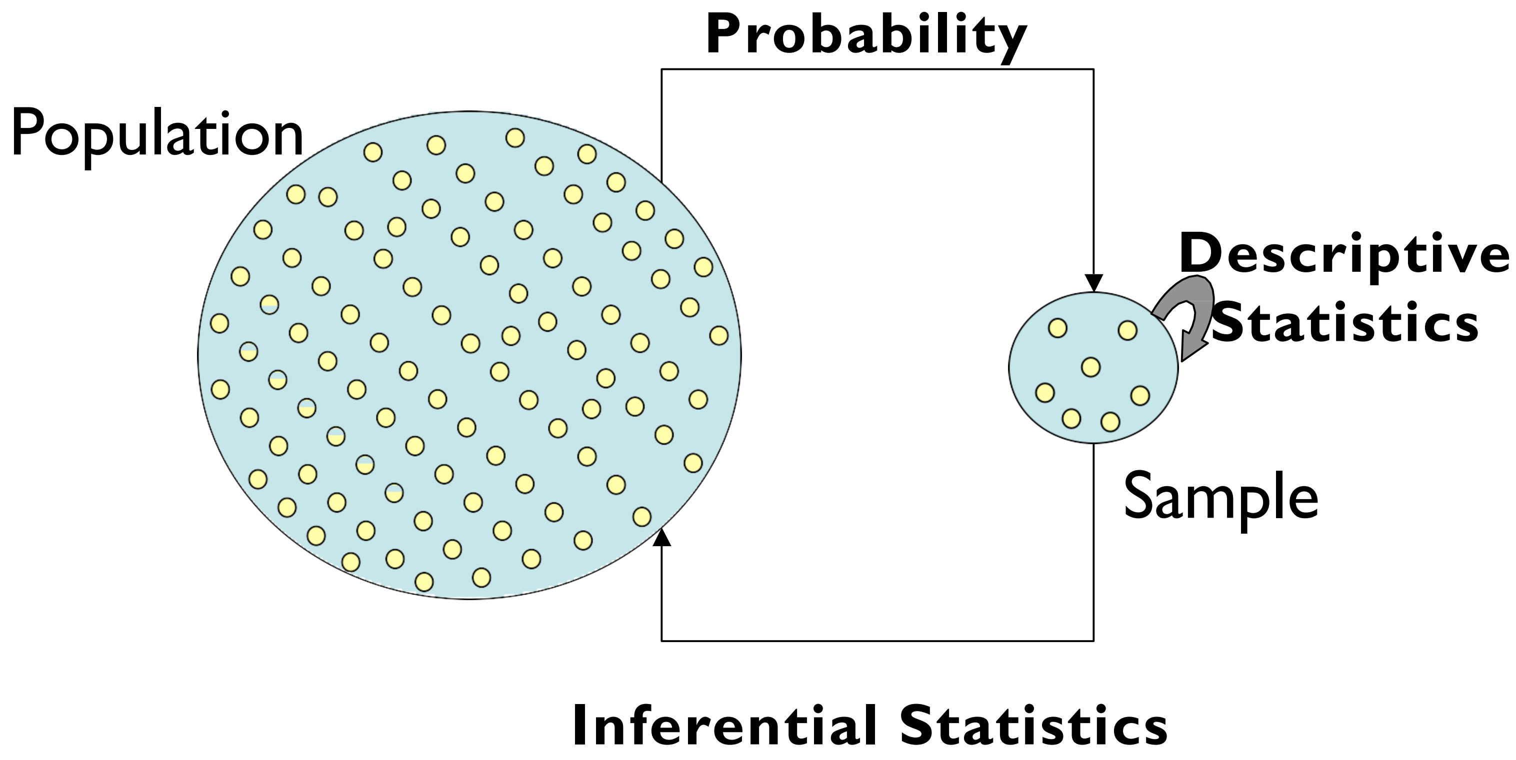
Lecture 7: Estimation and Statistical Inference

Descriptive vs. Inferential Statistics

- **Descriptive:** e.g., Median; describes data you have but can't be generalized beyond that

- **Inferential:** e.g., t-test, that enable inferences about the population beyond our data

“Central Dogma” of Statistics



Hypothesis

- In general, many scientific investigations start by expressing a hypothesis.
- For example, Mackowiak et al (1992) hypothesized that the average normal (i.e., for healthy people) body temperature is less than the widely accepted value of $98.6F$.
- If we denote the population mean of normal body temperature as μ, then we can express this hypothesis as $\mu < 98.6$.

