

UC San Diego

Econometrics 120A: Discussion Section

Week 2: Discrete random variables, Bernoulli & Binomial

Anita Sammarini & Lapo Bini

Department of Economics

Discrete random variables

Random Variables and Probability Distribution

Example: balls in a jar



- 5 orange
- 3 blue
- 2 purple

10 balls

We draw **one ball** at random. The outcome ω is a colour, not a number: the random variable X assigns a number to each colour.

❶ outcome ω
orange
blue
purple

Notation and definition

❶ **Sample space** Ω: the set of all possible outcomes ω of a random experiment.

Random Variables and Probability Distribution

Example: balls in a jar



- 5 orange
- 3 blue
- 2 purple

10 balls

We draw **one ball** at random. The outcome ω is a colour, not a number: the random variable X assigns a number to each colour.

❶ outcome ω	❷ $X(\omega)$
orange	$\longmapsto$ 0
blue	$\longmapsto$ 1
purple	$\longmapsto$ 2

Notation and definition

❶ **Sample space** Ω: the set of all possible outcomes ω of a random experiment.

❷ **Random variable** X: a function that assigns a real number to each outcome,

$$X : \Omega \to \mathbb{R}$$

$X(\omega)$ is the value assigned to the outcome $\omega \in \Omega$.

Random Variables and Probability Distribution

Example: balls in a jar



- 5 orange
- 3 blue
- 2 purple

10 balls

We draw **one ball** at random. The outcome ω is a colour, not a number: the random variable X assigns a number to each colour.

❶ outcome ω	❷ $X(\omega)$	❸ $P(X = x)$
orange	$\longmapsto$ 0	$5/10 = 0.5$
blue	$\longmapsto$ 1	$3/10 = 0.3$
purple	$\longmapsto$ 2	$2/10 = 0.2$

Notation and definition

❶ **Sample space** Ω: the set of all possible outcomes ω of a random experiment.

❷ **Random variable** X: a function that assigns a real number to each outcome,

$$X : \Omega \to \mathbb{R}$$

$X(\omega)$ is the value assigned to the outcome $\omega \in \Omega$.

❸ **Probability mass function**: the probability of each value x of X,

$$p(x) = P(X = x)$$

with $0 \leq p(x) \leq 1$ and $\sum_x p(x) = 1$.

Capital X is the random variable; lowercase x is one of the values it can take.

Discrete Probability Distribution

- **Probability Mass Function (PMF):** Defines the probability that a discrete random variable X equals a specific value.

$$P(X = x) = p(x)$$

The sum of probabilities for all possible values must equal 1:

$$\sum_{x} P(X = x) = 1$$

- **Cumulative Distribution Function (CDF):** Gives the probability that a discrete random variable is less than or equal to a certain value:

$$F(x) = P(X \leq x) = \sum_{x' \leq x} P(X = x')$$

Discrete Probability Distribution

Discrete Distribution

- **Mean:**

$$E[X] = \sum_x x \cdot P(X = x)$$

- **Variance:**

$$\text{Var}(X) = E[(X - E[X])^2] = \sum_x (x - E[X])^2 \cdot P(X = x)$$

- **Standard deviation:**

$$\sigma = \sqrt{\text{Var}(X)}$$

- **Examples:** Bernoulli, Binomial, and Poisson distributions.

Random variables and distribution questions

PS2 Question 2

Consider a random variable X with the distribution in the table below.

(a) Compute the mean.

