Statistics

CIS008-2 Logic and Foundations of Mathematics

David Goodwin

david.goodwin@perisic.com

University of Bedfordshire

11:00, Tuesday 20th March 2012

Outline

Outline

1. Statistical distributions
2. Hypothesis Testing
3. Statistical Tests

Binomial Distribution

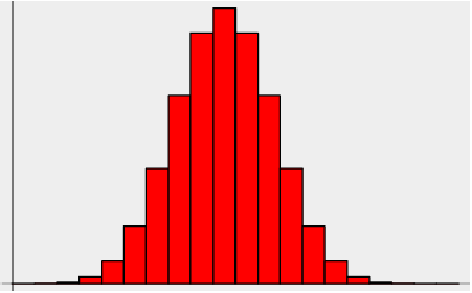
- If there are two possible outcomes of an event (e.g. success or failure), and the possibilities of the outcome are independent and constant, the distribution of probabilities is called a binomial distribution.
- Examples of binomial distribution are probabilities regarding the number of heads or tails when tossing a coin, or the probabilities regarding red or black when cutting and reshuffling a deck of cards.
- The binomial distribution gives the discrete probability distribution $P_p(n|N)$ of obtaining exactly n successes out of N trials (where the result of each trial is true with probability p and false with probability $q = 1 - p$).

Binomial Distribution

- The binomial distribution is therefore given by

$$
\begin{aligned}
P_p(n|N) &= \binom{N}{n} p^n q^{(N-n)} \\
&= \frac{N!}{n!(N-n)!} p^n (1-p)^{(N-n)}
\end{aligned}
$$

where $\binom{N}{n}$ is a binomial coefficient.



University of Bedfordshire

Normal Distribution

- A normal distribution in a variate X with mean μ and variance σ^2 is a statistic distribution with probability density function

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-mu)^2}{2\sigma^2}}$$

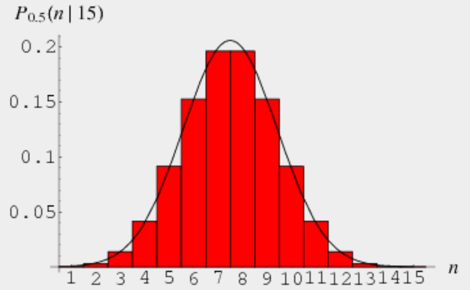
  on the domain $x \in (-\infty, \infty)$.

- While statisticians and mathematicians uniformly use the term "normal distribution" for this distribution, physicists sometimes call it a Gaussian distribution and, because of its curved flaring shape, social scientists refer to it as the "bell curve".

University of Bedfordshire

Normal and Binomial Distributions

- de Moivre developed the normal distribution as an approximation to the binomial distribution, and it was subsequently used by Laplace in 1783 to study measurement errors and by Gauss in 1809 in the analysis of astronomical data.



Example - Computer chip failure

Example

The probability that a given computer chip will fail is 0.02. Find the probability that exactly one of 10 given computer chips will fail, then find the probability that exactly 2 of 10 given computer chips will fail. What is the probability that all 10 will fail?

Example - Computer chip failure

Example

The probability that a given computer chip will fail is 0.02. Find the probability that exactly one of 10 given computer chips will fail, then find the probability that exactly 2 of 10 given computer chips will fail. What is the probability that all 10 will fail?

- The Binomial probability must be used, $P(n \mid N) = \frac{N!}{n!(N-n)!} p^n (1-p)^{(N-n)}$.

Example - Computer chip failure

Example

The probability that a given computer chip will fail is 0.02. Find the probability that exactly one of 10 given computer chips will fail, then find the probability that exactly 2 of 10 given computer chips will fail. What is the probability that all 10 will fail?

- The Binomial probability must be used, $P(n \mid N) = \frac{N!}{n!(N-n)!} p^n (1-p)^{(N-n)}$.
- $P(1 \mid 10) = \frac{10!}{1!(10-1)!} (0.02)^1 (1-0.02)^{(10-1)}$

Example - Computer chip failure

Example

The probability that a given computer chip will fail is 0.02. Find the probability that exactly one of 10 given computer chips will fail, then find the probability that exactly 2 of 10 given computer chips will fail. What is the probability that all 10 will fail?

