Biostatistics

Chapter 4 Hypothesis Testing II (假设检验)

Jing Li

jing.li@sjtu.edu.cn

http://cbb.sjtu.edu.cn/~jingli/courses/2017fall/bi372/

Dept of Bioinformatics & Biostatistics, SJTU

Review Questions (5 min)

- What's hypothesis test, and what's p-value ?

Review lecture 4 (I)

- Sampling Variability and confidence interval
- Hypothesis testing
- P-values

The standard deviation of change scores was 2.0 in both groups.

Results:

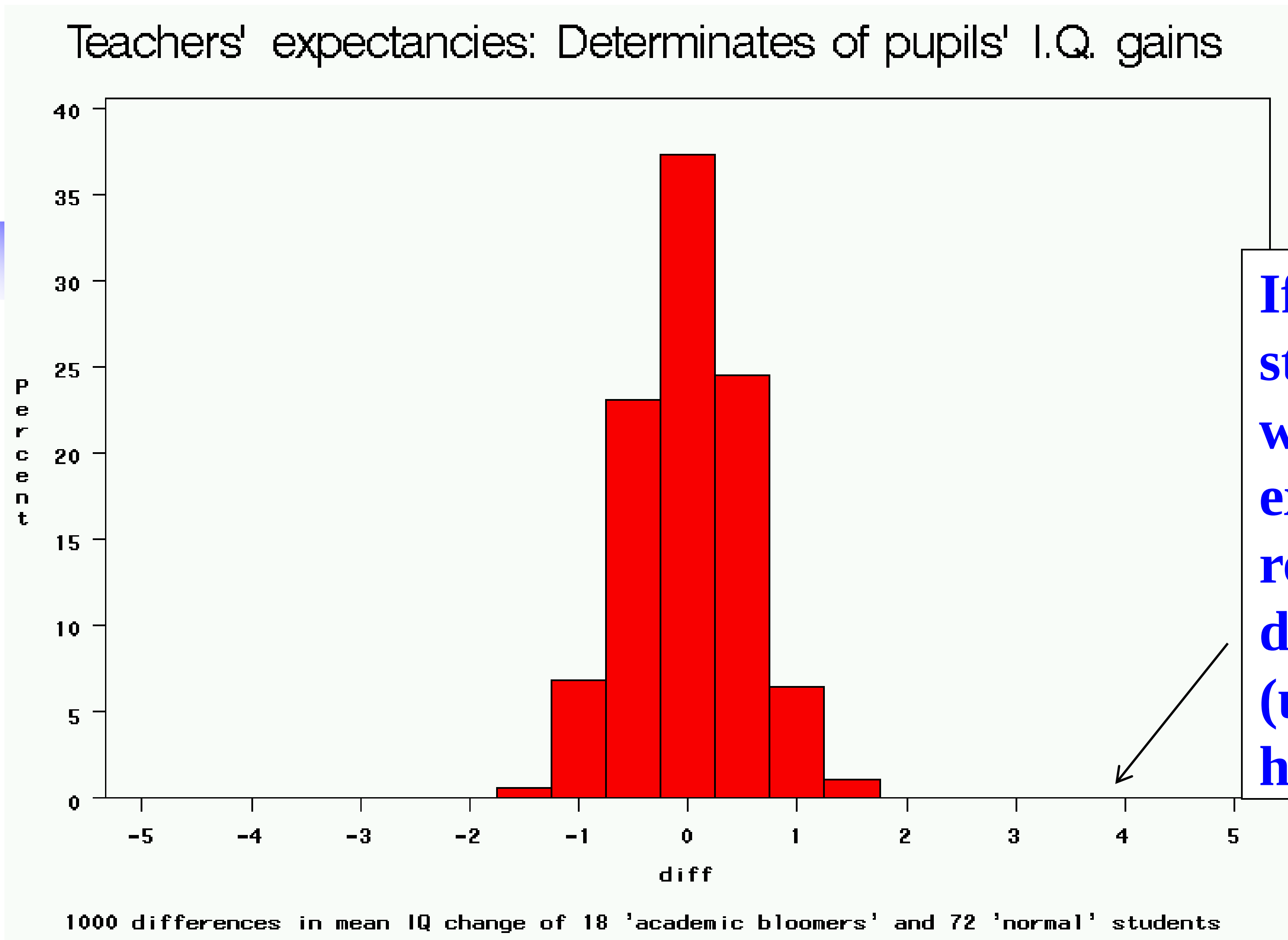
	"Academic bloomers" (n=18)	Controls (n=72)
Change in IQ score:	12.2 (2.0)	8.2 (2.0)

