

# Bring Your Own Device (BYOD) on PrairieLearn

**BIOE 310**

**Spring 2026** All Classes

**Computational Tools for Biological Data**

[Course in Catalog](#) [List of Terms offered](#) [Register for Classes](#)

Show Final Exam Schedule

Course	Section	CRN	Date	Day	Start Time	End Time	Room	Exam Type
BIOE 310	LEC	60965	05/12/2026	T	7:00 PM	10:00 PM	4029 Campus Instructional Facility 4031 Campus Instructional Facility	Extra Space

# Hypothesis testing: one sample

Is P53 gene expressed at a **lower level** in **cancer** patients than in **healthy** people?

- We are interested if a P53 gene expression is **lowered** in **population of cancer patients** compared to the **healthy population**.
- We know that mean gene expression in the **healthy population** is  $\mu_h = 50$  mRNAs/cell. We are interested in deciding whether or not the mean expression in **cancer population** is **lower than** in **healthy population**. Let's call hypothesis  $H_1$ . Here  $H_1$  is **one-sided**
- If we asked: cancer is **not equal** to healthy  $H_1$  would be a **two-sided hypothesis**
- Assume we have a sample of **100 cancer patients** with **sample mean  $\bar{x} = 48$  mRNAs/cell** and **standard deviation  $\sigma = 10$  mRNA/cell**
- Can we use our sample to reject the “business as usual” or **null hypothesis  $H_0$ : cancer = healthy** and select **one-sided hypothesis  $H_1$ : cancer < healthy**

# Two types of errors

	decide $H_0$	decide $H_1$
true $H_0$ probability	Correct action $1 - \alpha$	Type I error $\alpha$
true $H_1$ probability	Type II error $\beta$	Correct action power = $1 - \beta$

$$\alpha = P(\text{type I error}) = P(\text{reject } H_0 \text{ when } H_0 \text{ is true})$$

Sometimes the **type I error probability cutoff  $\alpha$**  is called the **significance level**, or the  **$\alpha$ -error cutoff**

**Instructions:** get the value of  $\alpha$  from your boss or PI (e.g., 5% or 1%)

Prob( $H_0$  is true given the sample data)  $< \alpha \rightarrow$  reject  $H_0$  and accept  $H_1$

Prob( $H_0$  is true given the sample data)  $> \alpha \rightarrow$  accept  $H_0$  and reject  $H_1$

Type II error is much harder to estimate. We will deal with it later

# P-Values of Hypothesis Tests

- **P-value**: what is the probability to get the observed value of sample mean of  $\bar{x} = 48$  mRNAs/cell (or even smaller) and  $\sigma = 10$  mRNAs/cell in a healthy population with  $\mu_h = 50$  mRNAs/cell
- If **P-value is small** – the null hypothesis is likely wrong and thus, the **probability of making a type I error** (incorrectly rejecting the null hypothesis) **is small**
- P-value answers the question: if I reject the null hypothesis  $H_0$  based on the sample, what is the probability that I am making a type I error?

# P-Value vs Type I error cutoff $\alpha$ in Hypothesis Testing

- Problem with using a predefined  $\alpha$ : you **don't know by how much you exceeded it**
- Another approach is to calculate **Prob( $H_0$  is true given the sample data)** referred to as **P-value**. It the smallest  $\alpha$  that would lead to rejection of null hypothesis
- You give your boss the P-value and let him/her decide if it is good enough
- Routinely with big datasets in genomics and systems biology P-values can be  $10^{-\text{large number} \sim 10-100}$ . This number is used to judge the quality of the hypothesis

