TEACHING STATISTICS

Summer 2010

Volume 32

Number 2

An International Journal for Teachers

Classroom Notes

Computing Corner

Curriculum Matters

Data Bank

Historical Perspective

Practical Activities

Book Reviews

News and Notes

WILEY-BLACKWELL

Hypothesis Testing Using the Films of The Three Stooges

KEYWORDS:
Teaching;
Data gathering;
t *test.*

Robert Gardner and Robert Davidson
East Tennessee State University, USA.
e-mail: gardnerr@etsu.edu
davidson@etsu.edu

Summary
The use of The Three Stooges' films as a source of data in an introductory statistics class is described. The Stooges' films are separated into three populations. Using these populations, students may conduct hypothesis tests with data they collect.

◆ INTRODUCTION ◆

The Three Stooges are probably the most widely recognized comedy team of all time. Even though their last work was done over 40 years ago, the Stooges are still prominent figures of popular culture and their films are broadcast even today. We propose that this makes them suitable for presentation in an introductory statistics class as a means of engaging students and reducing their anxiety over math and statistics. In the process, the Stooges' work is used as a source of data for application in statistical analysis. Students are given the chance to collect their own data, an opportunity not always provided in the traditional classroom setting.

The Three Stooges, paradoxically, consisted of six different members. At the core were Moe Howard and Larry Fine. The position of 'third stooge' was held by Curly Howard, Shemp Howard, Joe Besser and Curly-Joe DeRita. The trio of Moe–Larry–Curly made 97 films (also called 'two-reelers' – which averaged about 17 minutes each in length) between 1934 and 1947, Moe–Larry–Shemp made 77 films between 1947 and 1956 and Moe–Larry–Joe made 16 films between 1957 and 1959. Curly-Joe joined the team after the genre of two-reelers was over, and made several full-length movies as a Stooge. We concentrate on the two-reelers and treat them as three populations: the Moe–Larry–Curly population of size 97, the Moe–Larry–Shemp population of size 77 and the Moe–Larry–Joe population of size 16. See figure 1 for an example of the main title card from a film from each of the three populations considered in this project.

In this article we discuss, among other things:

- The motivation behind hypothesis statements.
- One- and two-sample t tests.
- Data gathering by students.

◆ POPULAR CULTURE IN THE MATH CLASSROOM ◆

The use of topics from popular culture is a common way to engage and motivate students in all academic areas, including the quantitative sciences. Greenwald and Nestler (2004) comment:

> Educators are extremely interested in using popular culture to enhance the teaching and learning of mathematics. Large audiences attending related talks at national mathematics meetings and colleges across the nation, including some that are standing room-only, provide evidence of that claim. . . . Capitalizing on student enjoyment of popular culture can alleviate math anxiety, energize shy and quiet students, and provide a creative introduction to an in-depth study of the related mathematics. (1–2)

Drs Greenwald and Nestler have concentrated on the contemporary animated television series *The Simpsons* and *Futurama*. Some writers of these shows have graduate math degrees and intentionally include math references in the stories. These references sometimes involve rather sophisticated ideas, including non-Euclidean geometry and Fermat's Last Theorem. This is interesting, but the complexity of these concepts may exacerbate math anxiety instead of alleviating it! On the other hand, no one could suffer from any sort of intellectual anxiety when viewing The Three Stooges!