12.2 points

8.2 points

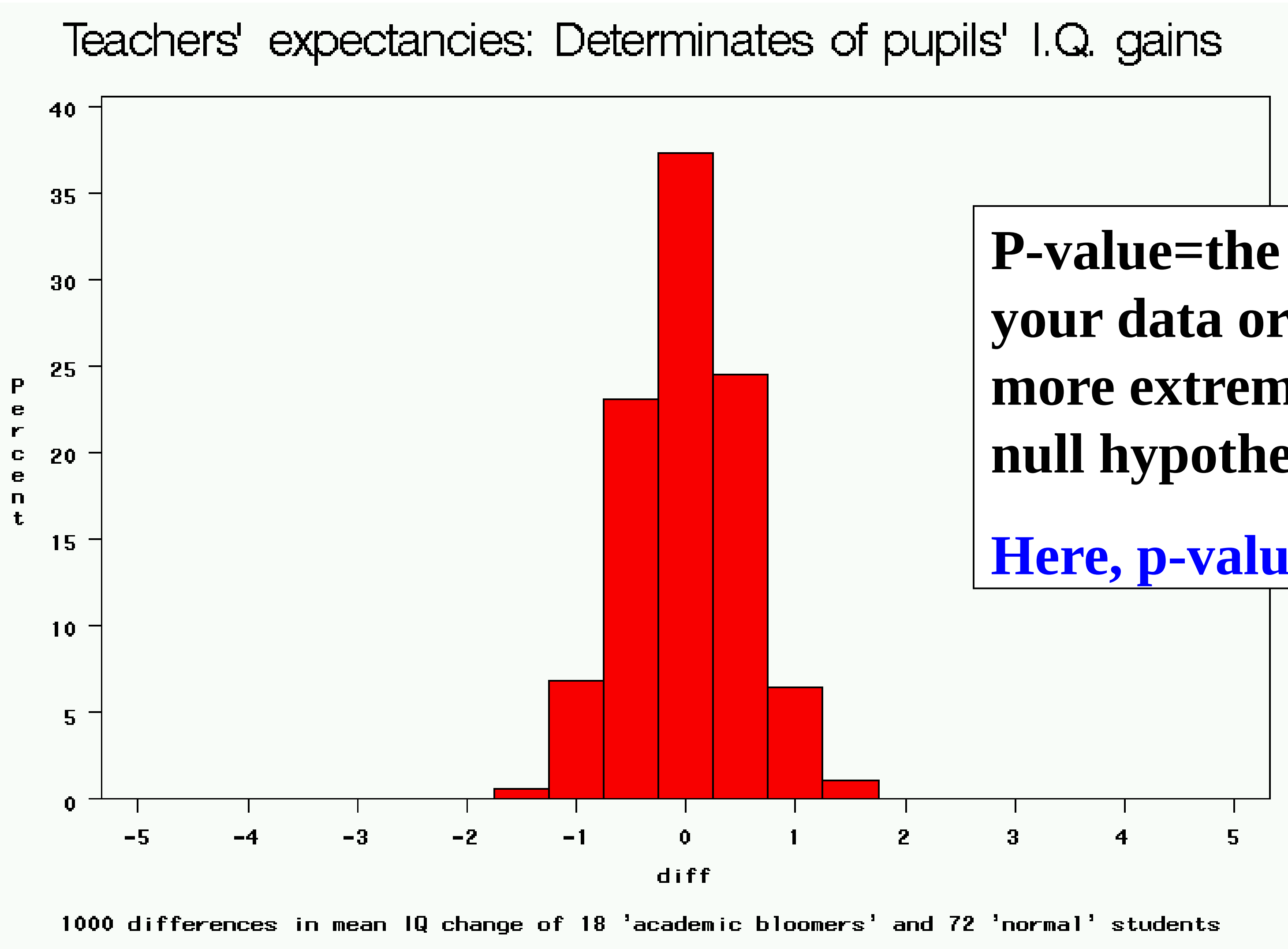
Difference=4 points

Getting the P-value from computer simulation…



If we ran this study 1000 times we wouldn't expect to get 1 result