Null and alternative hypotheses

- The null hypothesis usually reflects the "status quo" or "nothing of interest".

- In contrast, we refer to our hypothesis (i.e., the hypothesis we are investigating through a scientific study) as the **alternative hypothesis** and denote it as H_A.

- For hypothesis testing, we focus on the null hypothesis since it tends to be simpler.

The Null and Alternative Hypothesis

The null hypothesis, H_0:

- States the assumption (numerical) to be tested
- Begin with the assumption that the null hypothesis is TRUE
- Always contains the '=' sign

The alternative hypothesis, H_a:

- Is the opposite of the null hypothesis
- Challenges the status quo
- Never contains just the '=' sign
- Is generally the hypothesis that is believed to be true by the researcher

Null and alternative hypotheses

- Consider the body temperature example, where we want to examine the null hypothesis $H_0 : \mu = 98.6$ against the alternative hypothesis $H_A : \mu < 98.6$.
- To start, suppose that $\sigma^2 = 1$ is known.
- Further, suppose that we have randomly selected a sample of 25 healthy people from the population and measured their body temperature.

Hypothesis testing for the population mean

- To evaluate hypotheses regarding the population mean, we use the sample mean $\bar{X}$ as the test statistic.

$$\bar{X} \sim N(\mu, \sigma^2/n)$$

- For the above example,

$$\bar{X} \sim N(\mu, 1/25)$$

- If the null hypothesis is true, then

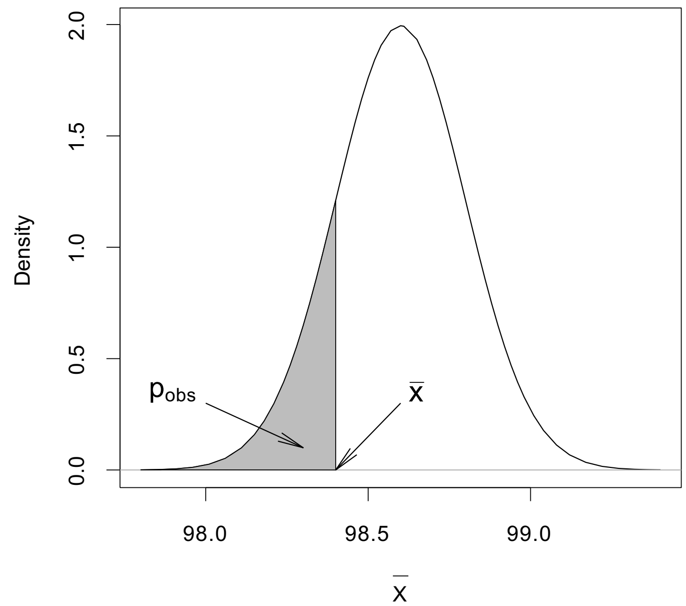
$$\bar{X} \sim N(98.6, 1/25)$$

Hypothesis testing for the population mean

- In reality, we have one value, $\bar{x}$, for the sample mean.
- We can use this value to quantify the evidence of departure from the null hypothesis.
- Suppose that from our sample of 25 people we find that the sample mean is $\bar{x} = 98.4$.

Hypothesis testing for the population mean

- To evaluate the null hypothesis $H_0 : \mu = 98.6$ versus the alternative $H_A : \mu < 98.6$, we use the lower tail probability of this value from the null distribution.



Observed significance level

- The *observed significance level* for a test is the probability of values as or more extreme than the observed value, based on the null distribution in the direction supporting the alternative hypothesis.

- This probability is also called the p-**value** and denoted p_{obs}.