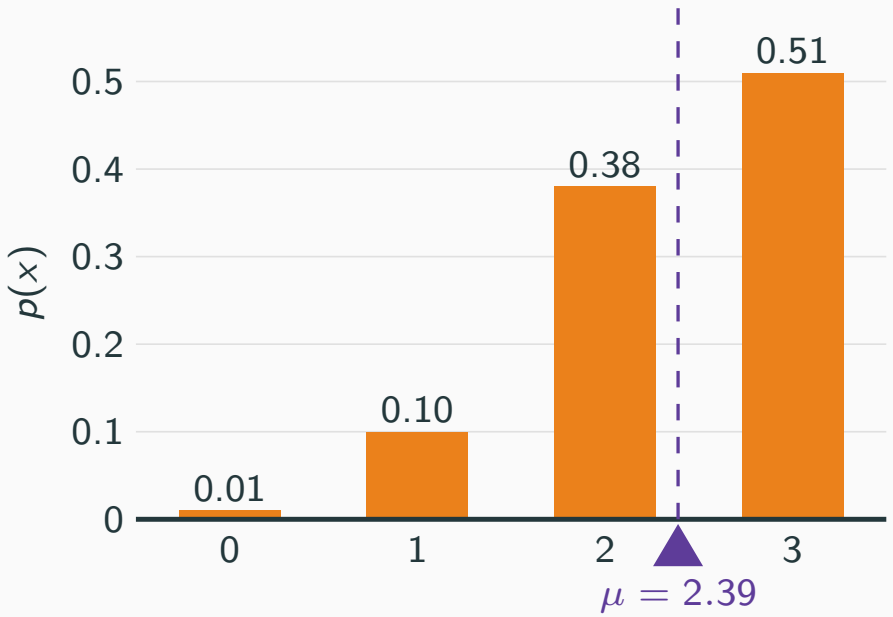
(b) Compute the variance and the standard deviation.

(c) What is $P(X \leq 2)$? And what is $P(X < 2)$?

(d) Draw the graph and show the mean is the balancing point and the standard deviation is a typical deviation.

X	p(x)
0	0.01
1	0.10
2	0.38
3	0.51

PS2 Question 2: the graph

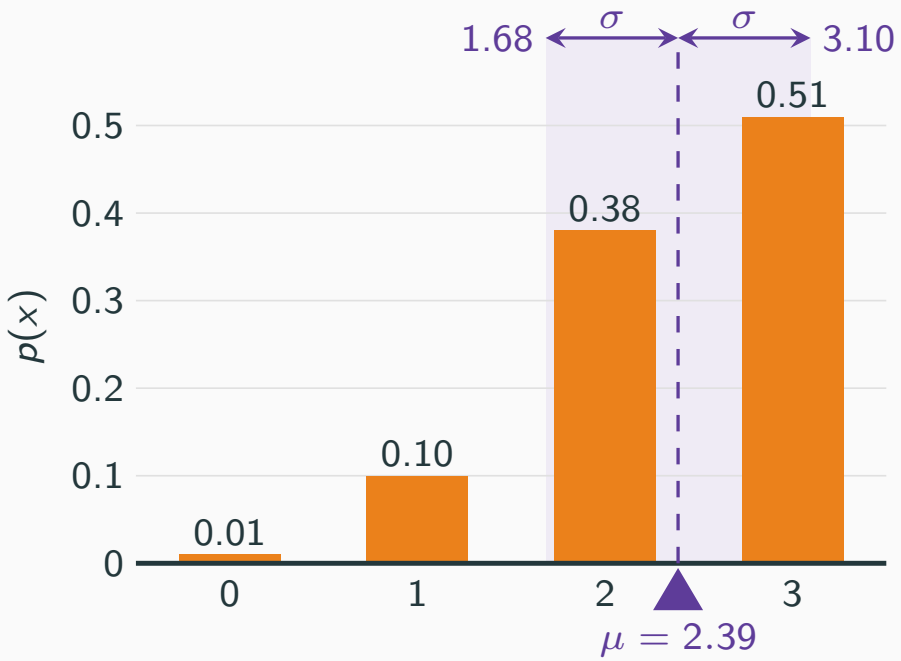


⇨ **The mean is the balancing point.** Put a fulcrum at $\mu = 2.39$: the large mass close by on the right balances the small masses further away on the left.

Distance from the mean $\times$ probability is the same on the two sides:

$$\underbrace{2.39\,(0.01) + 1.39\,(0.10) + 0.39\,(0.38)}_{\text{left of the mean}} = 0.311 = \underbrace{0.61\,(0.51)}_{\text{right of the mean}}$$

PS2 Question 2: the graph



⇨ **The mean is the balancing point.** Put a fulcrum at $\mu = 2.39$: the large mass close by on the right balances the small masses further away on the left.

⇨ **The standard deviation is a typical deviation.** Most outcomes are within about 1 of the mean, and $\sigma = 0.706$ is a number just below one.