as big as a difference of 4 (under the null hypothesis).

P-value

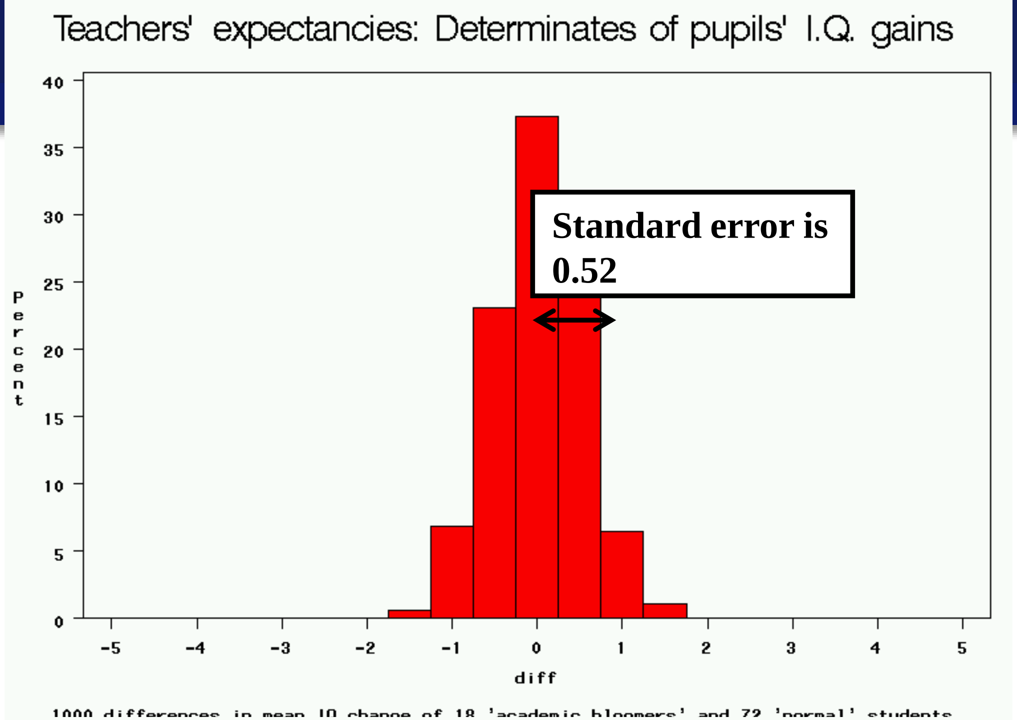


P-value=the probability of your data or something more extreme under the null hypothesis.

