Inference

Eli Gurarie

StatR 101 - Lecture 9
November 19, 2012

November 19, 2012

Probability Theory vs. Inference Statistics

If we *know* the specific probability distribution of a random variable, we can calculate the probability of various event

- e.g. if $Y \sim \text{Binomial}(n = 100, p = 0.5)$ we can calculate $\Pr(40 \leq Y \leq 60)$

Most of **statistics** is concerned with the *opposite* problem:

- e.g. if you know $Y \sim \text{Binomial}(n = 100, p)$ *where the value of* p *is unknown*, and you observed $Y = 32$, what can we say about p?

Definition: *Inference Statistics* refers to any procedure that gives us information about a probability distribution from an observed sample.

Probability Theory vs. Inference Statistics

If we *know* the specific probability distribution of a random variable, we can calculate the probability of various event

- e.g. if $Y \sim \text{Binomial}(n = 100, p = 0.5)$ we can calculate $\Pr(40 \leq Y \leq 60)$

Most of **statistics** is concerned with the *opposite* problem:

- e.g. if you know $Y \sim \text{Binomial}(n = 100, p)$ *where the value of* p *is unknown*, and you observed $Y = 32$, what can we say about p?

Definition: *Inference Statistics* refers to any procedure that gives us information about a probability distribution from an observed sample.

Probability Theory vs. Inference Statistics

If we *know* the specific probability distribution of a random variable, we can calculate the probability of various event

- e.g. if $Y \sim \text{Binomial}(n = 100, p = 0.5)$ we can calculate $\Pr(40 \leq Y \leq 60)$

Most of **statistics** is concerned with the *opposite* problem:

- e.g. if you know $Y \sim \text{Binomial}(n = 100, p)$ *where the value of* p *is unknown*, and you observed $Y = 32$, what can we say about p?

Definition: *Inference Statistics* refers to any procedure that gives us information about a probability distribution from an observed sample.

Inference Statistics

1. Point estimation
2. Hypothesis testing

Example: The spinning coin

A brother and sister want to settle a dispute over a toy. The brother suggests flipping a penny. The sister insists on spinning the penny (arguing: *"C'mon, what's the difference?"*). She calls "Tails", and wins the toy.

The brother later suspects spinning might not be "fair" (which is very important to him). So he performs an experiment and spins a coin 100 times, getting 60 tails.

He really wants to call his sister a cheater, but he's not sure if the 60 tails are just random fluctuation.

9M2869 [RM] © www.visualphotos.com

Example: Population to Sample

If we *know* that spinning a penny is "fair", then $p = 0.5$, and the probability that we would get, e.g. , more than 60 tails out of 100 spins is just:

$$
\begin{aligned}
\Pr(Y \geq 60) &= \sum_{i=60}^{100} f(x|n = 100, p = 0.5) \\
&= \texttt{1 - pbinom(59, size=100, p=0.5)} \\
&\approx \texttt{1-pnorm(59.5, mean=50, sd=sqrt(100*.5*.5))} \\
&= 0.0287
\end{aligned}
$$

Low, but not impossible.

Example: Sample to Population

Lets say the boy decides that spinnning coin is NOT fair. What then, is the actual probability p of a spinning penny landing tails?

His best guess is: $\widehat{p} = 60/100 = 0.6$. He knows (from CLT) that $\widehat{p}$ is a random variable with:

1. $\mathrm{E}(\widehat{p}) = p$
2. $\mathrm{Var}(\widehat{p}) = \frac{p(1-p)}{n}$
3. Distribution, pretty much normal.

So: $\hat{p} \sim \text{Normal}(\mu = 0.6, \sigma^2 = 0.6 \times 0.4/100)$

Example: Sample to Population

Lets say the boy decides that spinnning coin is NOT fair. What then, is the actual probability p of a spinning penny landing tails?

His best guess is: $\widehat{p} = 60/100 = 0.6$. He knows (from CLT) that $\widehat{p}$ is a random variable with:

1. $\mathrm{E}(\widehat{p}) = p$
2. $\mathrm{Var}(\widehat{p}) = \frac{p(1-p)}{n}$
3. Distribution, pretty much normal.

So: $\hat{p} \sim \text{Normal}(\mu = 0.6, \sigma^2 = 0.6 \times 0.4/100)$

Expectation of $\widehat{p}$

Assumption: $X \sim \text{Bernoulli}(p)$and X are i.i.d. (independent and identically distributed).

$$\begin{aligned}
\mathrm{E}(\widehat{p}) &= \mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\mathrm{E}\left(\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}(X_i) \\
&= \frac{1}{n}\sum_{i=1}^{n} p \\
&= \frac{1}{n} np = p
\end{aligned}$$

Expectation of $\widehat{p}$

Assumption: $X \sim \text{Bernoulli}(p)$and X are i.i.d. (independent and identically distributed).

$$\begin{aligned}
\mathsf{E}(\widehat{p}) &= \mathsf{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\mathsf{E}\left(\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathsf{E}(X_i) \\
&= \frac{1}{n}\sum_{i=1}^{n} p \\
&= \frac{1}{n} np = p
\end{aligned}$$

Variance of $\widehat{p}$

Assumption: $X \sim \text{Bernoulli}(p)$and X are i.i.d. (independent and identically distributed).

$$\begin{aligned}
\text{Var}\left(\widehat{p}\right) &= \text{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2}\text{Var}\left(\sum_{i=1}^{n} X_i\right) \\
&= \frac{n}{n^2}\text{Var}\left(X_i\right) \\
&= \frac{p(1-p)}{n}
\end{aligned}$$

Variance of $\widehat{p}$

Assumption: $X \sim \text{Bernoulli}(p)$and X are i.i.d. (independent and identically distributed).

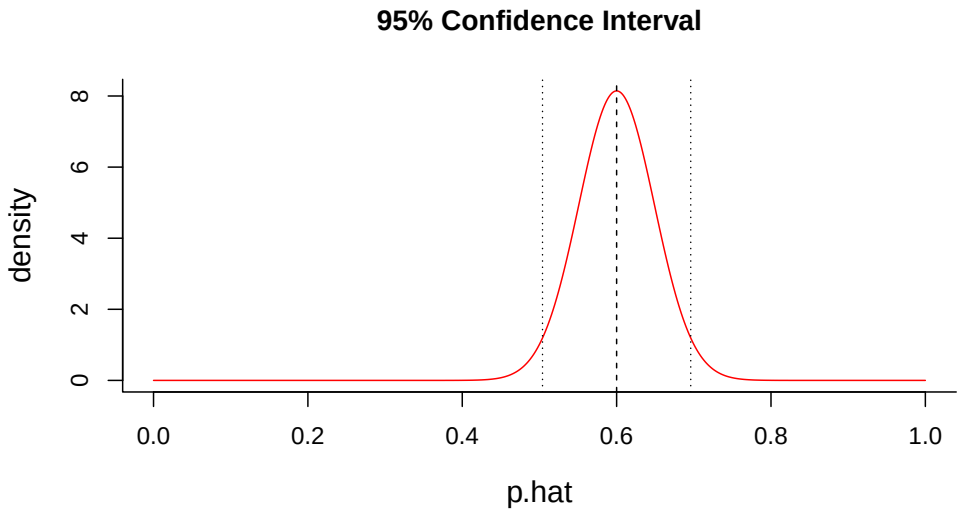
$$\begin{aligned}
\text{Var}\left(\widehat{p}\right) &= \text{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2}\text{Var}\left(\sum_{i=1}^{n} X_i\right) \\
&= \frac{n}{n^2}\text{Var}\left(X_i\right) \\
&= \frac{p(1-p)}{n}
\end{aligned}$$

Constructing a confidence interval

The boy wants to be 95% positive before he calls his sister a cheater. No problem! He has a good model for $\widehat{p} \sim \text{Normal}(\mu = 0.6, \sigma^2 = 0.0024)$.