Distance from the mean $\times$ probability is the same on the two sides:

$$\underbrace{2.39\,(0.01) + 1.39\,(0.10) + 0.39\,(0.38)}_{\text{left of the mean}} = 0.311 = \underbrace{0.61\,(0.51)}_{\text{right of the mean}}$$

Bernoulli random variable

Bernoulli random variable: examples

Many random experiments have only **two possible outcomes**. We call one of them a "success" and code it as $X = 1$; the other one is a "failure", coded as $X = 0$.

Experiment	**Success** $(X = 1)$	**Failure** $(X = 0)$	$p = P(X = 1)$
A coin flip	heads	tails	0.5
A ball from our jar	blue	not blue	$3/10 = 0.3$
A traffic light (Question 4)	you have to stop	you drive on	share of the time the light is red
A job opening (Question 5)	a woman is hired	a man is hired	0.5 with equal opportunity
A voter in a poll	in favour	against	unknown: the poll estimates it

"Success" is only the name of the outcome we count: it does not have to be good news.

Bernoulli Random Variable

Bernoulli Distribution

The Bernoulli distribution is a discrete probability distribution for a random variable which takes the value 1 with probability p (success) and the value 0 with probability $1 - p$ (failure).

- **Notation:** $X \sim \text{Bernoulli}(p)$
- **Parameter:** p (probability of success, $0 \leq p \leq 1$)
- **Probability Mass Function (PMF):**

$$P(X = x) = p^x(1 - p)^{1-x} \quad \text{for } x \in \{0, 1\}$$

so that $P(X = 1) = p^1(1 - p)^0 = p$ and $P(X = 0) = p^0(1 - p)^1 = 1 - p$.

- **Mean (Expected Value):**

$$E[X] = p$$

- **Variance:**

$$\text{Var}(X) = p(1 - p)$$

Bernoulli Random Variable

Bernoulli: where the mean and the variance come from

X takes only two values: 1 with probability p and 0 with probability $1 - p$. Plug them into the two formulas for a discrete distribution.

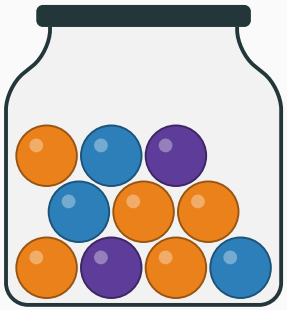
⇨ **Mean:** $E[X] =$

⇨ **Variance:** $\mathrm{Var}(X) =$

Binomial random variable

Binomial Random Variable

Binomial: an example



Draw a ball from our jar **3 times**, putting the ball back in the jar after every draw.

Each draw is a Bernoulli trial:

- ⇨ ● blue is a success ($p = 0.3$),
- ⇨ ○ not blue is a failure ($1 - p = 0.7$).

Let X be the **number of blue balls** in the 3 draws.

Binomial Random Variable

From the example to the general case

Repeat the same Bernoulli trial n times and count the successes. We need:

- ⇨ a **fixed number** of trials n,
- ⇨ only **two outcomes** in each trial: success (1) or failure (0),
- ⇨ the **same probability** of success p in every trial,
- ⇨ **independent** trials: the outcome of one trial does not change the others.

Then the number of successes

$$X = X_1 + X_2 + \cdots + X_n, \qquad X_i \sim \text{Bernoulli}(p),$$

is a **binomial random variable**: $X \sim \text{Binomial}(n, p)$.

In our example $n = 3$ and $p = 0.3$, so $X \sim \text{Binomial}(3, 0.3)$. Putting the ball back matters: without it, p would change from draw to draw and the draws would not be independent.

(Problem Set 2 also writes S for the number of successes and π for the probability of success.)

Binomial Random Variable

Binomial Distribution

The Binomial distribution models the number of successes in a fixed number of independent Bernoulli trials, each with the same probability of success.

- **Parameters:** n (number of trials, fixed) and p (probability of success)
- **Probability Mass Function (PMF):**

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{for } k = 0, 1, 2, \ldots, n$$

The binomial coefficient is defined as:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

- **Mean (Expected Value):**

$$E[X] = np$$

- **Variance:**

$$\text{Var}(X) = np(1-p)$$

Binomial Random Variable

The binomial coefficient

$\binom{n}{k}$, read "n choose k", is the **number of different ways to place k successes among n trials**.