- The Binomial probability must be used, $P(n \mid N) = \frac{N!}{n!(N-n)!} p^n (1-p)^{(N-n)}$.
- $P(1 \mid 10) = \frac{10!}{1!(10-1)!}(0.02)^1(1-0.02)^{(10-1)}$
- $P(1 \mid 10) = \frac{10!}{1!9!}(0.02)(0.98)^9 = 10 \times 0.02 \times (0.98)^9 = 0.167$ (3 s.f.)

Example - Computer chip failure

Example

The probability that a given computer chip will fail is 0.02. Find the probability that exactly one of 10 given computer chips will fail, then find the probability that exactly 2 of 10 given computer chips will fail. What is the probability that all 10 will fail?

- The Binomial probability must be used, $P(n \mid N) = \frac{N!}{n!(N-n)!} p^n (1-p)^{(N-n)}$.
- $P(1 \mid 10) = \frac{10!}{1!(10-1)!}(0.02)^1(1-0.02)^{(10-1)}$
- $P(1 \mid 10) = \frac{10!}{1!9!}(0.02)(0.98)^9 = 10 \times 0.02 \times (0.98)^9 = 0.167$ (3 s.f.)
- $P(2 \mid 10) = \frac{10!}{2!(10-2)!}(0.02)^2(1-0.02)^{(10-2)}$

Example - Computer chip failure

Example

The probability that a given computer chip will fail is 0.02. Find the probability that exactly one of 10 given computer chips will fail, then find the probability that exactly 2 of 10 given computer chips will fail. What is the probability that all 10 will fail?

- The Binomial probability must be used, $P(n \mid N) = \frac{N!}{n!(N-n)!} p^n (1-p)^{(N-n)}$.
- $P(1 \mid 10) = \frac{10!}{1!(10-1)!}(0.02)^1(1-0.02)^{(10-1)}$
- $P(1 \mid 10) = \frac{10!}{1!9!}(0.02)(0.98)^9 = 10 \times 0.02 \times (0.98)^9 = 0.167$ (3 s.f.)
- $P(2 \mid 10) = \frac{10!}{2!(10-2)!}(0.02)^2(1-0.02)^{(10-2)}$
- $P(2 \mid 10) = \frac{10!}{2!8!}(0.02)^2(0.98)^8 = \frac{10 \times 9}{2} \times 0.0004 \times (0.98)^8 = 0.0153$ (3 s.f.)

Example - Computer chip failure

Example

The probability that a given computer chip will fail is 0.02. Find the probability that exactly one of 10 given computer chips will fail, then find the probability that exactly 2 of 10 given computer chips will fail. What is the probability that all 10 will fail?

- The Binomial probability must be used, $P(n \mid N) = \frac{N!}{n!(N-n)!} p^n (1-p)^{(N-n)}$.
- $P(1 \mid 10) = \frac{10!}{1!(10-1)!}(0.02)^1(1-0.02)^{(10-1)}$
- $P(1 \mid 10) = \frac{10!}{1!9!}(0.02)(0.98)^9 = 10 \times 0.02 \times (0.98)^9 = 0.167$ (3 s.f.)
- $P(2 \mid 10) = \frac{10!}{2!(10-2)!}(0.02)^2(1-0.02)^{(10-2)}$
- $P(2 \mid 10) = \frac{10!}{2!8!}(0.02)^2(0.98)^8 = \frac{10 \times 9}{2} \times 0.0004 \times (0.98)^8 = 0.0153$ (3 s.f.)
- $P(10 \mid 10) = \frac{10!}{10!(10-10)!}(0.02)^1 0(1-0.02)^{(10-10)}$

Example - Computer chip failure

Example

The probability that a given computer chip will fail is 0.02. Find the probability that exactly one of 10 given computer chips will fail, then find the probability that exactly 2 of 10 given computer chips will fail. What is the probability that all 10 will fail?