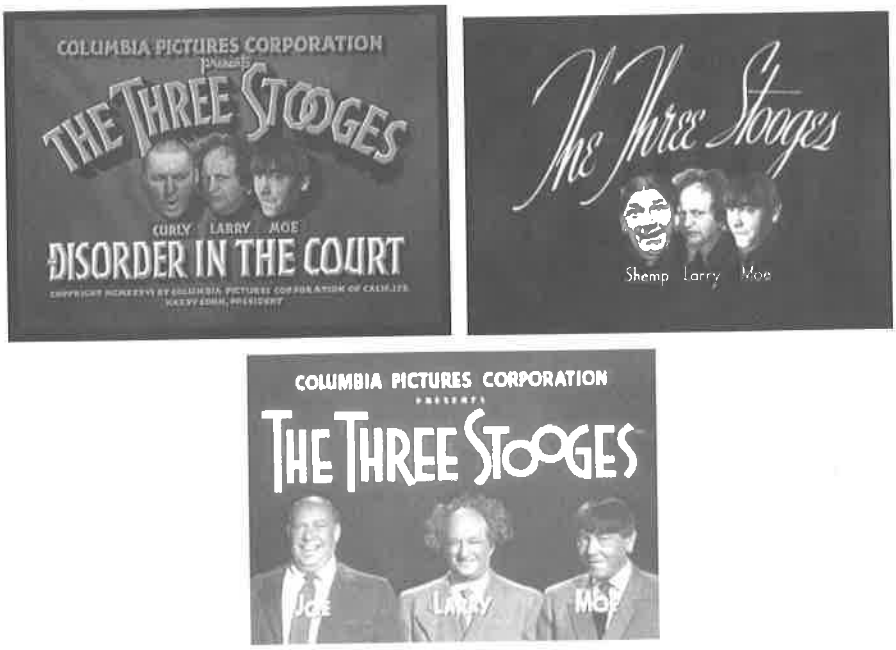

Fig 1. Main title cards of a film from each of the three populations (courtesy of Columbia Pictures)

THE THREE STOOGES AND TODAY'S STUDENTS

One might argue that today's students are unfamiliar with The Three Stooges. Fortunately, there are surveys that address exactly this issue. Zogby International conducted a poll of 1213 American adults by telephone in July 2006. One question asked for the names of The Three Stooges and another asked for the names of the three branches of the U.S. government. The results were that 73% were able to name the Stooges and only 42% were able to name the branches of government. Of course, the intent of this poll is not to reflect the high level of the American public's knowledge of the Stooges, but the evidence is there. Since the target student audience of *Teaching Statistics* is those of age 19 and less, 'American adults' is not the proper category of interest. A related poll was conducted by the National Constitution Center in 1998 (maybe a bit dated, but this is the data available) that asked the same questions of American teenagers. The results were similar, though the numbers smaller (in both categories), with 59% able to name the Stooges and 41% able to name the branches of government. These surveys involve the actual naming of the Stooges, versus simply recognizing them. This gives support to the contention that the Stooges hold a very high recognition value for today's high school and first-year college students. Even if today's students are unfamiliar with the Stooges, they can enjoy their films once they see them – four of the Stooges films are in the public domain and can be viewed online. One source is http://emol.org/movies/threestooges/index.html.

HYPOTHESIS TESTING ILLUSTRATED

For the sake of illustration, we compare the Moe–Larry–Curly population to the Moe–Larry–Shemp population. To simplify things, we choose to concentrate on the behaviour of Moe. Moe is, you might say, the 'boss' of the Stooges; he is the one most likely to institute a fit of slapstick slapping and eye pokes. We take the following as the null hypothesis:

> H_0: 'The average number of violent acts by Moe against Curly per episode is the same as the average number of violent acts by Moe against Shemp.'

Since we have no reason to expect that one mean is greater than the other, a two-tailed test with the following alternate hypothesis is used:

> H_a: 'The average number of violent acts by Moe against Curly per episode is different than the average number of violent acts by Moe against Shemp.'

With the presentation of this test, students:

- see an example of a hypothesis test before performing one themselves;

Shemp		*Curly*	
Episode Title (Episode #)	# acts	Episode Title (Episode #)	# acts
Shivering Sherlocks (104)	13	Uncivil Warriors (8)	27
Punchy Cowpunchers (120)	3	Whoops, I'm an Indian (18)	13
Love at First Bite (123)	20	Back to the Woods (23)	12
Three Arabian Nuts (129)	9	Three Missing Links (34)	9
Scrambled Brains (132)	11	How High is Up? (48)	38
Corny Casanovas (139)	17	Cookoo Cavaliers (51)	14
Cuckoo on a Choo Choo (143)	16	An Ache in Every Stake (57)	6
Knutzy Knights (156)	8	Sock-a-Bye Baby (66)	10
Shot in the Frontier (157)	2	A Bird in the Head (89)	11
Husbands Beware (167)	4	Uncivil Warbirds (90)	3
Shemp Average $\bar{x}_1$	10.3	Curly Average $\bar{x}_2$	14.3
Shemp Standard Deviation s_1	6.2191	Curly Standard Deviation s_2	10.4568

Table 1. The number of acts of violence by Moe towards Shemp and Curly in the random samples

- see an example of a two-tailed test; and
- get to see, as we will show, a test that is not statistically significant.