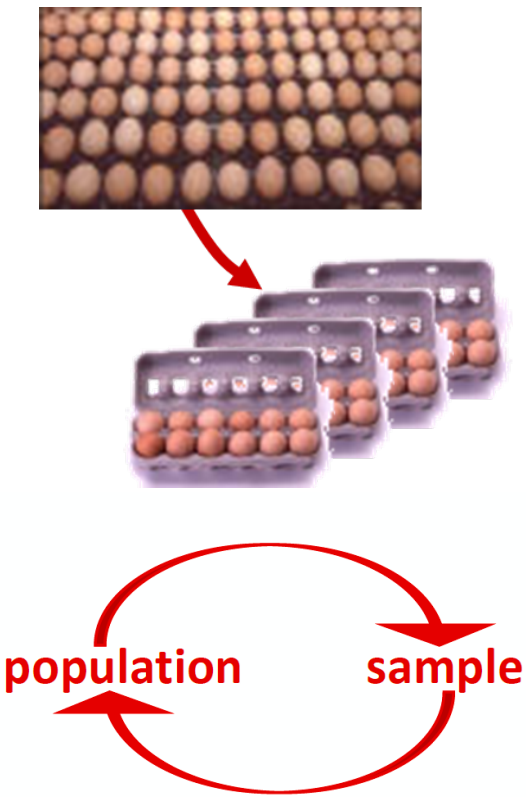
95% of that mass is concentrated between: $(\mu - 1.96\sigma, \mu + 1.96\sigma)$. (recall: the 1.96 is the 97.5% quantile .. `qnorm(0.975)` or `qnorm(1 - (1-.95)/2)`)

$$\widehat{p} = 0.60 \pm 1.96 \times \sqrt{.6 \times .4/100} = 0.60 \pm 0.9602 = \{50.4, 69.6\}$$



Example II: Population to sample

The weight of single eggs of the brown variety is normally distributed, with $N(\mu_X = 65 \text{ g}, \sigma_X = 5 \text{ g})$. You buy a carton of 12 brown eggs.



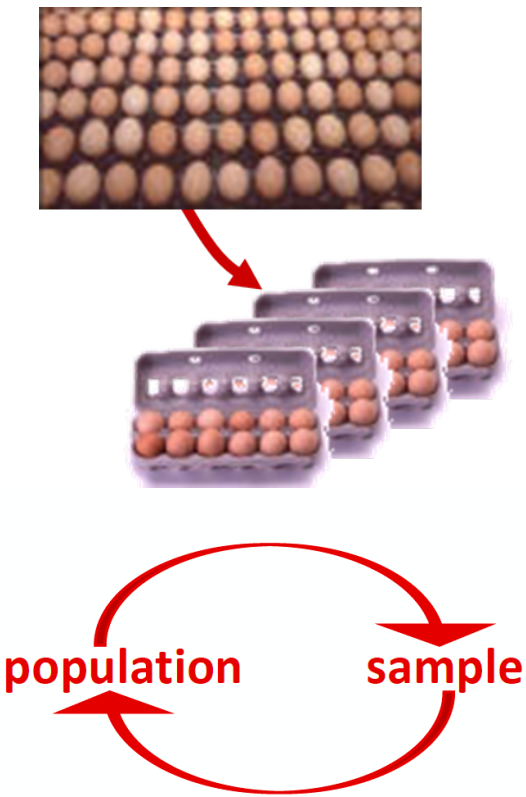
- What is the sampling distribution of $\overline{X}$?

$$\begin{aligned} \overline{X} &\sim N(\mu_{\bar{x}} = 65 \text{ g}, \sigma_{\bar{x}} = 5 \text{ g}/\sqrt{12}) \\ &\sim N(65, 1.443) \end{aligned}$$

- This is the distribution of the estimate of the average weight of an egg, as estimated from a single carton.
- *Note, we only predict the sampling distribution because we know the population mean.*

Example II: Population to sample

The weight of single eggs of the brown variety is normally distributed, with $N(\mu_X = 65 \text{ g}, \sigma_X = 5 \text{ g})$. You buy a carton of 12 brown eggs.



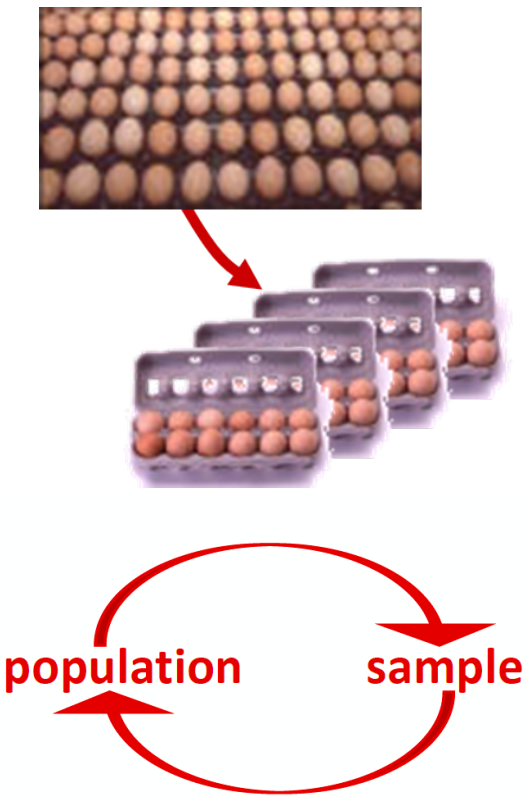
- What is the sampling distribution of $\overline{X}$?

$$\begin{aligned} \overline{X} &\sim N(\mu_{\bar{x}} = 65 \text{ g}, \sigma_{\bar{x}} = 5 \text{ g}/\sqrt{12}) \\ &\sim N(65, 1.443) \end{aligned}$$

- This is the distribution of the estimate of the average weight of an egg, as estimated from a single carton.
- *Note, we only predict the sampling distribution because we know the population mean.*

Example II: Population to sample

The weight of single eggs of the brown variety is normally distributed, with $N(\mu_X = 65 \text{ g}, \sigma_X = 5 \text{ g})$. You buy a carton of 12 brown eggs.



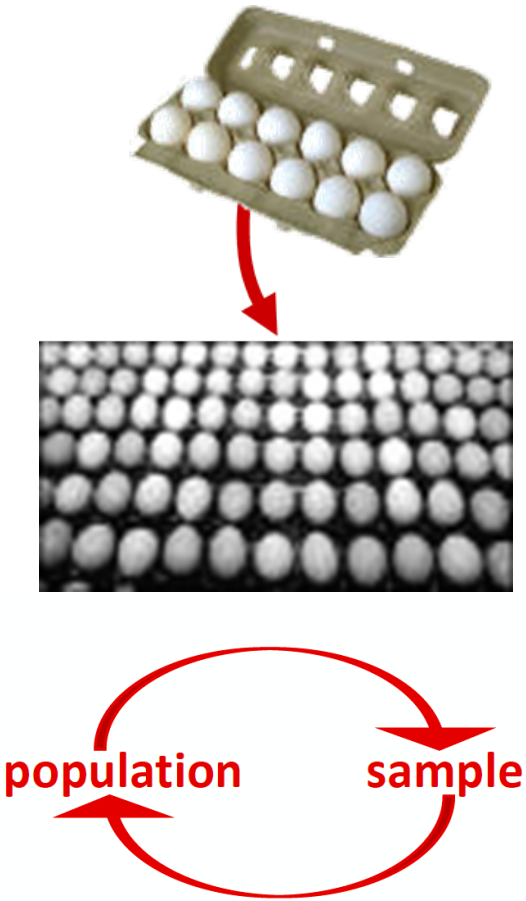
- What is the sampling distribution of $\overline{X}$?

$$\begin{aligned} \overline{X} &\sim N(\mu_{\bar{x}} = 65 \text{ g}, \sigma_{\bar{x}} = 5 \text{ g}/\sqrt{12}) \\ &\sim N(65, 1.443) \end{aligned}$$

- This is the distribution of the estimate of the average weight of an egg, as estimated from a single carton.
- *Note, we only predict the sampling distribution because we know the population mean.*

Example II: Sample to population

Say we don't know the true mean (but we magically know the true standard deviation: $\sigma_x = 5$ g). What can we infer about the **true mean** from a single carton?

