Power: Inference in Practice

Recall the simple conditions for inference

1. Simple Random Sample (SRS)
2. Population has a Normal distribution
3. σ is known

Condition 1: Where do the data come from?

When you use statistical inference, you are acting as if your data are a probability sample or come from a randomized experiment.

Sometimes we treat non-random samples as if they were random. We can do this if we believe that the non-randomness of the sample does not affect the outcome of interest

Condition 1: Where do the data come from?

A neurobiologist is interested in how visual perception can be fooled by optical illusions.

A sociologists is interested in attitudes toward the use of human subjects in experiments

They both use their students as a convenience sample. Can we act as if these students are a SRS? It is easier to make this argument for one of these research questions.

Which one? And why?

Condition 1: Where do the data come from?

Even if the data come from a randomized experiment, there could be issues that hinder the randomness - we covered some of these when we talked about study design:

- Non response
- Dropout ("Lost to follow-up", in RCT jargon)
- Social desirability

What were some of the biases mentioned in the article about political polling?

Condition 2: What is the shape of the population distribution?

- We have leeway around this condition. This is because the z-test (and others we'll learn) rely on the Normality of the sample mean which is guaranteed by the Central Limit Theorem when n is "large enough"
- The z-test is reasonably accurate for any symmetric distribution if the sample size is moderate
- If the distribution is skewed, need a large enough n for the z-test to work.

Condition 2: What is the shape of the population distribution?

- Examine the shape of the sample's distribution using a QQ plot (ideally) or a histogram. Use it to infer the shape of the population
- Difficult to infer much if there are too few observations.

Condition 2: What is the shape of the population distribution?

- Outliers can affect tests of non-resistant measures, like the mean.
- Double check if the outlier is "real" or an error. If it is real, we might use other methods that aren't sensitive to outliers.

(we may explore some non-parametric tests in part III if we have time)

Recap - confidence intervals and testing

How confidence intervals behave

Recall the form of a CI:

$$\bar{x} \pm z^* \frac{\sigma}{\sqrt{n}}$$

Where $z^* \frac{\sigma}{\sqrt{n}}$ is the margin of error.

How confidence intervals behave

The margin of error gets smaller when:

- z* is smaller (i.e., you change to a smaller confidence level). Thus, there is a trade-off between the confidence level and the margin of error.
- σ is smaller. You might be able to reduce σ if there is measurement error. Often times, the σ can't be reduced, it is just a characteristic of the population
- n is larger.

The margin of error only accounts for sampling error

- This is one of the most important points!
- If you're taking epidemiology, you've likely learned about other main errors: confounding, measurement error, and selection bias. These are systematic errors. The confidence interval does not account for systematic errors.

How hypothesis tests behave

- The p-value of a test is dependent on whether H_a is one-sided or two-sided.
- The p-value for a one sided test is one-half the p-value for the two-sided test of the same H_0
- The sample data is used as evidence. It should never be the basis for determining the direction of the alternative hypothesis.

How hypothesis tests behave

- Statistical significance depends on sample size (since sample size determines the standard error of the sampling mean)
- Recall the form of the z-test:

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{\text{magnitude of observed effect}}{\text{size of chance variation}} = \frac{\textit{signal}}{\textit{noise}}$$

- The numerator quantifies the distance between what you observe in the sample and the null hypothesized parameter.
- The denominator represents the size of chance variations from sample to sample

How hypothesis tests behave

- Statistical significance depends on:
    - The size of the observed effect ($\bar{x} - \mu$)
    - The variability of individuals in the population (σ)
    - The sample size (n)
    - Your criteria for rejection the null (α)

If you obtain a small p-value it is not necessarily because the effect size is large.

How hypothesis tests behave

Very tiny effects can be deemed statistically significant when you have enough data. This is a big problem in the age of big data, because you're almost guaranteed to obtain statistically significant results, no matter the effect size.

On the other hand, if your sample size is too small, you might not obtain statistical significance even if your effect size is large. Thus "an absence of evidence is not evidence of absence", or failing to reject the null hypothesis does not imply that the null hypothesis is true.

α β and types of error

α

- The significance level also called the p-value or α is the probability that we observe the estimate we observe or something more extreme when the null hypothesis is true
- This equates to the probability of of making a wrong decision and rejecting the null hypothesis when the null hypothesis is true.
- This is also called the type I error
- But there is another type of error

Type I error, and Type II error in hypothesis tests

- β is the chance of making a wrong decision when the alternative hypothesis is true.
- This is known as a type II error.