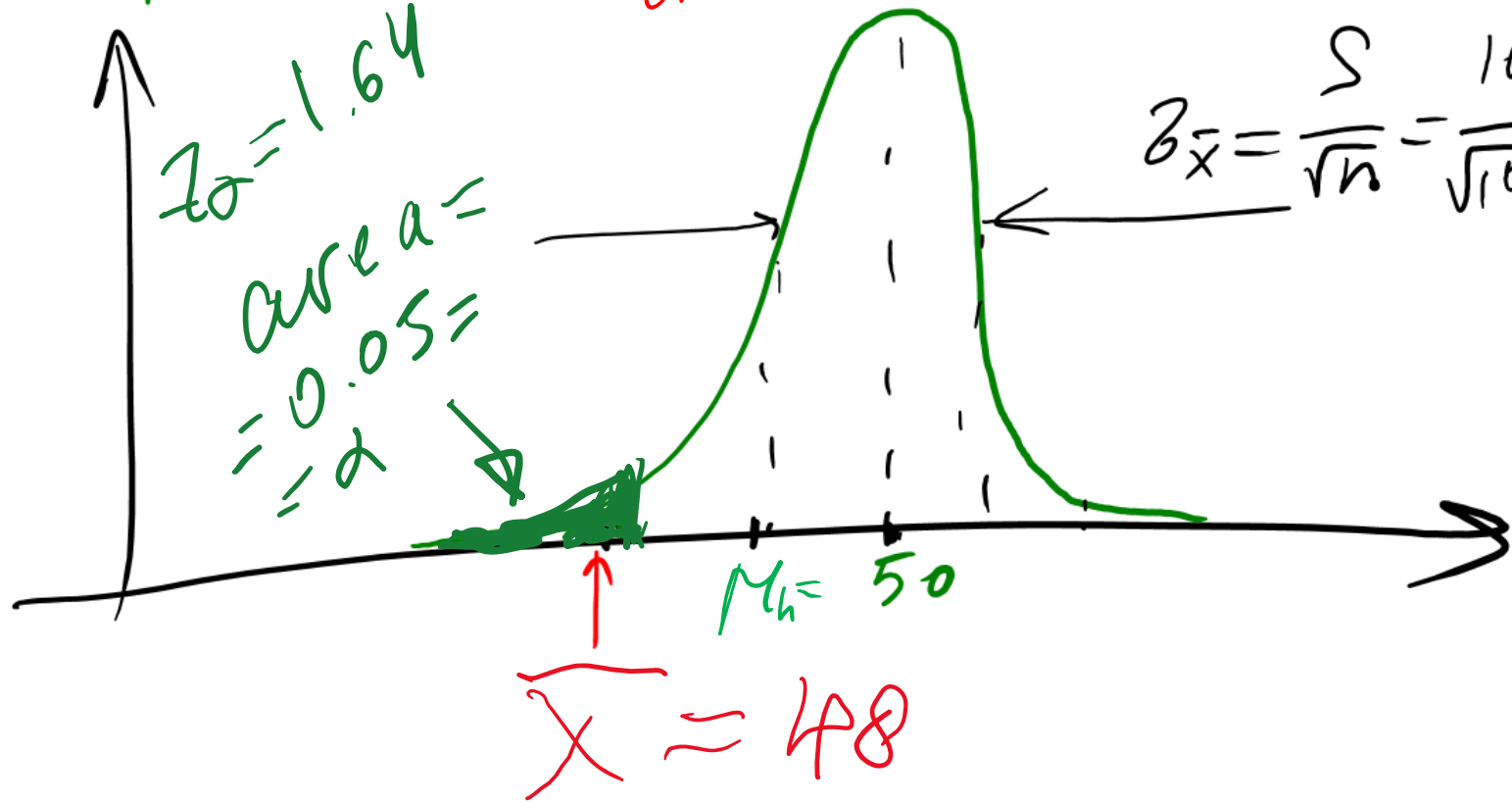


$$\mu_R = 50$$

$$H_0: \mu_C = \mu_R$$

$$n = 100, \bar{X} = 48, S = 10$$

One-sided hypothesis  $H_1: \mu_C < \mu_R$



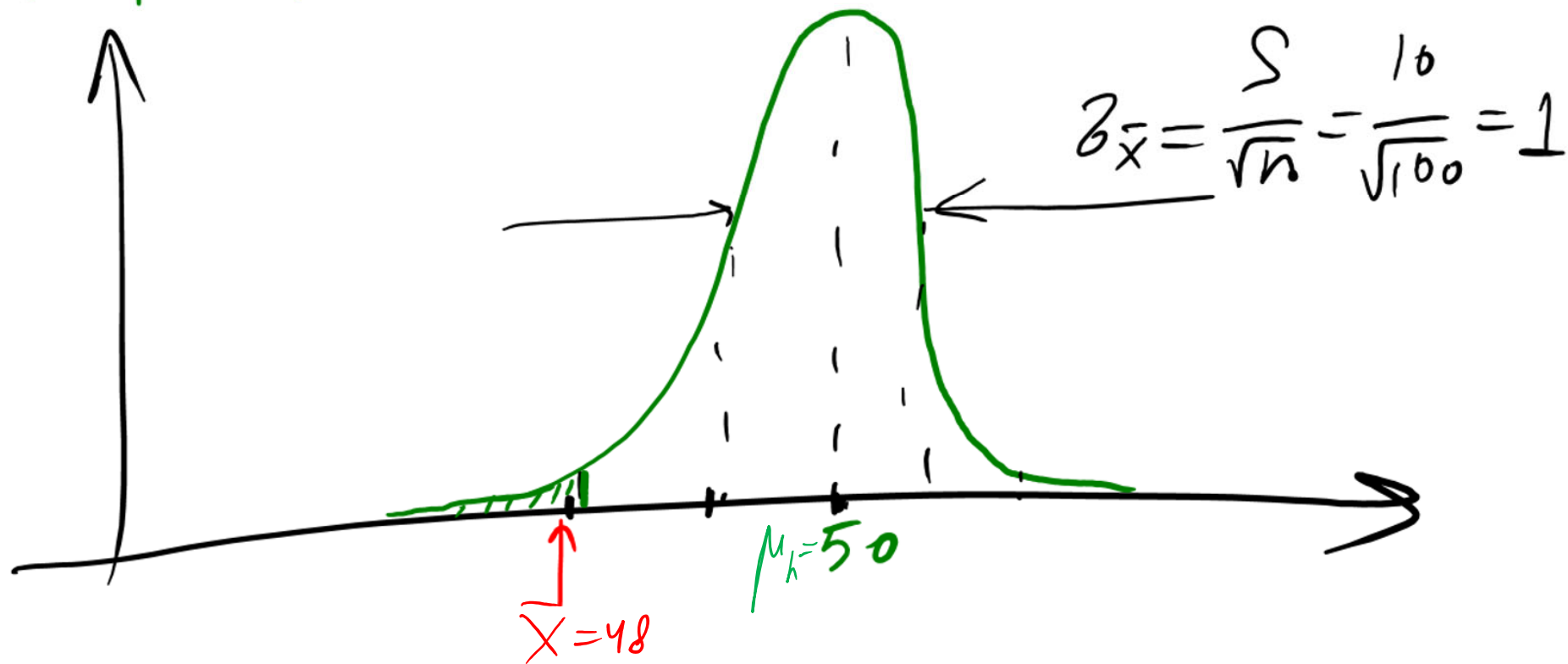
$$\text{P-value} = \text{Prob}(\bar{X}_n < 48 | H_0) =$$
$$\approx 2.5\%$$