- For the above example,

$$p_{\text{obs}} = P(\bar{X} \leq \bar{x} \mid H_0)$$

p-value

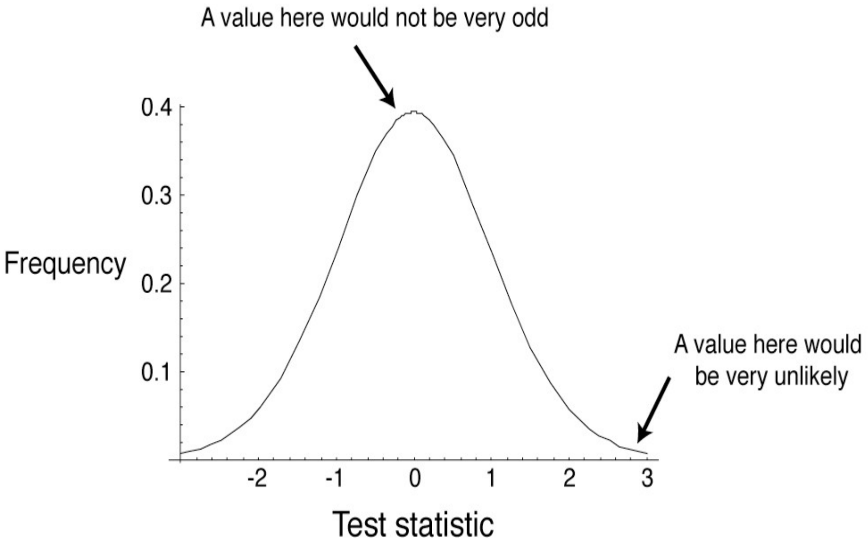
- A P-value is the probability of getting the data, or something as or more unusual, if the null hypothesis were true.

- When the *p*-value is small, say 0.05 for example, it is rare to find values as extreme as what we have observed.

- As the *p*-value increases, it indicates that there is a good chance to find more extreme values (for the test statistic) than what has been observed.

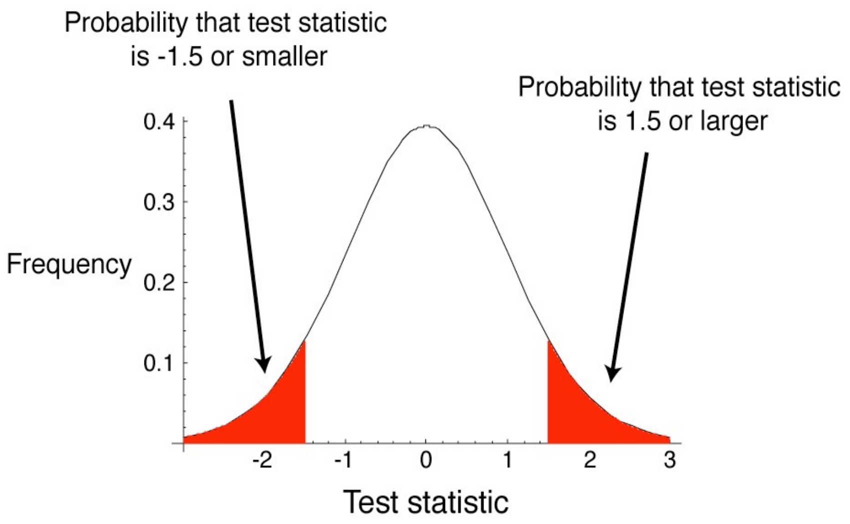
- Then, we would be more reluctant to reject the null hypothesis.
- A common **mistake** is to regard the *p*-value as the probability of null given the observed test statistic: $P(H_0|\bar{x})$.

A null hypothesis is specific;
an alternate hypothesis is not.



A test statistic summarizes the match between the data and the null hypothesis

P-value



Interpretation of *p*-value

- Calculate a test statistic in the sample data that is relevant to the hypothesis being tested
- After calculating a test statistic we convert this to a P-value by comparing its value to distribution of test statistic's under the null hypothesis
- Measure of how likely the test statistic value is under the null hypothesis

$$\text{P-value} \leq \alpha \Rightarrow \text{Reject } H_0 \text{ at level } \alpha$$

$$\text{P-value} > \alpha \Rightarrow \text{Do not reject } H_0 \text{ at level } \alpha$$

z-score

- In practice, it is more common to use the standardized version of the sample mean as our test statistic.

- We know that if a random variable is normally distributed (as it is the case for $\bar{X}$), subtracting the mean and dividing by standard deviation creates a new random variable with standard normal distribution,

  - $Z \sim N(0, 1)$.