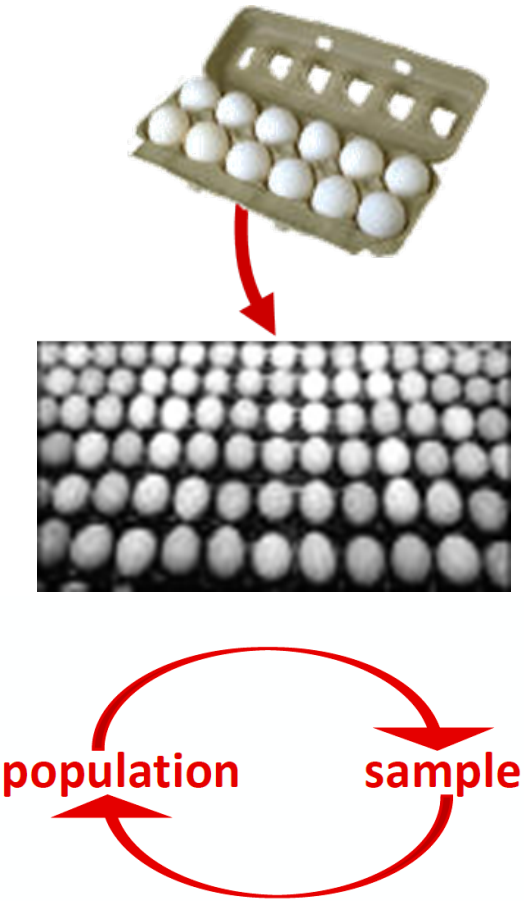


- We buy a carton, and weigh the eggs:
  - Total 770 g
  - Average weight: $\overline{X} = 64.2$ g.
  - Sample standard deviation $\sigma_{\overline{x}} = 5/\sqrt{12} = 1.44$.
- Build a confidence interval: We are 95% certain that the population mean μ lies within ± 1.96 standard deviations from the sample mean $\overline{x}$:

$$95\%\ \text{CI for } \mu\text{: } 64.2\text{g} \pm 2.89\ \text{g} = (61.3,\ 67.1)$$

Example II: Sample to population

Say we don't know the true mean (but we magically know the true standard deviation: $\sigma_x = 5$ g). What can we infer about the **true mean** from a single carton?

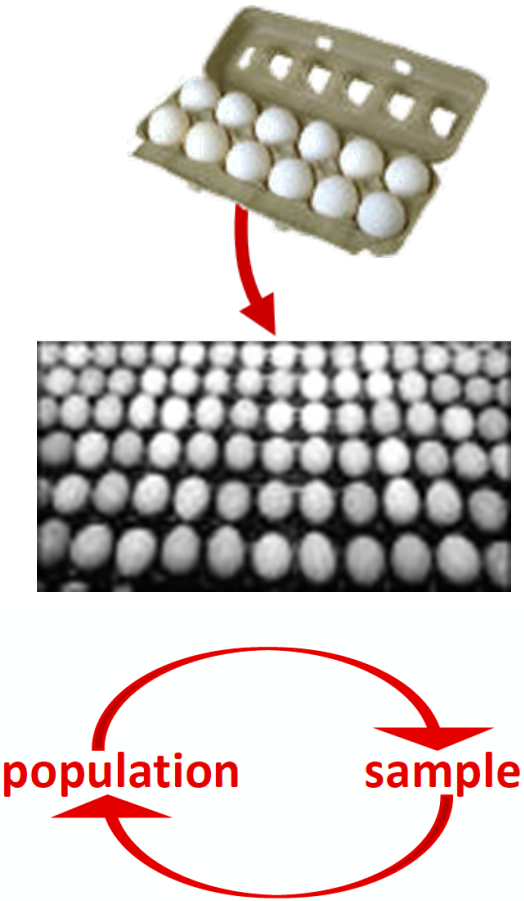


- We buy a carton, and weigh the eggs:
  - Total 770 g
  - Average weight: $\overline{X} = 64.2$ g.
  - Sample standard deviation $\sigma_{\overline{x}} = 5/\sqrt{12} = 1.44$.
- Build a confidence interval: We are 95% certain that the population mean μ lies within ± 1.96 standard deviations from the sample mean $\overline{x}$:

$$95\%\ \text{CI for } \mu\text{: } 64.2\text{g} \pm 2.89\ \text{g} = (61.3,\ 67.1)$$

Example II: Sample to population

Say we don't know the true mean (but we magically know the true standard deviation: $\sigma_x = 5$ g). What can we infer about the **true mean** from a single carton?

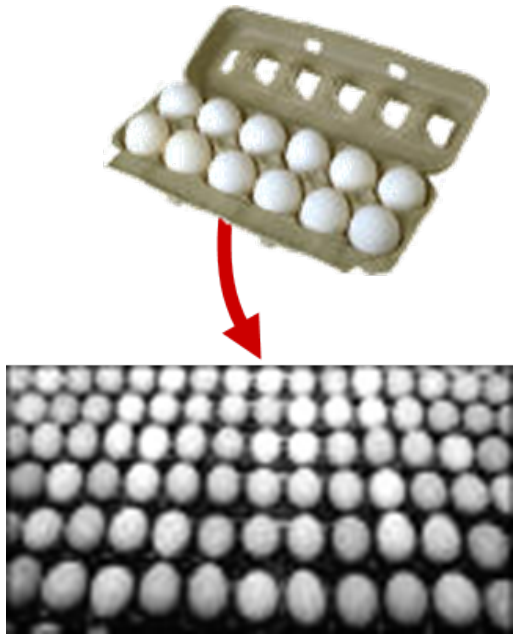


- We buy a carton, and weigh the eggs:
  - Total 770 g
  - Average weight: $\overline{X} = 64.2$ g.
  - Sample standard deviation $\sigma_{\overline{x}} = 5/\sqrt{12} = 1.44$.
- Build a confidence interval: We are 95% certain that the population mean μ lies within ± 1.96 standard deviations from the sample mean $\overline{x}$:

$$95\%\ \text{CI for}\ \mu\text{: } 64.2\text{g} \pm 2.89\ \text{g} = (61.3,\ 67.1)$$

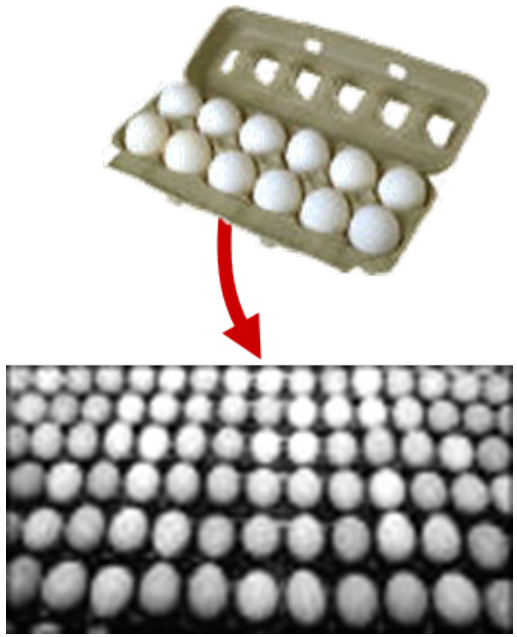
Thanks to statistical theory:

- We do not need to repeatedly sample the population to build up the distribution of the sample statistics
  - This would be lengthy and expensive!
- Instead we rely on a single SRS and the derived properties of the sampling distributions to make inferences to the population.
- This is not "free": we are substituting assumptions for data.



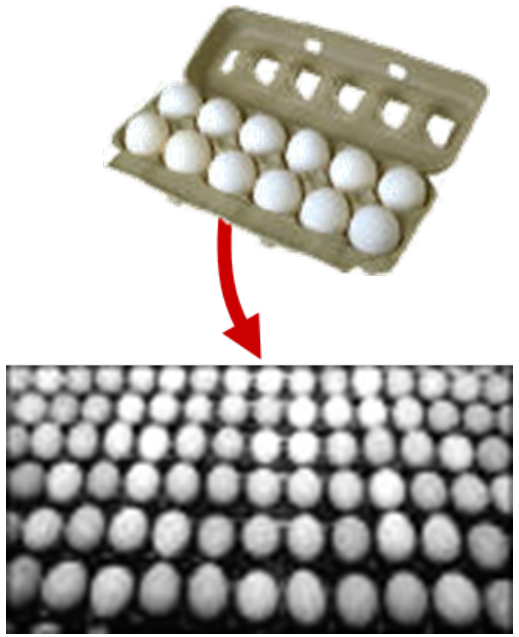
Thanks to statistical theory:

- We do not need to repeatedly sample the population to build up the distribution of the sample statistics
  - This would be lengthy and expensive!
- Instead we rely on a single SRS and the derived properties of the sampling distributions to make inferences to the population.
- This is not "free": we are substituting assumptions for data.



Thanks to statistical theory:

- We do not need to repeatedly sample the population to build up the distribution of the sample statistics
  - This would be lengthy and expensive!
- Instead we rely on a single SRS and the derived properties of the sampling distributions to make inferences to the population.
- This is not "free": we are substituting assumptions for data.



Assumptions:

- The data come from:
  - A randomized experimental design, or
  - A random sample
  - Are based on the population of interest
- The distribution of the sample statistic is known
  - Here we are assuming a Normal distribution (with known s.d.!).
    - We will **relax** these assumptions later.
  - But we can not relax the assumption that we know the PDF for the statistic!
    - Knowing the pdf allows us to quantify uncertainty about predictions.
- If these assumptions are not **met**, our inference is inaccurate.