Type I error, and Type II error in hypothesis tests

	H_a is true	H_0 is true
Reject H_0	Correct decision	Type I error (α)
Fail to reject H_0	Type II error (β)	Correct decision

This table should remind you of something we have seen before. . . .

Screening tests

	Have the disease	Do not have the disease
Test positive	True positive	False positive
Test negative	False negative	True negative

In the case of screening tests, we calculated sensitivity as the P(Test positive | Have the disease)

in the case of hypothesis testing we are often interested in power

Power

- The power is the chance of making the correct decision when the alternative hypothesis is true.
- Thus, it is the complement of β
- Power $= 1 - \beta$

	H_a is true	H_0 is true
Reject H_0	Correct decision	Type I error (α)
Fail to reject H_0	Type II error (β)	Correct decision

Beta and Power

Type II error or β is a conditional probability

$\beta =$ P(fail to reject $H_0 | H_0$ is false)

Power $= 1 - \beta =$ P(reject $H_0 | H_0$ is false)

However, there are an infinite number of possible values that μ could assume that are not $= \mu_0$

Thus we must choose a value at which to evaluate the β and power for an alternative hypothesis. . .

When we evaluate β we do so at a single such value μ_1

Calculating power in R

Example of calculating power

Suppose you have a known standard deviation $\sigma = 1$. $H_0 : \mu = 0$ vs. $H_a : \mu > 0.8$ and choose $\alpha = 0.05$. Calculate the power when $n = 10$.

You can calculate the minimum z-value required to reject H_0:

```
qnorm(p = 0.05, mean = 0, sd = 1/sqrt(10), lower.tail = F)
```

```
## [1] 0.5201484
```

So for any z-test with this value or higher, you will reject H_0 in favor of H_a.

This is often called Z_α

Example of calculating power

Now suppose that H_a is true. The test will reject H_0 about what percent of the time when H_a is true? To calculate this probability, we take the value from the previous calculation and calculate the *probability* above its value under H_a:

```
pnorm(q = 0.5201484, mean = 0.8, 1/sqrt(10), lower.tail = F)
```

```
## [1] 0.8119132
```

Example of calculating power, illustrated

Size of the difference

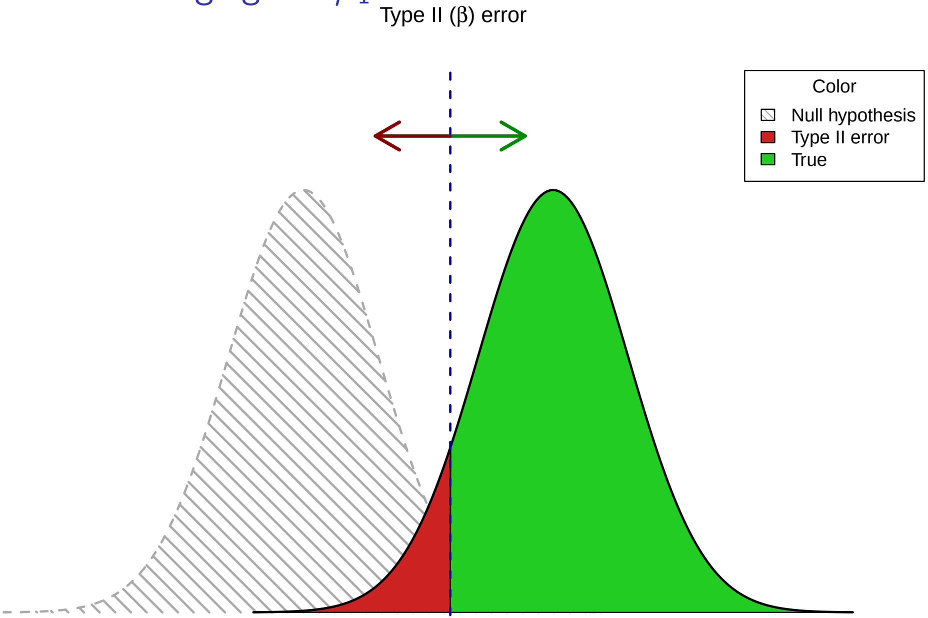
Effect of changing the μ_1

Imagine we our H_0 is a standard normal (mean=0, SD=1) and we set our α at 0.05.

