Comparing the lifetimes of two brands of batteries

Peter K. Dunn
University of the Sunshine Coast

Journal of Statistics Education Volume 21, Number 1 (2013),
www.amstat.org/publications/jse/v21n1/dunn.pdf

Key Words: Real data; *t*-tests; Hypothesis tests; Experimental design; Practical significance.

Abstract

In this paper, we report a case study that illustrates the importance in interpreting the results from statistical tests, and shows the difference between practical importance and statistical significance. This case study presents three sets of data concerning the performance of two brands of batteries. The data are easy to describe and understand, familiar to students, and allow a range of analyses, from simple to more complex. The data were the basis of a claim made in an advertisement, and this claim is re-assessed using the data to show that the company undersold the performance of their batteries from a statistical point-of-view in one of the three tests. However, a challenge is how such a conclusion can be communicated to the public succinctly but correctly.

1. Introduction

Most readers will be aware of the value and importance of using real data in classes, which has been reinforced by the American Statistical Association's Guidelines for Assessment and Instruction in Statistics Education (GAISE; http://www.amstat.org/education/gaise) in their *College Report* (Aliaga, Cobb, Cuff, Garfield, Gould, Lock, Moore, Rossman, Stephenson, Utts, Velleman, and Witmer 2010). Many books collate and make available sets of real data (for example, Hand, Daly, Lunn, McConway, and Ostrowski 1996; Chatterjee, Handcock, and Simonoff 1995; Peck, Haugh, and Goodman 1998; Peck, et al. 2006), and websites exist to make such data easy to use (such as OzDASL, Smyth 2011; the *JSE* Data Archive; DASL 1996).

In this article, three related datasets are presented that emerge from a situation that is very simple to state, easy to understand, familiar to students, yet allow a variety of statistical issues to be discussed. The focus of this paper is on the second of the three tests. The data and their background are presented in Section 2, and the claims and the data for Test 2 presented in Section 3. In Section 4, the claims of Test 2 are re-examined, and other uses of the data are briefly considered in Section 5. Tests 1 and 3 are then introduced in Section 6.

2. The data

I was alerted to the data used in this paper through an advertising catalogue that appeared in our letterbox for ALDI supermarkets (http://www.aldi.com.au). The advertisement claimed "that Ultracell AA Alkaline batteries [the ALDI brand batteries] outperformed the Energizer Max AA Alkaline batteries in 2 of 3 battery life tests". Their webpage directed the reader to the independent report (which includes the data) upon which the claim is made (Lindström 2011). The data have been extract from linked report:
http://www.amstat.org/publications/jse/v21n1/dunn/RE11_021_ALDI_Stores_Primary_battery_testing.pdf

The Ultracell batteries are substantially cheaper than the Energizer batteries. A four-pack of the Ultracell Max cost AUD$2.49 from ALDI online, and a four pack of Energizer Max from Woolworths online cost AUD$5.97 on special, usually AUD$8.01 (data from 05 September 2012).

As the advertising indicates, three types of tests were conducted comparing the lifetime of two brands of batteries (where lifetime is measured in time, or the number of pulses), using fresh batteries starting with a notional 1.5 volts. Both brands of batteries are alkaline batteries, designated as LR6 batteries. Details of the tests appear in the full report (Lindström 2011). In the first test, batteries were loaded with a camera flash, using 1000 m A loaded for 10 s/min for one hour per day. The number of "pulses" to reach pre-defined voltage levels was recorded. In the second test, the batteries were loaded using a 250 m A electronic game for one hour per day. The time taken to reach pre-defined voltage levels was recorded. In the third test, batteries were loaded with a digital camera drawing 1500 m W / 650 m W, for 2s every 28 s, for 5 min/h for 24 hours/day. The number of "pulses" to reach pre-defined voltage levels was recorded. Tests were started in March 2011, and nine batteries of each brand were used for each test.

A separate data file exists for each test. Each data file contains 108 observations (nine batteries for two brands for six pre-defined voltage levels), and are described in Appendix 1. No missing values are present. The second dataset is the focus of this paper, but brief comments are made on the other two datasets in Section 6.

> *Helpful Hint: Before showing the data to the students, ask them how long they think an AA battery would last, on average, from their own experience, and how much variation they might expect in battery life from one battery to another.*

3. The claims

The claim made in the ALDI advertising (that "Ultracell AA Alkaline batteries outperformed the Energizer Max AA Alkaline batteries in 2 of 3 battery life tests") is based on the summary results displayed in Table 1 (based on Table 6.1 of Lindström (2011)).

Table 1: The summary of the results from the three tests on the two brands of batteries. Figures in **bold** indicate the superior results.

			Ultracell		Energizer	
Test	End voltage	Units	Mean	Range	Mean	Range
1	0.9 V	pulses	**574**	**540** to **584**	403	276 to 467
2	0.9 V	hours	8.24	**7.93** to 8.35	**8.28**	7.88 to **8.49**
3	1.05 V	pulses	**99**	**89** to **117**	70	64 to 77

Clearly, based on means alone, the Ultracell batteries are superior than Energizer batteries in two tests, as claimed. However, any robust indication of variability is clearly missing from the ALDI advertisement, and any indication of whether the differences shown in the table are *statistically* significant is also missing. In any case, the Test 2 results, regardless of statistical significance, show modest *practical* difference: In over 8 hours of use, the mean difference in time to reach 0.9 V is 0.04 hours, or about 2.5 minutes.

> *Helpful Hint: As an introduction to the data, present the summary table to students (or have them calculate these summaries themselves), and ask students to write their own one-sentence summary of the summary table that could be used in advertisements (before showing the students the actual ALDI claim). Then, the ALDI claim can be presented, and students asked if the ALDI claims appear to be supported by the results in this table.*

> *Helpful Hint: Once the table and claims have been presented, talk about the issues of practical significance and statistical significance. Subsequently, talk about whether the differences observed may be the result of chance or actual difference in the brands.*

The main focus in this paper is the data from Test 2 (Table 2; Figures 1, 2 and 3), which records the time taken for AA batteries to discharge to specified end voltages. This is the only test where ALDI does not claim that their batteries are superior to the Energizer batteries.