$$\mu_h = 50$$

$$H_0: \mu_c = \mu_h$$

$$n = 100, \bar{X} = 48, S = 10$$

$$H_1: \mu_c < \mu_h$$

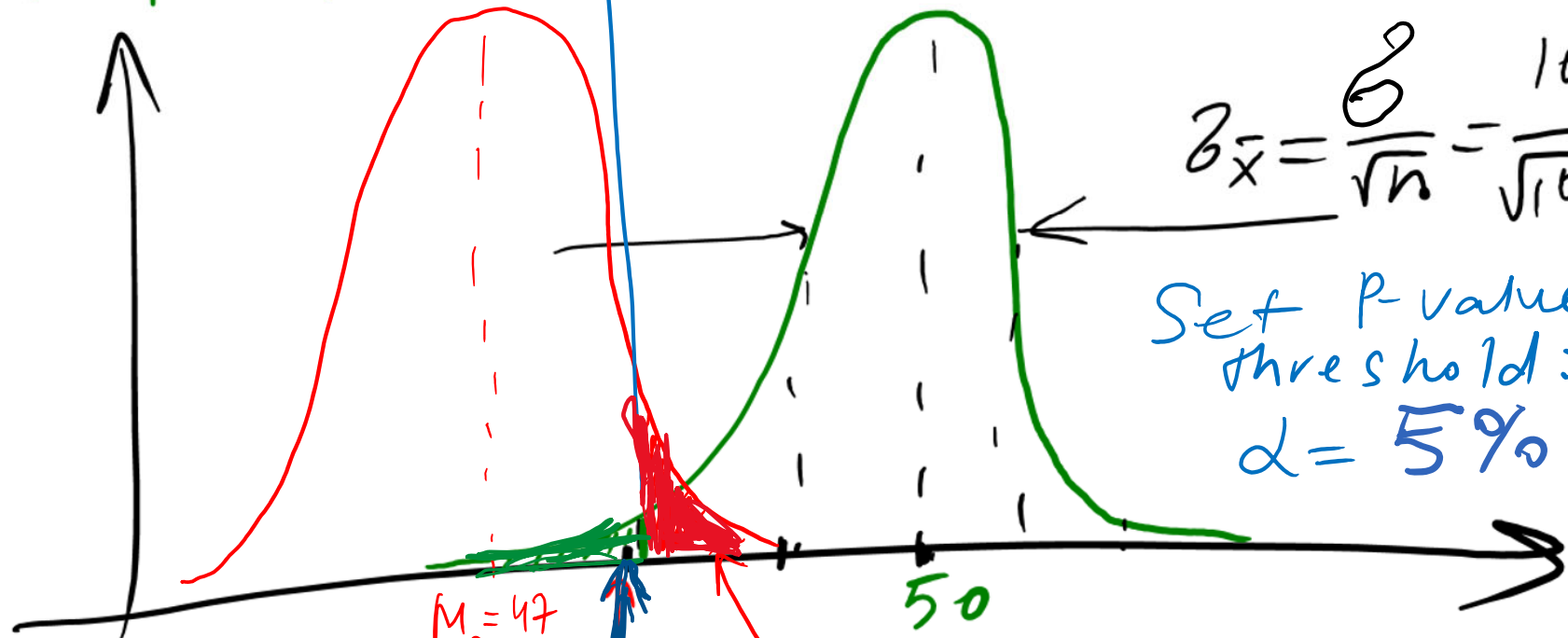


$$\mu_h = 50$$

$$H_0: \mu_c = \mu_h$$

$$n = 100, \bar{X} = 48, \sigma = 10$$

$$H_1: \mu_c < \mu_h$$



$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{100}} = 1$$

Set P-value threshold:  
 $\alpha = 5\%$

$$\mu_h - z_{\alpha} \sigma_{\bar{x}} = 50 - 1.64 = 48.36$$

$$\beta = P(\text{Accept } H_0 \mid H_1 \text{ is true}) =$$

$$\alpha = 1 - \Phi(1.64) = 5\%$$

Type II error

$$\int_{48.36}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-47)^2}{2}\right) dx =$$

$$= 1 - \Phi(1.36) = 8.8\%$$

# Generalizations

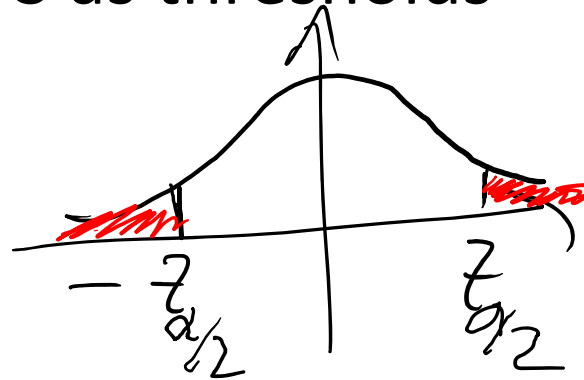
- What if  $H_1$  is a two-sided hypothesis?

- A: P-value is  $2(1-\Phi(|Z|))$ , where  $Z=(\bar{X}-\mu_0)/[S/\sqrt{n}]$

Compare it to: For one sided  $\mu_1 > \mu_0$  it is  $1-\Phi(Z)$

For one sided  $\mu_1 < \mu_0$  it is  $\Phi(Z)$

- If  $\alpha$  is given, use  $\mu_0 \pm z_{\alpha/2} * S$  as thresholds to reject the null hypothesis



- What if the sample size  $n$  is small (say  $n < 10$ ):

- A: Use t-distribution with  $n-1$  degrees of freedom for 2-sided  $P\text{-value} = 2(1 - \text{CDF\_Tdist}(|T|))$

where  $T = (\bar{X} - \mu_0) / [S / \sqrt{n}]$ .

- For a given  $\alpha$  use  $\mu_0 \pm t_{\alpha/2, n-1} T$  to reject the null hypothesis

# Type II Error and Choice of Sample Size

Assume you know the minimum  $\delta = |\mu_1 - \mu_0|$  that you care about.

What is the minimal sample you should use to separate  $H_0$  and  $H_1$  hypotheses if your tolerance to type I and type II errors is  $\alpha$  and  $\beta$  ?