If the true mean of our sampled population is 1.7 standard deviations above the μ_0,

what does our β look like?

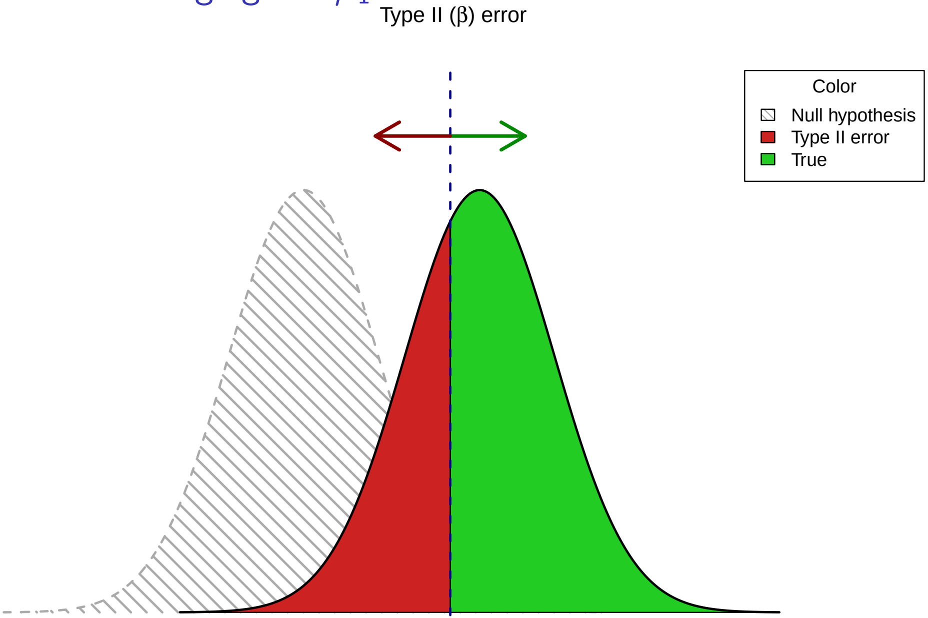
Effect of changing the μ_1



Effect of changing the μ_1

What happens if the "true" mean is closer to the Null?

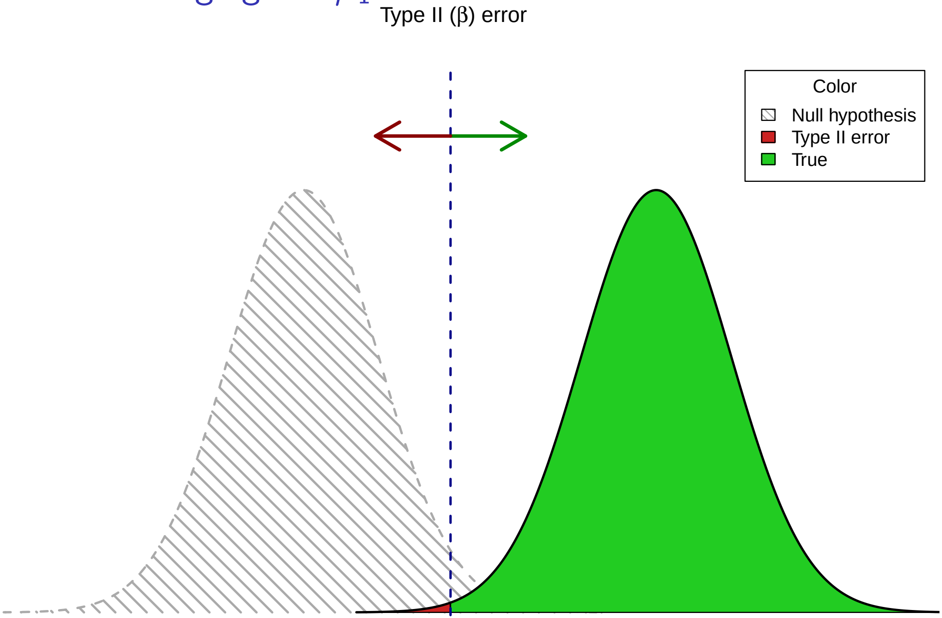
Effect of changing the μ_1



Effect of changing the μ_1

What happens if the "true" mean is further from the null?

