good way to stabilize the dropping tool.

7. The helicopter should be clipped to the edge of the clothespin.

8. Before dropping a helicopter, try to assure that its rotors are horizontal.

9. Randomization may be done with a calculator or random number table, but a deck of cards and coin are recommended to emphasize the difference in the randomization procedure for the two designs.

10. All trials should be counted unless there is an obvious problem in the trial—for example, the helicopter is faulty and does not spin or there is a problem with the drop. Trials in which there is an obvious timer error (*e.g.*, timer forgot to turn the timer off when the helicopter hit the ground) should be repeated.

Commentary

1. Population of Inference:

It is appropriate after doing this experiment to discuss with students whether or not the original question ("Do long-rotor helicopters take a different length of time to fall, on average, than short-rotor helicopters?") was answered. In fact it was not answered by Procedures One or Two, since they generalize only to the descent times for two *particular* helicopters. In order to generalize to descent times for *all* long- and short-rotor helicopters the experiment must be based on random samples from the two populations.

What are the populations? One is the descent times for long-rotor helicopters and the other descent times of the short-rotor helicopters. Consider the long-rotor population. The helicopters in Procedure Three were probably made from paper in the same package. Strictly speaking, we can only draw inference to that specific package of paper. However, given the manufacturing controls used in paper production, we probably feel comfortable drawing inference to descent times for all helicopters made with that type of paper. If the paper was colored, but the same weight and brand that we had used in the study, we might still feel comfortable in drawing inference. We might be mistaken, but based on our understanding of the materials, we do not think color should make a difference. However, if a different weight of paper is used or if the paper is linen instead of copier paper, we have definitely moved beyond the scope of inference for our study. In other words, the new population *would* be different and we do not think that the difference between mean times for the long- and short-rotor helicopters estimated in our study would be a valid estimate of the difference for the new population. A new study would be needed if that difference were of interest. Similar considerations should be given to the environment the study is conducted in—for example the droppers, the timers, physical conditions like humidity, etc. While trying to be as careful as possible, there are limits on the practicality of getting a random sample of all possible paper helicopters.

2. Experimental Design: Two-Sample versus Paired

Two-Sample Design (Procedure 1)

In a two-sample design such as that described in Procedure One, the two samples are randomly chosen from two populations that we are interested in comparing on the basis of some response of interest. The way in which the first sample is chosen is independent of the way in which the second sample is chosen. In our example, the samples were actually sample drops from the theoretically infinite population of all possible drops. The helicopters were dropped in a random order so the order that short-rotor helicopters were dropped did not in any way depend on the order of the long-rotor helicopters—it would have been possible to have several short- or long-rotor helicopters dropped in a row. This type of design is called a completely randomized design and is unblocked.

In our study we have two populations (descent times for long- and short-rotor helicopters) and we want to compare the means of these populations. Therefore, we took an SRS from each for inclusion in the study. The two sample design is also used when you want to compare the difference in two treatments for the same population. In this case, an SRS is drawn from the population and the treatments are randomly assigned to the population units. If possible, sample sizes should be approximately equal because the two sample *t*-procedures are more robust against non-normality when this is true. (Yates, Moore, and McCabe, p. 631)

Paired design (Procedure 2)

Each observation from one population (treatment) is matched with an observation from the other population (treatment). In our study, each pair of drops included a descent time for both a short- and long-rotor helicopter. When pairing we would like to pair using a factor that accounts for some of the variability in the response. Paired samples often provide more information than independent samples because extraneous effects are "screened out". In our experiment it turned out that the order may have made a difference (due to perhaps the timer growing weary and sloppy towards the end of the experiment). The evidence of a lack of independence between descent times in a pair suggests that some variability has been removed using a paired design. Often scientists can anticipate whether a factor will make a difference or not. This type of design is called a "blocked design".