For this test, the end-point of testing is 0.9 volts (though data are available up to 0.8 volts), as recommended by the International Electrotechnical Commission standard IEC 60086-2, Ed. 11.0, 2006-12.

Table 2. The data from Test 2: The times taken (in decimal hours) for nine batteries of two brands (E: Energizer; U: Ultracell) to reach specified voltage end-points. The figures in bold are the superior lifetime summary averages.

	1.3 volts		**1.2 volts**		**1.1 volts**		**1.0 volts**		**0.9 volts**		**0.8 volts**	
	E	**U**	**E**	**U**	**E**	**U**	**E**	**U**	**E**	**U**	**E**	**U**
	1.40	1.56	2.86	3.57	5.71	5.76	7.58	7.50	8.45	8.35	8.86	8.76
	1.39	1.54	2.77	3.55	5.64	5.73	7.46	7.48	8.34	8.35	8.65	8.81
	1.35	1.53	2.71	3.55	5.63	5.74	7.46	7.47	8.35	8.32	8.74	8.81
	1.38	1.54	2.81	3.54	5.78	5.71	7.59	7.48	8.49	8.32	8.91	8.70
	1.35	1.54	2.65	3.54	5.63	5.72	7.46	7.48	8.33	8.31	8.72	8.73
	1.36	1.47	2.73	3.51	5.70	5.72	7.52	7.41	8.41	8.28	8.85	8.76
	1.31	1.49	2.48	3.54	4.65	5.71	6.83	7.47	7.88	7.99	8.52	8.68
	1.26	1.54	2.44	3.54	4.67	5.68	6.89	6.96	7.94	7.93	8.62	8.64
	1.37	1.50	2.76	3.56	5.57	5.74	7.45	7.48	8.32	8.34	8.68	8.79
Mean	1.35	**1.52**	2.69	**3.54**	5.44	**5.73**	7.36	**7.41**	**8.28**	8.24	8.73	**8.74**
Median	1.36	**1.54**	2.73	**3.54**	5.63	**5.72**	7.46	**7.48**	**8.34**	8.32	8.72	**8.76**
Std dev	0.04	0.03	0.14	0.02	0.45	0.02	0.29	0.17	0.22	0.16	0.13	0.06

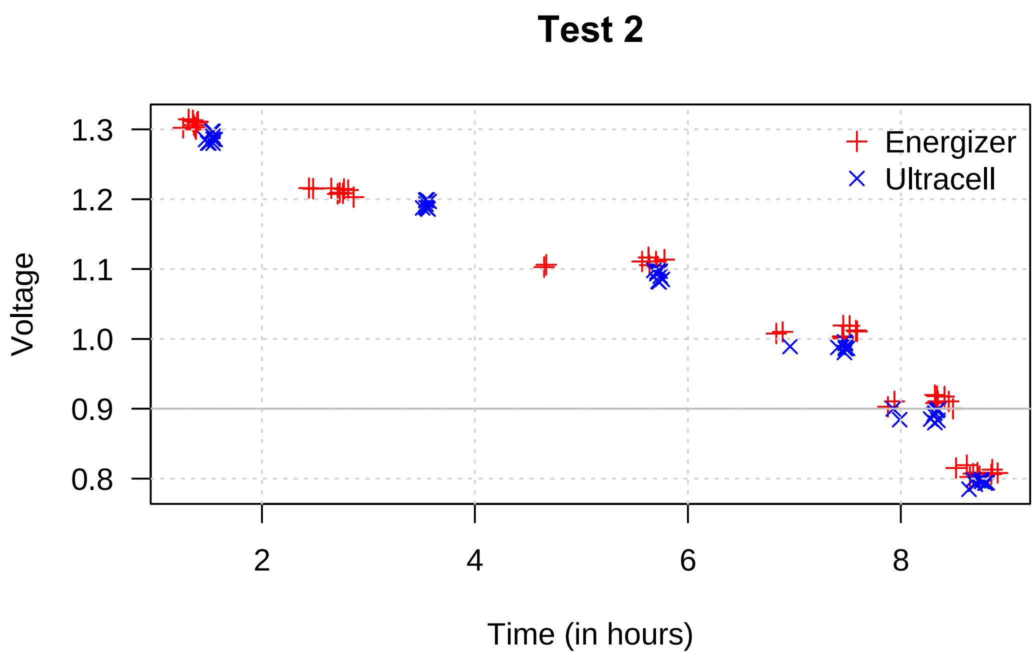

Figure 1: A plot of the data from Test 2. The data from the two brands have been shifted slightly and jittered in the vertical direction to avoid overplotting. The solid, gray horizontal line corresponds to the standard end-point for the test: 0.9 volts.