Effect of changing the μ_1



Example From Pagano and Gauvreau "Principles of Biostatistics"

Example: from Pagano text ch. 10

Suppose we have a sample of 25 20 to 74 year old males from the United States. We know that 20-24 year old males have a mean serum cholesterol level of 180 mg/100ml. Since cholesterol tends to increase with age, we would expect the mean cholesterol in our sample (of 20-74 year olds) to be higher than 180 mg/100ml. We also assume the standard deviation in the population is 46 mg/100ml.

This is a one-sided hypothesis with $\alpha = 0.05$, we can use r to get the critical value in Z

```
qnorm(0.95, 0,1)
```

```
## [1] 1.644854
```

so H_0 would be rejected at $Z \geq 1.645$

At what mean value of cholesterol would we decide to reject the null?

Example: from Pagano text ch. 10

$$Z = \frac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

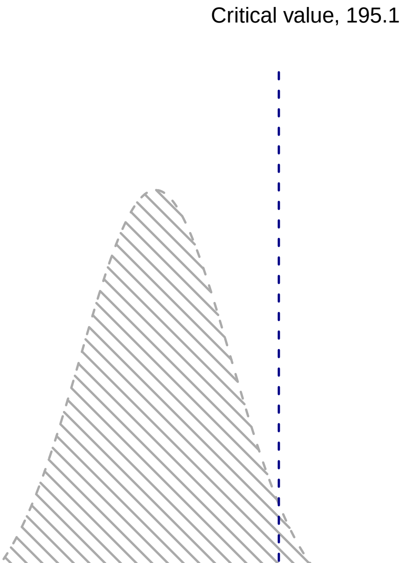
$$1.645 = \frac{\overline{x} - 180}{\frac{46}{\sqrt{25}}}$$

Solve this for $\overline{x}$

$$\bar{x} = 1.645 \times \frac{46}{\sqrt{25}} + 180 = 195.1$$

Example: from Pagano text ch. 10

So here we have our null distribution with the value at which we reject the null



Example: from Pagano text ch. 10

We must choose a value at which to evaluate β. Here we will choose an alternate hypothesis that $\mu_1 = 211$. Since we know a sample mean less than 195.1 causes us to fail to reject H_0 we need to calculate the proportion of a distribution centered at 211mg/100ml that would be below this value.

$$Z = \frac{195.1 - 211}{\frac{46}{\sqrt{25}}}$$

$$Z = -1.73$$

Example: from Pagano text ch. 10

Using R to calculate the probability,

```
pnorm(-1.73, 0,1)
```

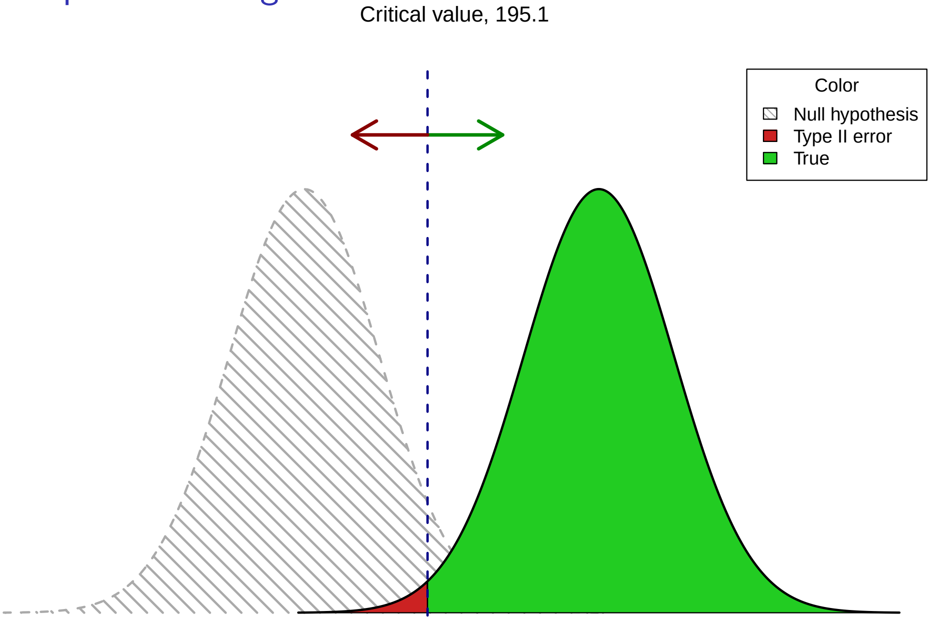
```
## [1] 0.04181514
```

Thus β P(do not reject null(180)|Null is false (true population mean is 211)) is $\sim$0.042

Remember that Power is 1-β = P(reject null | null is false)

In this example, Power is 1-0.042 = 0.958

Example: from Pagano text ch. 10



Size of the difference

So we have we learned about power and β ?

