Continuing Judicial Education

Standard Deviation

Melissa Nathanson

JUSTICE BLACKMUN'S NOTE TO Justice Stewart conveying the formula for computing standard deviation (recently published in the *Green Bag*[1]) had its genesis in a footnote prepared by Blackmun clerk Richard A. Meserve for the opinion of the Court in *Castaneda v. Partida*.[2] The sole question before the Court in that case was whether the State had successfully rebutted Partida's prima facie showing of discrimination against Mexican-Americans in the Texas grand jury selection process. Five justices decided the State had not done so. Justice Brennan, the senior associate justice in the majority, assigned the opinion to Blackmun. Diane Wood, who, like Meserve, was clerking for Blackmun that term, composed a preliminary draft. After reviewing Wood's draft, Meserve suggested that the addition of a statistical analysis would "bolster the opinion enormously," and got to work preparing one.[3]

Melissa Nathanson is working on a biography of Justice Harry A. Blackmun.

1 Standard Judicial Deviation as of June 20, 1977, 22 Green Bag 2d 253-55 (2019).

2 430 U.S. 482, 496 n.17 (1977).

3 Meserve to Blackmun, January 31, 1977, Container 246. This and other selected internal Court documents from the Harry A. Blackmun Papers, Manuscript Division, Library of Congress, Washington, D.C., are reproduced in the Appendix.

Meserve was particularly well suited for the job. Like Blackmun, he had majored in mathematics as an undergraduate. Unlike Blackmun, he had simultaneously completed a major in physics. He went on to earn a law degree from Blackmun's alma mater (Harvard) and a Ph.D. in applied physics from Stanford before beginning his clerkship. (Meserve later served as Chairman of the Nuclear Regulatory Commission in the Clinton and Bush administrations, and is now a Senior Of Counsel with Covington & Burling LLP. Wood is currently the Chief Judge of the U.S. Court of Appeals for the Seventh Circuit).

In making out his prima facie case, Partida had presented data from the 1970 federal census and the grand jury records of Hidalgo County. Meserve's statistical analysis of the data showed that out of the 870 individuals summoned to serve as grand jurors during an 11-year period, 688 would be expected to be Mexican-Americans if the selection had been truly random. In fact, the actual number of Mexican-American grand jurors was 339. As Meserve explained in his footnote, "in any given drawing some fluctuation from the expected number is predicted. The important point, however, is that the statistical model shows that the results of a random drawing are likely to fall in the vicinity of the expected value."[4] The measure of the predicted fluctuation is called the standard deviation. The standard deviation for the expected number of Mexican-American grand jurors was 12. The standard deviation for the actual number of Mexican-American grand jurors was 29. The difference was so large that Meserve was unable to locate a table estimating the probability of it occurring by chance. Nothing less than an extraordinarily powerful computer would suffice for the job. Luckily, Meserve had a friend at Argonne National Laboratory. The friend turned the problem over to the lab's supercomputers. The answer: the probability that a random selection of grand jurors from the total population of Hidalgo County would result in only 339 Mexican-American grand jurors was less than one in 10^{140}.[5]

Blackmun circulated his draft opinion with Meserve's footnote on February 7, 1977. Justice Brennan, Justice Marshall, and Justice Stevens joined immediately. Justice White said he wanted to see the dissent before making up his mind. Chief Justice Burger had assigned the dissent to Justice Powell,

---

4 *Castaneda*, 430 U.S. at 496 n.17.

5 Meserve to Blackmun, February 4, 1977, Container 246.

who circulated his draft on February 18. After he saw Powell's draft, White joined Blackmun's opinion. Burger circulated a memorandum to the Conference saying he agreed with Powell's dissent, but saw an additional flaw in Partida's case. He attacked Partida's data (and Blackmun's draft opinion) on a variety of grounds, including that the population number from the census included children, undocumented immigrants, and those not literate in the English language, none of whom were eligible for jury service. Blackmun asked Meserve to run another statistical analysis assuming Burger's contentions were correct. Meserve reported back that, even if they were, the likelihood of drawing 339 Mexican-American grand jurors as a matter of pure chance was still vanishingly remote: one in 10^{50}. Burger's memorandum became a separate dissent.[6] Blackmun responded by adding a new paragraph to one of his footnotes.[7]