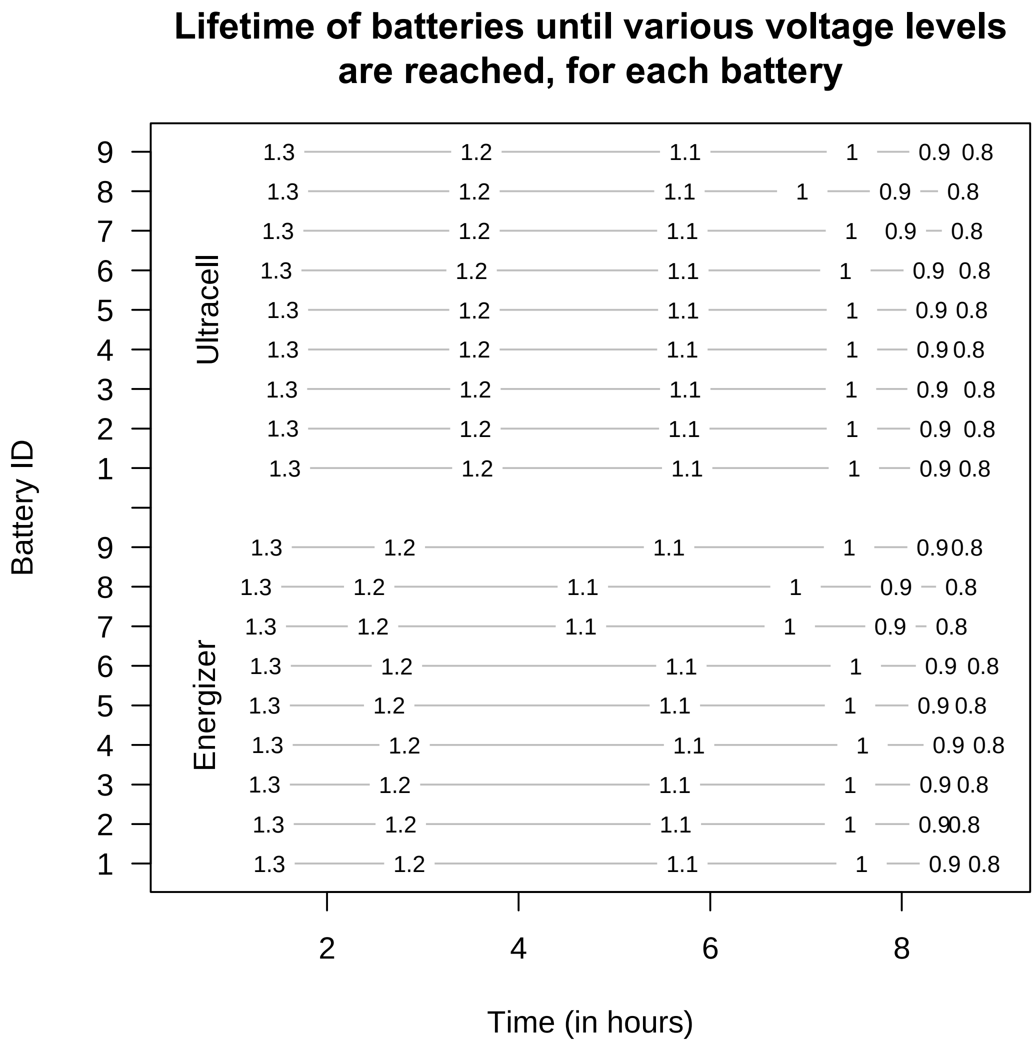

Figure 2: A plot of the data from Test 2. The plot shows the times for each individual battery to reach each voltage end-point.

4. Addressing the Test 2 claims

The purpose behind the data collection was to compare the lifetime of the batteries, to determine which brand lasted longer. However, deciding which brand of battery has a superior lifetime is not as simple as may appear initially. For example, the mean decay times of the Ultracell batteries are greater (perhaps not statistically) at every voltage level, except at 0.9; however, the standard tests use 0.9 volts as the only test end-point at which the decisions are made. Does this fairly represent the comparison between the brands of batteries?

Suppose that we are only concerned with the 0.9 volts end-point (though similar issues arise if a different end-point is chosen). To test the hypothesis of no difference between the mean times to decay to 0.9 volts (H_0: $\mu_U = \mu_E$) against the two-tailed alternative (H_1: $\mu_U \neq \mu_E$), a standard (two-tailed, unequal variances) *t*-test could be used. However, a variety of alternative methods could be used to compare the battery lifetimes; for example, a non-parametric test for comparing the medians (H_0: $median_U = median_E$ against H_1: $median_U \neq median_E$) may be preferred since the sample sizes are small ($n = 9$ for both samples) and the distributions are not symmetric (Figure 3). To do so, a two-way table can be constructed (Table 3) of Brand against whether the observations are below the overall median (which is 8.325 hours) or not, and use Fisher's exact test (H_0: The number of batteries less than the median is the same for both brands, against H_1: The number of batteries less than the median is *not* the same for both brands). However, this test is known to have very low power in small samples (Freidlin and Gastwirth 2000). A Mann-Whitney (also known as a Wilcoxon) test could be considered (H_0: The distributions are the same, against H_1: The distributions are not the same), but may not test the hypothesis of interest unless we also assume that the distributions are the same apart from a shift in location.

Table 3. The data from Test 2 arranged for Fisher's exact test to compare the proportion of batteries from each Brand that exceed the overall median time (8.325 hours) to reach 0.9 volts.

	Above or equal to overall median time	Below overall median time	Total
Energizer	6		
Ultracell	3	6	9
Total	9	9	18

Other alternatives are to use a bootstrap approach to compare the medians or means (we use a test for comparing medians in what follows), or to use permutation tests (though these require distributions with similar variance; see Chapter 5 of Good and Hardin (2006)). In short, many avenues of analysis are available. Regardless of the chosen method, however, the conclusion remains the same (Table 4): No evidence exists that the Ultracell batteries have an average time less than the Energizer batteries to reach the 0.9 volts end-point. (We have used the bias-corrected, accelerated (BCa) confidence intervals; see Efron and Tibshirani (1993).)

Helpful Hint: The discussion above about the type of test to use could be used as a directed discussion in more advanced classes, where the pros and cons of each test could be studied and compared.