There are several ways that data may be paired:

a) Measurements are made on the same group of units or subjects before and after a treatment
b) Naturally occurring pairs are used--such as twins or husbands and wives
c) Pairs are constructed by matching on factors with an effect that might otherwise obscure differences (or the lack of them). In our study, we were interested in comparing the mean difference in descent times for long- and

short-rotor helicopters. Each pair of drops included a short- and long-rotor drop so the *order* in which the drops were made would not be a factor. (Devore & Peck, 3rd Ed., p. 374)

3. Assumptions:

Independent two sample t-test (Procedures 1 and 3)

a) Independent simple random samples from the two populations of interest
b) Both populations are normally distributed
c) Analysis depends on whether or not the population variances are equal

Matched pairs (or paired-differences) t-test (Procedure 2)

a) Simple random sample of differences based on a valid randomization process
b) Differences are normally distributed—if each population is normally distributed this will imply that the differences are normally distributed but it is also possible that the differences will be normally distributed even when the population distributions are not. Thus the assumption that the differences are normally distributed is less restrictive than the assumption that both populations are normally distributed (Wackerly, Mendenhall & Scheaffer, p. 558).
c) Does not require the assumption of equal population variance (Wackerly, Mendenhall & Scheaffer, p. 558)

The two sample *t*-test and the *t*-test for a single mean (as in the paired design) are robust relative to the assumption of normality and relative to equal variance if the sample sizes are the same. (P. 443, Wackerly, Mendenhall, Schaeffer)

4. Sample Size:

Detecting small mean differences requires a larger sample than does detecting large mean differences. Because we expected the descent times of the long- and short-rotor helicopters to be noticeably different, we used a small sample size for convenience.

According to Yates, Moore & McCabe (p.606,631), a sample size of 40 differences (paired data) or 40 total drops (two sample data) is large enough so that, barring extreme outliers or extreme skewness, one need not be concerned about normality.

5. Variance:

Examine the variability in differences versus the variation of long and shorts. In our data, the sum of the variances of the two samples is greater than the variance of the differences. This indicates that a paired design may be more appropriate than a completely randomized design because pairing may reduce the variability (Wackerly, Mendall & Schaeffer, p. 558).

In this case, it was only after conducting the experiments that we could know that the variability was reduced by pairing. An excellent question is how to determine which design is better in advance of conducting an experiment. One option is to conduct a pilot study with a small sample. Another option is to consider by what factor the trials may be paired (in this experiment, the order of the drops), and to make an educated guess regarding whether that factor would contribute a significant amount of variability to the data. If so, then it may be “filtered” out by pairing. If not, pairing would needlessly reduce degrees of freedom without decreasing variability, leading to reduced power. Scientists’ experience often tells them what should be blocked on for reduction of response variability.

In either case, once the experimental design is chosen and the data are gathered, the analysis *must* match the design.

6. AP Statistics Topics Covered:

Topic IA: Interpreting graphical displays of univariate data
1. Center and spread
2. Outliers and other unusual features
3. Shape

Topic IB: Summarizing distributions of univariate data
1. Measuring center
2. Measuring spread
3. Using Boxplots

Topic IIC: Planning and conducting experiments
1. Characteristics of a well-designed and well-conducted experiment
2. Completely randomized design
3. Matched pairs design

Topic IVC: Special case of normally distributed data
1. Two sample (independent and matched pairs) t procedures

Suggestions for Other Studies

Paper helicopter activities are both interesting to the students and logistically possible to conduct in a classroom environment. The following activities can also be used to study both two sample and paired designs in a classroom environment. Teachers can develop activities and questions similar to the ones in the helicopter activity. The following activities are suitable as individual as well as group activities.

1. Compare the speed of dominant and non-dominant hands. Students are to work as quickly and accurately as possible with both their dominant and non-dominant hands to see if they can detect a difference in the mean rates. Students should fill each circle (see attached data collection sheets pages 30 and 31), in order, with an “x” (or some other symbol) and try to fill in as many circles as possible in the allotted time. The timer will tell them when to start and when to stop. They should find the number of circles per second they are able to complete. You can use all the given times or any subset of them. Several two sample and paired designs can be used with this activity.