If we increase the size of the difference we are looking for, what happens to power?

What happens to β as we increase the size of the difference?

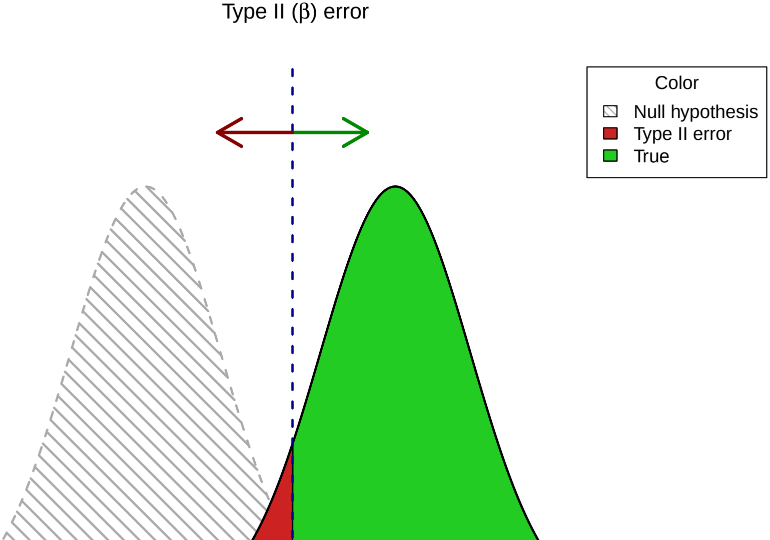
Size of the sample

Remember what happens to our test statistic as sample size increases

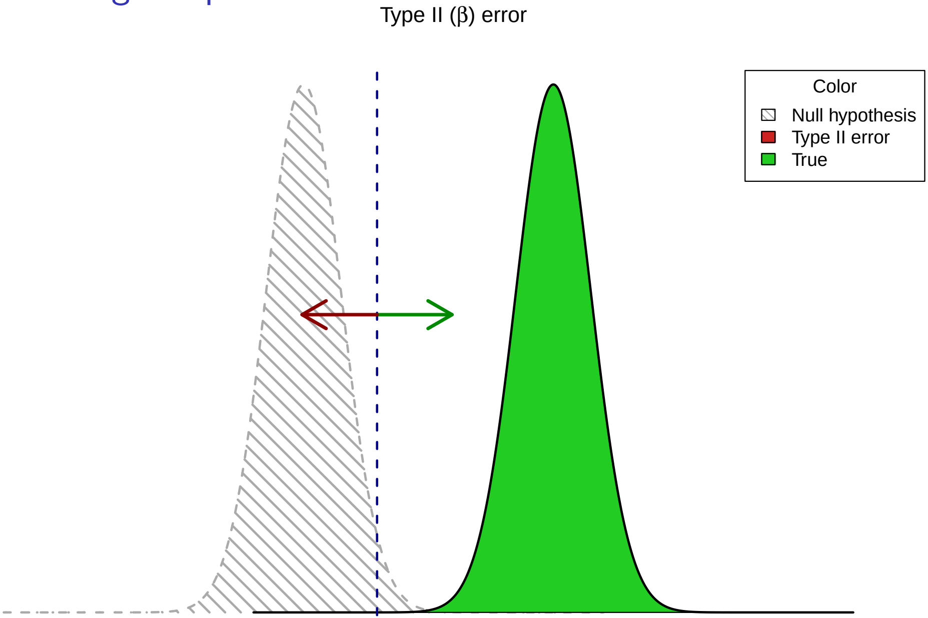
$$Z = \frac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

Increasing sample size

If we look at this graph from earlier in the slides, what do we expect to change if we increase our sample size from 25 to 100?



Increasing sample size



Sample Size

So how does increasing sample size impact power and β

If we increase sample size we expect β to ?

And we expect power to ?

How might this affect our decisions when we are planning a study?