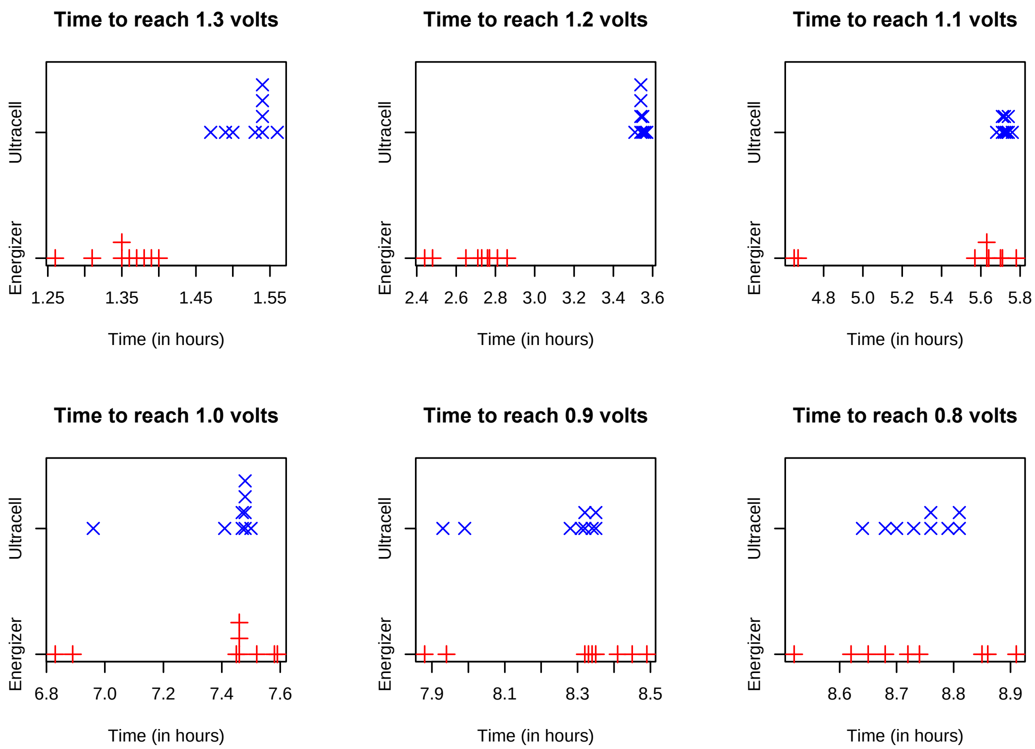

Figure 3: Dot plots of the data from Test 2, showing the non-symmetry of the distributions in many cases.

Table 4. The data from Test 2 for comparing the time for the two brands of batteries to reach the end-point of 0.9 volts. The hypotheses are given in the text. In all cases, a positive difference means that the parameter of interest for the Energizer batteries is larger than that for the Ultracell batteries. The results for the Mann-Whitney test are approximate because of the presence of ties.

Method	*P*-value	Confidence interval
t-test	0.70	–0.16 to 0.23
Mann-Whitney test	0.27	–0.03 to 0.17
Median test	0.35	NA
Bootstrap (BCa; 5 000 samples)	0.89	–0.40 to 0.10
Permutation test for means	0.64	–0.08 to 0.11

Despite this, ALDI only claims that the Ultracell batteries outperform Energizer batteries in two tests; they fail to note that no evidence exists that Energizer batteries outperform Ultracell batteries in any test. They have understated the evidence in the data, and perhaps misunderstood or ignored the role of sampling error.

> *Helpful Hint: Ask students if the ALDI claim is correct, and then if the claim could be worded better for marketing purposes (for example, "There is no evidence that Energizer batteries do not last longer on average in any of the three standard tests").*

Students could then discuss what the implications might be after making this claim if a (non-statistically literate) member of the public saw the results in Table 1 (where the Energizer mean time is greater than the Ultracell mean time). Ask the students how they would explain this to this layperson.

Of course, the times for both brands can be compared statistically for *every* voltage level, not just 0.9 volts. In other words, it may be useful to test whether one brand of battery is superior to the other, in some overall sense and not just at 0.9 volts.

Various tests suggest themselves again, including multiple *t*-tests, bringing the associated problem of multiple testing, and one-way ANOVA. Both of these options ignore the non-independence of the observations, which are (of course) an example of repeated measures (Fitzmaurice, Laird and Ware 2004; Weiss 2005). The output from fitting a repeated measures model in R (noting that the experiment is balanced, and treating the voltage levels as factors) is shown below:

```
> t1 <- aov(Time ~ factor(Voltage)*Brand + Error(Battery),
            data = batteries2)
> summary(t1)

Error: Battery
          Df Sum Sq Mean Sq F value Pr(>F)
Residuals  1  0.598   0.598

Error: Within
                      Df Sum Sq Mean Sq F value   Pr(>F)
factor(Voltage)        5  782.1  156.42 5216.05  < 2e-16 ***
Brand                  1    1.3    1.35   44.91 1.46e-09 ***
factor(Voltage):Brand  5    2.4    0.49   16.31 1.36e-11 ***
Residuals             95    2.8    0.03
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

`Brand` is significant in this model (and depends on the voltage level), suggesting strong evidence that indeed the two brands of battery are different in a general sense, and that the Ultracell batteries are actually superior. ALDI has indeed undersold itself.

Potential Pitfall: Treating the voltage levels as factors is not necessary, of course. However, treating the voltage as quantitative needs care, as the relationship between time and voltage is non-linear, so the naïve approach (which many students will use) is inadequate:

```
> t2 <- aov(Time ~ Voltage*Brand + Error(Battery),
            data = batteries2)
```

This approach, while incorrect, leads to a P-value of 0.080 for the `Brand`, which is quite different from the P-value when treating the voltage as a factor.

An alternative approach to assessing which brand of battery is superior in a general sense is to consider how the voltage may decay over time, and compare these decay parameters for both brands. A naïve but not unreasonable model for the voltage decay is the exponential decay model: $V(t) = V_0 \exp(t/k)$, where k (< 0) is the decay constant for the battery brand, $V(t)$ is the voltage at time t, and V_0 is the voltage at time zero (notionally 1.5 volts). Under this model, the relationship between log $V(t)$ and t will be linear; however, a plot of the data shows that this is certainly not the case, at least over the whole range of the data (Figure 4). This is consistent with the specification sheet provided about the Energizer batteries from the manufacturer (see the figures at the end of the specification sheets available at: http://data.energizer.com/PDFs/E91.pdf, accessed 06 September 2012).

> *Helpful Hint: The instructor may wish to make these specification sheets available for students so that they can compare their graphs with those from the manufacturer, and comment. Of interest is that the specification sheet figure equivalent to Test 2 uses the same test parameters as those that gave rise to the data for Test 2 (250 mA for 1 hour per day), so the graphs are comparable. (Note: The specification results were conducted at 21°C, and the ALDI test results at 20 ± 2°C.)*

> *Helpful Hint: An interesting exercise is to provide a copy of the equivalent figure to Figure 1 that appears in the specification sheet, and have students plot the data on that provided curve to see how closely the data follow the manufacturer's test curve.*

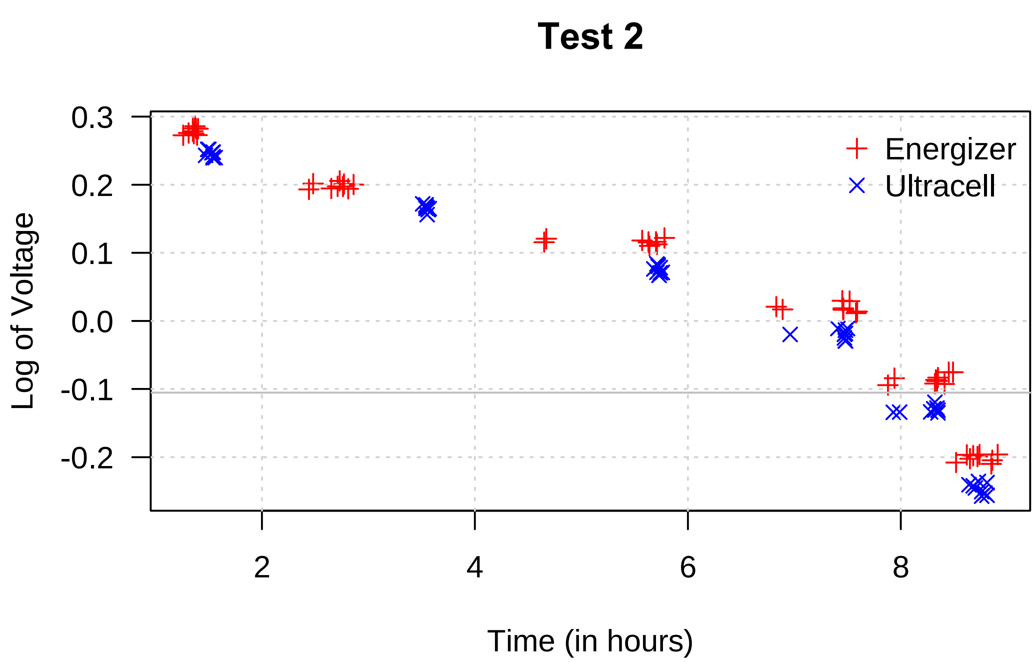

Figure 4: The natural logarithm of voltage plotted against time for the data from Test 2. The data from the two brands have been shifted slightly and jittered in the vertical direction to avoid overplotting. The solid, gray horizontal line corresponds to the standard end-point for the test: 0.9 volts

5. Other classroom uses and observations

We have discussed ways of using the Test 2 data to answer the question of interest that gave rise to the data. However, the data have other potential classroom uses also. For example, the data may be used for simple tasks such as one-sample summary statistics (by examining the data from just one voltage end point for example) and graphs (for example, boxplots comparing the two brands at a specified voltage end-point). Two-sample tests could be conducted, as already explained. Statistical models could be developed for modelling the relationship between the voltage at specified times beyond the simple exponential decay model presented earlier. Splines could also be fitted to the data. The data are also an example of survival data and could be analysed as such.

The data can also be used to demonstrate the identification of outliers (for example, see the Energizer times at 1.1 volts, or the Ultracell times at 1.0 volts), and to discuss the variation in the data (for example, the standard deviation is always greater for Energizer batteries than for Ultracell, and the IQR always greater apart from 1.3 volts). Formal hypothesis tests comparing variances could be considered.

> *Helpful Hint: Students could be asked to compare the measures of variation (standard deviation and IQR): What does it mean in this context that the Energizer times are almost always more variable than the Ultracell times, and is this important? What extra information does this provide beyond means?*

6. The other two datasets

We close by making some brief observations on the other two datasets, firstly the data from Test 1 (Figure 5). Clear differences are evident between the two brands of batteries (Table 5), in both mean/median number of pulses and the variation in the number of pulses, and noticeably so at the standard end-point of 0.9 volts.