The 5-4 decision for Partida came down in March. A month later, the Court heard the final oral argument of the term in *Hazelwood School District v. United States*,[8] which concerned discrimination in the employment of teachers. As in *Castaneda*, quantitative data had been presented to establish a prima facie case. Clerks in the chambers of Justice Stewart, who would write the Court's opinion vacating and remanding the decision below for further fact-finding, and Justice Stevens, who would write the sole dissenting opinion, were sufficiently impressed with what had emerged from Justice Blackmun's chambers in *Castaneda* that they consulted Meserve as to the proper conclusions to be drawn from the *Hazelwood* data. Once again, Argonne National Laboratory produced the necessary computations.[9] Meserve suggested technical tweaks to opinion drafts.[10] In a memorandum concerning one of these changes, Stewart advised the other justices that he planned to add the following sentence to one of his footnotes:

---

6 *Castaneda*, 430 U.S. at 504 (Burger, C.J. dissenting).

7 *Id*. at 488 n.8 (paragraph 4); Blackmun to the Conference, March 18, 1977, Container 246.

8 433 U.S. 299 (1977).

9 Meserve to Blackmun, June 17, 1977; Meserve, Binomial Model – *Hazelwood* Data, June 16, 1977, Container 251. Meserve produced a more detailed explanation of the binomial model in a seven-page memorandum the following week, which appears to be a recapitulation of oral explanations he provided during the opinion-writing process. Meserve, Re: *Hazelwood* and *Castaneda* Binomial Model, June 23, 1977, Container 246.

10 Meserve to Blackmun, June 21, 1977, Container 251.

> A more precise method of analyzing these statistics confirms the results of the standard deviation analysis. See F. Mosteller, R. Rourke, & G. Thomas, Probability with Statistical Applications 494 (2d ed. 1970).[11]

He concluded his memo with an entreaty: "Please do not ask me to explain it."[12] There can be little doubt that this was what prompted Blackmun's tongue-in-cheek reply enclosing the formula for standard deviation "to straighten out any confusion that may exist among all of us."[13]

A coda of sorts came in 1984, three years after Justice Stewart's retirement. Justice Blackmun was "troubled" to read in Justice Rehnquist's draft opinion in *Richardson v. United States*[14] that 170 years had passed since Justice Story's opinion in *United States v. Perez*.[15] "I know I have never been able to add, subtract, multiply, or divide," Blackmun wrote to Rehnquist,

> "but if that numeral is not corrected, I shall threaten you with an erudite footnote similar to footnote 9 of the recent *Tully* opinion [*Westinghouse Electric Corp. v. Tully*, 466 U.S. 388, 400 n. 9 (1984) (setting forth hypothetical examples demonstrating that similarly situated corporations, each operating a wholly owned Domestic International Sales Corporation ("DISC"), would face different tax assessments in New York State depending on the location from which the DISC shipped its exports)] or, heaven forbid, even similar to the infamous footnote 17 of *Castaneda v. Partida*."[16]

Rehnquist promptly replied that Blackmun's threat "overbore" him, and agreed to make the correction. "The mistake in my circulation," he explained, "originated from the fact that, being a traditionalist, I still use the Julian calendar rather than the Gregorian calendar and was computing the passage of time on the basis of the former."[17]

---

[11] Stewart to the Conference, June 15, 1977, Container 251.

[12] *Id.*

[13] Blackmun to Stewart, June 20, 1977, Container 251.

[14] 468 U.S. 317 (1984).

[15] 9 Wheat. 579 (1824).

[16] Blackmun to Rehnquist, May 1, 1984, Container 404.

[17] Rehnquist to Blackmun, May 4, 1984, Container 404.

Appendix

Selected internal Court documents
(in chronological order)
from the
Harry A. Blackmun Papers, Manuscript Division,
Library of Congress, Washington, D.C.