   Some questions that can be looked at:
   a) If you take a sample of rates for your dominant hand and a sample of rates for your non-dominant hand, is the mean rate for *your* dominant hand higher than the mean rate for *your* non-dominant hand?
   b) Is the mean rate for dominant hands generally higher than the mean rate for non-dominant hands?
   c) Is the mean rate for the dominant hand higher for males than for females?
   d) Do females tend to have a higher mean difference in rates than males?
   e) Are there differences between the rates for left-handed and right-handed subjects?

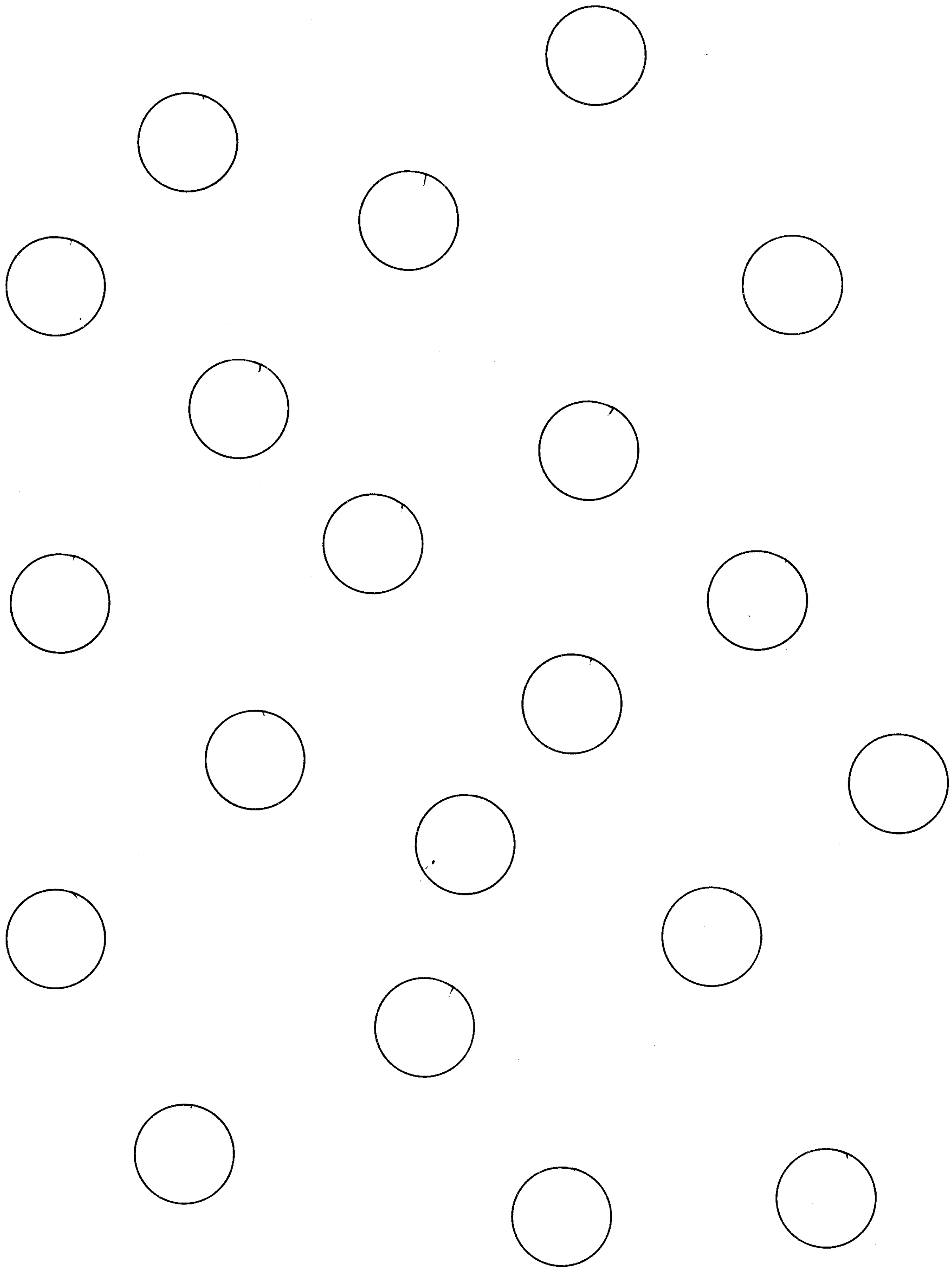
2. Compare pulse rates both before and after exercise. Many different types of exercise can be done depending on your students. This will be a paired design since measurements are made on the same group of subjects before and after the treatment (exercise). Even two minutes of intense walking around will generally raise the pulse rates of students in a class significantly.