First, random samples of size 10 from each of the two populations are drawn. We the instructors watch (and rewatch) the films in the sample until we come to a consensus on the number of relevant acts. This yields the data given in table 1.

The t statistic is calculated as

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

This t statistic has an approximate t distribution with 15 degrees of freedom giving rise to a p-value of about 0.315. (The instructor should discuss the needed assumptions here. See, for example, Moore (2009) on the robustness of the two-sample t test.) Therefore, we fail to reject the null hypothesis at, say, the 0.05 level.

This example provides the opportunity to discuss what it means to decide to 'fail to reject the null hypothesis' in a statistical test. Students can see that it is not being claimed that the average is the same for the Shemp and Curly data. Instead, the data shows *no significant difference* between the averages of these two populations.

◆ A MOTIVATED HYPOTHESIS ◆

We had no real motivation for the hypothesis tested in the previous section – we were merely illustrating the mechanics of a test. In this section, we give motivation for a specific related hypothesis.

It is widely held among Stooge fans that the Joe Besser episodes were the least dynamic. To motivate a hypothesis along these lines, we expose students to a couple of quotes concerning Joe Besser's Stooges years. An article, 'Pure slap shtick', by Richard von Busack (1997) includes the quote:

> To Besser's eternal shame, he had it written into his contract that he would not be subject to slapping or bodily harm.

In addition, the Mackinac Media DVD *The Men Behind the Mayhem* (2004) has an interview of Joe in which he says:

> Moe and Larry - they were the best. I enjoyed every minute of it with them. In fact, to show you how wonderful they were, I never liked to be hit with anything. And Larry used to say to me 'don't worry Joe, I'll take it.' Now that's the kind of guys they were.

Based on these quotes, students are quick to hypothesize that Joe is the recipient of less violence than the other Stooges. Curly is probably the most popular Stooge. Therefore, the treatment of Moe towards Curly versus the treatment of Moe towards Joe is a reasonable comparison to study. For this study, the null hypothesis is:

> H_0: 'The average number of violent acts by Moe against Curly per episode is the same as the average number of violent acts by Moe against Joe.'

Since less Moe violence aimed at Joe is expected, a one-tailed test is performed with alternate hypothesis:

> H_a: 'The average number of violent acts by Moe against Curly per episode is greater than the average number of violent acts by Moe against Joe.'

Episode Title (Episode #)	*# acts*
Hoofs and Goofs (175)	1
Muscle Up a Little Closer (176)	2
A Merry Mix Up (177)	1
Space Ship Sappy (178)	3
Guns a Poppin! (179)	2
Horsing Around (180)	1
Rusty Romeos (181)	3
Outer Space Jitters (182)	1
Quiz Whizz (183)	4
Fifi Blows Her Top (184)	3
Pies and Guys (185)	8
Sweet and Hot (186)	0
Flying Saucer Daffy (187)	8
Oil's Well That Ends Well (188)	4
Triple Crossed (189)	2
Sappy Bull Fighters (190)	4
Joe average μ	2.9375

Table 2. The number of acts of violence by Moe towards Joe in the population of Joe Besser shorts

STUDENT GATHERING OF DATA AND HYPOTHESIS TESTING

Since the Moe–Larry–Joe population is small (size 16), it is reasonable to simply watch all of the episodes and determine the relevant parameter. Our results are given in table 2. This gives the average for the Joe population as $\mu = 2.9375$. However, since the Moe–Larry–Curly population is relatively large (size 97), a random sample is taken and a statistical analysis is performed.

To involve students in data collection, we show parts of the Moe–Larry–Curly sample and parts of the Moe–Larry–Joe population that are dynamic, entertaining and 'typical Stooges.' About 12 minutes of video from three different episodes is presented to the students. This allows enough time per episode to get some flavour of the story, while not spending too much time in class watching videos. While watching, students count the number of relevant acts. After each part, the class is polled to see what numbers were gathered. There is rarely universal agreement on the collected numbers and this leads to a discussion of possible reasons for disagreement. For example, is a distinction drawn between a violent act that is intentional and one that is unintentional? If Moe shoves Joe and then Joe falls into a cactus, should we count this as one act of violence or two? This gives the students a flavour of the difficulty in gathering data from the 'real world.' After viewing the videos, the data gathered by the students is added to the data we collected earlier. In this way, a large sample size can be attained without spending an inordinate amount of time in class collecting data.