- The Binomial probability must be used, $P(n \mid N) = \frac{N!}{n!(N-n)!} p^n (1-p)^{(N-n)}$.
- $P(1 \mid 10) = \frac{10!}{1!(10-1)!}(0.02)^1(1-0.02)^{(10-1)}$
- $P(1 \mid 10) = \frac{10!}{1!9!}(0.02)(0.98)^9 = 10 \times 0.02 \times (0.98)^9 = 0.167$ (3 s.f.)
- $P(2 \mid 10) = \frac{10!}{2!(10-2)!}(0.02)^2(1-0.02)^{(10-2)}$
- $P(2 \mid 10) = \frac{10!}{2!8!}(0.02)^2(0.98)^8 = \frac{10 \times 9}{2} \times 0.0004 \times (0.98)^8 = 0.0153$ (3 s.f.)
- $P(10 \mid 10) = \frac{10!}{10!(10-10)!}(0.02)^1 0(1-0.02)^{(10-10)}$
- $P(10 \mid 10) = \frac{10!}{10!0!}(0.02)^1 0(0.98)^0 = 0.02^1 0 = 1.024 \times 10^{-17}$

Example - Capacity of a lift

Example

Suppose you must establish regulations concerning the maximum number of people who can occupy a lift. You know that the total weight of 8 people chosen at random follows a normal distibution with a mean of 550kg and standard deviation of 150kg. What is the probability that the total weight of 8 people exceeds 600kg?

Example - Capacity of a lift

Example

Suppose you must establish regulations concerning the maximum number of people who can occupy a lift. You know that the total weight of 8 people chosen at random follows a normal distibution with a mean of 550kg and standard deviation of 150kg. What is the probability that the total weight of 8 people exceeds 600kg?

- The normal probability must be used, $P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-mu)^2}{2\sigma^2}}$.

Example - Capacity of a lift

Example

Suppose you must establish regulations concerning the maximum number of people who can occupy a lift. You know that the total weight of 8 people chosen at random follows a normal distibution with a mean of 550kg and standard deviation of 150kg. What is the probability that the total weight of 8 people exceeds 600kg?

- The normal probability must be used, $P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-mu)^2}{2\sigma^2}}$.
- $P(600) = \frac{1}{150\sqrt{2\pi}} e^{-\frac{(600-550)^2}{2\times 150^2}}$.

University of Bedfordshire

EXAMPLE - CAPACITY OF A LIFT

EXAMPLE

Suppose you must establish regulations concerning the maximum number of people who can occupy a lift. You know that the total weight of 8 people chosen at random follows a normal distibution with a mean of 550kg and standard deviation of 150kg. What is the probability that the total weight of 8 people exceeds 600kg?

- The normal probability must be used, $P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-mu)^2}{2\sigma^2}}$.
- $P(600) = \frac{1}{150\sqrt{2\pi}} e^{-\frac{(600-550)^2}{2\times 150^2}}$.
- $P(600) = 0.00270 e^{-\frac{(50)^2}{2\times 150^2}}$.

Example - Capacity of a lift

Example

Suppose you must establish regulations concerning the maximum number of people who can occupy a lift. You know that the total weight of 8 people chosen at random follows a normal distibution with a mean of 550kg and standard deviation of 150kg. What is the probability that the total weight of 8 people exceeds 600kg?

- The normal probability must be used, $P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-mu)^2}{2\sigma^2}}$.
- $P(600) = \frac{1}{150\sqrt{2\pi}} e^{-\frac{(600-550)^2}{2\times 150^2}}$.
- $P(600) = 0.00270e^{-\frac{(50)^2}{2\times 150^2}}$.
- $P(600) = 0.00270e^{-\frac{2500}{45000}} = 0.00270e^{-0.0556} = 0.00255$ (3 s.f.). This is the probability that the total weight of 8 people is exactly 600 kg.

Example - Capacity of a lift

Example

Suppose you must establish regulations concerning the maximum number of people who can occupy a lift. You know that the total weight of 8 people chosen at random follows a normal distibution with a mean of 550kg and standard deviation of 150kg. What is the probability that the total weight of 8 people exceeds 600kg?