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!} \qquad \text{where} \quad n! = n \times (n-1) \times \cdots \times 2 \times 1 \quad \text{and} \quad 0! = 1$$

Binomial Random Variable

The binomial coefficient

$\binom{n}{k}$, read "n choose k", is the **number of different ways to place k successes among n trials**.

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!} \qquad \text{where} \quad n! = n \times (n-1) \times \cdots \times 2 \times 1 \quad \text{and} \quad 0! = 1$$

⇨ **3 draws, 1 blue:** $\binom{3}{1} = \frac{3!}{1!\,2!} = \frac{3 \times 2 \times 1}{1 \times (2 \times 1)} = 3$

The blue ball can be the first, the second or the third draw: ●○○ ○●○ ○○●

Binomial Random Variable

The binomial coefficient

$\binom{n}{k}$, read "n choose k", is the **number of different ways to place k successes among n trials**.

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!} \qquad \text{where} \quad n! = n \times (n-1) \times \cdots \times 2 \times 1 \quad \text{and} \quad 0! = 1$$

⇨ **3 draws, 1 blue:** $\binom{3}{1} = \frac{3!}{1!\,2!} = \frac{3 \times 2 \times 1}{1 \times (2 \times 1)} = 3$

The blue ball can be the first, the second or the third draw: 

⇨ **10 draws, 2 blue:** $\binom{10}{2} = \frac{10!}{2!\,8!} = \frac{10 \times 9}{2 \times 1} = 45$

Too many sequences to list: this is why we use the formula.

Binomial Random Variable

The binomial coefficient

$\binom{n}{k}$, read "n choose k", is the **number of different ways to place k successes among n trials**.

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!} \qquad \text{where} \quad n! = n \times (n-1) \times \cdots \times 2 \times 1 \quad \text{and} \quad 0! = 1$$

⇨ **3 draws, 1 blue:** $\binom{3}{1} = \frac{3!}{1!\,2!} = \frac{3 \times 2 \times 1}{1 \times (2 \times 1)} = 3$

The blue ball can be the first, the second or the third draw: ●○○ ○●○ ○○●

⇨ **10 draws, 2 blue:** $\binom{10}{2} = \frac{10!}{2!\,8!} = \frac{10 \times 9}{2 \times 1} = 45$

Too many sequences to list: this is why we use the formula.

⇨ **Useful cases:** $\binom{n}{0} = \binom{n}{n} = 1, \qquad \binom{n}{1} = n, \qquad \binom{n}{k} = \binom{n}{n-k}$

On a scientific calculator this is the **nCr** key.

Binomial distribution table

The table gives $P(X = k)$ for $X \sim \text{Binomial}(n, p)$: ❶ find the table for your n, ❷ go to the column of your p, ❸ go to the row of your k.

$n = 10$	p									
k	**0.05**	**0.10**	**0.15**	**0.20**	**0.25**	**0.30**	**0.35**	**0.40**	**0.45**	**0.50**
0	0.599	0.349	0.197	0.107	0.056	0.028	0.013	0.006	0.003	0.001
1	0.315	0.387	0.347	0.268	0.188	0.121	0.072	0.040	0.021	0.010
2	0.075	0.194	0.276	0.302	0.282	0.233	0.176	0.121	0.076	0.044
3	0.010	0.057	0.130	0.201	0.250	0.267	0.252	0.215	0.166	0.117
4	0.001	0.011	0.040	0.088	0.146	0.200	0.238	0.251	0.238	0.205
5	0.000	0.001	0.008	0.026	0.058	0.103	0.154	0.201	0.234	0.246
6	0.000	0.000	0.001	0.006	0.016	0.037	0.069	0.111	0.160	0.205
7	0.000	0.000	0.000	0.001	0.003	0.009	0.021	0.042	0.075	0.117
8	0.000	0.000	0.000	0.000	0.000	0.001	0.004	0.011	0.023	0.044
9	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.004	0.010
10	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001

Binomial distribution table

The table gives $P(X = k)$ for $X \sim \text{Binomial}(n, p)$: ❶ find the table for your n, ❷ go to the column of your p, ❸ go to the row of your k.