Assumptions:

- The data come from:
  - A randomized experimental design, or
  - A random sample
  - Are based on the population of interest
- The distribution of the sample statistic is known
  - Here we are assuming a Normal distribution (with known s.d.!).
    - We will **relax** these assumptions later.
  - But we can not relax the assumption that we know the PDF for the statistic!
    - Knowing the pdf allows us to quantify uncertainty about predictions.
- If these assumptions are not **met**, our inference is inaccurate.

Rule of thumb about μ and σ

- When estimating sample mean $\overline{X}$, the margin of error is proportional to $1/\sqrt{n}$
- So ... to *halve* the margin of error, you increase n by $2^2 = 4$
- In general, to reduce the margin of error by a factor of k, you need to multiply the sample size by k^2

Rule of thumb about μ and σ

- When estimating sample mean $\overline{X}$, the margin of error is proportional to $1/\sqrt{n}$
- So ... to *halve* the margin of error, you increase n by $2^2 = 4$
- In general, to reduce the margin of error by a factor of k, you need to multiply the sample size by k^2

Dogs of a different size

- Dogs come in different sizes.
- There is an average dog size μ.
- There is a standard deviation of dog size (say $\sigma = 20cm$).

Point Estimate of Dog Length

Question: What is the average length of a dog - with a 95% confidence interval?

Data: You have 4 (randomly selected) dogs, and their average length is 85 cm.

$$\begin{aligned}\widehat{\mu} &= \overline{X} \pm z_c \frac{\sigma_x}{\sqrt{n}} \\ &= 85 \pm 1.96 \times \frac{20 \text{ cm}}{\sqrt{4}} \\ &= 85 \pm 1.96 \times 10 \text{ cm} = 85 \pm 20\end{aligned}$$

Our best estimate for the average dog is 85 cm, with 95% confidence that the true mean lies between 65 and 105 cm.

Point Estimate of Dog Length

Question: What is the average length of a dog - with a 95% confidence interval?

Data: You have 4 (randomly selected) dogs, and their average length is 85 cm.

$$\begin{aligned}
\widehat{\mu} &= \overline{X} \pm z_c \frac{\sigma_x}{\sqrt{n}} \\
&= 85 \pm 1.96 \times \frac{20 \text{ cm}}{\sqrt{4}} \\
&= 85 \pm 1.96 \times 10 \text{ cm} = 85 \pm 20
\end{aligned}$$

Our best estimate for the average dog is 85 cm, with 95% confidence that the true mean lies between 65 and 105 cm.

Point Estimate of Dog Length

Question: What is the average length of a dog - with a 95% confidence interval?

Data: You have 4 (randomly selected) dogs, and their average length is 85 cm.

$$\begin{aligned}
\widehat{\mu} &= \overline{X} \pm z_c \frac{\sigma_x}{\sqrt{n}} \\
&= 85 \pm 1.96 \times \frac{20 \text{ cm}}{\sqrt{4}} \\
&= 85 \pm 1.96 \times 10 \text{ cm} = 85 \pm 20
\end{aligned}$$

Our best estimate for the average dog is 85 cm, with 95% confidence that the true mean lies between 65 and 105 cm.

Hypothesis testing

- We know that the global population of domestic dogs has mean length $\mu = 100$ cm.
- We sampled 25 Sri Lankan strays - and found that they are 92 cm long on average (but their standard deviation is still $\sigma = 20$ cm).

Question: Are Sri Lankan stray dogs smaller than the average domestic dog?

Hypothesis testing

1. State the hypothesis:
   "Sri Lankan stray dogs are smaller than the average domestic dog."
2. State the hypothesis mathematically:
   $$H_A : \mu_{stray} < 100$$
3. State the hypothesis you want to disprove:
   $$H_0 : \mu_{stray} = 100$$

- H_0 is the **null hypothesis**
- H_A is the **alternative hypothesis**
- The strange thing about hypothesis testing, is that we usually test the **null hypothesis** ... and hope that it's wrong!

Hypothesis testing

1. State the hypothesis:
   "Sri Lankan stray dogs are smaller than the average domestic dog."
2. State the hypothesis mathematically:
   $$H_A : \mu_{stray} < 100$$
3. State the hypothesis you want to disprove:
   $$H_0 : \mu_{stray} = 100$$

- H_0 is the **null hypothesis**
- H_A is the **alternative hypothesis**
- The strange thing about hypothesis testing, is that we usually test the **null hypothesis** ... and hope that it's wrong!

Hypothesis testing

1. State the hypothesis:
   "Sri Lankan stray dogs are smaller than the average domestic dog."
2. State the hypothesis mathematically:
   $$H_A : \mu_{stray} < 100$$
3. State the hypothesis you want to disprove:
   $$H_0 : \mu_{stray} = 100$$

- H_0 is the **null hypothesis**
- H_A is the **alternative hypothesis**
- The strange thing about hypothesis testing, is that we usually test the **null hypothesis** ... and hope that it's wrong!

Hypothesis testing

1. State the hypothesis:
   "Sri Lankan stray dogs are smaller than the average domestic dog."
2. State the hypothesis mathematically:
   $$H_A : \mu_{stray} < 100$$
3. State the hypothesis you want to disprove:
   $$H_0 : \mu_{stray} = 100$$

- H_0 is the **null hypothesis**
- H_A is the **alternative hypothesis**
- The strange thing about hypothesis testing, is that we usually test the **null hypothesis** ... and hope that it's wrong!

Hypothesis testing in a bunch of easy (ha!) steps

1. Formulate null hypothesis:

$$H_0 : \mu_{stray} = 100$$

2. Rewrite the null hypothesis in terms of a **test statistic**:

$$z_{test} = \frac{\overline{X} - 100}{\sigma_{\bar{x}}} = \frac{92 - 100}{20/\sqrt{25}} = -\frac{8}{4} = -2$$

3. Under the assumption of the null hypothesis, we know the distribution of the test statistic:

$$z_{test} \sim N(0, 1)$$

4. We calculate the probability that another measurement would be "more extreme" than the test statistic. This number is called the P-**value**

$$P(Z < z_{test}) = P(Z < -2) = 0.0228$$

5. We have some **critical value** (α-level) that is our **level of statistical significance**. Usually it is 5%. If the P-value is less than α then we **reject the null-hypothesis**. If not, then we **fail to reject the null hypothesis**.

$$0.0228 < 0.05$$

6. Based on the hypothesis test, we **accept** or **reject** the null-hypothesis.
Here, we reject the null hypothesis that Sri Lankan stray dogs are the same size as the average dog, and conclude that they ARE smaller.

Hypothesis testing in a bunch of easy (ha!) steps

1. Formulate null hypothesis:

$$H_0 : \mu_{stray} = 100$$

2. Rewrite the null hypothesis in terms of a **test statistic**:

$$z_{test} = \frac{\overline{X} - 100}{\sigma_{\bar{x}}} = \frac{92 - 100}{20/\sqrt{25}} = -\frac{8}{4} = -2$$

3. Under the assumption of the null hypothesis, we know the distribution of the test statistic:

$$z_{test} \sim N(0, 1)$$

4. We calculate the probability that another measurement would be "more extreme" than the test statistic. This number is called the P-**value**

$$P(Z < z_{test}) = P(Z < -2) = 0.0228$$

5. We have some **critical value** (α-level) that is our **level of statistical significance**. Usually it is 5%. If the P-value is less than α then we **reject the null-hypothesis**. If not, then we **fail to reject the null hypothesis**.

$$0.0228 < 0.05$$

6. Based on the hypothesis test, we **accept** or **reject** the null-hypothesis.
Here, we reject the null hypothesis that Sri Lankan stray dogs are the same size as the average dog, and conclude that they ARE smaller.

Hypothesis testing in a bunch of easy (ha!) steps

1. Formulate null hypothesis:

$$H_0 : \mu_{stray} = 100$$

2. Rewrite the null hypothesis in terms of a **test statistic**:

$$z_{test} = \frac{\overline{X} - 100}{\sigma_{\bar{x}}} = \frac{92 - 100}{20/\sqrt{25}} = -\frac{8}{4} = -2$$

3. Under the assumption of the null hypothesis, we know the distribution of the test statistic:

$$z_{test} \sim N(0, 1)$$

4. We calculate the probability that another measurement would be "more extreme" than the test statistic. This number is called the P-**value**

$$P(Z < z_{test}) = P(Z < -2) = 0.0228$$

5. We have some **critical value** (α-level) that is our **level of statistical significance**. Usually it is 5%. If the P-value is less than α then we **reject the null-hypothesis**. If not, then we **fail to reject the null hypothesis**.

$$0.0228 < 0.05$$

6. Based on the hypothesis test, we **accept** or **reject** the null-hypothesis.
Here, we reject the null hypothesis that Sri Lankan stray dogs are the same size as the average dog, and conclude that they ARE smaller.