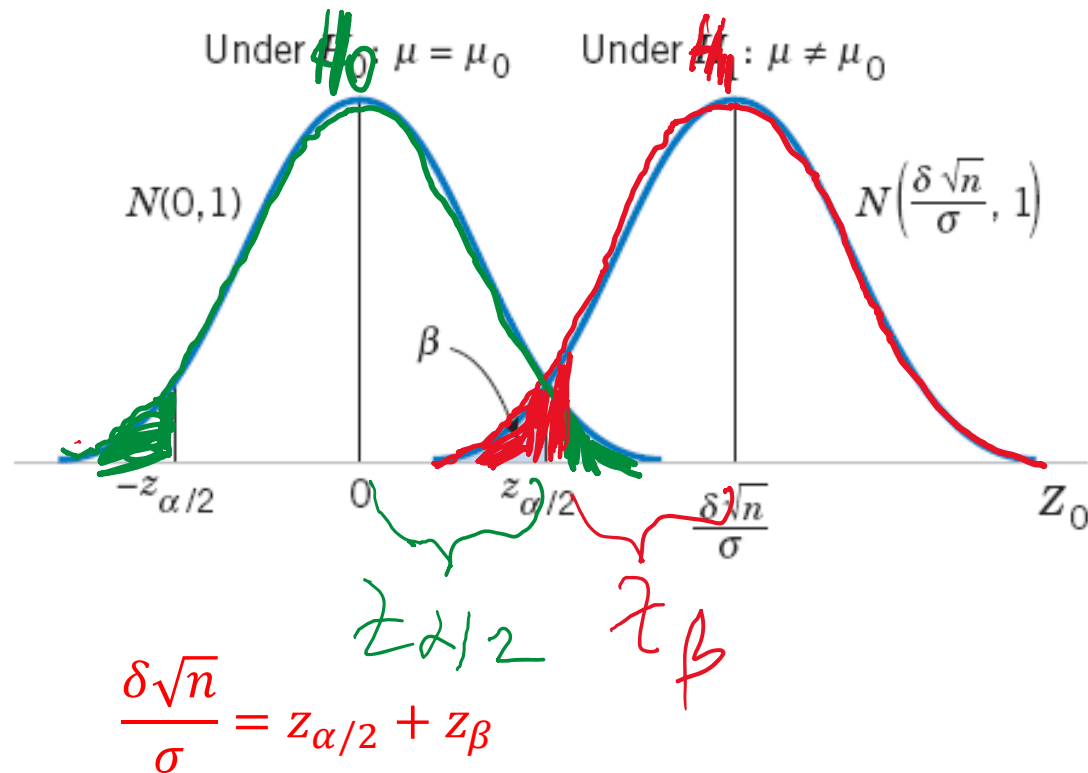


Figure 9-9 The distribution of  $Z_0$  under  $H_0$  and  $H_1$ .

$$n \approx \frac{(z_{\alpha/2} + z_{\beta})^2 \sigma^2}{\delta^2} \quad \text{where} \quad \delta = \mu - \mu_0 \quad (9-22)$$

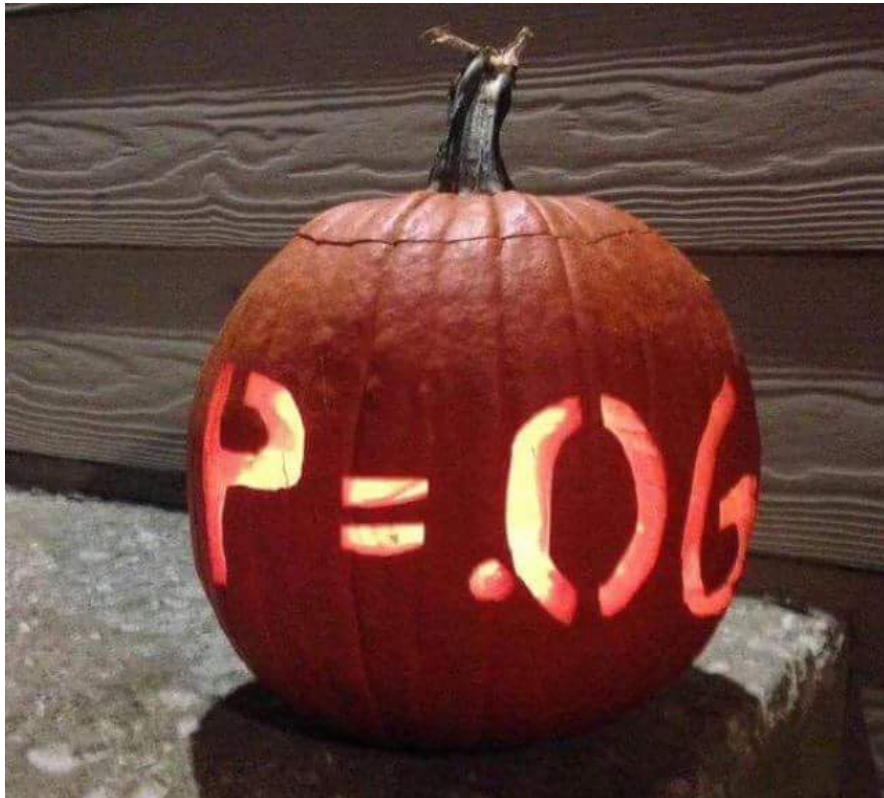
# Standard notation to indicate P-value with

**\* , \*\* , \*\*\***

Table 11.1: A commonly adopted convention for reporting  $p$  values: in many places it is conventional to report one of four different things (e.g.,  $p < .05$ ) as shown below. I've included the "significance stars" notation (i.e., a \* indicates  $p < .05$ ) because you sometimes see this notation produced by statistical software. It's also worth noting that some people will write *n.s.* (not significant) rather than  $p > .05$ .