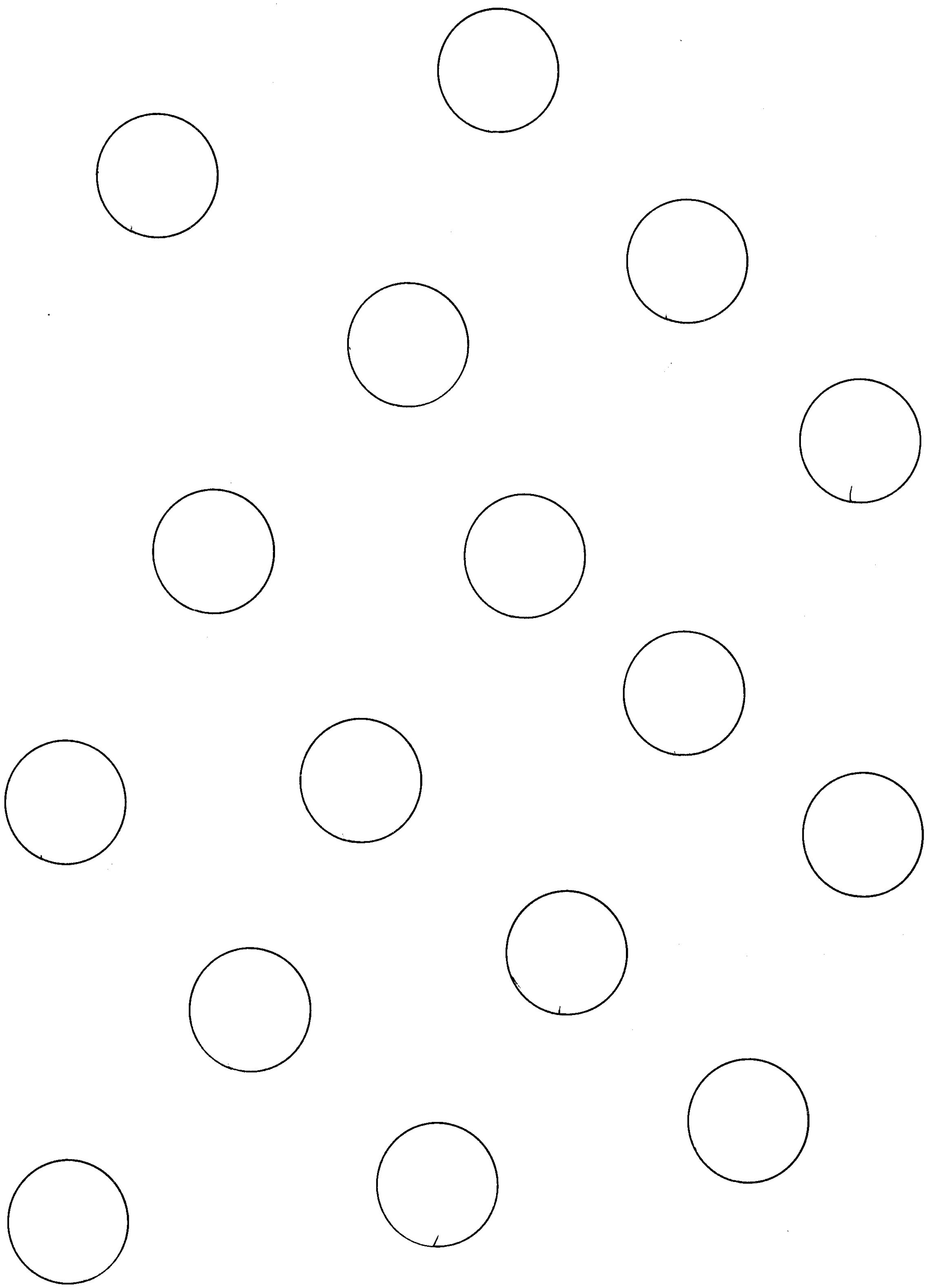
3. Compare the number of beans that can be picked up with the right vs. left hand. Find the biggest, cheapest beans available. Students will grasp and hold (for about 7 sec) a handful of beans—with right hand and left hand. Then count the number of beans held.