Hypothesis testing in a bunch of easy (ha!) steps

1. Formulate null hypothesis:

$$H_0 : \mu_{stray} = 100$$

2. Rewrite the null hypothesis in terms of a **test statistic**:

$$z_{test} = \frac{\overline{X} - 100}{\sigma_{\bar{x}}} = \frac{92 - 100}{20/\sqrt{25}} = -\frac{8}{4} = -2$$

3. Under the assumption of the null hypothesis, we know the distribution of the test statistic:

$$z_{test} \sim N(0, 1)$$

4. We calculate the probability that another measurement would be "more extreme" than the test statistic. This number is called the *P*-**value**

$$P(Z < z_{test}) = P(Z < -2) = 0.0228$$

5. We have some **critical value** (α-level) that is our **level of statistical significance**. Usually it is 5%. If the *P*-value is less than α then we **reject the null-hypothesis**. If not, then we **fail to reject the null hypothesis**.

$$0.0228 < 0.05$$

6. Based on the hypothesis test, we **accept** or **reject** the null-hypothesis.
Here, we reject the null hypothesis that Sri Lankan stray dogs are the same size as the average dog, and conclude that they ARE smaller.

Hypothesis testing in a bunch of easy (ha!) steps

1. Formulate null hypothesis:

$$H_0 : \mu_{stray} = 100$$

2. Rewrite the null hypothesis in terms of a **test statistic**:

$$z_{test} = \frac{\overline{X} - 100}{\sigma_{\bar{x}}} = \frac{92 - 100}{20/\sqrt{25}} = -\frac{8}{4} = -2$$

3. Under the assumption of the null hypothesis, we know the distribution of the test statistic:

$$z_{test} \sim N(0, 1)$$

4. We calculate the probability that another measurement would be "more extreme" than the test statistic. This number is called the P-**value**

$$P(Z < z_{test}) = P(Z < -2) = 0.0228$$

5. We have some **critical value** (α-level) that is our **level of statistical significance**. Usually it is 5%. If the P-value is less than α then we **reject the null-hypothesis**. If not, then we **fail to reject the null hypothesis**.

$$0.0228 < 0.05$$

6. Based on the hypothesis test, we **accept** or **reject** the null-hypothesis. Here, we reject the null hypothesis that Sri Lankan stray dogs are the same size as the average dog, and conclude that they ARE smaller.

Hypothesis testing in a bunch of easy (ha!) steps

1. Formulate null hypothesis:

$$H_0 : \mu_{stray} = 100$$

2. Rewrite the null hypothesis in terms of a **test statistic**:

$$z_{test} = \frac{\overline{X} - 100}{\sigma_{\bar{x}}} = \frac{92 - 100}{20/\sqrt{25}} = -\frac{8}{4} = -2$$

3. Under the assumption of the null hypothesis, we know the distribution of the test statistic:

$$z_{test} \sim N(0, 1)$$

4. We calculate the probability that another measurement would be "more extreme" than the test statistic. This number is called the *P*-**value**

$$P(Z < z_{test}) = P(Z < -2) = 0.0228$$

5. We have some **critical value** (α-level) that is our **level of statistical significance**. Usually it is 5%. If the *P*-value is less than α then we **reject the null-hypothesis**. If not, then we **fail to reject the null hypothesis**.

$$0.0228 < 0.05$$

6. Based on the hypothesis test, we **accept** or **reject** the null-hypothesis.
Here, we reject the null hypothesis that Sri Lankan stray dogs are the same size as the average dog, and conclude that they ARE smaller.

Hypothesis testing

- Maybe the most important piece of this is the **test statistic**.

$$z_{test} = \frac{\overline{X}_{stray} - 100}{\sigma_{\bar{x}}} \sim N(0, 1)$$

- If we *did* sample from the true distribution of dogs ($\mu = 100, \sigma = 20$), then z_{test} would be an r.v. with a standard normal distribution (with typical values between -1.96 and 1.96). But OUR value was a (teeny) bit more extreme:
- The probability that we would get a **more extreme** value from the **null distribution** is quite low (0.0228). That probability is the *P*-**level**.

Hypothesis testing

- Maybe the most important piece of this is the **test statistic**.

$$z_{test} = \frac{\overline{X}_{stray} - 100}{\sigma_{\bar{x}}} \sim N(0, 1)$$

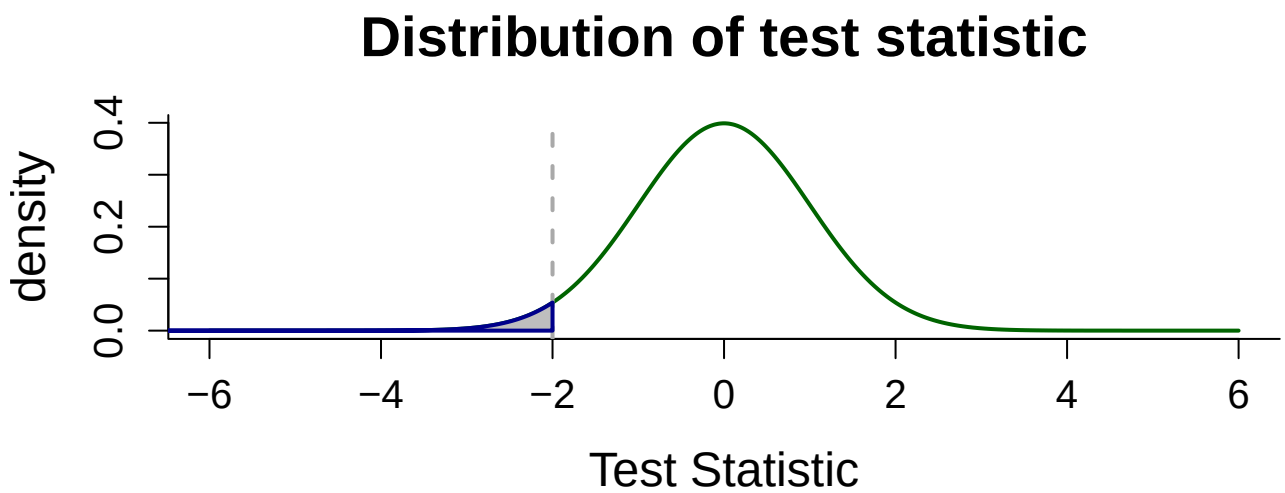
- If we *did* sample from the true distribution of dogs ($\mu = 100, \sigma = 20$), then z_{test} would be an r.v. with a standard normal distribution (with typical values between -1.96 and 1.96). But OUR value was a (teeny) bit more extreme:
- The probability that we would get a **more extreme** value from the **null distribution** is quite low (0.0228). That probability is the *P*-**level**.

Hypothesis testing

- Maybe the most important piece of this is the **test statistic**.

$$z_{test} = \frac{\overline{X}_{stray} - 100}{\sigma_{\bar{x}}} \sim N(0, 1)$$

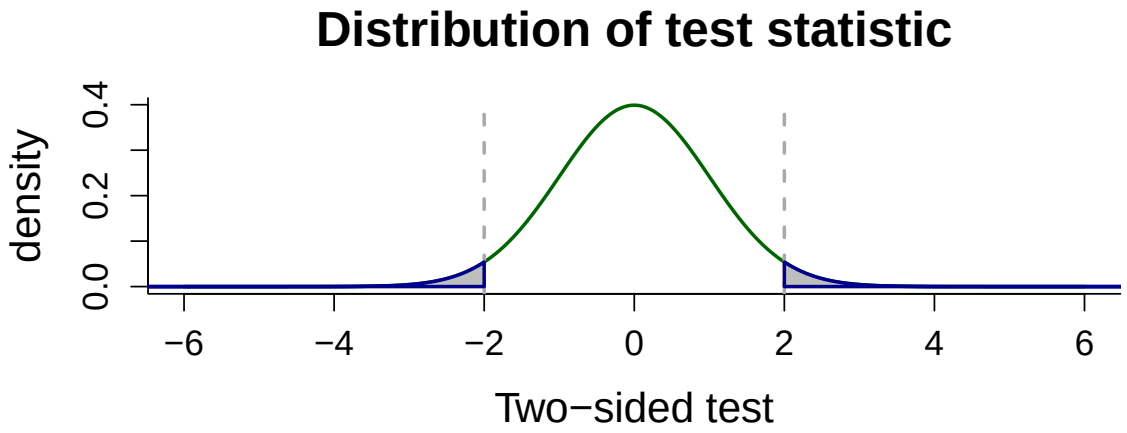
- If we *did* sample from the true distribution of dogs ($\mu = 100, \sigma = 20$), then z_{test} would be an r.v. with a standard normal distribution (with typical values between -1.96 and 1.96). But OUR value was a (teeny) bit more extreme:



- The probability that we would get a **more extreme** value from the **null distribution** is quite low (0.0228). That probability is the *P*-**level**.