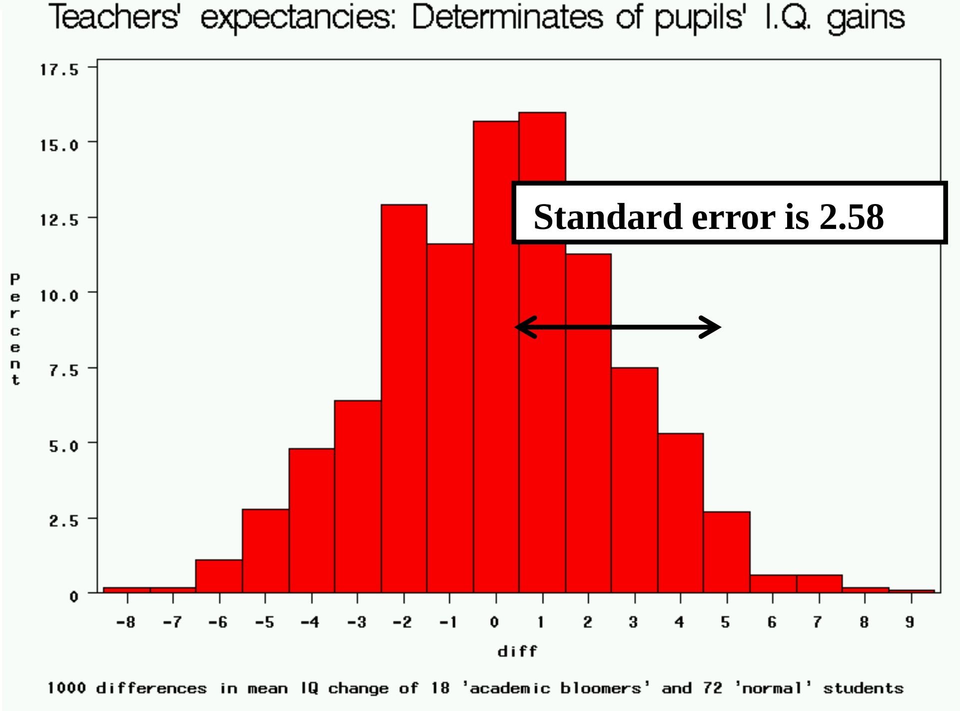
Here, p-value<.0001

What if our standard deviation had been higher?

- The standard deviation for change scores in both treatment and control was 2.0. What if change scores had been much more variable—say a standard deviation of 10.0?