Usual notation	Signif. stars	English translation	The null is...
$p > .05$		The test wasn't significant	Retained
$p < .05$	*	The test was significant at $\alpha = .05$ but not at $\alpha = .01$ or $\alpha = .001$ .	Rejected
$p < .01$	**	The test was significant at $\alpha = .05$ and $\alpha = .01$ but not at $\alpha = .001$ .	Rejected
$p < .001$	***	The test was significant at all levels	Rejected

---



Happy  
Halloween!  
(belated)

Credit: Trust me,  
I'm a "Biologist"  
Facebook community

<u>P-VALUE</u>	<u>INTERPRETATION</u>
0.001	HIGHLY SIGNIFICANT
0.01	
0.02	
0.03	
0.04	SIGNIFICANT
0.049	
0.050	OH CRAP. REDO CALCULATIONS.
0.051	ON THE EDGE OF SIGNIFICANCE
0.06	
0.07	HIGHLY SUGGESTIVE, SIGNIFICANT AT THE P<0.10 LEVEL
0.08	
0.09	
0.099	HEY, LOOK AT THIS INTERESTING SUBGROUP ANALYSIS
$\geq 0.1$	

Credit: XKCD  
comics

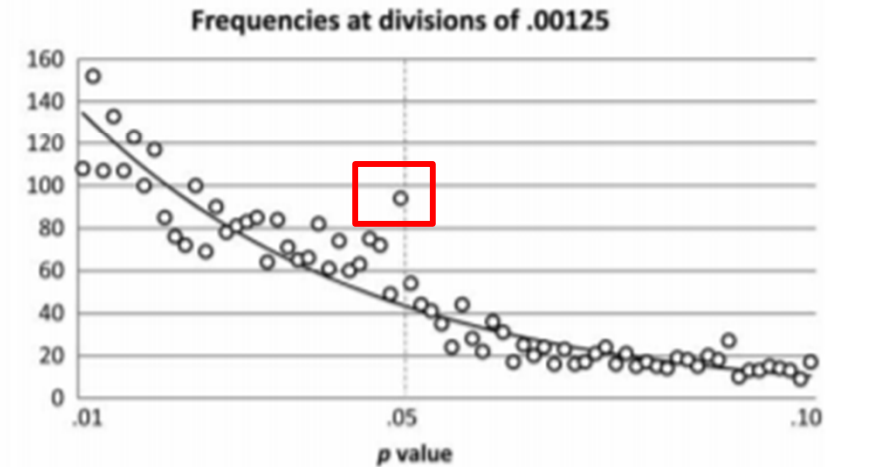
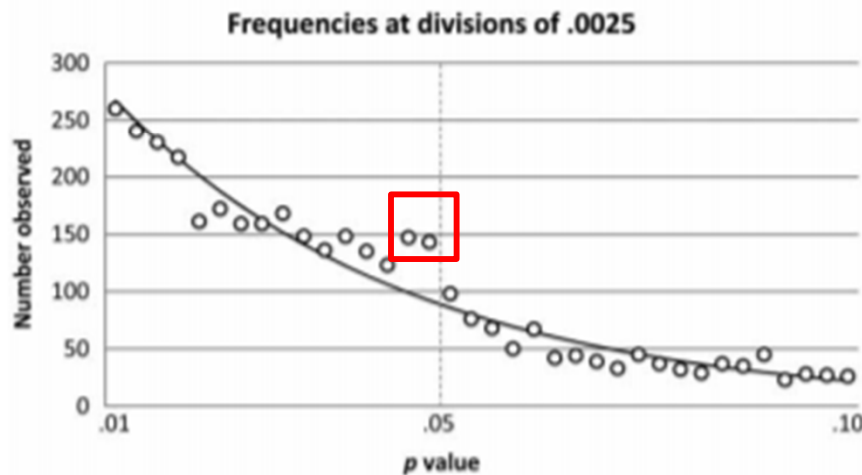
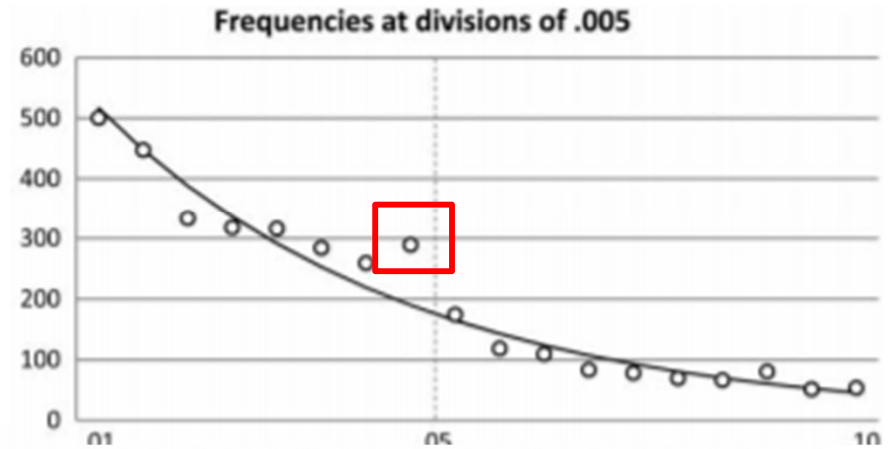
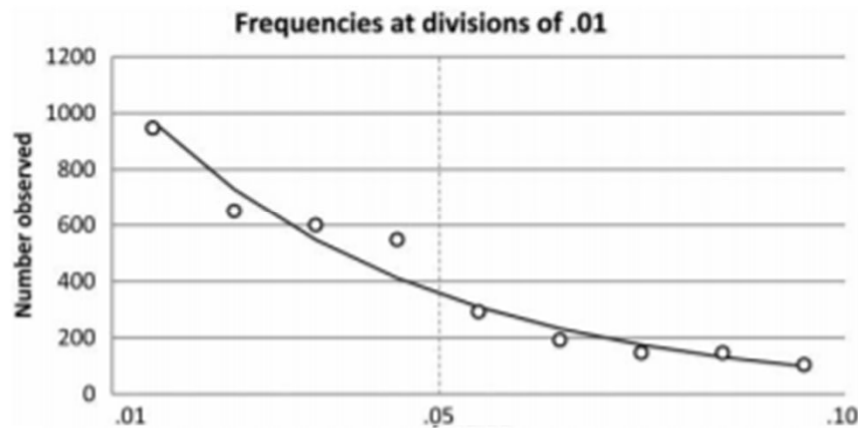
# A peculiar prevalence of $p$ values just below .05