<u>Re</u>: <u>Castaneda</u> v. <u>Partida</u>, No. 75-1552

I requested the Library to find several statistics books that I hoped would be helpful in drafting the proposed FN in <u>Castaneda</u>. Some reference to a set of tables will be necessary in calculating the likelihood that the result in this case would have occurred purely by chance. (Or, alternatively, some computer time.) Although I placed the order right after breakfast, none of the books has yet appeared.

I hope this will not hold things up. My feeling is that the statistical analysis will serve to bolster the opinion enormously.

RM 1/31/77 4 P.M.

Re: FN 17, *Castaneda* v. *Partida*

The draft gives the probability of the 11-year data occurring by chance as less than 10^{-140} and of the $2^1/_2$-year data occurring by chance as less than 10^{-25}. Neither proposition has a citation. As it happens, the probability is so extremely slight that there is no table that I could find that even estimated the probability in this range.

The results are from an exact calculation performed by a friend at Argonne Nat'l Laboratory at my request. He has a Ph.D. in physics from Stanford and I am confident did a careful and accurate job. The actual calculation is rather straightforward; it involves the computation of the sum

$$\sum_{\ell=0}^{x} \frac{N!}{\ell!\,(N-\ell)!} (.791)^{\ell} (.209)^{N-\ell}$$

where, for the 11-year data, N=870 and x=339. The only difficulty comes from the extremely small magnitude of each term in the sum -- the computer cannot comfortably accomodate a number smaller than 10^{-100}. Perhaps, in the absence of a cite, it would be appropriate to say "*A detailed calculation reveals that* the likelihood that . . . is less than 1 in 10^{140}." (Underlined material is new.)

I am still awaiting some materials from the Library of Congress. I have been assured that the material will be retrieved but apparently the books are buried deep in the catacombs. When this material appears I will insert some additional authority.

I'm sorry this FN is proving to be such a problem.

RM 2/4/77

Supreme Court of the United States
Washington, D. C. 20543

CHAMBERS OF
JUSTICE HARRY A. BLACKMUN

March 18, 1977

MEMORANDUM TO THE CONFERENCE

Re: No. 75-1552 - Castaneda v. Partida

The following new paragraph should be inserted before the final paragraph in footnote 8, on page 6.

"The suggestion is made in the dissenting opinion of the Chief Justice, post, that reliance on eligible population figures and allowance for literacy would defeat respondent's prima facie showing of discrimination. But the 65% to 39% disparity between Mexican-Americans over the age of 25 who have some schooling and Mexican-Americans represented on the grand jury venires takes both of the Chief Justice's concerns into account. Statistical analysis, which is described in more detail in n. 17, infra, indicates that the discrepancy is significant. If one assumes that Mexican-Americans constitute only 65% of the jury pool, then a detailed calculation reveals that the likelihood that so substantial a discrepancy would occur by chance is less than 1 in 10^{50}."

H.A.B.

HAB

Supreme Court of the United States
Washington, D. C. 20543

CHAMBERS OF
JUSTICE POTTER STEWART

June 15, 1977

MEMORANDUM TO THE CONFERENCE

Re: No. 76-255, Hazelwood School District v. United States

I plan to add the following sentence at the end of the first paragraph of footnote 16:

> A more precise method of analyzing these statistics confirms the results of the standard deviation analysis. See F. Mosteller, R. Rourke, & G. Thomas, Probability with Statistical Applications 494 (2d ed. 1970).

Please do not ask me to explain it.

P.S.

P.S.

Binomial Model -- Hazelwood Data

	p=0.057	p=0.154
Sum 10/282	6.91 D-02	1.22 D-10
Sum 5/123	2.92 D-01	6.44 D-05
Sum 15/405	4.56 D-02	2.95 D-14

I believe the proper conclusions to be drawn from these results are: (1) If the population being sampled is characterized by p=0.154, then a statistician would reject the hypothesis that the selection was random. (2) If the population being sampled is characterized by p=0.057, then a statistician would conclude that the results are, at best, inconclusive as to whether the selection was random. The results hover at the rejection level for the random hypothesis. Certainly the results do not show the converse, i.e., that the selection was random.