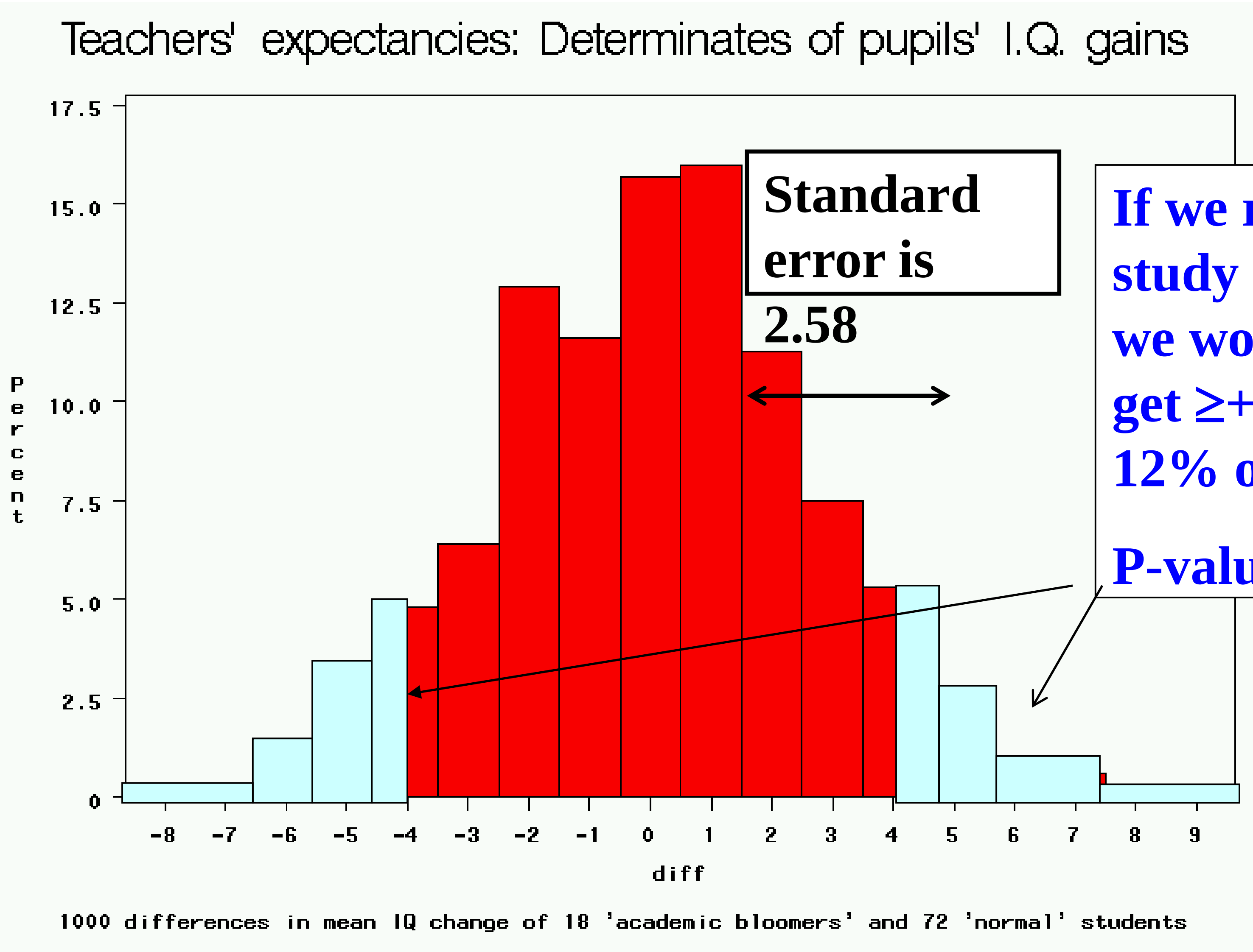


Std. dev in change scores = 2.0



Std. dev in change scores = 10.0

With a std. dev. of 10.0…



If we ran this study 1000 times, we would expect to get ≥+4.0 or ≤–4.0 12% of the time.

P-value=.12

What would a 4.0 difference mean (std. dev=10)?

- Is it "statistically significant"? NO

No evidence of an effect ≠ Evidence of no effect.

Hypothesis testing summary

- Null hypothesis: the hypothesis of no effect (usually the opposite of what you hope to prove). The straw man you are trying to shoot down.
    - Example: antidepressants *have no effect on* suicide risk

- P-value: the probability of your observed data if the null hypothesis is true.
    - If the p-value is low enough (i.e., if our data are very unlikely given the null hypothesis), this is evidence that the null hypothesis is wrong.
    - If p-value is low enough (typically <.05), we reject the null hypothesis and conclude that antidepressants do have an effect.

Steps for hypothesis testing

1) State null and alternative hypotheses
2) State type of test and alpha level
3) Determine and calculate appropriate test statistic
4) Calculate p-value
5) Decide whether to reject or not reject the null hypothesis
   - ✓ NEVER accept null !!!
6) Write conclusion

Error and power

- **Type I error rate (or significance level)**: the probability of finding an effect that isn't real (false positive).
  - *If we require **p-value<.05** for statistical significance, this means that 1/20 times we will find a positive result just by chance.*
- **Type II error rate**: the probability of missing an effect (false negative).
- **Statistical power**: the probability of finding an effect if it is there (the probability of not making a type II error).
  - *When we design studies, we typically aim for a power of 80% (allowing a false negative rate, or type II error rate, of 20%).*

Type I and Type II Error

Your Statistical Decision	True state of null hypothesis	
	H_0 True (e.g., antidepressants do not increase suicide risk)	**H_0 False** (e.g., antidepressants do increase suicide risk)
Reject H_0 (e.g., conclude antidepressants increase suicide risk)	*Type I error (α)*	Correct
Do not reject H_0 (e.g., conclude there is insufficient evidence that antidepresssants increase suicide risk)	Correct	*Type II Error (β)*

Group discussion

Type I error (α) **vs** *Type II Error (β)*

Type I and Type II Error

Your Statistical Decision	True state of null hypothesis	
	H_0 True	H_0 False
Reject H_0	*Type I error (α)* *Big mistake* **去真**	*Correct*
Do not reject H_0	*Correct*	*Type II Error (β)* *Minor mistake* **存伪**

Statistical Power

- Statistical power is the probability of finding an effect if it's real.

$$\text{Power} = 1 - \beta$$

Can we quantify how much power we have for given sample sizes?