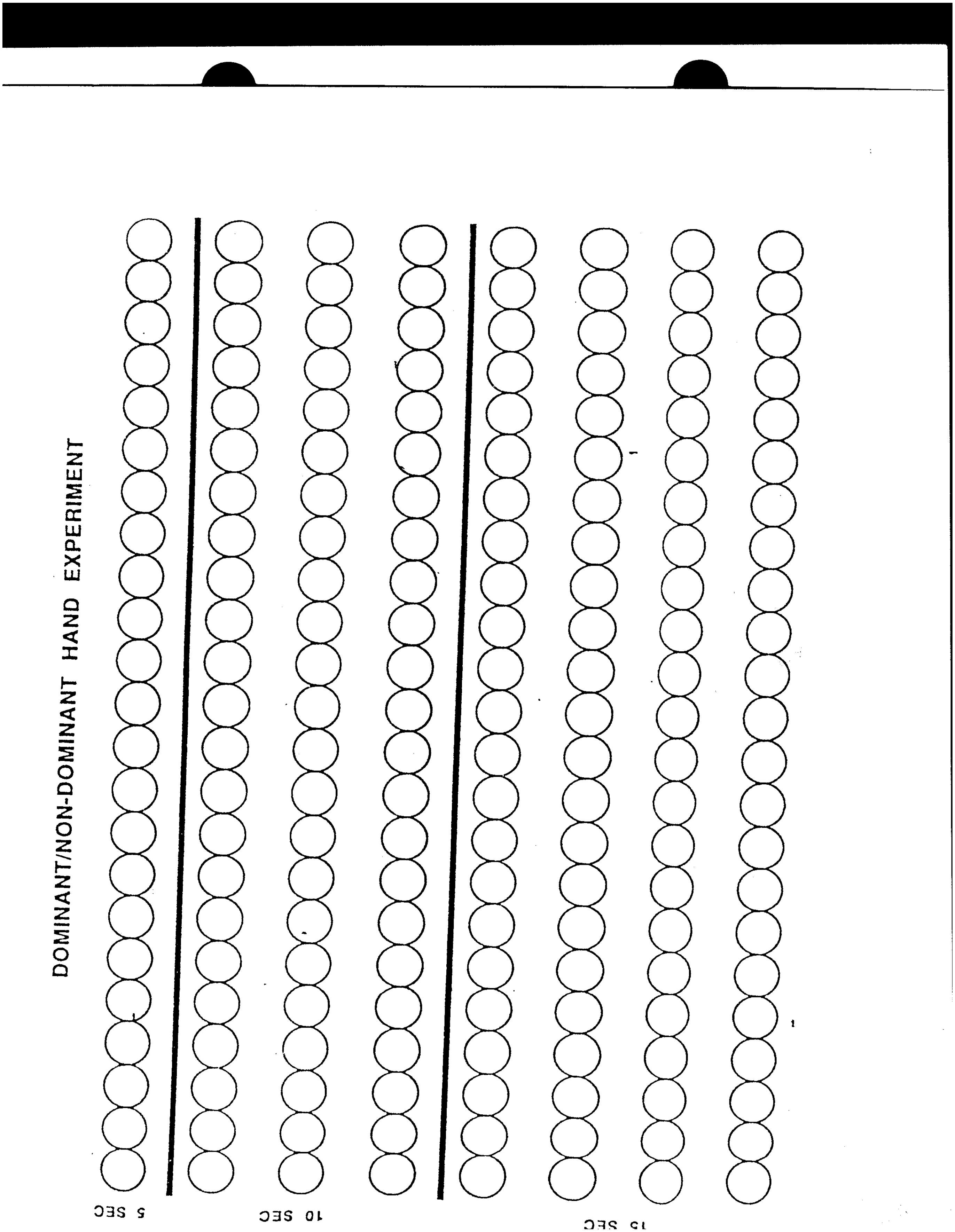
4. Compare the dexterity of the dominant versus the non-dominant hand. This activity requires a supply of reasonably sized beans and a target sheet (pages 29 and 30). Students must place beans in the circles until the whole sheet is filled. The time taken to complete this task is recorded. Students should practice and discuss the exact procedure to be used before beginning the experiment.