Hypothesis testing: Two sided

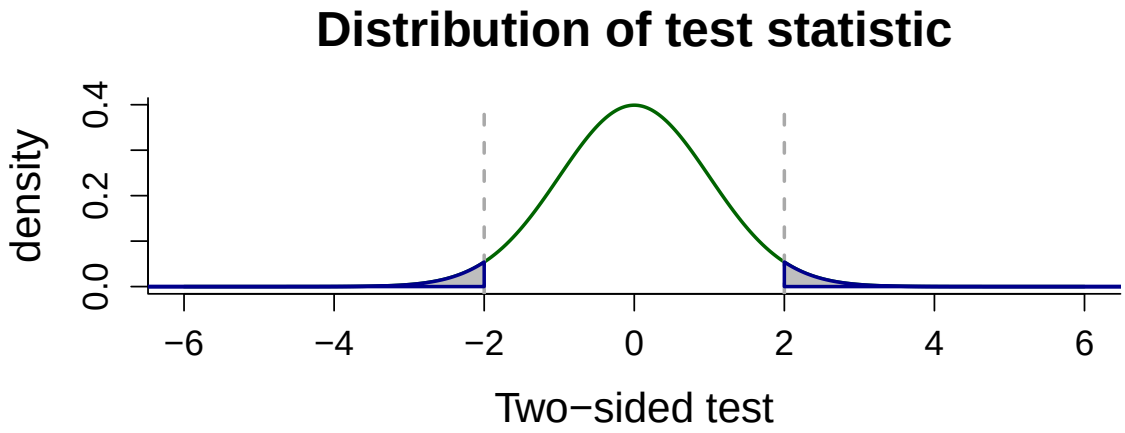
- Different hypothesis: $H_0 : \mu_{stray} = 100$, $H_A : \mu_{stray} \neq 100$



- In a two-sided test, we see if there is a possibility that we could "draw" a number that extreme in either direction from the null distribution.
  - *If* there were *no differences* between the dogs and we picked 25, would we see a number as "extreme" as 92 compared to 100 (i.e. 92 or less, or 108 and more)?
- Here: the P-value is $2 \times 0.0228 = 0.0456$. This is still less than the common significance threshold of $\alpha = 0.05$, but it is quite close.
  - Sometimes people talk about a result being "marginally significant"
  - Remember, there is still almost a 1 in 20 chance that a random draw of 25 dogs would lead to a difference this big!

Hypothesis testing: Two sided

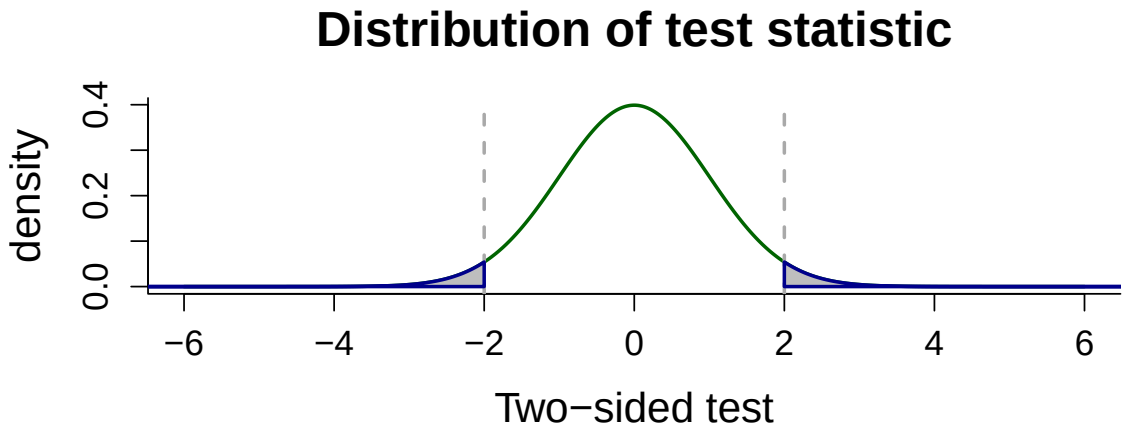
- Different hypothesis: $H_0 : \mu_{stray} = 100$, $H_A : \mu_{stray} \neq 100$



- In a two-sided test, we see if there is a possibility that we could "draw" a number that extreme in either direction from the null distribution.
  - *If* there were *no differences* between the dogs and we picked 25, would we see a number as "extreme" as 92 compared to 100 (i.e. 92 or less, or 108 and more)?
- Here: the P-value is $2 \times 0.0228 = 0.0456$. This is still less than the common significance threshold of $\alpha = 0.05$, but it is quite close.
  - Sometimes people talk about a result being "marginally significant"
  - Remember, there is still almost a 1 in 20 chance that a random draw of 25 dogs would lead to a difference this big!

Hypothesis testing: Two sided

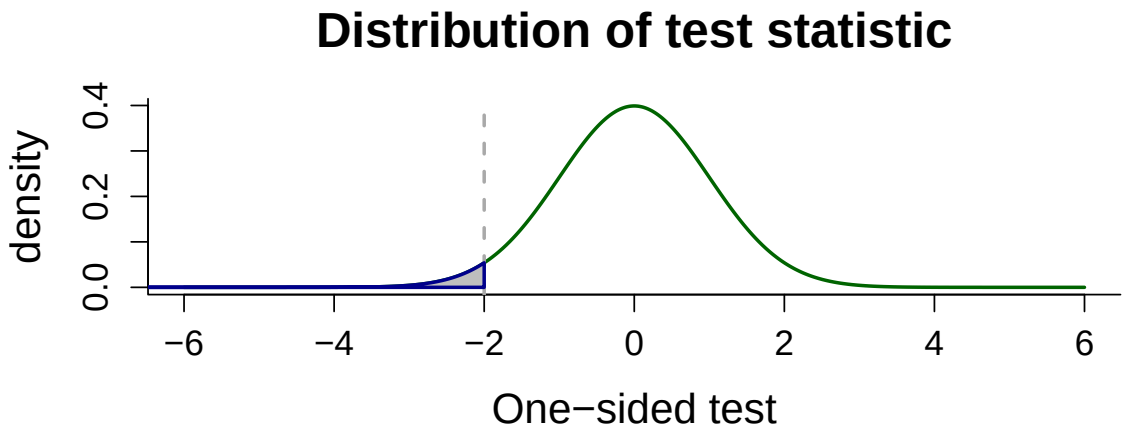
- Different hypothesis: $H_0 : \mu_{stray} = 100$, $H_A : \mu_{stray} \neq 100$



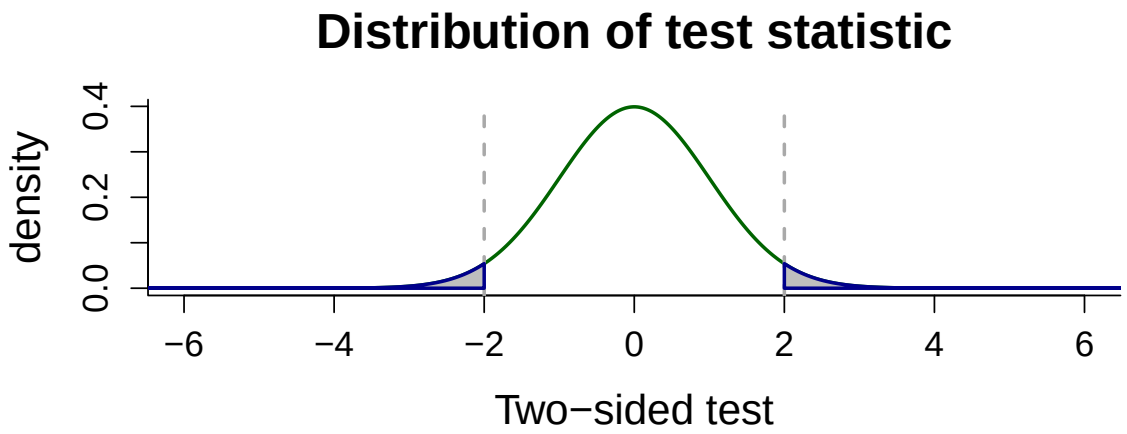
- In a two-sided test, we see if there is a possibility that we could "draw" a number that extreme in either direction from the null distribution.
  - *If* there were *no differences* between the dogs and we picked 25, would we see a number as "extreme" as 92 compared to 100 (i.e. 92 or less, or 108 and more)?
- Here: the P-value is $2 \times 0.0228 = 0.0456$. This is still less than the common significance threshold of $\alpha = 0.05$, but it is quite close.
  - Sometimes people talk about a result being "marginally significant"
  - Remember, there is still almost a 1 in 20 chance that a random draw of 25 dogs would lead to a difference this big!