- The normal probability must be used, $P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-mu)^2}{2\sigma^2}}$.
- $P(600) = \frac{1}{150\sqrt{2\pi}} e^{-\frac{(600-550)^2}{2\times 150^2}}$.
- $P(600) = 0.00270e^{-\frac{(50)^2}{2\times 150^2}}$.
- $P(600) = 0.00270e^{-\frac{2500}{45000}} = 0.00270e^{-0.0556} = 0.00255$ (3 s.f.). This is the probability that the total weight of 8 people is exactly 600 kg.
- We require all eventualities greater than 600kg.

Example - Capacity of a lift

Example

Suppose you must establish regulations concerning the maximum number of people who can occupy a lift. You know that the total weight of 8 people chosen at random follows a normal distibution with a mean of 550kg and standard deviation of 150kg. What is the probability that the total weight of 8 people exceeds 600kg?

- The normal probability must be used, $P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-mu)^2}{2\sigma^2}}$.
- $P(600) = \frac{1}{150\sqrt{2\pi}} e^{-\frac{(600-550)^2}{2\times 150^2}}$.
- $P(600) = 0.00270 e^{-\frac{(50)^2}{2\times 150^2}}$.
- $P(600) = 0.00270 e^{-\frac{2500}{45000}} = 0.00270 e^{-0.0556} = 0.00255$ (3 s.f.). This is the probability that the total weight of 8 people is exactly 600 kg.
- We require all eventualities greater than 600kg.
- To do this we need the area under the curve that is greater than 600kg.

University of Bedfordshire

Example - Capacity of a lift

Example

Suppose you must establish regulations concerning the maximum number of people who can occupy a lift. You know that the total weight of 8 people chosen at random follows a normal distibution with a mean of 550kg and standard deviation of 150kg. What is the probability that the total weight of 8 people exceeds 600kg?

- The normal probability must be used, $P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-mu)^2}{2\sigma^2}}$.
- $P(600) = \frac{1}{150\sqrt{2\pi}} e^{-\frac{(600-550)^2}{2\times150^2}}$.
- $P(600) = 0.00270 e^{-\frac{(50)^2}{2\times150^2}}$.
- $P(600) = 0.00270 e^{-\frac{2500}{45000}} = 0.00270 e^{-0.0556} = 0.00255$ (3 s.f.). This is the probability that the total weight of 8 people is exactly 600 kg.
- We require all eventualities greater than 600kg.
- To do this we need the area under the curve that is greater than 600kg.
- We can either integrate our function between these limits, or use a normal distribution look-up table.

Example - Capacity of a lift

- can be found in public domain normal distribution tables:

z	.00	.01	.02	.03	.04	.05
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088

Example - Capacity of a lift

- can be found in public domain normal distribution tables:

z	.00	.01	.02	.03	.04	.05
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088

- tables are defined by a variable $z = \frac{x-\mu}{\sigma}$.

Example - Capacity of a lift

- can be found in public domain normal distribution tables:

z	.00	.01	.02	.03	.04	.05
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088

- tables are defined by a variable $z = \frac{x-\mu}{\sigma}$.
- we have $z = \frac{1}{3}$ so we round down to $z = 0.3$ being the first digit of the z-value, and this will be represented as rows of the table. The second digit of the z-value is $z = 0.03$ which is represented as columns in a normal distribution table.

Example - Capacity of a lift

- can be found in public domain normal distribution tables:

z	.00	.01	.02	.03	.04	.05
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088

- tables are defined by a variable $z = \frac{x-\mu}{\sigma}$.
- we have $z = \frac{1}{3}$ so we round down to $z = 0.3$ being the first digit of the z-value, and this will be represented as rows of the table. The second digit of the z-value is $z = 0.03$ which is represented as columns in a normal distribution table.
- We simply read off the value in the 0.3-row and 0.03-column, being 0.6293.

University of Bedfordshire

Example - Capacity of a lift

- can be found in public domain normal distribution tables:

z	.00	.01	.02	.03	.04	.05
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088

- tables are defined by a variable $z = \frac{x-\mu}{\sigma}$.
- we have $z = \frac{1}{3}$ so we round down to $z = 0.3$ being the first digit of the z-value, and this will be represented as rows of the table. The second digit of the z-value is $z = 0.03$ which is represented as columns in a normal distribution table.
- We simply read off the value in the 0.3-row and 0.03-column, being 0.6293.
- This is the value for the probability that 8 people are *less than* 600kg, and we require *more than* 600kg.