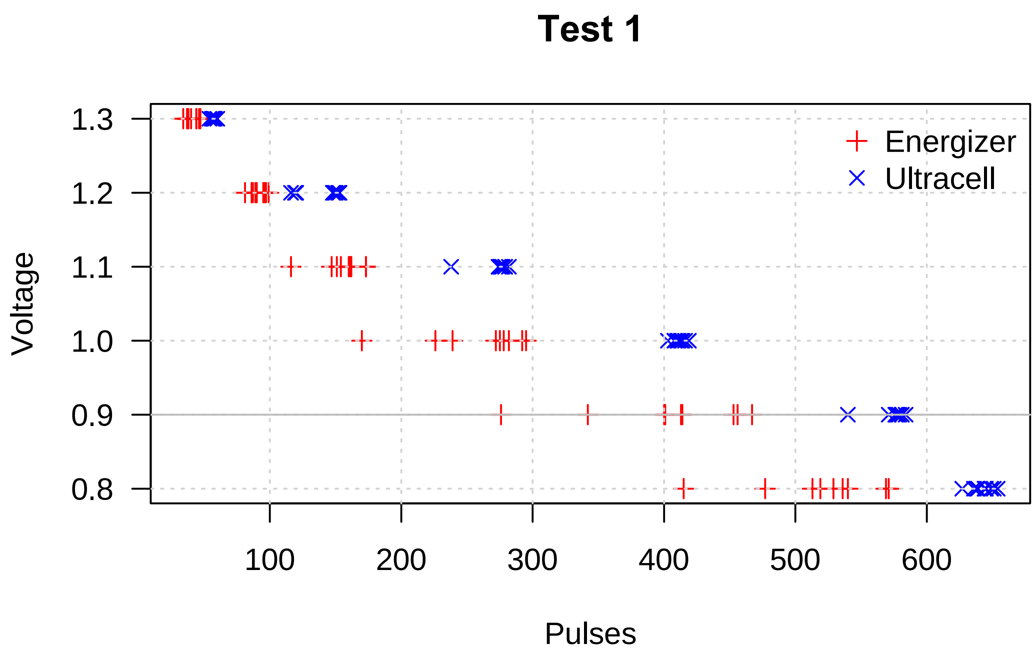

Figure 5: A plot of the data from Test 1. The solid, gray horizontal line corresponds to the standard end-point for the test: 0.9 volts.

Table 5. The data from Test 1 for comparing the number of pulses for the two brands of batteries to reach the end-point of 0.9 volts. The hypotheses are given in the text. In all cases, a positive difference means that the parameter of interest for the Energizer batteries is larger than that for the Ultracell batteries. The results for the Mann-Whitney test are approximate because of the presence of ties.

Method	*P*-value	Confidence interval for difference (in pulses)
t-test	< 0.001	–218 to –124
Mann-Whitney test	<0.001	–198 to –123
Median test	<0.001	NA
Bootstrap (BCa; 5 000 samples)	<0.001	–206 to –122
Permutation test	<0.001	–108 to –62

These comments apply equally to the Test 3 results (where the standard end-point is 1.05 volts; Figure 6). In Test 3, the decision is quite easy to make (Table 6): The data strongly suggest that the Ultracell batteries last longer than the more expensive Energizer batteries. Furthermore, the Ultracell batteries appear less variable (under the test conditions).

Given that the Ultracell batteries are at least as good as the Energizer batteries in all three tests—and are sometimes substantially superior—and that they are cheaper to purchase, the decision of which batteries to purchase (all other things being equal) seems clear. A further extension to the analysis, then, is to ask the question: At what per-unit cost would the Energizer batteries be more economical?

Helpful Hint: As a final question for students, consider asking them to write a one-sentence summary to use in advertising, on the basis of the results of the analysis. The students could even be asked to design an advertisement to sell the ALDI batteries on the basis of these results.

Table 6. The data from Test 3 for comparing the number of pulses for the two brands of batteries to reach the end-point of 1.05 volts. The hypotheses are given in the text. In all cases, a positive difference means that the parameter of interest for the Energizer batteries is larger than that for the Ultracell batteries. The results for the Mann-Whitney test are approximate because of the presence of ties.

Method	*P*-value	Confidence interval for difference (in pulses)
t-test	< 0.001	–35 to –22
Mann-Whitney test	<0.001	–33 to –21
Median test	<0.001	NA
Bootstrap (BCa; 5 000 samples)	<0.001	–32 to –21
Permutation test	<0.001	–18 to –11

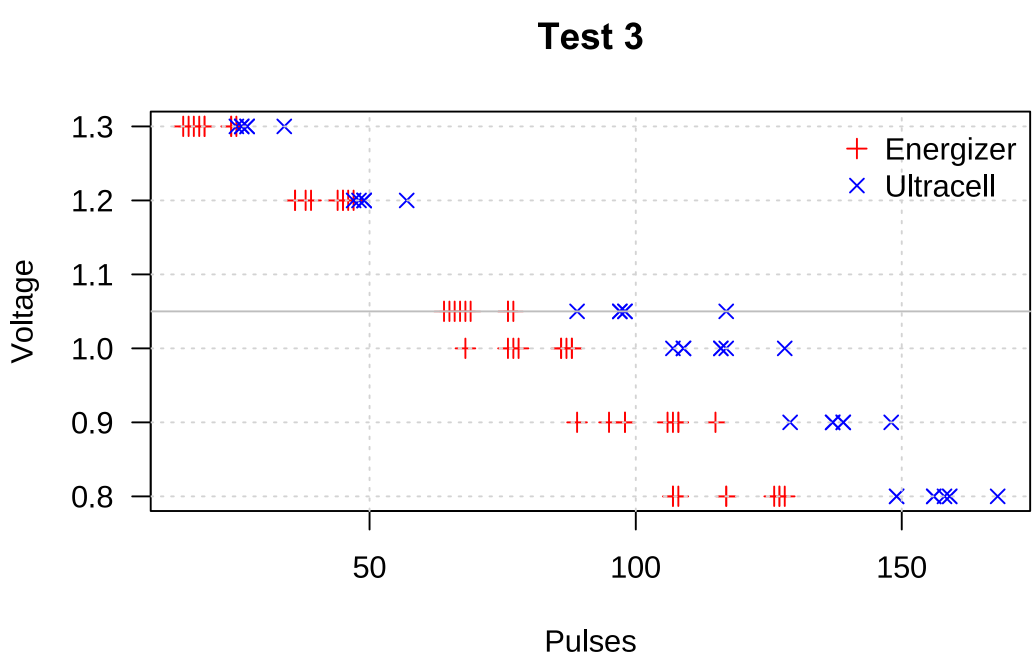

Figure 6: A plot of the data from Test 3. The solid, gray horizontal line corresponds to the standard end-point for the test: 1.05 volts.

7. Conclusions

In this paper, we have presented three sets of simple, real data. The data can be analysed to determine if the advertised claims are supported by the data. The claims are clearly supported in

Tests 1 and 3; for Test 2, the evidence suggests that ALDI have undersold the performance of their batteries. Other uses of the data are also suggested.

Appendix A: Data coding

The following table explains the variables appearing in the three datasets `batteries1`, `batteries2` and `batteries3`.

Data file variable	Description	Details
`Brand`	The brand of the battery	Either `Energizer` or `Ultracell`
`Voltage`	The voltage levels of interest	The values are 1.3, 1.2, 1.1 (for Tests 1 and 2), 1.05 (for Test 3), 1.0, 0.9 and 0.8 volts
`Time` (in `batteries2` only)	The time taken to reach the pre-defined voltage levels	The time is given in decimal hours
`Pulses` (in `batteries1` and `batteries3`)	The number of pulses taken to reach the pre-defined voltage levels	The pulses are discrete counts
`Battery`	An identifier	The integers `1` to `9`

Appendix B: R code for analysis

To load the data (as downloaded from the JSE Data Archive), use these commands:

```
batteries1 <- read.csv("batteries1.csv", header=TRUE)
batteries2 <- read.csv("batteries2.csv", header=TRUE)
batteries3 <- read.csv("batteries3.csv", header=TRUE)
```

The summaries in Table 2 are found using these commands:

```
with(batteries2, tapply(Time, list(Brand, Voltage), mean))
with(batteries2, tapply(Time, list(Brand, Voltage), median))
with(batteries2, tapply(Time, list(Brand, Voltage), sd))
```

Figure 1 is produced using these commands:

```
offset <- c(0.01, -0.01)
plot( Voltage+jitter(ifelse(Brand=="Energizer",offset[1], offset[2]),
amount=0.01)~Time,
      main="Test 2",
      pch=ifelse(Brand=="Energizer",3,4),
      col=ifelse(Brand=="Energizer","red","blue"),
      xlab="Time (in hours)", ylab="Voltage",
```

```
      las=1,
      data=batteries2
)
grid()
abline(h=0.9, col="gray")
legend("topright",pch=c(3,4), col=c("red","blue"),
       legend=c("Energizer","Ultracell"), bty="n")
```

Figure 2 is produced using these commands:

```
plot(c(0.5,9), c(1,19),
   type="n",
   xlab="Time (in hours)",
   ylab="Battery ID",
   axes=FALSE,
   main="Lifetime of batteries until various voltage
         levels\nare reached, for each battery"
)
axis(side=1)
axis(side=2,
     at=c( 1:19),
     labels=c(as.character(1:9), "", as.character(1:9)),
     las=1
)
box()

text(0.75, 5, "Energizer", srt=90)
text(0.75, 15, "Ultracell", srt=90)
text.labels <- rev( levels(as.factor(batteries2$Voltage)) )

attach(batteries2)
for (i in (1:9)){
   lines( Time[Brand=="Energizer" & Battery==i], rep(i,6),
          col="gray" ,  lty=1 )
   points(Time[Brand=="Energizer" & Battery==i], rep(i,6),
          col="white", cex=3.5, pch=19 )
   text(Time[Brand=="Energizer" & Battery==i], rep(i,6),
        labels=text.labels, cex=0.75 )
}
for (i in (11:19)){
   lines( Time[Brand=="Ultracell" & Battery==(i-10)], rep(i,6),
          col="gray",  lty=1 )
   points(Time[Brand=="Ultracell" & Battery==(i-10)], rep(i,6),
          col="white", cex=3.5, pch=19  )
   text(Time[Brand=="Ultracell" & Battery==(i-10)], rep(i,6),
        labels=text.labels, cex=0.75 )
}
detach(batteries2)
```

Table 3 is produced using these commands:

```
attach(batteries2)
```

```
AllTimes <- c( Time[Voltage==0.9 & Brand=="Energizer"],
               Time[Voltage==0.9 & Brand=="Ultracell"] )
Groups <- c( rep("Energizer", 9), rep("Ultracell", 9) )
xtabs(~ Groups + (AllTimes <  median(AllTimes)) )
detach(batteries2)
```

Figure 3 is produced using these commands:

```
par(mfrow=c(2,3))          # Establish 2x3 grid for plots
stripchart(Time~Brand, data=subset(batteries2, Voltage==1.3),
      method="stack", pch=c(3,4), col=c("red","blue"), cex=1.5,
      xlab="Time (in hours)", main="Time to reach 1.3 volts" )
stripchart(Time~Brand, data=subset(batteries2, Voltage==1.2),
      method="stack", col=c("red","blue"), cex=1.5, pch=c(3,4),
      xlab="Time (in hours)", main="Time to reach 1.2 volts" )
stripchart(Time~Brand, data=subset(batteries2, Voltage==1.1),
      method="stack", pch=c(3,4), col=c("red","blue"), cex=1.5,
      xlab="Time (in hours)", main="Time to reach 1.1 volts" )
stripchart(Time~Brand, data=subset(batteries2, Voltage==1.0),
      method="stack", pch=c(3,4), col=c("red","blue"), cex=1.5,
      xlab="Time (in hours)", main="Time to reach 1.0 volts" )
stripchart(Time~Brand, data=subset(batteries2, Voltage==0.9),
      method="stack", pch=c(3,4), col=c("red","blue"), cex=1.5,
      xlab="Time (in hours)", main="Time to reach 0.9 volts" )
stripchart(Time~Brand, data=subset(batteries2, Voltage==0.8),
      method="stack", pch=c(3,4), col=c("red","blue"), cex=1.5,
      xlab="Time (in hours)", main="Time to reach 0.8 volts" )
par(mfrow=c(1,1))
```

Table 4 is produced using these commands; Table 5 and 6 are produced similarly and the code not shown:

```
# T for a difference at 0.9 V
p.t <- t.test(Time ~ Brand,
     data=subset(batteries2, Voltage==0.9))

# Wilcoxon test for a difference at 0.9 V
p.wilcox <- wilcox.test(Time ~ Brand,
     data=subset(batteries2, Voltage==0.9), conf.int=TRUE)

# Median test for a difference at 0.9 V
median.test <- function(x,y){
   z <- c(x,y)
   g <- rep(1:2, c(length(x),length(y)))
   m <- median(z)
   fisher.test(z<m,g)$p.value
}
p.median <- with( batteries2,
   median.test( Time[Voltage==0.9 & Brand=="Ultracell"],
   Time[Voltage==0.9 & Brand=="Energizer"])
)
```

```
# Bootstrap test for a difference at 0.9 V
med.diff <- function(dataset, ind){
   y     <- dataset$Time[ind]
   group <- dataset$Brand[ind]

   med.E <- median(y[group=="Energizer"]) # E
   med.U <- median(y[group=="Ultracell"]) # U
   return(med.E - med.U)
}

library(boot)  ### LOAD THE  boot  LIBRARY
out <- boot(batteries2, med.diff, R=5000)

ci.boot <- boot.ci(out)
p.boot <- sum(out$t >= 0)/out$R


# Permutation test
library(lmPerm) ### LOAD THE  lmPerm  LIBRARY
out.perm <- aovp(Time~Brand,
                 data=subset(batteries2, Voltage==0.9))

