# Binomial tests

Jori Mur

# Overview

- Own research
- Counts and Proportions
- Binomial Setting
- Binomial Distributions
- Finding binomial probabilities
- Sign test

- Book: Moore and McCabe, Introduction to the practice of statistics

# Question Answering

Who is the president of France?
When did Marilyn Monroe die?
Where was Adolf Hitler born?
Who is Silvio Berlusconi?

# Patterns

[Person] is the president of [Country]
[Person], the president of [Country]
[Person] was born in [Country]

[Person] died in [Year]

# Answering Questions

| Name | Birth place |
| --- | --- |
| Wim Kok | Bergambacht |
| Gerard Reve | Amsterdam |
| ... | ... |

| Name | Function |
| --- | --- |
| M. Jackson | popstar |
| Elisabeth II | queen of England |
| ... | ... |

# Counts and Proportions (1)

- Random sample of questions which have to be answered $(n)$
- Answers are correct or incorrect
- Count: number of correct answers $(X)$
- Sample proportion: $\hat{p} = \frac{X}{n}$

In my case: $n = 220, X = 125, \hat{p} = \frac{125}{220} = 0.57$

# Sampling distribution

A statistic from a random sample or randomized experiment is a random variable. The sampling distribution of this variable is the distribution of its values for all possible samples.

The probability distribution of the statistic is its sampling distribution

# Population distribution

The population distribution of a variable is the distribution of its values for all members of the population.

The population distribution is also the probability distribution of the variable when we randomly choose one individual from the population.

# Example

- Length of women between ages 18 and 24
- Distribution is normal, mean = 64.5 inches and standard deviation = 2.5 inches.
- Select a woman at random. Her height is $X$.
- Repeated sampling: $N(64.5, 2.5)$.

# Binomial setting

Binomial setting

- There are a fixed number $n$ of observations
- The $n$ observations are all independent
- Each observation falls into one of just two categories.
- The probability of success is the same for each observation

# Binomial distribution for sample counts

The distribution of the count $X$ of successes in the binomial setting is called the binomial distribution with parameters $n$ and $p$. $X$ is $B(n, p)$.

# Counts and proportions (2)

- $X$ is a count. It takes a value between $0$ and $n$. It has a binomial distribution.
- $\hat{p}$ is the sample proportion. It takes a value between $0$ and $1$. It does not have a binomial distribution. To do probability calculations, restate $\hat{p}$ in $X$.

# Recognizing binomial settings

- Tossing a coin 10 times. How many times do we see heads?
- Dealing 10 cards. How many times do we see a red card?
- Answering questions. How many questions are answered correctly?

# Binomial Mean and Standard deviation

The mean and standard deviation of a binomial count $X$ and a sample proportion of successes $\hat{p} = \frac{X}{n}$ are:

$$\mu_X = np \qquad \mu_{\hat{p}} = p$$

$$\sigma_X = \sqrt{np(1-p)} \qquad \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

The sample proportion $\hat{p}$ is unbiased estimator of the population proportion $p$.

# Binomial formula

Example

Each child born to a particular set of parents has probability 0.25 of having blood type O. If these parents have 5 children, what is the probability that exactly 2 of them have type O blood?

# Binomial formula

$n = 5$, $X = 2$, $p = 0.25$.
We want $P(X = 2)$.

P(OOØØØ)=P(O)P(O)P(Ø)P(Ø)P(Ø)

$$= (0.25)(0.25)(0.75)(0.75)(0.75)$$
$$= (0.25)^2(0.75)^3$$

# Binomial formula

$n = 5$, $X = 2$, $p = 0.25$.
We want $P(X = 2)$.

P(OOØØØ)=P(O)P(O)P(Ø)P(Ø)P(Ø)

$$= (0.25)(0.25)(0.75)(0.75)(0.75)$$
$$= (0.25)^2(0.75)^3$$

OOØØØ OØOØØ OØØOØ OØØØO ØOOØØ

ØOØOØ ØOØØO ØØOOØ ØØOØO ØØØOO

# Binomial formula

$n = 5$, $X = 2$, $p = 0.25$.
We want $P(X = 2)$.

P(OOØØØ)=P(O)P(O)P(Ø)P(Ø)P(Ø)

$$= (0.25)(0.25)(0.75)(0.75)(0.75)$$
$$= (0.25)^2(0.75)^3$$

OOØØØ OØOØØ OØØOØ OØØØO ØOOØØ
ØOØOØ ØOØØO ØØOOØ ØØOØO ØØØOO

$P(X = 2) = 10(0.25)^2(0.75)^3 = 0.26$

# Binomial formula

$k$: number of successes
P(OOØØØ)$= (0.25)^2(0.75)^3$

$p^k(1-p)^{n-k}$

$\binom{n}{k} = \frac{n!}{k!(n-k)!}$ $\quad$ $n! = n$x$(n-1)$x$(n-2)$x...x2x1

$P(X=k) = \binom{n}{k}p^k(1-p)^{n-k}$

# Binomial formula on own data

[Pronoun] is the president of [Country]
[Pronoun] was born in [Country]
The [Definite Noun] died in [Year]

# Binomial formula on own data

Remember: The sample proportion $\hat{p}$ is unbiased estimator of the population proportion $p$.
Same set of questions: $n = 220$
Number of correct questions: $X = 131$
For $p$ we take $\hat{p}$: $p = \hat{p} = 0.57$

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

$$P(X = 131) = \binom{220}{131} 0.57^{131} (1 - 0.57)^{220-131} = 0.041$$

The probability to answer 131 question correct having the $B(220, 0.57)$ distribution is 4.1%.

# Paired sign test on own data

Ignore pairs with difference 0; the number of trials $n$ is the count of the remaining pairs. The test statistic is the count $X$ of pairs with a positive difference. $P$-values for $X$ are based on the binomial distribution $B(n, 1/2)$.

simple pattens: 125 correct
Added coreference patterns: 131 correct
8 differences, 7 improved, 1 did more poorly.

Is this evidence for an improved result?

# Paired sign test on own data

$H_0 : p = 0.5$ no effect
$H_a : p > 0.5$ positive effect

$$\sum_{k=X}^{n} \binom{n}{k} p^k (1-p)^{n-k}$$

$$\binom{8}{7} 0.5^7 (1-0.5)^{8-7} + \binom{8}{8} 0.5^8 (1-0.5)^{8-8} =$$
$$8 \text{x} 0.5^7 \text{x} 0.5 + (0.5)^8 = 9(0.5)^8 = 0.035$$

The probability to get this result assuming that there was no difference in the performance of the system is 3.5%. That means we can reject $H_0$.