Example - Capacity of a lift

- can be found in public domain normal distribution tables:

z	.00	.01	.02	.03	.04	.05
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088

- tables are defined by a variable $z = \frac{x-\mu}{\sigma}$.
- we have $z = \frac{1}{3}$ so we round down to $z = 0.3$ being the first digit of the z-value, and this will be represented as rows of the table. The second digit of the z-value is $z = 0.03$ which is represented as columns in a normal distribution table.
- We simply read off the value in the 0.3-row and 0.03-column, being 0.6293.
- This is the value for the probability that 8 people are *less than* 600kg, and we require *more than* 600kg.
- We simple use $P(600) = 1 - 0.6293 = 0.3707$

University of Bedfordshire

Outline

University of Bedfordshire

Hypothesis

- A hypothesis is a proposition that is consistent with known data, but has been neither verified nor shown to be false.
- In statistics, a hypothesis (sometimes called a statistical hypothesis) refers to a statement on which hypothesis testing will be based. Particularly important statistical hypotheses include the null hypothesis and alternative hypothesis.
- In symbolic logic, a hypothesis is the first part of an implication (with the second part being known as the predicate).
- In general mathematical usage, "hypothesis" is roughly synonymous with "conjecture".

Hypothesis Testing

Hypothesis testing is the use of statistics to determine the probability that a given hypothesis is true. The usual process of hypothesis testing consists of four steps.

1. Formulate the null hypothesis H_0 (commonly, that the observations are the result of pure chance) and the alternative hypothesis H_a (commonly, that the observations show a real effect combined with a component of chance variation).
2. Identify a test statistic that can be used to assess the truth of the null hypothesis.
3. Compute the P-value, which is the probability that a test statistic at least as significant as the one observed would be obtained assuming that the null hypothesis were true. The smaller the P-value, the stronger the evidence against the null hypothesis.
4. Compare the p-value to an acceptable significance value alpha (sometimes called an alpha value). If $p \leq \alpha$, that the observed effect is statistically significant, the null hypothesis is ruled out, and the alternative hypothesis is valid.

Definitions

Null Hypothesis A null hypothesis is a statistical hypothesis that is tested for possible rejection under the assumption that it is true (usually that observations are the result of chance). The hypothesis contrary to the null hypothesis, usually that the observations are the result of a real effect, is known as the alternative hypothesis.

Alternative Hypothesis The alternative hypothesis is the hypothesis used in hypothesis testing that is contrary to the null hypothesis. It is usually taken to be that the observations are the result of a real effect (with some amount of chance variation superposed).

Definitions

P-Value The probability that a variate would assume a value greater than or equal to the observed value strictly by chance: $P(z \geq z_{(\text{observed})})$.

Alpha Value An alpha value is a number $0 \leq \alpha \leq 1$ such that $P(z \geq z_{(\text{observed})}) \leq \alpha$ is considered "significant", where P is a P-value.

Type I Error An error in a statistical test which occurs when a false hypothesis is accepted (a false positive in terms of the null hypothesis)

Type II Error An error in a statistical test which occurs when a true hypothesis is rejected (a false negative in terms of the null hypothesis).

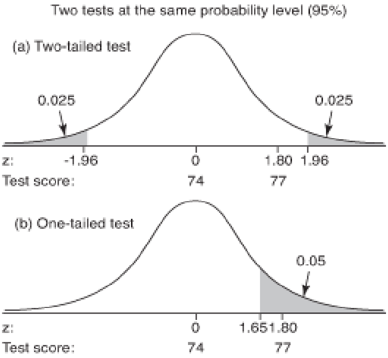
One- and Two-Tailed tests

- When the hypothesis states the direction of the difference or relationship, then use a one-tailed probability.
- For example, a one-tailed test would be used to test these null hypotheses:
  1. Females will not score significantly higher than males on an IQ test.
  2. Blue collar workers are will not buy significantly more product than white collar workers.
  3. Superman is not significantly stronger than the average person.
- In each case, the null hypothesis (indirectly) predicts the direction of the difference.