E. J. Masicampo<sup>1</sup>, and Daniel R. Lalande<sup>2</sup>

<sup>1</sup>Department of Psychology, Wake Forest University, Winston-Salem, NC, USA

<sup>2</sup>Department of Health Sciences, Université du Québec à Chicoutimi, Chicoutimi, QC, Canada

## MASICAMPO AND LALANDE



Credit: XKCD  
comics

# WHY ARE THERE SLAVES IN THE BIBLE

WHY DO TWINS HAVE DIFFERENT FINGERPRINTS  
WHY ARE AMERICANS AFRAID OF DRAGONS

WHY IS HTTPS CROSSED OUT IN RED  
WHY IS THERE A LINE THROUGH HTTPS  
WHY IS THERE A RED LINE THROUGH HTTPS ON FACEBOOK  
WHY IS HTTPS IMPORTANT

# QUESTIONS FOUND IN GOOGLE AUTOCOMPLETE



WHY ARE THERE WEEKS  
WHY DO I FEEL DIZZY

WHY DO WHALES JUMP  
WHY ARE WITCHES GREEN  
WHY ARE THERE MIRRORS ABOVE BEDS  
WHY DO I SAY UH  
WHY IS SEA SALT BETTER  
WHY ARE THERE TREES IN THE MIDDLE OF FIELDS  
WHY IS THERE NOT A POKEMON MMO  
WHY IS THERE LAUGHING IN TV SHOWS  
WHY ARE THERE DOORS ON THE FREEWAY  
WHY ARE THERE SO MANY SVCHOST.EXE RUNNING  
WHY AREN'T THERE ANY COUNTRIES IN ANTARCTICA  
WHY ARE THERE SCARY SOUNDS IN MINECRAFT  
WHY IS THERE KICKING IN MY STOMACH  
WHY ARE THERE TWO SLASHES AFTER HTTP  
WHY ARE THERE CELEBRITIES  
WHY DO SNAKES EXIST  
WHY DO OYSTERS HAVE PEARLS  
WHY ARE DUCKS CALLED DUCKS  
WHY DO THEY CALL IT THE CLAP  
WHY ARE KYLE AND CARTMAN FRIENDS  
WHY IS THERE AN ARROW ON AANG'S HEAD  
WHY ARE TEXT MESSAGES BLUE  
WHY ARE THERE MUSTACHES ON CLOTHES  
WHY ARE THERE MUSTACHES ON CARS  
WHY ARE THERE MUSTACHES EVERYWHERE  
WHY ARE THERE SO MANY BIRDS IN OHIO  
WHY IS THERE SO MUCH RAIN IN OHIO  
WHY IS OHIO WEATHER SO WEIRD

WHY DO IGUANAS DIE

WHY AREN'T ECONOMISTS RICH  
WHY DO AMERICANS CALL IT SOCCER  
WHY ARE MY EARS RINGING  
WHY ARE THERE SO MANY AVENGERS  
WHY ARE THE AVENGERS FIGHTING THE X MEN  
WHY IS WOLVERINE NOT IN THE AVENGERS

WHY ARE THERE DINOSAUR GHOSTS  
WHY IS EARTH TILTED  
WHY IS SPACE BLACK  
WHY IS OUTER SPACE SO COLD  
WHY ARE THERE PYRAMIDS ON THE MOON  
WHY IS NASA SHUTTING DOWN



WHY IS THERE AN OWL IN MY BACKYARD  
WHY IS THERE AN OWL OUTSIDE MY WINDOW  
WHY IS THERE AN OWL ON THE DOLLAR BILL  
WHY DO OWLS ATTACK PEOPLE  
WHY ARE AK 47s SO EXPENSIVE  
WHY ARE THERE HELICOPTERS CIRCLING MY HOUSE  
WHY ARE THERE GODS  
WHY ARE THERE TWO SPOCKS

WHY ARE CIGARETTES LEGAL  
WHY ARE THERE DUCKS IN MY POOL  
WHY IS JESUS WHITE  
WHY IS THERE LIQUID IN MY EAR  
WHY DO Q TIPS FEEL GOOD  
WHY DO GOOD PEOPLE DIE

WHY ARE THERE SO MANY CROWS IN ROCHESTER, MN  
WHY IS PSYCHIC WEAK TO BUG  
WHY DO CHILDREN GET CANCER  
WHY IS POSEIDON ANGRY WITH ODYSSEUS  
WHY IS THERE ICE IN SPACE  
WHY ARE DOGS AFRAID OF FIREWORKS  
WHY IS THERE NO KING IN ENGLAND

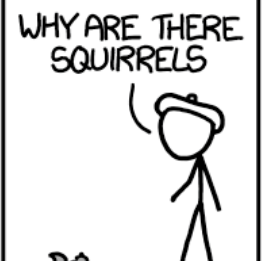
WHY ARE THERE BRIDESMAIDS  
WHY DO DYING PEOPLE REACH UP  
WHY AREN'T THERE VARICOSE ARTERIES  
WHY ARE OLD KUNGONS DIFFERENT