5. Compare mean jumping distance of origami frogs (Activity Based Statistics, Schaeffer)

Activity Templates



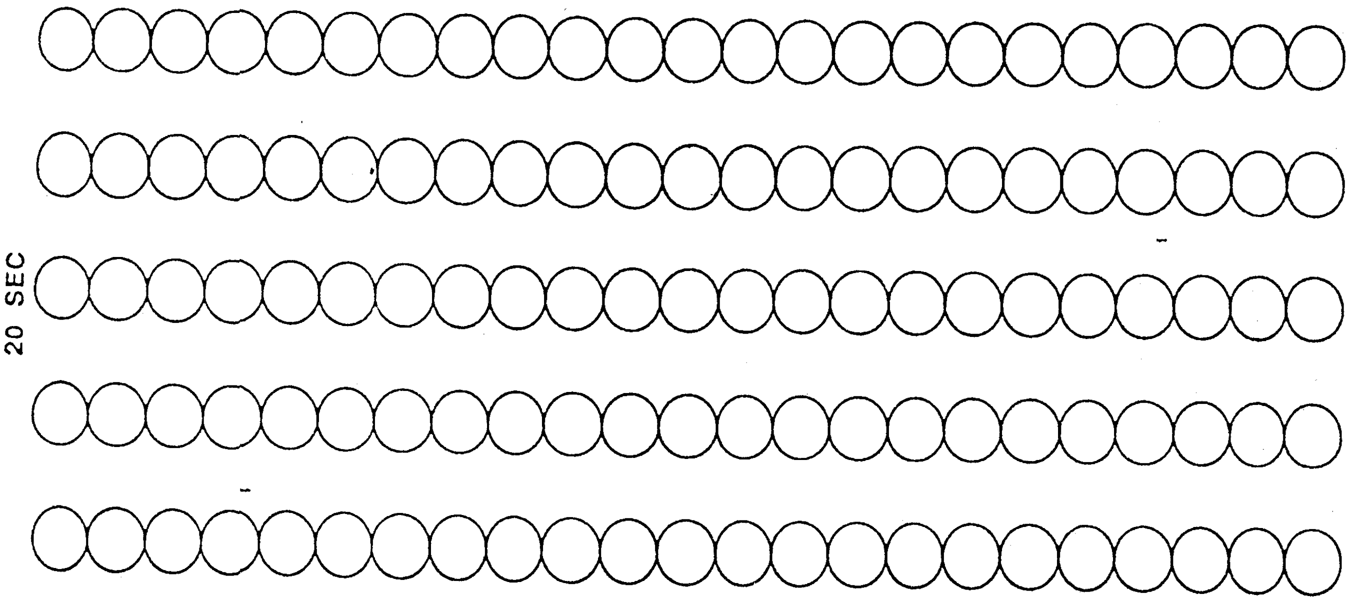



DOMINANT/NON-DOMINANT HAND EXPERIMENT
5 SEC
10 SEC
15 SEC

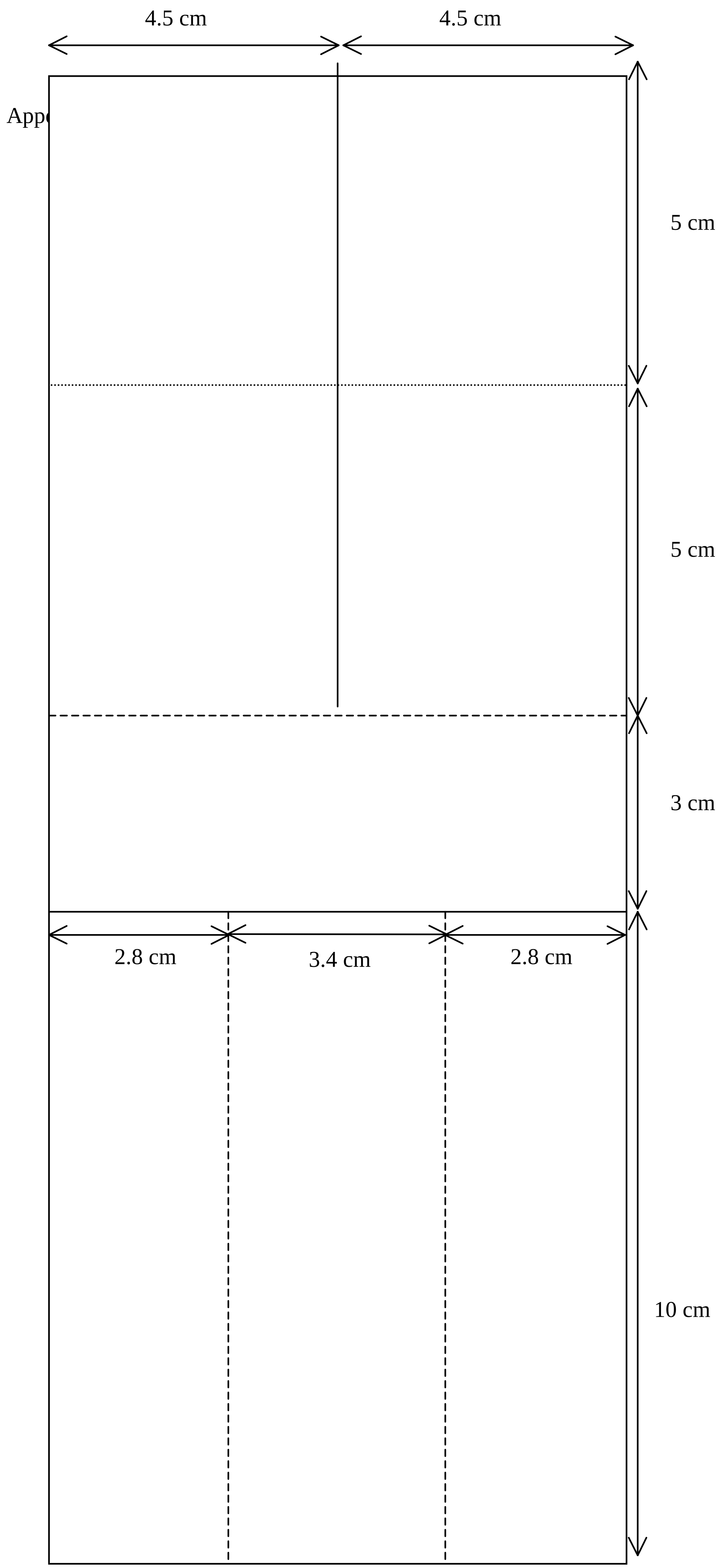
-DOMINANT/NON-DOMINANT HAND EXPERIMENT

PAGE 2

20 SEC



PAPER HELICOTPER DESIGN
Original Design by George Box



The following instructions use the George Box helicopter template on the previous page

1. Cut out the rectangular shape of the helicopter on the solid lines.

2. Cut one-third of the way in from each side of the helicopter to the vertical dashed lines on the solid line.

3. Fold both sides toward the center creating the base. The base can be stapled at the top and bottom. Try to be consistent about where the staples are placed. Use a paper clip to add some weight to the body

4. For long-rotor helicopters, cut down from the top along the solid center line to the horizontal dashed line.

5. For short-rotor helicopters, proceed as in step 4, but cut the rotors off along the horizontal line marked.

6. Fold the rotors in opposite directions.