Sample Size and phase of clinical trials

From FDA.com

	Size	Length of study	Purpose
I	20-100	Months	Safety and dosage
II	up to several hundred	Months to $\sim$ 2 years	Efficacy and side effects
III	~300 to 3,000	1-4 years	Efficacy and adverse effects
IV	Thousands being treated	Ongoing	Safety and efficacy

What would you guess here about the size of the difference that would cause you to stop a trial at each of these phases?

How do we use what we have learned about power to plan our studies?

Calculating sample size

Calculating Sample Size

While you may be asked after the fact what power you had to detect a difference in your study, when you are planning a study you would usually want to figure out what sample size would be necessary to provide a specified power for a one sided Z test.

Sample size when conducting a hypothesis test

To think about sample size for a z-test, four things matter:

- Significance level α : How much protection do we want against getting a statistical significant results from our sample when there really is no effect in the population?
- Effect size: How large an effect in the population is important in practice?
- Power $(1-\beta)$: How confident do we want to be that our study will detect an effect of the size we think is important? I.e., what is the probability of rejecting H_0 when the alternative hypothesis is true?
- variability in the population: Remember that the underlying variability in our population affects the variability of our sample mean

Example

For example, using our previous study of mean serum cholesterol levels, if we remember that we assumed the following:

$H_0 : \mu \leq 180mg/100ml$

α: 0.01

σ: 46

If the true population mean is as large as 211 and we want to risk only a 5% chance of failing to reject the null, so $\beta = 0.05$ and power would be $= 1 - \beta = 0.95$

Calculating Sample Size

We start by finding the Z value at which we would reject H_0 at $\alpha = 0.01$

We call this value Z_α

```
qnorm(0.01, lower=FALSE)
```

```
## [1] 2.326348
```

Solve for $\overline{x}$

$$2.32 = \frac{\overline{x} - 180}{\frac{46}{\sqrt{n}}}$$

$$\overline{x} = 180 + 2.32(\frac{46}{\sqrt{n}})$$

Calculating Sample Size

Next we find the Z value at which we would reject H_A at $\beta = 0.05$

We call this value Z_β

```
qnorm(0.05)
```

```
## [1] -1.644854
```

Solve for $\overline{x}$

$$-1.645 = \frac{\overline{x} - 211}{\frac{46}{\sqrt{n}}}$$

$$\overline{x} = 211 - 1.645(\frac{46}{\sqrt{n}})$$

Calculating Sample Size

Because the value of $\overline{x}$ represents the same cutpoint for each of these, we can set the two equations equal to each other and solve for n.

$$180 + 2.32\left(\frac{46}{\sqrt{n}}\right) = 211 - 1.645\left(\frac{46}{\sqrt{n}}\right)$$

$$\sqrt{n}(180 - 211) = (-2.32 + (-1.645)) * 46$$

$$n = \left(\frac{(-2.32 - 1.645) * (46)}{(180 - 211)}\right)^2$$

$$n = 34.6$$

As we cannot include 0.6 of a person, the convention is to round up. So we would need 35 people in our sample.

Calculating Sample Size

The more general way to express this sample size formula for a one sided Z test is

$$n = \left(\frac{(Z_{\alpha} + Z_{\beta}) * (\sigma)}{(\mu_1 - \mu_0)} \right)^2$$

For a two sided Z test this would be

$$n = \left(\frac{(Z_{(\frac{\alpha}{2})} + Z_{\beta}) * (\sigma)}{(\mu_1 - \mu_0)} \right)^2$$

Determining sample size for a margin of error

Selecting an appropriate sample size

How big does the sample size need to be to give a certain size of confidence interval?

Suppose you want your margin of error to equal m. What sample size do you need to obtain a margin of error of m?

You can re-frame the sample size formula for a two sided hypothesis test

$$n = \left(\frac{(Z_{(\frac{\alpha}{2})} + Z_{\beta}) * (\sigma)}{(\mu_1 - \mu_0)} \right)^2$$

to a margin of error

$$n = \left(\frac{Z^* \sigma}{m} \right)^2$$

Example of calculation sample size

Body temperature has a known $\sigma = 0.6$ degrees F. We want to estimate the mean body temperature μ for healthy adults within ± 0.05 F with 95% confidence. How many healthy adults must we measure?

$$n = \left(\frac{z^* \sigma}{m}\right)^2$$

$$n = \left(\frac{1.96 \times 0.6}{0.05}\right)^2 = 553.2$$

We must recruit 554 (round up!) healthy adults for this study.

Parting Humor

From xkcd.com