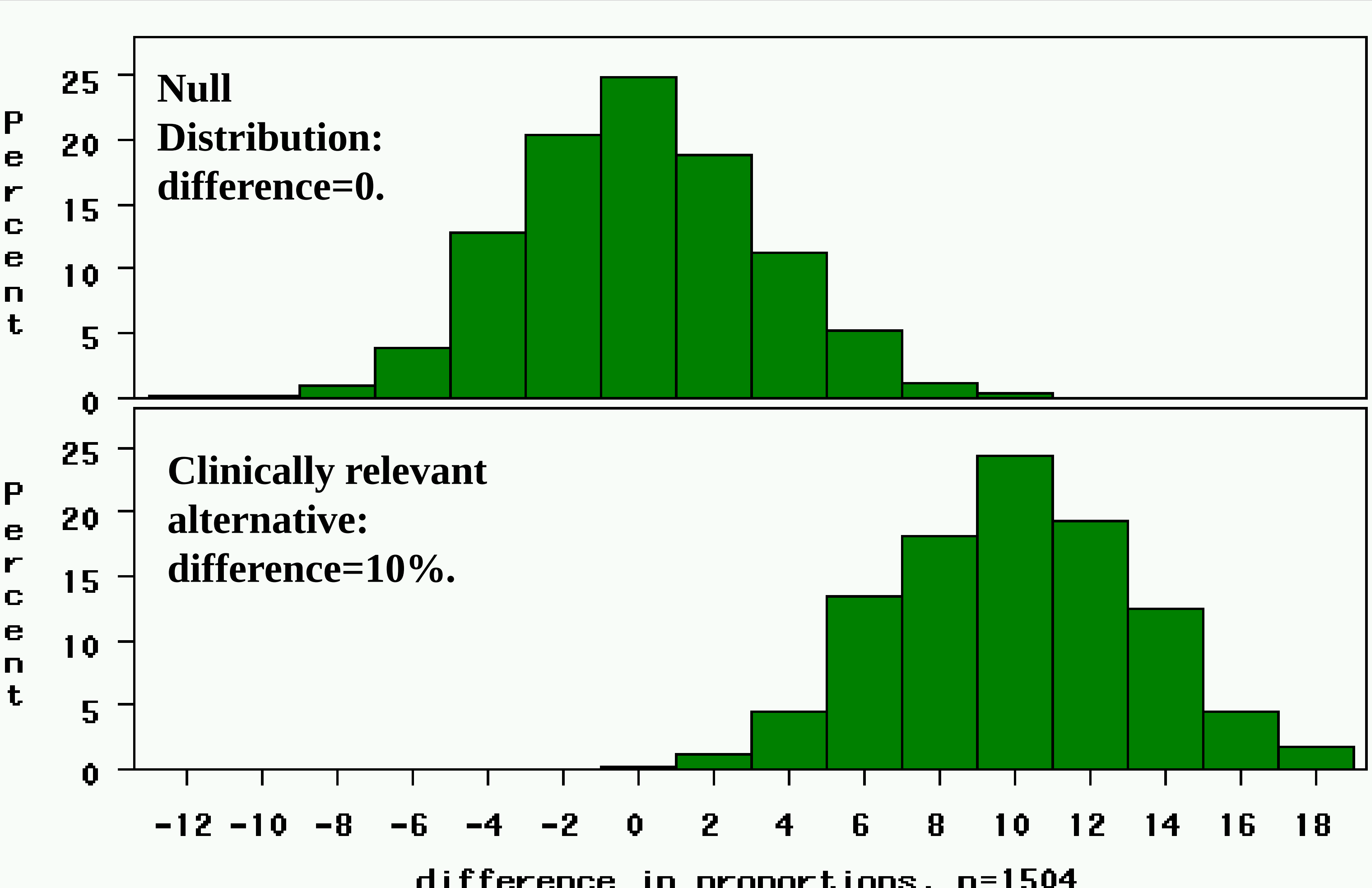
Case-control study 1

Are antidepressants a risk factor for suicide attempts in children and adolescents?

