Appendix: Activity

Chris Franklin

A Case of Possible Discrimination

Statisticians are often asked to look at data from situations where an individual or individuals believe that discrimination has taken place. A well-known study of possible discrimination was reported in the *Journal of Applied Psychology*. The scenario of this study is given below.

SCENARIO

In 1972, 48 male bank supervisors were each given the same personnel file and asked to judge whether the person should be promoted to a branch manager job that was described as "routine' or whether the person's file should be held and other applicants interviewed. The files were identical except that half of the supervisors had files showing the person was male while the other half had files showing the person was female. Of the 48 files reviewed, 35 were promoted. (B.Rosen and T. Jerdee (1974), "Influence of sex role stereotypes on personnel decisions," *J. Applied Psychology*, 59:9-14.)

PRELMINARY QUESTIONS

1. Suppose there was no discrimination involved in the promotions. Enter the expected numbers of males promoted and females promoted for this case in Table1.

	PROMOTED	NOT PROMOTED	TOTAL
MALE			24
FEMALE			24
TOTAL	35	13	48

Table 1 No discrimination

2. Suppose there was strong evidence of discrimination against the women in those recommended for promotion. Create a table that would show this case.

	PROMOTED	NOT PROMOTED	TOTAL
MALE			24
FEMALE			24
TOTAL	35	13	48

Table 2 Strong case of discrimination against the women

3. Suppose the evidence of discrimination against the women falls into a 'gray area'; i.e., a case where discrimination against the women is not clearly obvious without further investigation. Create a table that would show this case.

	PROMOTED	NOT PROMOTED	TOTAL
MALE			24
FEMALE			24
TOTAL	35	13	48

Table 3. 'Gray' case of discrimination against the women

Returning to the study reported earlier in the activity scenario, the results were reported that of the 24 "male" files, 21 were recommended for promotion. Of the 24 "female" files, 14 were recommended for promotion.

4. Enter the data from the actual discrimination study in Table 4.

	PROMOTED	NOT PROMOTED	TOTAL
MALE			
FEMALE			
TOTAL			

Table 4. Actual discrimination study

5. What percentage of males was recommended for promotion? What percentage of females was recommended for promotion?

6. Without exploring the data any further, do you think there was discrimination by the bank supervisors against the females? How certain are you?

7. Could the smaller number of recommended female applicants for promotion be attributed to chance? What is your sense of how likely the smaller number of recommended females could have occurred by chance?

8. Suppose that the bank supervisors looked at files of actual female and male applicants. Assume that all of the applicants were identical with regard to their qualifications and use the same results as before (21 males and14 females). If a lawyer retained by the female applicants hired you as a statistical consultant, how would you consider obtaining evidence to make a decision about whether the observed results were due to chance variation or if the observed results were due to discrimination against the women?

Statisticians would formalize the overriding question by giving two statements, a ***null hypothesis*** *which represents no discrimination so that any departure from Table 1 is due* *solely to a chance process, and an* ***alternative hypothesis*** *which represents discrimination against the women.*

9. What initial thoughts do you have about the manner in which the experiment (study) was conducted? What would need to be assumed about the study in order to infer that gender is the cause of the apparent differences?

[PSSM quote: Students should formulate questions that can be addressed with data and collect, organize, and display relevant data to answer them.]

SIMULATION OF THE DISCRIMINATION CASE

Using a deck of cards, let 24 black cards represent the males, and 24 red cards represent the females (remove 2 red cards and 2 black cards from the deck). This will simulate the 48 folders, half of which were labeled male and the other half female.

1: Shuffle the 48 cards at least 6 or 7 times to insure that the cards counted out are from a random process.

2: Count out the top 35 cards. These cards represent the applicants recommended for promotion to bank manager. The simulation could be conducted more efficiently by dealing out 13 “not promoted” cards, which would be equivalent to dealing out 35 “promoted” cards.

3: Out of the 35 cards, count the number of black cards (representing the males).

4: On the number line provided, Figure 1, create a dot plot by placing a blackened circle or X above the number of black cards counted. The range of values for possible black cards is 11 to 24.

5: Repeat steps 1 – 4 nineteen more times for a total of 20 simulations.

FIGURE 1

DOT PLOT TO BE USED TO GRAPH THE 20 SIMULATED RESULTS

11 12 13 14 15 16 17 18 19 20 21 22 23 24

Number of Men Promoted

[PSSM quote: Students should use simulations to explore the variability of sample statistics from a known population and to construct sampling distributions.]

A sampling distribution is the distribution of possible values of a statistic for repeated samples of the same size from a population. For the scenario under consideration, the number of black cards (number of males promoted) from each of the simulations is the statistic.

6: Using the results (the counts) plotted on the number line, estimate the chances that 21 or more black cards (males) out of 35 will be selected if the selection process is random; that is, if there is no discrimination against the women in the selection process.