Next, the collected data is analyzed. This time, the null hypothesis to be tested is that the mean of the Curly population is the same as the mean of the Joe population (which we know to be $\mu = 2.9375$). The formula for the t statistic is

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}}$$

If the same sample from the Moe-Larry-Curly population as used above is used here, then $\bar{x} = 14.3$, $s = 10.4568$ and $n = 10$. The number of degrees of freedom is $df = n - 1 = 9$. The t statistic is then $t = 3.4362$, and so for a one-tailed test, the p-value is about 0.0037. (This p-value should be viewed as approximate as it is a bit questionable as to whether the needed assumptions are satisfied.) Therefore, the statistical test reveals significant evidence that there is more violence by Moe towards Curly on average than there is violence by Moe towards Joe (as suspected).

Notice that the example in the earlier section and the problem in this section differ in a number of ways. For the earlier Curly/Shemp example,

- a two-tailed test was performed;
- the test involved sampling from two populations; and
- the test was not statistically significant.

For the Curly/Joe problem,

- a one-tailed test was performed;
- one population was sampled while the mean of the other population was measured outright; and
- the test was statistically significant.

These differences between the two tests give students exposure to a variety of nuances of statistical tests. In particular, we did not sample the Joe population and there was no statistical analysis involving the parameters of that population. This gives an opportunity to discuss sample mean $\bar{x}$ versus population mean μ. In our experience, students are quick to head for statistical analysis whenever given numbers, with reckless disregard for whether the numbers are from a sample or from the entire population; this aspect of the test helps address this misconception.

DISCUSSION

The tests described raise a number of questions in terms of the gathering and analysis of the data. As mentioned, the precise definition of 'act of violence' is not so easy to tie down. Students get exposure to dealing with the vagaries of categorizing and defining parameters to be studied in a statistical test. Since we decided to count a parameter *per episode*, it is reasonable to ask if the episodes are all of the same length. In fact, they are not and lengths range from 16 to 19 minutes. Students can be asked how they would adjust for this. The samples are assumed to be independent (in the two-sample problem), so we might ask if this is the case. In fact, a number of the Stooge stories were remade in later years and sometimes footage was even recycled. Therefore, a topic of discussion in the classroom is how this might affect the counts of violent acts and the subsequent statistical analysis. Finally, the use of a t test assumes that the populations are normally distributed. This topic is certainly worthy of mention in class, since the samples may be examined to shed light on the validity of this assumption.

The number of Three Stooges films is large, so there are a number of ways to partition them into populations for analysis:

- by writer;
- by director;
- by decade,
- prewar/war/postwar periods; and
- Curly/Shemp/Joe etc.

Statistical tests between these populations are, of course, also possible. Three Stooges films can act as a source for a number of statistical summaries as well as tests. The number of times Moe is violent to Curly, for example, can act as a variable for which we can calculate mean, median, mode, standard deviation, quartiles etc. In terms of more sophisticated statistical material, one could analyze these populations with nonparametric statistical tests that make limited assumptions on the distribution of the population. Since the films span 25 years, one could also perform time series analysis or regression to see if there is any trend in the level of violence over time.

◆ CONCLUSION ◆

We believe using this activity will help motivate students to comprehend and develop a proficiency in data analysis. While using this teaching strategy in several statistical classes, student surveys seem to justify this claim. In a traditional statistics classroom, active participation in collecting and analyzing data is limited (and not very interesting). Using the Stooge shorts will actively involve the students in the world of statistics and make the classroom more inquiry-based.

References

Greenwald, S. and Nestler, A. (2004). Using popular culture in the mathematics and mathematics education classroom. *PRIMUS*, **14**(1), 1–4.

The Men Behind the Mayhem. DVD. Directed by Paul E. Gierucki. Littleton, Colorado: Mackinac Media, 2004.

Constitution Center Press Release. (1998). *Constitution Center 1998 poll information.* http://constitutioncenter.org/Files/stooges.pdf (accessed 12 August 2009).

Moore, D. (2009). *The Basic Practice of Statistics*, 5th edn. New York: W. H. Freeman.

von Busack, R. (1997). Pure slap shtick. *Metro Santa Cruz*, January 16–22. http://www.ratical.org/ratville/3stooges75yrs.html (accessed 12 August 2009).

Business Wire. (2006). *Zogby 2006 poll results.* http://www.zogby.com/Soundbites/ReadClips.cfm?ID=13498 (accessed 12 August 2009).