Hypothesis testing

- One-sided test: $H_0 : \mu_{stray} = 100$, $H_A : \mu_{stray} < 100$



- Two-sided test: $H_0 : \mu_{stray} = 100$, $H_A : \mu_{stray} \neq 100$



Null hypothesis: H_0

- ..is the probability that your observation is an artifact of *randomness* and not of the *structure* that you are interested in.

Test statistic: z_{test}

- ...is a way of converting the **null hypothesis** into a random variable, THE DISTIBUTION OF WHICH YOU KNOW.

***P*-value P.**

- ... is the probability that your test statistic would be MORE EXTREME under total randomness.

Significance level: α

- ... is the threshold at which we determine a "significant difference",
- i.e. it tells you the probability that you will REJECT the null hypothesis even if it's TRUE. (If $\alpha = 0.05$, then one out of twenty times you will say: "Yes! It is a significant result!" ... but be wrong.)
- Usually (historically) $\alpha = 5\%$, or 1% (depending on how conservative you want to be).

Null hypothesis: H_0

- ..is the probability that your observation is an artifact of *randomness* and not of the *structure* that you are interested in.

Test statistic: z_{test}

- ...is a way of converting the **null hypothesis** into a random variable, THE DISTIBUTION OF WHICH YOU KNOW.

***P*-value P.**

- ... is the probability that your test statistic would be MORE EXTREME under total randomness.

Significance level: α

- ... is the threshold at which we determine a "significant difference",
- i.e. it tells you the probability that you will REJECT the null hypothesis even if it's TRUE. (If $\alpha = 0.05$, then one out of twenty times you will say: "Yes! It is a significant result!" ... but be wrong.)
- Usually (historically) $\alpha = 5\%$, or 1% (depending on how conservative you want to be).

Null hypothesis: H_0

- ..is the probability that your observation is an artifact of *randomness* and not of the *structure* that you are interested in.

Test statistic: z_{test}

- ...is a way of converting the **null hypothesis** into a random variable, THE DISTIBUTION OF WHICH YOU KNOW.

***P*-value P.**

- ... is the probability that your test statistic would be MORE EXTREME under total randomness.

Significance level: α

- ... is the threshold at which we determine a "significant difference",
- i.e. it tells you the probability that you will REJECT the null hypothesis even if it's TRUE. (If $\alpha = 0.05$, then one out of twenty times you will say: "Yes! It is a significant result!" ... but be wrong.)
- Usually (historically) $\alpha = 5\%$, or 1% (depending on how conservative you want to be).

Null hypothesis: H_0

- ..is the probability that your observation is an artifact of *randomness* and not of the *structure* that you are interested in.

Test statistic: z_{test}

- ...is a way of converting the **null hypothesis** into a random variable, THE DISTIBUTION OF WHICH YOU KNOW.

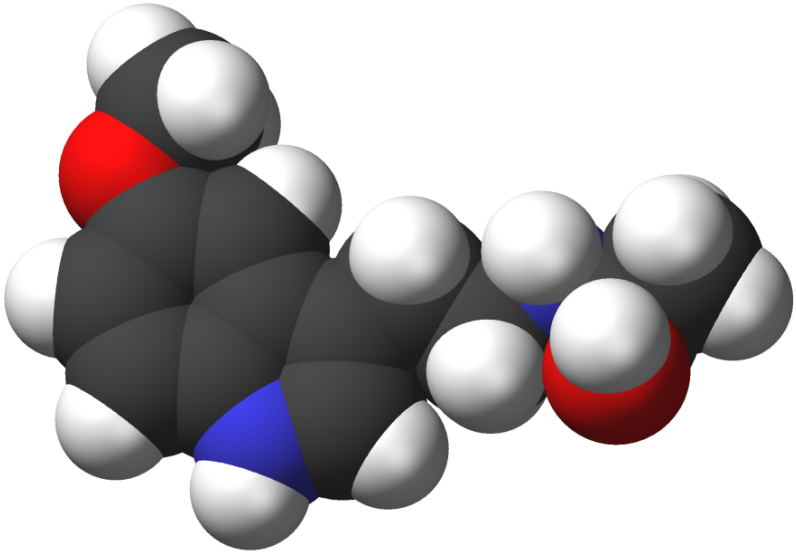
***P*-value P.**

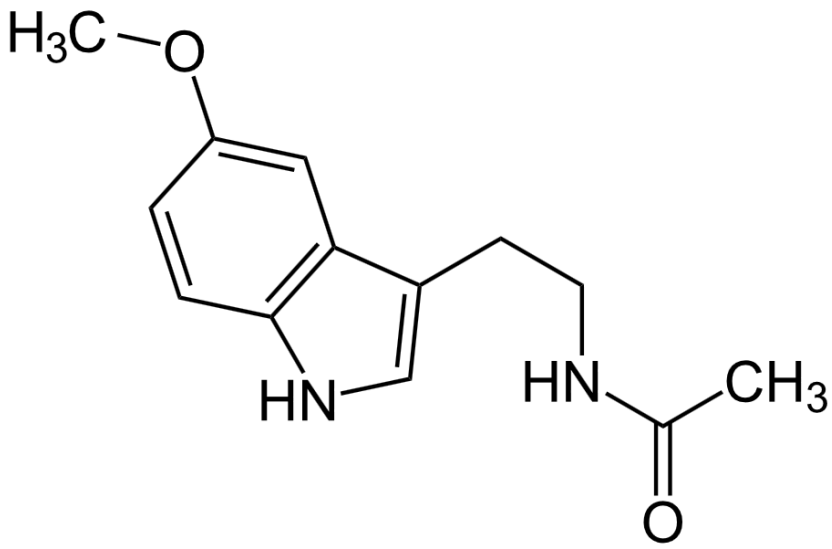
- ... is the probability that your test statistic would be MORE EXTREME under total randomness.

Significance level: α

- ... is the threshold at which we determine a "significant difference",
- i.e. it tells you the probability that you will REJECT the null hypothesis even if it's TRUE. (If $\alpha = 0.05$, then one out of twenty times you will say: "Yes! It is a significant result!" ... but be wrong.)
- Usually (historically) $\alpha = 5\%$, or 1% (depending on how conservative you want to be).

Example with melatonin





Melatonin

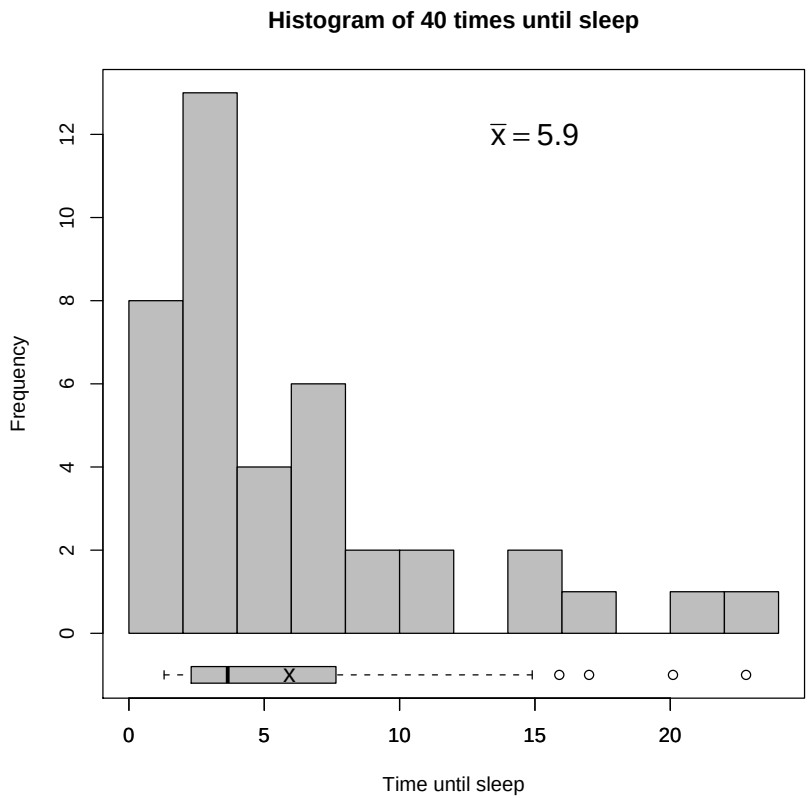
From Wikipedia, the free encyclopedia

Not to be confused with Melanin or Melanotan.