Two-Tailed tests

- A two-tailed test would be used to test these null hypotheses:
  1. There will be no significant difference in IQ scores between males and females.
  2. There will be no significant difference in the amount of product purchased between blue collar and white collar workers.
  3. There is no significant difference in strength between Superman and the average person.
- The one-tailed probability is exactly half the value of the two-tailed probability.

One- and Two-Tailed tests



Hypothesis testing - Example

Example

Candidate Jones is one of two candidates running for mayor of Central City. A random polling of 672 registered voters finds that 323 (48% of those polled) will vote for him. It is reasonable to assume that the race for mayor is one of either of the two outcomes being more or less equally likely? (That is, if p is the proportion of the population who will vote for Jones, is it reasonable to assume that p = 0.5?)

- The population involved is the registered voters in Central City. As worded, this is a 2-tail situation.
- We are testing the null hypothesis $H_0 : p = 0.5$ against the alternate hypothesis $H_a : p \neq 0.5$.
- This is a binomial distribution with $N = 672$ and $p = 0.5$.

Hypothesis testing - Example

- Let $p(hat)$ is the proportion of registered voters in a sample of size 672 who will vote for Jones, then if Ho is true, the distribution of p(hat) is approximately normal, the mean of the distribution of p(hat) is 0.5 and the standard deviation of p(hat) is $\sqrt{(0.5)(0.5)/672} = 0.019287$, or about 1.93%.
- The mean of our specific sample is 0.48. If H_0 is true, the associated z score is $(0.475 - 0.5)/0.0193 = -1.3$. At the 2-tail, 5% level of significance, the critical z scores are $z > 1.96$ or $z < -1.96$. There is not enough evidence to reject H_0 at the 5% level, although if a significance level of 1% were used, then H_0 would have to be considered in more detail.

Hypothesis testing - Example

Example

I roll a single die 1000 times and obtain a "6" on 204 rolls. Is there significant evidence to suggest that the die is not fair?

- The population involved consists of the proportion of 6's obtained when a die is rolled 1000 times.
- This population consists of 1001 proportions: $0/1000, 1/1000, 2/1000, \ldots, 999/1000, 1000/1000$. If p is the proportion of 6's obtained, then our null and alternate hypotheses are, respectively $H_0 : p = 1/6$ $H_a : p \neq 1/6$
- This is a 2-tail situation. If H_0 is true, we have a binomial setting with $p = 1/6$ and $N = 1000$.

University of Bedfordshire

Hypothesis testing - Examples using normal distributions

- If p(hat) is a sample proportion of 6's, then the distribution of p(hat is approximately normal, the mean of the distribution of p(hat) is 1/6 and the standard deviation is $\sqrt{(1/6)(5/6)/1000} = 0.011785$, or approximately 0.012.
- Our sample proportion is $p(hat) = 204/1000 = 0.204$, or 20.4%. If H_0 is true, the probability of obtaining this result is 0.00107, or approximately (1/10)%.
- There is strong evidence to reject H_0, since it is highly unlikely that this result would be obtained if H_0 is true.

Outline

Statistical Test

A statistical test is one used to determine the statistical significance of an observation. Two main types of error can occur:

1. A type I error occurs when a false negative result is obtained in terms of the null hypothesis by obtaining a false positive measurement.
2. A type II error occurs when a false positive result is obtained in terms of the null hypothesis by obtaining a false negative measurement.

The probability that a statistical test will be positive for a true statistic is sometimes called the test's sensitivity, and the probability that a test will be negative for a negative statistic is sometimes called the specificity. The following table summarizes the names given to the various combinations of the actual state of affairs and observed test results.

true positive result	sensitivity
false negative result	1-sensitivity
true negative result	specificity
false positive result	1-specificity

Paired t-Test

- Given two paired sets X_i and Y_i of n measured values, the paired t-test determines whether they differ from each other in a significant way under the assumptions that the paired differences are independent and identically normally distributed.
- To apply the test, let

$$\hat{X}_i = (X_i - \bar{X})$$
$$\hat{Y}_i = (Y_i - \bar{Y})$$

be the standard deviations

- Then define the test t by

$$t = (\bar{X} - \bar{Y})\sqrt{\frac{n(n-1)}{\sum_{i=1}^{n}(\hat{X}_i - \hat{Y}_i)^2}}$$