- We refer to the standardized value of the observed test statistic as the z-**score**,

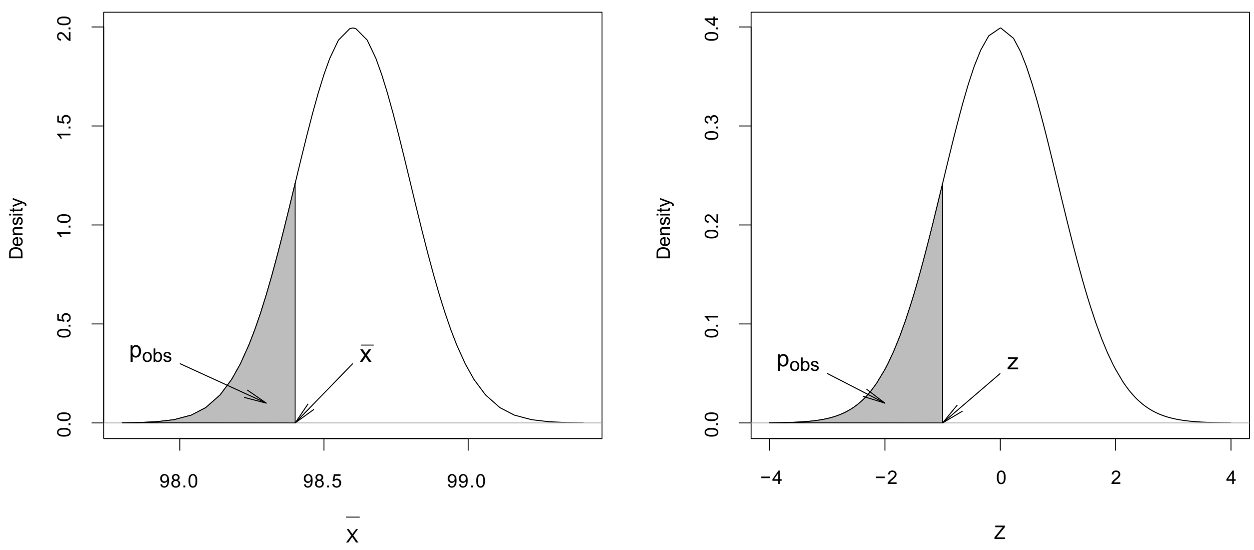
$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}},$$

$$= \frac{98.4 - 98.6}{0.2} = -1.$$

z-test

- We refer to the corresponding hypothesis test of the population mean as the z **-test**.

- In a z -test, instead of comparing the observed sample mean $\bar{x}$ to the population mean according to the null hypothesis, we compare the z -score to 0.



One-sided vs. two-sided hypothesis testing

- Hypothesis tests can be one or two sided (tailed)

- One tailed tests are directional:

$$H_0: \mu_1 - \mu_2 \leq 0$$

$$H_A: \mu_1 - \mu_2 > 0$$

- Two tailed tests are not directional:

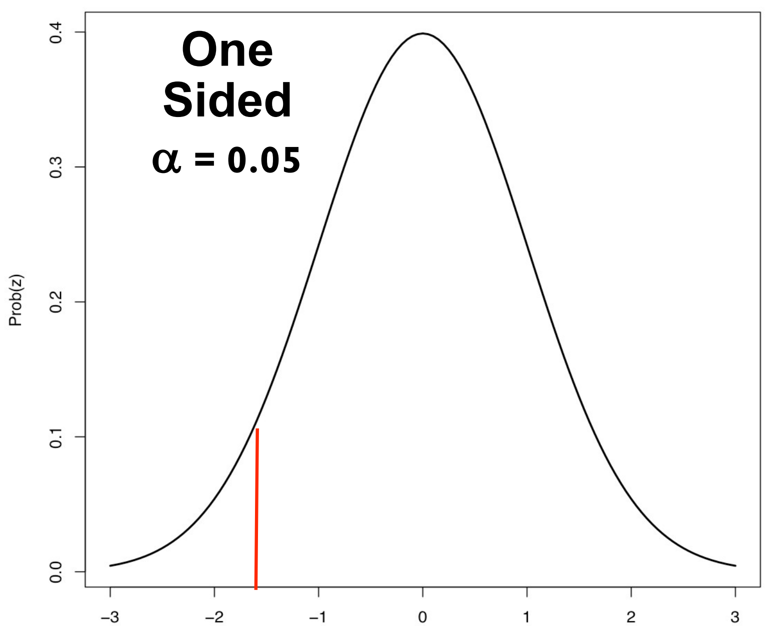
$$H_0: \mu_1 - \mu_2 = 0$$

$$H_A: \mu_1 - \mu_2 \neq 0$$

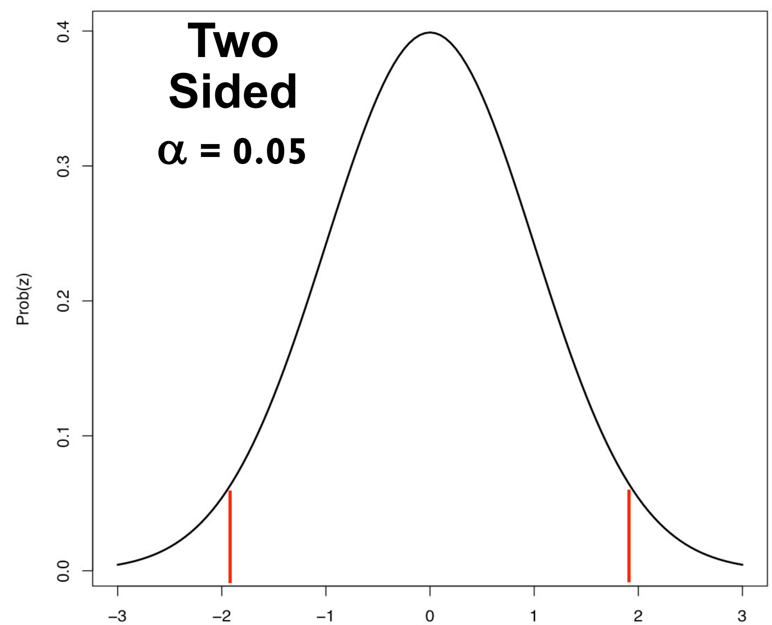
When To Reject H_0

Level of significance, α: Specified before an experiment to define rejection region

Rejection region: set of all test statistic values for which H_0 will be rejected



Critical Value = -1.64



Critical Values = -1.96 and +1.96

One-sided vs. two-sided hypothesis testing

- The alternative hypothesis $H_A : \mu < 98.6$ or $H_A : \mu > 98.6$ are called *one-sided* alternatives.
- For these hypotheses, $p_{\mathrm{obs}} = P(Z \leq z)$ and $p_{\mathrm{obs}} = P(Z \geq z)$ respectively.
- In contrast, the alternative hypothesis $H_A : \mu \mathrel{!=} 98.6$ is *two-sided*.
- For the above three alternatives, the null hypothesis is the same, $H_0 : \mu = 98.6$
- In this case, $p_{\mathrm{obs}} = 2 \times P(Z \geq |z|)$.

Hypothesis testing using t-tests

- So far, we have assumed that the population variance σ^2 is known.

- In reality, σ^2 is almost always unknown, and we need to estimate it from the data.

- As before, we estimate σ^2 using the sample variance S^2.

- Similar to our approach for finding confidence intervals, we account for this additional source of uncertainty by using the t-distribution with $n - 1$ degrees of freedom instead of the standard normal distribution.

- The hypothesis testing procedure is then called the **t-test**.

Hypothesis testing using t-tests

- Using the observed values of $\bar{X}$ and S, the observed value of the test statistic is obtained as follows:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}.$$

- We refer to t as the t-**score**.

- Then,

$$\begin{aligned}
&\text{if } H_A : \mu < \mu_0, \quad p_{\text{obs}} = P(T \le t),\\
&\text{if } H_A : \mu > \mu_0, \quad p_{\text{obs}} = P(T \ge t),\\
&\text{if } H_A : \mu \ne \mu_0, \quad p_{\text{obs}} = 2 \times P(T \ge |t|),
\end{aligned}$$

- Here, T has a t-distribution with $n - 1$ degrees of freedom, and t is our observed t-score.

t-test

- The assumption that the sampling distribution will be normally distributed holds for large samples but not for small samples

- Sample size is large, use z-test
- t-test is used when sample size is small
  - Statistical concept of t-distribution

t-distribution

- Sampling distribution based on small samples will be symmetrical (bell shaped) but not necessarily normal

- Spread of these symmetrical distributions is determined by the specific sample size. The smaller the sample size, the wider the spread, and hence the bigger the standard error

- These symmetrical distributions are known as
  - student's t-distribution or simply, t-distribution

- The t-distribution approaches the normal distribution when sample size tends to infinity

If sample data is normally distributed, t is **almost** normally distributed, but not quite because of the presence of $\bar{\sigma}$.

Family of t-distributions

Hypothesis testing: an example

Does a red shirt help win wrestling?

The experiment and the results

- Does red influence the outcome of wrestling, taekwondo, and boxing?
  - 16 of 20 rounds had more red-shirted than blue-shirted winners in these sports in the 2004 Olympics
  - Shirt color was randomly assigned

Stating the hypotheses

H_0: Red- and blue-shirted athletes are equally likely to win (*proportion* = 0.5).

