Journalist Sandy Tolan reported in the book *Hank and Me* that in 1999 there were fewer minorities coaching at third base than at first base. Tolan argued that since third base is the more challenging of the two positions and typically leads to more managing responsibilities, this discrepancy constitutes evidence of discrimination against minority coaches. Of the 60 base coaches in Major League Baseball that year (30 at first base; 30 at third base), 21 were minorities. Of these 21, only 6 coached at third base.

1. Use the given information to complete the following contingency table:

	Caucasian Coach	Minority Coach	Total
First Base Coach	_____	_____	_____
Third Base Coach	_____	_____	_____
Total	_____	_____	60

2. Is the base coached independent of the racial status of a coach? Defend your answer using probability calculations.

3. Calculate the proportion of Caucasian coaches who coach third base. Then calculate the proportion of minority coaches who coach third base. We can compare the two populations by calculating:

$$\text{relative rate} = \frac{\text{proportion of successes in Group \#1}}{\text{proportion of successes in Group \#2}}$$

The relative rate tells us how more likely it is for Caucasian coaches to coach third base than it is for minority coaches. Calculate this relative rate statistic. Does this statistic support Sandy Tolan's claim of discrimination against minority coaches?
Note: It is often standard to put the group with the lower rate of successes in the denominator.

4. Use hypergeometric probabilities to calculate the likelihood of observing data as extreme as we actually observed (Only 6 minorities coached at third base. If we assume race played no factor in coaching assignments, how likely is it that we would have observed 6 or fewer minorities coaching at third base?). Interpret this probability. Do you conclude that this provides evidence of discrimination against minority coaches? Explain.

Once you have completed these exercises on your own, show them to me and I will give you the answers.

Answers:

1.

	Caucasian Coach	Minority Coach	Total
First Base Coach	15	15	30
Third Base Coach	24	6	30
Total	39	21	60

2. No, they are not independent. For example, let's look at (Caucasian) and (1st base coach). If they were independent, then: $P(\text{Caucasian} \cap \text{1st base}) = \text{P(Caucasian)P(1st base)}$

   I can see from the table that: $P(\text{Caucasian} \cap \text{1st base}) = \frac{15}{60} = 0.25$.

   I can calculate: $\text{P(Caucasian)P(1st base)} = \left(\frac{39}{60}\right)\left(\frac{30}{60}\right) = 0.325$

   Since these values are not equal, $P(\text{Caucasian} \cap \text{1st base}) \neq \text{P(Caucasian)P(1st base)}$, race and coaching assignment are not independent.

   I could have also found this by showing: $\left[P(\text{Caucasian} \mid \text{1st base}) = \frac{15}{30}\right] \neq \left[\text{P(Caucasian)} = \frac{39}{60}\right]$

3. We define a *success* as an individual coaching third base. Therefore, we can calculate:

   $P(\text{Coaches 3rd base} \mid \text{Caucasian}) = \frac{24}{39} = 0.6154$ and $P(\text{Coaches 3rd base} \mid \text{Minority}) = \frac{6}{21} = 0.2857$

   The relative rate is simply the ratio of these probabilities: $\frac{0.6154}{0.2857} = 2.15$.

   This means that Caucasians are 2.15 times as likely to coach third base as minority coaches. This supports Sandy Tolan's claim.

4. We want to calculate $P(\text{observing 6 or fewer minorities coaching 3rd base})$ or $P(X \le 6)$.
This is the same as calculating: $P(X=0)+P(X=1)+P(X=2)+P(X=3)+P(X=4)+P(X=5)+P(X=6)$.

I know I need to choose 30 third-base coaches from the 60 individuals (39 Caucasian and 21 Minority).

Using hypergeometric probabilities:

$$P(X=0)=\frac{\binom{21}{0}\binom{39}{30}}{\binom{60}{30}}=0.00000000179 \qquad P(X=1)=\frac{\binom{21}{1}\binom{39}{29}}{\binom{60}{30}}=0.000000113$$

$$P(X=2)=\frac{\binom{21}{2}\binom{39}{28}}{\binom{60}{30}}=0.000002976 \qquad P(X=3)=\frac{\binom{21}{3}\binom{39}{27}}{\binom{60}{30}}=0.0000439807$$

$$P(X=4)=\frac{\binom{21}{4}\binom{39}{26}}{\binom{60}{30}}=0.00041105 \qquad P(X=5)=\frac{\binom{21}{5}\binom{39}{25}}{\binom{60}{30}}=0.0025955$$

$$P(X=6)=\frac{\binom{21}{6}\binom{39}{24}}{\binom{60}{30}}=0.0115355$$

The sum of these probabilities is 0.014589 (or approximately 1.5%).

This is another example of a p-value. The following statement interprets this p-value:

If we assume race has no impact on whether an individual is assigned to first or third base, then the likelihood of observing only 6 (or fewer) minority third base coaches is 0.015.

Since this probability is low (below a cut-off we would have established prior to running the study), we would conclude that our assumption (that race has no impact on coaching placement) is incorrect. Therefore, we would conclude that race does have an impact on coaching placement decisions.