The binomial sums were performed by a professional physicist who is employed by Argonne National Laboratory. Although I am confident they are correct, probably reference to these results in other than general terms in an opinion would be inappropriate.

R.A.Meserve 6/16/77

Re: Hazelwood School District v. United States, No. 76-255

I was consulted by both Justice Stewart's law clerk and Justice Stevens' law clerk as to the proper conclusions to be drawn from the data in Hazelwood. Since I now understand that Mr. Justice Stevens is going to make use of this data in his dissent, I thought I should inform you as to the information that I gave both chambers. The document that is attached, which I gave to both, was accompanied with some verbal explanation of the problem.

I believe that a proper statistical analysis will require some revision of the majority opinion. The majority opinion now suggests that if the relevant population figure was 5.7%, then the disparity between the percentage of those hired (3.7%) and 5.7% may be sufficiently small to rebut the prima facie case. Draft at 11. As the attached document shows, this conclusion is not supported by the statistics. If the drawing were random, the actual results/or worse would occur by chance on the order of 5% of the time. The data is at best inconclusive and hardly rebuts the prima facie case. The revisions, as I understand it, will emphasize that the data is suggestive of discrimination, but inconclusive; that the CA should have remanded for fuller consideration of the statistics; and that applicant flow is particularly relevant and should/also be examined. In my view your vote in the case can be justified.

RM 6/17/77

June 20, 1977

Dear Potter:

I am advised that the enclosure is the formula for a standard deviation. This ought to straighten out any confusion that may exist among all of us.

Sincerely,

Mr. Justice Stewart

cc: Mr. Justice Brennan

$P(x)$

$$\bar{x} = \sum x \, P(x)$$

$$\sigma^2 = \sum (x - \bar{x})^2 \, P(x)$$

Re: No. 76-255 Hazelwood School District v. United States

I have read Justice Stevens' FN 5. It is essentially correct although I have suggested some slight changes in language to his clerk.

RM 6/21/77

$\#$ 75-1552

Re: Hazelwood and Castaneda
76-255 75-1552

BINOMIAL MODEL

The Problem: Imagine a very large urn that is filled with millions of marbles. The marbles are either white or black. The portion of all the marbles that are black is given by a parameter we call "p". (For example, if p=0.20, then 20% or the marbles are black.) The portion of the marbles that are white is given by "q" and obviously q=1-p. (In the example above q=0.80, meaning 80% of the marbles are white.) N marbles are drawn from the urn by a blindfolded man and we wish to know the chance that "x" of the marbles that are drawn will be black. (For example, we draw 20 marbles from the urn and we wish to know the chance that 8 of them will be black.) Obviously, in such a drawing it is possible that we could draw anywhere from zero to N black marbles.

1. The Mathematics of the Model.

If we reach into the urn and pull out one marble, the probability that it will be black is p. If we draw from the urn again, the probability of drawing a black marble in this second try is again p.* The probability of drawing two black marbles in a row is given by the product of these probabilities -- p X p=p^2. Now imagine that we have drawn x black marbles in a row, and then (N-x) white marbles. The probability of drawing the marbles in this sequence is:

$$\underbrace{p \times p \times p \cdots \times p}_{x} \times \underbrace{q \times q \cdots \times q}_{N-x} = p^x q^{N-x}$$

---

* If we draw one black marble and do not replace it, then obviously the ratio of white to black marbles remaining in the urn has changed slightly. So long as the urn contains millions of of marbles, this effect is insignificant.

The probability of drawing precisely x black marbles and (N-x) white marbles in another sequence is exactly the same since for any other sequence the order of the p's and q's is merely rearranged. Thus, the probability of drawing x black marbles and (N-x) white marbles, regardless of sequence, is given by the term above multiplied by the total number of distinct sequences. (A different sequence, for example would include drawing a white marble first, then the x black marbles, and then the remaining white marbles.)