Significance

- Let $\delta \equiv z \geq z_{(\text{observed})}$. A value $0 \leq \alpha \leq 1$ such that $P(\delta) \leq \alpha$ is considered "significant" (i.e., is not simply due to chance) is known as an alpha value. The probability that a variate would assume a value greater than or equal to the observed value strictly by chance, $P(\delta)$, is known as a P-value.
- Depending on the type of data and conventional practices of a given field of study, a variety of different alpha values may be used. One commonly used terminology takes $P(\delta) \geq 5\%$ as "not significant", $1\% < P(\delta) < 5\%$, as "significant", and $P(\delta) < 1\%$ as "highly significant".
- A significance test is a test for determining the probability that a given result could not have occurred by chance (its significance).

Coincidence

- A confidence interval is an interval in which a measurement or trial falls corresponding to a given probability.
- Usually, the confidence interval of interest is symmetrically placed around the mean, so a 50% confidence interval for a symmetric probability density function would be the interval $[-a, a]$ such that

$$\frac{1}{2} = \int_{-a}^{a} P(x)dx$$

- For a normal distribution, the probability that a measurement falls within n standard deviations ($n\sigma$) of the mean μ is given by

$$P(\mu - n\sigma < x < \mu + n\sigma) = 1/(\sigma\sqrt{2\pi}) \int_{\mu - n\sigma}^{\mu + n\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$
$$= \frac{2}{(\sigma\sqrt{2\pi}} \int_{\mu}^{\mu + n\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

COINCIDENCE

- Now let $u = \frac{x-\mu}{\sqrt{2}\sigma}$, so $du = \frac{dx}{\sqrt{2}\sigma}$. Then

$$P(\mu - n\sigma < x < \mu + n\sigma) = \frac{2}{\sigma\sqrt{2\pi}}\sqrt{2}\sigma \int_0^{\frac{n}{\sqrt{2}}} e^{-u^2} du = \frac{2}{\sqrt{\pi}} \int_0^{\frac{n}{\sqrt{2}}} e^{-u^2} du = erf\left(\frac{n}{\sqrt{2}}\right)$$

  where $erf(x)$ is the so-called errror function.

- Conversely, to find the probability-P confidence interval centered about the mean for a normal distribution in units of σ,

$$n = \sqrt{2} erf^{-1}(P)$$

- $erf^{-1}(x)$ is the inverse error function. $erf(z)$ is the "error function" encountered in integrating the normal distribution (which is a normalized form of the Gaussian function). It is an entire function defined by

$$erf(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt$$

University of Bedfordshire

Correlations and correlation coefficients

- The expectation value of the product of two random variables is called the correlation of the two variables.
- Item the correlation is a measure of how correlated two variables are.
- For a measure of how mutually dependent two variables are we divide the correlation by the square root of the product of the variances

$$C_{XY} \equiv \frac{\langle XY \rangle}{\sqrt{V(X)V(Y)}}$$

where C_{XY} is called the correlation coefficient of X and Y.

- If the means of X and Y are not zero, we can remove these when calculating the correlation coefficient and preserve its properties, we can find in general the correlation coefficient as

$$C_{XY} \equiv \frac{\langle (X - \langle X \rangle)(Y - \langle Y \rangle) \rangle}{\sqrt{V(X)V(Y)}} = \frac{\langle XY \rangle - \langle X \rangle \langle Y \rangle}{\sqrt{V(X)V(Y)}}$$

Paired t test - Example

- Both variables should be normally distributed.
- Hypothesis: H_0 There is no significant difference between the means of the two variables. H_a There is a significant difference between the means of the two variables.
- Following is sample output of a paired samples T test. We compared the mean test scores before (pre-test) and after (post-test) the subjects completed a test preparation course. We want to see if our test preparation course improved people's score on the test.
- First, we see the descriptive statistics for both variables.

		μ	N	σ	std. error mean
Pair	Pre-test scores	2.433	12	0.261	0.07521
1	Post-test scores	2.533	12	0.281	0.08103

where we have used the standard error in the mean, defined here as $\sigma/\sqrt{N}$.