$n = 10$	p									
k	**0.05**	**0.10**	**0.15**	**0.20**	**0.25**	**0.30**	**0.35**	**0.40**	**0.45**	**0.50**
0	0.599	0.349	0.197	0.107	0.056	0.028	0.013	0.006	0.003	0.001
1	0.315	0.387	0.347	0.268	0.188	**0.121**	0.072	0.040	0.021	0.010
2	0.075	0.194	0.276	0.302	0.282	0.233	0.176	0.121	0.076	0.044
3	0.010	0.057	0.130	0.201	0.250	0.267	0.252	0.215	0.166	0.117
4	0.001	0.011	0.040	0.088	0.146	0.200	0.238	0.251	0.238	0.205
5	0.000	0.001	0.008	0.026	0.058	0.103	0.154	0.201	0.234	0.246
6	0.000	0.000	0.001	0.006	0.016	0.037	0.069	0.111	0.160	0.205
7	0.000	0.000	0.000	0.001	0.003	0.009	0.021	0.042	0.075	0.117
8	0.000	0.000	0.000	0.000	0.000	0.001	0.004	0.011	0.023	0.044
9	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.004	0.010
10	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001

Example: 10 draws from our jar and X = number of blue balls, so $n = 10$ and $p = 0.30$. The probability of exactly one blue ball is $P(X = 1) = 0.121$.

Binomial Random Variable

Cumulative Binomial distribution table

This table gives $P(X \leq k)$: it adds up the previous table from 0 to k. Always check which of the two you are reading.

$n = 10$	p									
k	**0.05**	**0.10**	**0.15**	**0.20**	**0.25**	**0.30**	**0.35**	**0.40**	**0.45**	**0.50**
0	0.599	0.349	0.197	0.107	0.056	0.028	0.013	0.006	0.003	0.001
1	0.914	0.736	0.544	0.376	0.244	0.149	0.086	0.046	0.023	0.011
2	0.988	0.930	0.820	0.678	0.526	0.383	0.262	0.167	0.100	0.055
3	0.999	0.987	0.950	0.879	0.776	0.650	0.514	0.382	0.266	0.172
4	1.000	0.998	0.990	0.967	0.922	0.850	0.751	0.633	0.504	0.377
5	1.000	1.000	0.999	0.994	0.980	0.953	0.905	0.834	0.738	0.623
6	1.000	1.000	1.000	0.999	0.996	0.989	0.974	0.945	0.898	0.828
7	1.000	1.000	1.000	1.000	1.000	0.998	0.995	0.988	0.973	0.945
8	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.998	0.995	0.989
9	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999
10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Binomial Random Variable

Cumulative Binomial distribution table

This table gives $P(X \leq k)$: it adds up the previous table from 0 to k. Always check which of the two you are reading.

$n = 10$	p									
k	**0.05**	**0.10**	**0.15**	**0.20**	**0.25**	**0.30**	**0.35**	**0.40**	**0.45**	**0.50**
0	0.599	0.349	0.197	0.107	0.056	0.028	0.013	0.006	0.003	0.001
1	0.914	0.736	0.544	0.376	0.244	**0.149**	0.086	0.046	0.023	0.011
2	0.988	0.930	0.820	0.678	0.526	0.383	0.262	0.167	0.100	0.055
3	0.999	0.987	0.950	0.879	0.776	0.650	0.514	0.382	0.266	0.172
4	1.000	0.998	0.990	0.967	0.922	0.850	0.751	0.633	0.504	0.377
5	1.000	1.000	0.999	0.994	0.980	0.953	0.905	0.834	0.738	0.623
6	1.000	1.000	1.000	0.999	0.996	0.989	0.974	0.945	0.898	0.828
7	1.000	1.000	1.000	1.000	1.000	0.998	0.995	0.988	0.973	0.945
8	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.998	0.995	0.989
9	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999
10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Example: at most one blue ball in 10 draws: $P(X \leq 1) = 0.149 \quad (= 0.028 + 0.121)$.