How many ways can we rearrange the order of drawing N marbles? We can select any of the N marbles to be first, and then any of the (N-1) remaining marbles can be drawn second, and so on. Thus the total number of sequences for drawing N marbles is:

$$N \times (N-1) \times (N-2) \times \ldots \times 1 \equiv N!$$

(That is, we define N! to mean the peculiar product given above.) However, this calculation overestimates the number of distinct sequences in our problem since each of the black (white) marbles is identical and we can switch one for another without changing the sequence. Thus the number given above must be divided by the number of ways we can rearrange the set of white marbles among themselves and the set of black marbles among themselves. Thus, the total number of ways we can draw N marbles, of which x are black and (N-x) are white, is:

$$\frac{N!}{x!\ (N-x)!}$$

Then, combining our results, the probability of observing x black marbles in a drawing of a total of N marbles is:

$$P(x) = \frac{N!}{x!\,(N-x)!} \; p^x q^{N-x}$$

This formula defines what is known as the binomial distribution. The formula is derived in every elementary statistics textbook. See, e.g., P. Hoel, Introduction to Mathematical Statistics 58-63 (4th ed 1971); F. Mosteller, R. Rourke, G. Thomas, Probability of Statistical Applications 123-143 (2d ed. 1970).

Example. Suppose we toss a coin 6 times. What is the probability of observing two or fewer "heads."

A moments reflection reveals that the problem is analogous to the drawing of marbles from a large urn and so the model derived above applies. The probability of tossing a "head" is one-half, and thus p=0.5. Application of the formula above yields the following results:

Number of Heads	Probability
0	0.016
1	0.094
2	0.234
3	0.312
4	0.234
5	0.094
6	0.016

-4-

The sum of all the probabilities equals 1, as it must. As may be seen there is something less than a 2% chance of tossing either no "heads" or 6 "heads."* The most probable result, as expected, is 3 heads. The probability of drawing two or fewer heads is the sum of the terms:

$$P(x \leq 2) = P(0) + P(1) + P(2) \equiv \sum_{x=0}^{2} P(x) = 0.344$$

2. Application to the Hazelwood Problem.

Of course, the selection of teachers is hardly a random process. A school district properly surveys the available applicants and selects those best qualified. An examination of the relevant labor market, however, enables us to estimate the proportion of those who are among the best qualified who are black. If the selection from the labor pool is made without regard to race, then the problem is mathematically identical to drawing black marbles from an urn. Assuming a selection process that is "blind" as to race, the probability that a school district will "draw" x blacks in a selection of N teachers is given by the binomial distribution calculated above.

In Hazelwood over a two-year period the district hired 15 blacks out of 405 teachers. Since the total size of the drawing was 405 teachers, then N in our formula is 405. Using the assumption most favorable to the district -- that the relevant labor market is 5.7% black -- p is equal to 0.057. Then applying our formula, the probability of drawing 15 or

* Note the general property that an outcome becomes increasingly less probable the greater its departure from the expected result of three "heads."

fewer blacks is

$$P(x \leq 15) = P(0) + P(1) \cdots + P(15) = \sum_{x=0}^{15} \frac{405!}{x!\,(405-x)!} (0.057)^{x} (0.943)^{405-x}$$

The actual computation of this sum is arduous by hand, but it may be readily calculated on a computer. (The total calculation takes about one second of computer time on a reasonably sized computer.) The result is:

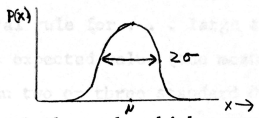
$$P(x \leq 15) = 0.0456$$

There thus is about a 5% likelihood that the school district could have happened to draw as few blacks as it did by chance.

Caveat: The crucial assumption here is that the labor market during the two-year period was 5.7% black. In light of the substantial increase in the competition for teaching jobs, it is clear that the labor market is not stable. Thus the historical figure (5.7%) may not accurately reflect the racial composition of the labor market in the two-year period. Analysis of the applicant flow data, if available, using techniques far more sophisticated than those discussed here, would seem the appropriate way to deal with the problem. See generally Note, Beyond the Prima Facie Case in Employment Discrimination Law: Statistical Proof and Rebuttal, 89 Harv. L. Rev. 387 (1975).