WHY ARE THERE TINY SPIDERS IN MY HOUSE  
WHY DO SPIDERS COME INSIDE  
WHY ARE THERE HUGE SPIDERS IN MY HOUSE  
WHY ARE THERE LOTS OF SPIDERS IN MY HOUSE  
WHY ARE THERE SPIDERS IN MY ROOM  
WHY ARE THERE SO MANY SPIDERS IN MY ROOM  
WHY DO SPIDER BITES ITCH  
WHY IS DYING SO SCARY

WHY IS THERE HELL IF GOD FORGIVES  
WHY IS THERE NO GPS IN LAPTOPS  
WHY DO KNEES CLICK  
WHY AREN'T THERE E GRADES  
WHY IS ISOLATION BAD  
WHY DO BOYS LIKE ME  
WHY DON'T BOYS LIKE ME  
WHY IS THERE ALWAYS A JAVA UPDATE  
WHY ARE THERE RED DOTS ON MY THIGHS  
WHY IS LYING GOOD



WHY ARE THERE FEMALE MR NIMES  
WHY IS MT VESUVIUS THERE  
WHY DO THEY SAY T MINUS  
WHY ARE THERE OBELISKS  
WHY ARE WRESTLERS ALWAYS WET  
WHY ARE OCEANS BECOMING MORE ACIDIC  
WHY IS ARWEN DYING  
WHY AREN'T MY QUAIL LAYING EGGS  
WHY AREN'T MY QUAIL EGGS HATCHING  
WHY AREN'T THERE ANY FOREIGN MILITARY BASES IN AMERICA

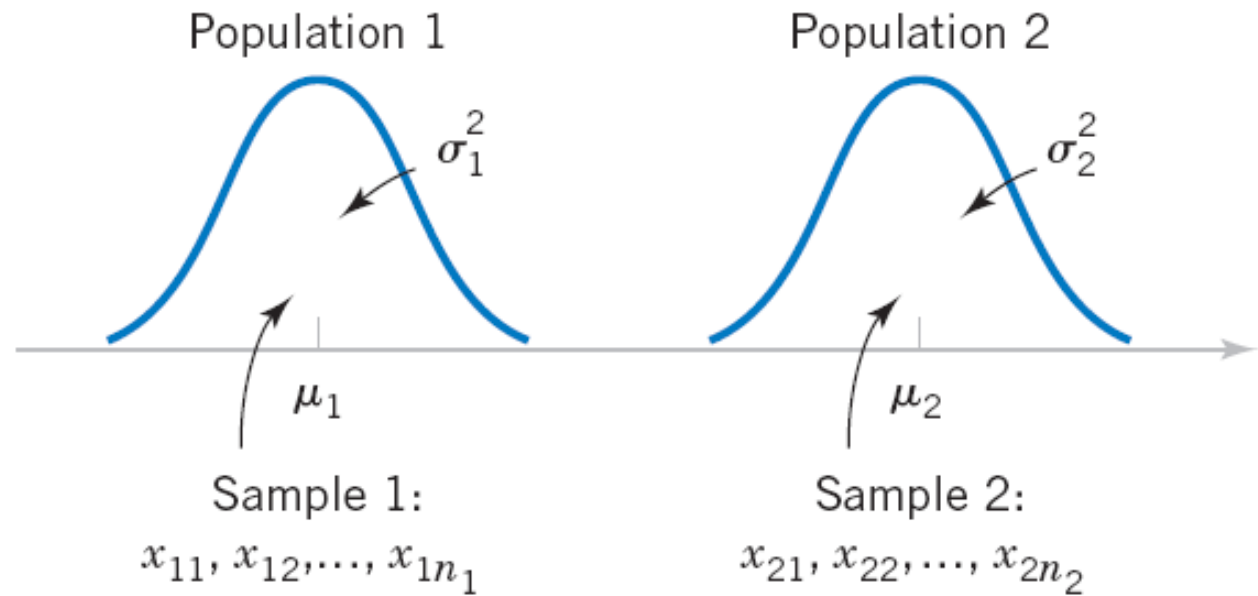


WHY IS PROGRAMMING SO HARD  
WHY IS THERE A 0 OHM RESISTOR  
WHY DO AMERICANS HATE SOCCER  
WHY DO RHYMES SOUND GOOD  
WHY DO TREES DIE  
WHY IS THERE NO SOUND ON CNN  
WHY AREN'T POKEMON REAL  
WHY AREN'T BULLETS SHARP  
WHY DO DREAMS SEEM SO REAL

WHY IS GPS FREE

# Hypothesis testing: two samples

## 10-2: Inference for a Difference in Means of Two Normal Distributions, Variances Known



**Figure 10-1** Two independent populations.

**Figure 10-1** Two independent populations.

## 10-2: Inference for a Difference in Means of Two Normal Distributions, Variances Known

### Assumptions

1.  $X_{11}, X_{12}, \dots, X_{1n_1}$  is a random sample from population 1.
2.  $X_{21}, X_{22}, \dots, X_{2n_2}$  is a random sample from population 2.
3. The two populations represented by  $X_1$  and  $X_2$  are independent.
4. Both populations are normal.

$$E(\bar{X}_1 - \bar{X}_2) = E(\bar{X}_1) - E(\bar{X}_2) = \mu_1 - \mu_2$$

$$V(\bar{X}_1 - \bar{X}_2) = V(\bar{X}_1) + V(\bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

## 10-2: Inference for a Difference in Means of Two Normal Distributions, Variances Known

The quantity

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad (10-1)$$

has a  $N(0, 1)$  distribution.