Melatonin /ˌmɛləˈtoʊnɪn/, also known chemically as ***N*-acetyl-5-methoxytryptamine**,[1] is a naturally occurring compound found in animals, plants and microbes.[2][3] In animals, circulating levels of the hormone melatonin vary in a daily cycle, thereby allowing the entrainment of the circadian rhythms of several biological functions.[4]

Example with melatonin

- Time to fall asleep for humans: $\mu = 15$ min, $\sigma = 10$ min.
- Mean time to fall asleep for 40 people dosed with melatonin: 5.9 minutes.



Example with melatonin

- **Null Hypothesis** H_0: $\mu_M = \mu = 15$
- **Alternative Hypothesis** H_a: $\mu_M < \mu$
- If H_0 is true (melatonin has no effect on time to sleep)

$$X_1, X_2, ..., X_{40} \sim \text{Some Distribution}(\mu = 15, \sigma = 10)$$

- But, by CLT

$$\overline{X} \sim \mathcal{N}\left(\mu = 15, \sigma^2_{\overline{X}} = \frac{100}{40}\right)$$
$$\mathcal{N}(\mu = 15, \sigma^2 = 2.5)$$

- **Test statistic:** $z_{test} = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}} = \frac{5.9 - 15}{\sqrt{(2.5)}} = -5.755$

  Under H_0, $z_{test} \sim \mathcal{N}(0, 1)$. Thus:

$$P(Z \leq z_{test}) = \int_{-\infty}^{z_{test}} \phi(x)dx = \texttt{pnorm(z.test)}$$

- This is equivalent to directly calculating:

$$P(\overline{X} \leq 5.9) = \texttt{pnorm(5.9, mean=15, sd=sqrt(2.5)}$$
$$= \texttt{4.323243e-09} \approx 0$$

Anatomy of a test: One-sided vs. two-sided

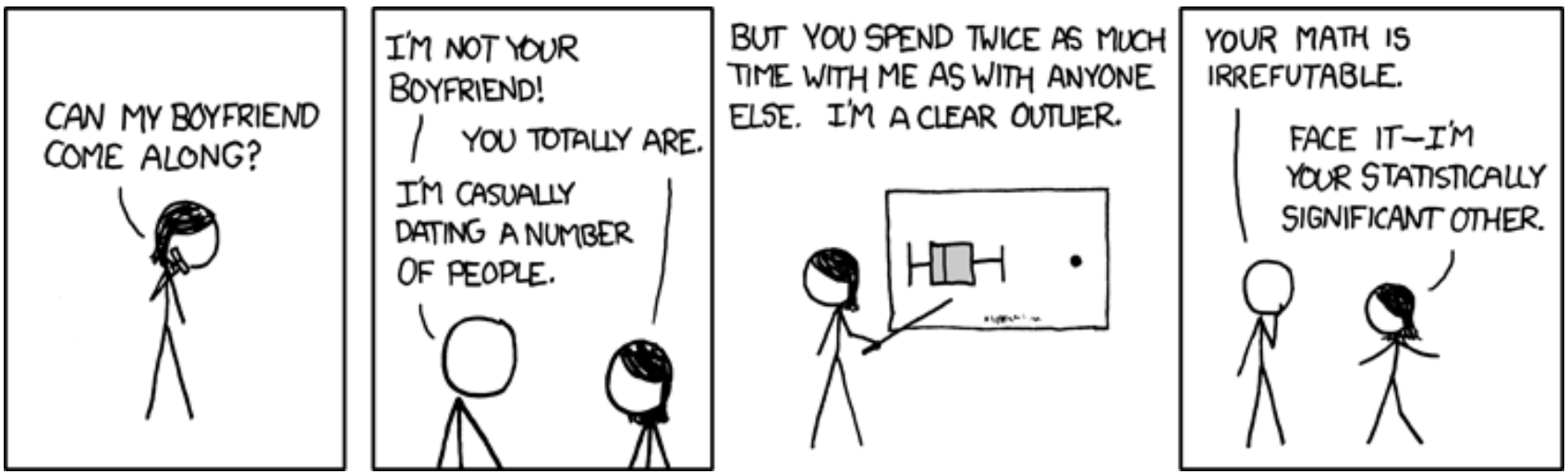
Question:	Is μ different from μ_0?	Is μ smaller than μ_0?
Type of test:	Two-sided Z-test	One-sided Z-test
Data:	$\overline{X}, n$	
Assumptions:	σ_x is known	
H_0:	$\mu = \mu_0$	
H_A:	$\mu \neq \mu_0$	$\mu < \mu_0$
Test statistic:	$z_{test} = \frac{\overline{X} - \mu_0}{\sigma_{\bar{x}}} = \frac{\overline{X} - \mu_0}{\sigma_x / \sqrt{n}}$	
Distribution:	$N(0, 1)$	
P-value:	$2\,\mathrm{P}(Z > \lvert z_{test} \rvert)$	$\mathrm{P}(Z < z_{test})$
α-level	arbitrary!	

The rest of this course is all about making adjustments to this table. Different questions and different assumptions lead to different test statistics. Different test statistics have different distributions under H_0. But the basic strategy is the same.

Comments on Hypothesis Testing

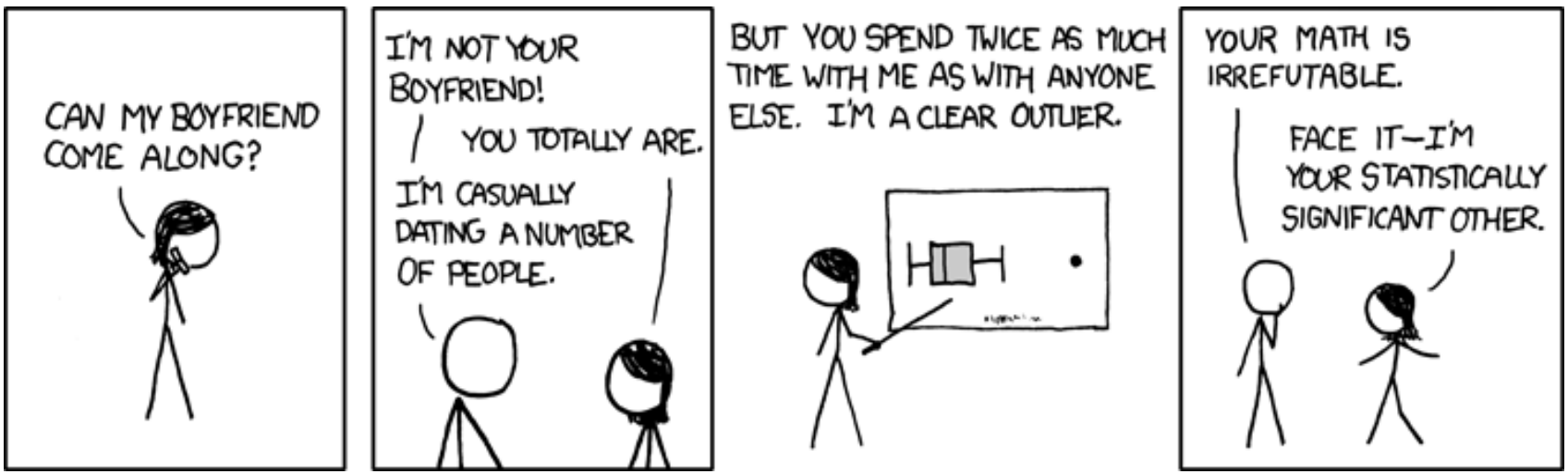
- Hypothesis testing is a *conservative framework* that allows us to apply the *scientific method* rigorously.
- The use of the **null hypothesis** put the burden on having enough *evidence* to conclude there is a real effect, and not random noise.
- It tells you how to untangle the the conflicting effects of:
  - **Observed Differences** (**effect size**)
    - big differences are good, but not good enough alone.
  - and **Sample Sizes**
    - big sample sizes are good, but not good enough alone.
- **Significance** does NOT mean "Large Difference" - it means that there is a small probability - statistically - that you would observe that difference (big OR small) from a random population.
- Hypothesis tests are often mis-applied, and often criticized. But, in their defense:
  - They are useful in a larger context of statistical modeling of multiple explanatory factors and parsimonious model selection.
  - They are useful in formalizing decision-making, decision science, decision trees.

Final thought on statistical significance...



But with $n = 1$, can you really be statistically significant!?

Final thought on statistical significance...



But with $n = 1$, can you really be statistically significant!?