- The post-test mean scores are higher.

University of Bedfordshire

Paired t test - Example

- we see the correlation between the two variables:

		N	correlation	Significance
Pair 1	Pre-test scores & Post-test scores	12	0.829	0.001

- There is a strong positive correlation. People who did well on the pre-test also did well on the post-test.

Paired t test - Example

- We then find the paired differences of the two. The difference of the mean is easy to find, and now we have the correlation, we can find the standard deviation.

		μ	σ	std. error mean	95% confidence interval
Pair 1	Pre-test scores - Post-test scores	-0.1	0.16	0.04608	-2.01 to 0.00137

- The T value is calculated to be -2.171, we have 11 degrees of freedom and our significance is 0.053.
- We see that the significance value is approaching significance, but it is not a significant difference. There is no difference between pre- and post-test scores. Our test preparation course did not help!

T-test - Example

Usually, social researchers use the T-Test to test for significant differences between means observed for two independent groups, such as Democrats v. Republicans, or men v. women, etc. These groups are independent in the sense that cases in one group are not matched with cases in the other group. But occasionally, researchers will want to determine whether there is a significant change in the scores of the same cases on the same variables over time. In this instance, the standard T-Test for Independent Samples does not apply.

Example

Consider the situation of voting turnout in American states in elections since 1980, % turnout in presidential elections by states

Election	N	Mean	Std. Deviation
1980	51	55.7	7.3
1984	51	54.6	6.5
1988	51	52.1	6.4
1992	51	57.6	7.4
1996	51	48.9	7.4
2000	51	53.8	6.9

T-test - Example

Different stories could be told for each of these elections, but let's concentrate on the turnout data for 1980 and 1984. The 1980 election, between incumbent president Jimmy Carter and challenger Ronald Reagan also had a third candidate, John Anderson, who had been a Republican party leader but who ran as an Independent.
Reagan was elected in 1980 and ran for re-election in 1984 against Walter Mondale, but there was no third party candidate in 1984. The mean voting turnout in 1984, when calculated across all the states, dropped compared with 1980. Some say that turnout dropped because there was no third party candidate. Others say that the observed differene between means of only 1.1% points (55.7-54.6) could have been attributable to chance.

T-test - Example

Perhaps we could do a standard T-Test to check this out. Let's compute the T-Test for Independent Samples:

$$t = (\bar{X} - \bar{Y})\sqrt{\frac{n(n-1)}{\sum_{i=1}^{n}(\hat{X}_i - \hat{Y}_i)^2}}$$

$$= \frac{\bar{X} - \bar{Y}}{\sigma_{(\bar{X}-\bar{Y})}} = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_X^2}{n_X} + \frac{s_Y^2}{n_Y}}} = \frac{55.7 - 54.6}{\sqrt{\frac{7.3}{\sqrt{51}} + \frac{6.5}{\sqrt{51}}}} = 0.78$$

Treating the 51 states as "Independent" in computing the T-Test produces a test statistic (t) less than one. A test statistic this small falls far short of significance at the customary 0.05 significance level, so it suggests that the observed difference in states' voting turnout between 1980 and 1984 is unlikely to have occurred by chance.

But suppose that some systematic process was going on. Suppose that most states tended to demonstrate a slightly lower voting turnout rate between 1980 and 1984, perhaps dropping by about one point. Perhaps this systematic change is lost by only calculating the means for each year.

T-test - Example

In fact, because the states are matched on repeated measures, we must use the Paired Samples T-Test, which produces this very different result:

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	% turnout in 1980 election	55.739	51	7.295	1.022
	% turnout in 1984 election	54.612	51	6.541	0.916

Paired Samples Test: % turnout in 1980 election - % turnout in 1984 election

Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval	t	df	Sig. (2-tailed)
1.127	2.522	0.353	0.418 to 1.837	3.193	50	0.002

When states are matched on their turnout levels in 1980 and 1984, a paired samples T Test shows that the changes from one year to the next were small, but systematic. Such systematic shifts downward in voting turnout would have occurrred fewer than 1 time out of 100, if the shifts were purely due to change variation.