# 10-2: Inference for a Difference in Means of Two Normal Distributions, Variances Known

## 10-2.1 Hypothesis Tests for a Difference in Means, Variances Known

usually  $\Delta_0 = 0$

Null hypothesis:  $H_0: \mu_1 - \mu_2 = \Delta_0$

Test statistic: 
$$Z_0 = \frac{\bar{X}_1 - \bar{X}_2 - \Delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad (10-2)$$

Alternative Hypotheses	P-Value	Rejection Criterion For for Fixed-Level Tests
$H_1: \mu_1 - \mu_2 \neq \Delta_0$	Probability above $ z_0 $ and probability below $- z_0 $ , $P = 2[1 - \Phi( z_0 )]$	$z_0 > z_{\alpha/2}$ or $z_0 < -z_{\alpha/2}$
$H_1: \mu_1 - \mu_2 > \Delta_0$	Probability above $z_0$ , $P = 1 - \Phi(z_0)$	$z_0 > z_\alpha$
$H_1: \mu_1 - \mu_2 < \Delta_0$	Probability below $z_0$ , $P = \Phi(z_0)$	$z_0 < -z_\alpha$

## 10-2.1 Hypotheses Tests on the Difference in Means, Variances Unknown

### Case 2: $\sigma_1^2 \neq \sigma_2^2$

If  $H_0: \mu_1 - \mu_2 = \Delta_0$  is true, the statistic

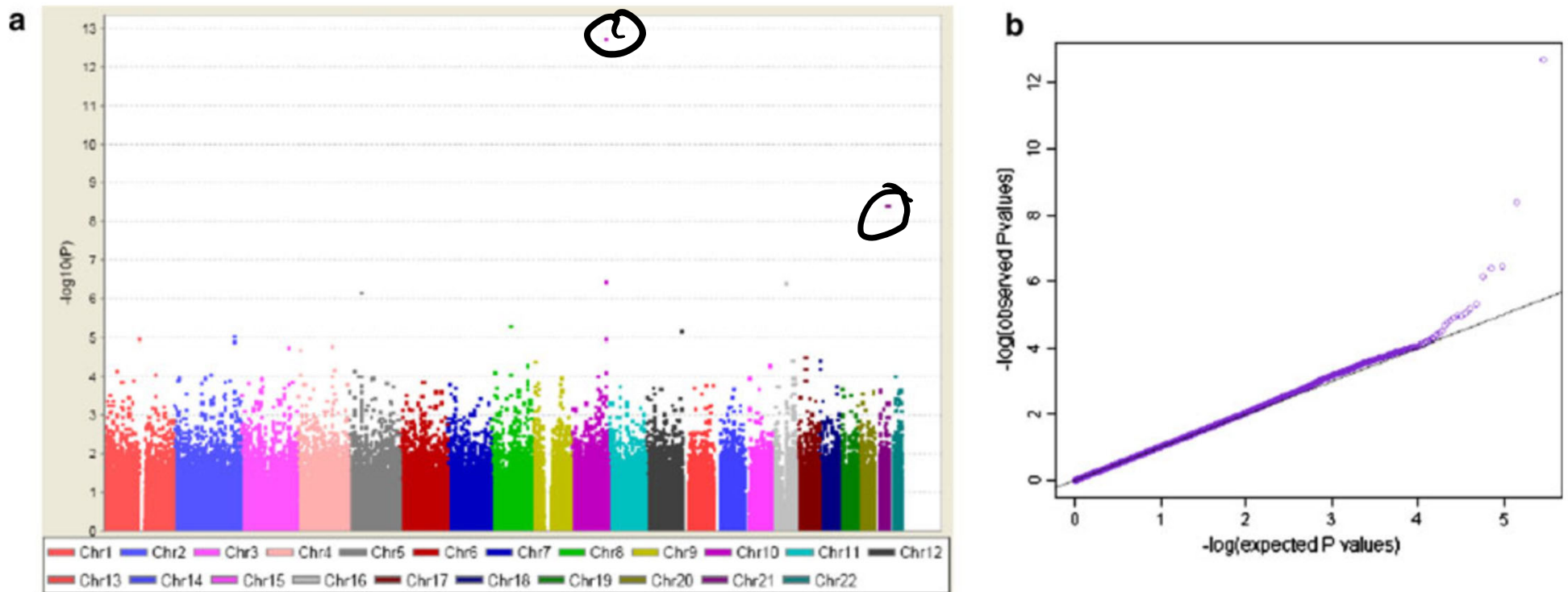
$$T_0^* = \frac{\bar{X}_1 - \bar{X}_2 - \Delta_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \quad (10-15)$$

is distributed as **t-distribution** with degrees of freedom given by

$$v = n_1 + n_2 - 2,$$

or more generally

# Manhattan plot for Genome-Wide Association Study (GWAS)



**Fig. 1** Genome-wide association results comparing 2,702 cases and 5,726 controls: **a** Manhattan and **b** quantile–quantile plots of  $-\log_{10}$  transformed  $P$  values of 285,984 SNPs genotyped

Li J., et al. A combined analysis of genome-wide association studies in breast cancer. *Breast Cancer Res Treat.* 2011;126: 717–727

# Multiple null hypotheses: Bonferroni correction

- What if you have **m independent null hypotheses**?  
Say you have **m=25,000 genes** in a genome?
- What is the probability that **at least one** of the **null-hypotheses** will be shown to be **false** at significance threshold  $\alpha_1$ ?
- Answer:  
Family-Wise Error Rate  
or  **$FWER=1-(1-\alpha_1)^m \approx m\alpha_1$**
- If  $m=20$  and  $\alpha_1=0.05$ ,  
 **$FWER=0.6415$**
- If you want to get  **$FWER < \alpha$** , use  
 **$\alpha_1 = \alpha/m$**

**Carlo Emilio Bonferroni**  
(1892 –1960)  
Italian mathematician  
who worked on  
probability theory.