H_A: Red- and blue-shirted athletes are not equally likely to win (*proportion* ! 0.5).

Estimating the value

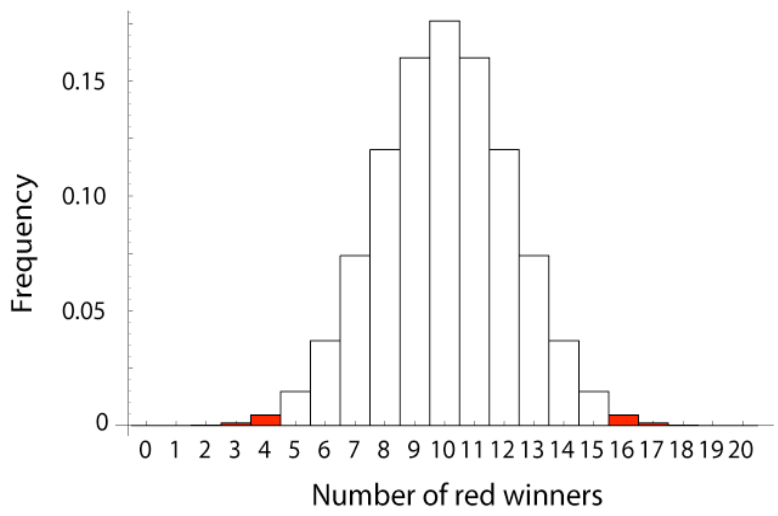
- 16 of 20 is a proportion of *proportion* = 0.8
- This is a discrepancy of 0.3 from the proportion proposed by the null hypothesis, *proportion* = 0.5

Is this discrepancy by chance alone?:

Estimating the probability of such an extreme result

- The *null distribution* for a test statistic is the probability distribution of alternative outcomes when a random sample is taken from a population corresponding to the null expectation.

The null distribution of the *sample proportion*



Calculating the *P*-value from the null distribution

The *P*-value is calculated as

P = 2 ∗ [Pr(16) + Pr(17) + Pr(18) + Pr(19) + Pr(20)] = 0.012.

Significance for the red shirt example

- $P = 0.012$
- $P < p$, so we can reject the null hypothesis

- Athletes in red shirts were more likely to win.

Larger samples give more information

- A larger sample will tend to give and estimate with a smaller confidence interval
- A larger sample will give more power to reject a false null hypothesis

Hypothesis testing: another example

Do dogs resemble their owners?

Common wisdom holds that dogs resemble their owners. Is this true?

- 41 dog owners approached in parks; photos taken of dog and owner separately
- Photo of owner and dog, along with another photo of dog, shown to students to match

Roy, M.M., & Christenfeld, N.J.S. (2004). Do dogs resemble their owners? *Psychological Science*, **15**, 361–363

Hypotheses

H_0: The proportion of correct matches is *proportion* = 0.5.

H_A: The proportion of correct matches is different from *proportion* = 0.5.

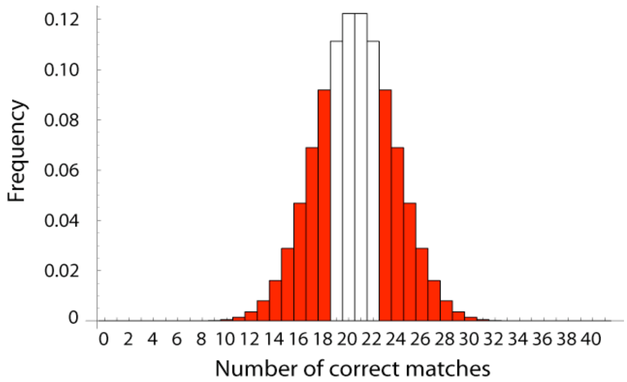
Data

Of 41 matches, 23 were correct and 18 were incorrect.

Estimating the proportion

$$sample\ proportion = \frac{23}{41} = 0.56$$

Null distribution for dog/owner resemblance



The *P*-value:

P =0.53.

We do not reject the null hypothesis that dogs do not resemble their owners.

Errors in Hypothesis Testing

Decision \ Actual Situation "Truth"	H_0 True	H_0 False
Do Not Reject H_0	Correct Decision $1 - \alpha$	Incorrect Decision Type II Error β
Reject H_0	Incorrect Decision Type I Error α	Correct Decision $1 - \beta$