# Display results
p.t
p.median
p.wilcox
p.boot
ci.boot
summary(out.perm)
confint(out.perm)
```

Figure 4 is produced using these commands:

```
offset <- c(0.01, -0.01)
logV <- log(batteries2$Voltage +
   jitter(ifelse(Brand=="Energizer", offset[1], offset[2]),
          amount=0.01)) +
   ifelse(Brand=="Energizer", offset[1], offset[2])
plot(logV~batteries2$Time,
      main="Test 2",
      pch=ifelse(Brand=="Energizer",3,4),
      col=ifelse(Brand=="Energizer","red","blue"),
      xlab="Time (in hours)", ylab="Log of Voltage",
      las=1,
      data=batteries2
)
grid()
abline(h=log(0.9), col="gray")
legend("topright",pch=c(3,4), col=c("red","blue"),
legend=c("Energizer","Ultracell"), bty="n")
```

The repeated measures analyses are conducted using these commands:
```
t1 <- aov(Time ~ factor(Voltage)*Brand + Error(Battery),
      data = batteries2)
summary(t1)
model.tables(t1,"means")


t2 <- aov(Time ~ Voltage*Brand + Error(Battery),
      data = batteries2)
      ### Assumes linear relationship between Time and Voltage
summary(t2)
```

Figure 5 is produced using these commands:
```
plot( Voltage~Pulses,
      main="Test 1",
      pch=ifelse(Brand=="Energizer",3,4),
      col=ifelse(Brand=="Energizer","red","blue"),
      xlab="Pulses", ylab="Voltage",
      las=1,
      data=batteries1)
grid()
abline(h=0.9, col="gray")
legend("topright",pch=c(3,4), col=c("red","blue"),
legend=c("Energizer","Ultracell"), bty="n")
```

Figure 6 is produced using these commands:
```
plot( Voltage~Pulses,
      main="Test 3",
      pch=ifelse(Brand=="Energizer",3,4),
      col=ifelse(Brand=="Energizer","red","blue"),
      xlab="Pulses", ylab="Voltage",
      las=1,
      data=batteries3)
grid()
abline(h=1.05, col="gray")
legend("topright",pch=c(3,4), col=c("red","blue"),
legend=c("Energizer","Ultracell"), bty="n")
```

Appendix C: Battery Life Data

This Battery Life Data designed experiment included 108 observations with 4 variables in each data file.

The dataset designated *batteries1.csv* is available as a comma-separated value Excel file:
http://www.amstat.org/publications/jse/v21n1/dunn/batteries1.csv

The dataset designated *batteries2.csv* is available as a comma-separated value Excel file:
http://www.amstat.org/publications/jse/v21n1/dunn/batteries2.csv

The dataset designated *batteries3.csv* is available as a comma separated value Excel file:
http://www.amstat.org/publications/jse/v21n1/dunn/batteries3.csv

A documentation file for the data set can be accessed in a .pdf file at:
http://www.amstat.org/publications/jse/v21n1/dunn/batterylife.pdf

---

Acknowledgements

The contributions and suggestions of the reviewers are gratefully acknowledged.

References

Aliaga, M., Cobb, G., Cuff, C., Garfield, J., Gould, R., Lock, R., Moore, T., Rossman, A., Stephenson, B., Utts, J., Velleman, P., and Witmer, J. (2010), "Guidelines For Assessment and Instruction in Statistics Education: College Report." Technical report, American Statistical Association.

Chatterjee, S., Handcock, M. S., and Simonoff, J. S. (1995), *A Casebook for a First Course in Statistics and Data Analysis,* New York: John Wiley and Sons.

Data and Story Library, The (1996), Accessed 13 February 2012, from the StatLib Web site: http://lib9stat.cmu.edu/DASL/.

Efron, B., Tibshirani, R. J. (1993), *An Introduction to the Bootstrap*, New York: Chapman and Hall.

Fitzmaurice, G. M., Laird, N. M., and Ware, J. H. (2004), *Applied Longitudinal Analysis*. Wiley Series in Probability and Statistics. Wiley.

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., and Scheaffer, R. (2007), "Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report: A Pre-K–12 Curriculum Framework." Technical report, American Statistical Association.

Freidlin, B. and Gastwirth, J. L. (2000), "Should the Median Test be Retired from General Use?" *The American Statistician*, 54(3):161–164.

Good, P. and Hardin, J. W. (2006), *Common Errors in Statistics (and How to Avoid Them)*, Second Edition, New Jersey: Wiley-Interscience.

Hand, D. J., Daly, F., Lunn, A. D., McConway, K. Y., and Ostrowski, E. (1996), *A Handbook of Small Data Sets,* London: Chapman and Hall.

JSE Data Archive (1993), Accessed 13 February 2012, from http://www.amstat.org/publications/jse/jse_data_archive.htm.

Lindström, P. (2011), "Test report for primary battery testing for ALDI Stores Australia. "Technical report, Intertek.

Peck, R., Haugh, L. D., and Goodman, A. (1998), *Statistical Case Studies: A Collaboration between Academe and Industry*, Philadelphia: American Statistical Association and the Society for Industrial and Applied Mathematics.

Peck, R., Casella, G., Cobb, G., Hoerl, R., Nolan, D., Starbuck, R., and Stern, H. (2006), *Statistics: A Guide to the Unknown,* Fourth Edition, Thomson Brooks/Cole.

R Development Core Team (2011), R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Smyth, G. K. (2011), "OzDASL: Australasian Data and Story Library (OzDASL)" [online]. Accessed 10 September 2012, from http://www.statsci.org/data.

Weiss, R. E. (2005). *Modeling Longitudinal Data*. Springer.

Peter K. Dunn
Faculty of Science, Health, Education and Engineering
University of the Sunshine Coast
Locked Bag 4
Maroochydore DC Queensland 4558
Australia
pdunn2@usc.edu.au