*The probability of observing 21 or more black cards if the selection process is due to randomness or chance variation is called the **p-value.***

7: Look at the dot plot and comment on the shape, center, and variability of distribution of the counts by answering the following.
- (a) Is the distribution somewhat symmetric, pulled (skewed) to the right, or pulled to the left?
- (b) Do you observe any unusual observation(s)?
- (c) Where on the dot plot is the lower 50% of the observations?
- (d) Estimate the mean of the distribution representing the number of black cards obtained out of the 20 simulations.
- (e) What values occurred most often?
- (f) Finally, using the dot plot, comment on the spread of the data.

[PSSM quote: Students should select and use appropriate statistical methods to analyze data for univariate measurement data, be able to display the distribution, describe its shape, and select and calculate summary statistics.]

8. Is the behavior of this distribution what you might expect? Why or why not?

9. Think about the question of possible discrimination against the women. Based on the exploration just made, does there appear to be evidence to support the claim that selecting 21 males (black cards) for promotion out of 35 candidates was not due to chance variation?; i.e., how does your data compare with that of the original study?

10. How do your results compare with the results of your classmates?

USING PROBABILITY THEORY

In attempting to answer the question, " Is the difference between the number of observed promoted males and the number of expected male due to chance variation or is this difference a real difference?", simulation provided an estimated probability of 21 or more males being promoted out of 35 promotions. This probability can also be found by the application of probability theory. The stated question of interest leads to a situation that results in " success-failure" outcomes. A group of N population objects is classified into two subgroups, where one group of size k is the success group and the other group of size N-k is the failure group. The number of objects selected from the population is denoted as n. Since a male being promoted was defined as the variable of interest, a male being promoted is a success; a female being promoted is a failure. For this discrimination activity, $N = 48$ applicants, $k = 24$ males, and N-$k = 24$ females. Of the 48 applicants, n=35 were selected for promotion. A random variable X is defined that represents the <u>number of successes</u> out of n; that is, the number of males selected out of the 35 promoted. The probability of a success occurring on any given promotion is $k/N = 24/48 = 0.50$. The expected value for the number of males to be promoted out of the 35 would equal 35(0.50) = 17.5.

[The expected value of a random variable X is another name given to the mean of the random variable X.]

[PSSM quote: All students should be able to compute and interpret that expected value of random variables in simple cases.]

The selection of 35 of the 48 applicants for promotion is viewed as sampling without replacement; that is, once an applicant is selected for promotion, the applicant is not returned to the population of applicants, which does not provide the opportunity for another evaluation and promotion. Mathematically, the probability of selecting 21 or more males for promotion out of the 35 promoted can be found as follows. Let $N = k + (N\text{-}k)$, n = number of objects selected from the population of size N, and X = number of successes out of n. The desired probability is

$$P(X \geq 21) = P(X=21)+P(X=22)+P(X=23)+P(X=24).$$

To find the probability that exactly X of the males were promoted, we must first find the number of ways to select x males from $k = 24$ males in the population and the number of ways to select $(n\text{-}x)$ females from $(N\text{-}k)$ females in the population. This can be found by evaluating the combinations $\binom{k}{x}$ and $\binom{N-k}{n-x}$. By the multiplication principle, the product $\binom{k}{x}\binom{N-k}{n-x}$ equals the number of ways of promoting x males and n-x females. The combination $\binom{N}{n}$ gives the total number of ways of promoting n applicants out of N.

Thus, $P(X=x) = \binom{k}{x}\binom{N-k}{n-x} / \binom{N}{n}$.

[The random variable X follows a hypergeometric distribution.]

[PSSM quote: Students should be able to represent and analyze mathematical situations using algebraic symbols.]

Evaluating for

$P(X{=}21) = \binom{24}{21}\binom{24}{14} / \binom{48}{35}$ = [24!/(21! 3!)] [24!/(14! 10!) / 48!(35!13!) = 0.021

$P(X{=}22) = \binom{24}{22}\binom{24}{13} / \binom{48}{35} = 0.004$

$P(X{=}23) = \binom{24}{23}\binom{24}{12} / \binom{48}{35} = 0.000$

$P(X{=}24) = \binom{24}{24}\binom{24}{11} / \binom{48}{35} = \binom{24}{24}\binom{24}{11} / \binom{48}{35} = 0.000.$

Therefore, $P(X \geq 21) = 0.021 + 0.004 + 0.000 + 0.000 = 0.025$.

This mathematical probability is a long-term probability. It is the percentage of time that an outcome would be expected to occur in the long run if a simulation or experiment was replicated many times. Observe how close the mathematical probability of 0.025 is to the simulated probability of 0.028 from the class data sets considered in Sample Distribution 4 and Sample Distributions 5.

[Refer to a high school mathematics book that covers combinations for additional details on how to evaluate a combination and evaluating factorials.]