3. The Normal Approximation to the Binomial Distribution.

As mentioned above, the actual computation of the binomial sums for large N becomes intractible without the use of a computer. Fortunately, in this regime (and with p not too small or too large) the binomial sum is well approximated by what is known as the normal distribution. See, e.g., P. Hoel, supra at 79-86; F. Mosteller, R. Rourke, G. Thomas, supra at 270-291. Suffice it to say, the normal distribution is the most important function in statistics and forms the backbone of all statistical analysis. Tables for it and extensive discussion of its properties can be found in every statistics textbook. The normal distribution may be plotted as follows:



As may be seek the peak, which corresponds to the most probable result, is at x= μ , and the characteristic width is given by a parameter known as σ. As it happens,

$$\mu = \sum_{\text{all } x} x\, P(x)$$

The parameter μ is called the "mean" and gives the expected value of the parameter we are measuring. For example, in the coin tossing example, the mean of 6 tosses is 3 "heads". Also

$$\sigma^2 = \sum_{\text{all } x} (x-\mu)^2\, P(x)$$

The parameter σ is called the "standard deviation" and, as may be seen, it gives a measure of the expected fluctuations about the mean (the "width" of the probability distribution).

For a binomial model it may be shown that

$$\mu = Np$$
$$\sigma = \sqrt{Npq}$$

As As it happens, 95% of the probability of the normal distribution lies with $\pm 2\sigma$ of μ and about 99% of the probability lies within $\pm 3\sigma$. Hence the rule applied in Castaneda and Hazelwood -- "as a general rule for . . . large samples, if the difference between the expected value (the mean) and the observed number is greater than two or three standard deviations, then the hypothesis that the . . . drawing was random would be suspect" -- reflects the conclusion that if so extreme a departure from the expected value as is observed would happen by chance less than about 5% of the time, we will reject the hypothesis that the outcome did happen by chance. The rule embodies the standard test for statistical significance.

R.A. Meserve 6/23/77

Supreme Court of the United States
Washington, D. C. 20543

CHAMBERS OF
JUSTICE HARRY A. BLACKMUN

May 1, 1984

Re: No. 82-2113 - Richardson v. United States

Dear Bill:

Would you consider dropping a footnote at the end of the third line from the bottom of page 8 of your last circulation to the following effect?

> Of course, a trial court's finding of insufficient evidence also is the equivalent of an acquittal, see Hudson v. Louisiana, 450 U.S. 40, 44-45, n. 5 (1981), but Burks was not necessary to establish that principle. See Burks, 437 U.S., at 11, citing Fong Foo v. United States, 369 U.S. 141 (1962); Kepner v. United States, 195 U.S. 100 (1904).

I realize you already have a court for this case, but I would feel easier if a footnote to this effect were inserted. You would then have my vote, provided you satisfy my next observation.

I am troubled by the numeral in the second line of page 7. I know I have never been able to add, subtract, multiply, or divide, but if that numeral is not corrected, I shall threaten you with an erudite footnote similar to footnote 9 of the recent Tully opinion or, heaven forbid, even similar to the infamous footnote 17 of Castaneda v. Partida, 430 U.S. 482, 496 (1977). I am surprised that none of these bright law clerks and Justices around here have not made a like observation since your first circulation of April 24.

Sincerely,

HAB.

Justice Rehnquist

cc: The Conference

Supreme Court of the United States
Washington, D. C. 20543

CHAMBERS OF
JUSTICE WILLIAM H. REHNQUIST

May 4, 1984

Re: No. 82-2113 Richardson v. United States

Dear Harry:

Both of the changes to my circulating opinion in this case which you propose in your letter of May 1st are agreeable to me. With respect to the second, your threat of another erudite footnote overbore me. The mistake in my circulation originated from the fact that, being a traditionalist, I still use the Julian calendar rather than the Gregorian calendar and was computing the passage of time on the basis of the former.

Sincerely,

WHR

Justice Blackmun

cc: The Conference