Binomial Random Variable

Cumulative Binomial distribution table

This table gives $P(X \leq k)$: it adds up the previous table from 0 to k. Always check which of the two you are reading.

$n = 10$	p									
k	**0.05**	**0.10**	**0.15**	**0.20**	**0.25**	**0.30**	**0.35**	**0.40**	**0.45**	**0.50**
0	0.599	0.349	0.197	0.107	0.056	0.028	0.013	0.006	0.003	0.001
1	0.914	0.736	0.544	0.376	0.244	**0.149**	0.086	0.046	0.023	0.011
2	0.988	0.930	0.820	0.678	0.526	0.383	0.262	0.167	0.100	0.055
3	0.999	0.987	0.950	0.879	0.776	0.650	0.514	0.382	0.266	0.172
4	1.000	0.998	0.990	0.967	0.922	0.850	0.751	0.633	0.504	0.377
5	1.000	1.000	0.999	0.994	0.980	0.953	0.905	0.834	0.738	0.623
6	1.000	1.000	1.000	0.999	0.996	0.989	0.974	0.945	0.898	0.828
7	1.000	1.000	1.000	1.000	1.000	0.998	0.995	0.988	0.973	0.945
8	1.000	1.000	1.000	1.000	1.000	1.000	0.999	0.998	0.995	0.989
9	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999
10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Example: at most one blue ball in 10 draws: $P(X \leq 1) = 0.149 \quad (= 0.028 + 0.121)$.

From this table: $P(X < k) = P(X \leq k-1)$, $\quad P(X \geq k) = 1 - P(X \leq k-1)$, $P(X = k) = P(X \leq k) - P(X \leq k-1)$.

Binomial Questions

PS2 Question 4

In a city the traffic lights on the main road are set up so that traffic lights are green for **55 seconds**, red for **48 seconds**, and amber for **17 seconds** (the city is in California, so no one stops on amber).

(a) What is the chance that you will have to stop at a particular traffic light?

(b) Suppose that there are 10 traffic lights in the main street. What is the expected amount of time you are likely to be stopped at traffic lights if you drive all the way down the main street?

(c) What is the chance that you only will need to stop less than twice?

Binomial Questions

PS2 Question 4

In a city the traffic lights on the main road are set up so that traffic lights are green for **55 seconds**, red for **48 seconds**, and amber for **17 seconds** (the city is in California, so no one stops on amber).

(a) What is the chance that you will have to stop at a particular traffic light?

Binomial Questions

PS2 Question 4

In a city the traffic lights on the main road are set up so that traffic lights are green for **55 seconds**, red for **48 seconds**, and amber for **17 seconds** (the city is in California, so no one stops on amber).

(b) Suppose that there are 10 traffic lights in the main street. What is the expected amount of time you are likely to be stopped at traffic lights if you drive all the way down the main street?

Binomial Questions

PS2 Question 4

In a city the traffic lights on the main road are set up so that traffic lights are green for **55 seconds**, red for **48 seconds**, and amber for **17 seconds** (the city is in California, so no one stops on amber).

(c) What is the chance that you only will need to stop less than twice?

Binomial Questions

PS2 Question 5

A firm states that it has an equal opportunity hiring policy. There are **ten job openings**.

(a) Calculate the probability that three or less women would be hired if on average we expect all applicants to be similarly qualified, and hence the probability that any particular individual is hired is equal to that of any other individual.

(b) What is the chance that exactly an even number of men and women are hired?

(c) A newspaper reporter notes that less women than men were hired. Is this evidence of discrimination? Why or why not?

Binomial Questions

PS2 Question 5

A firm states that it has an equal opportunity hiring policy. There are **ten job openings**.

(a) Calculate the probability that three or less women would be hired if on average we expect all applicants to be similarly qualified, and hence the probability that any particular individual is hired is equal to that of any other individual.

Binomial Questions

PS2 Question 5

A firm states that it has an equal opportunity hiring policy. There are **ten job openings**.

(b) What is the chance that exactly an even number of men and women are hired?

Binomial Questions

PS2 Question 5

A firm states that it has an equal opportunity hiring policy. There are **ten job openings**.

(c) A newspaper reporter notes that less women than men were hired. Is this evidence of discrimination? Why or why not?