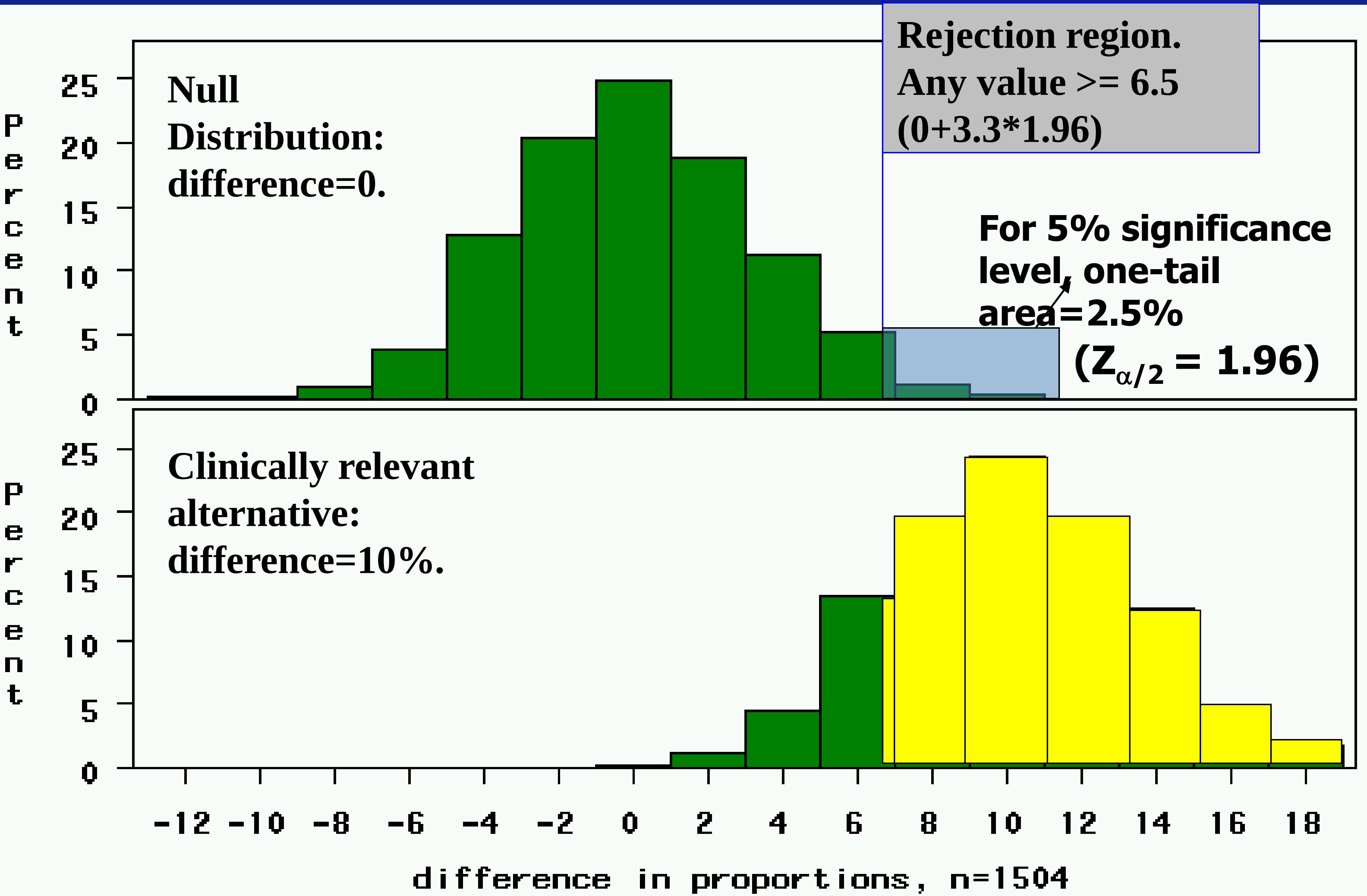
	Cases (n=263)	Controls (n=1241)
	120 (46%)	448 (36%)
	46%	**36%**

Difference=10%

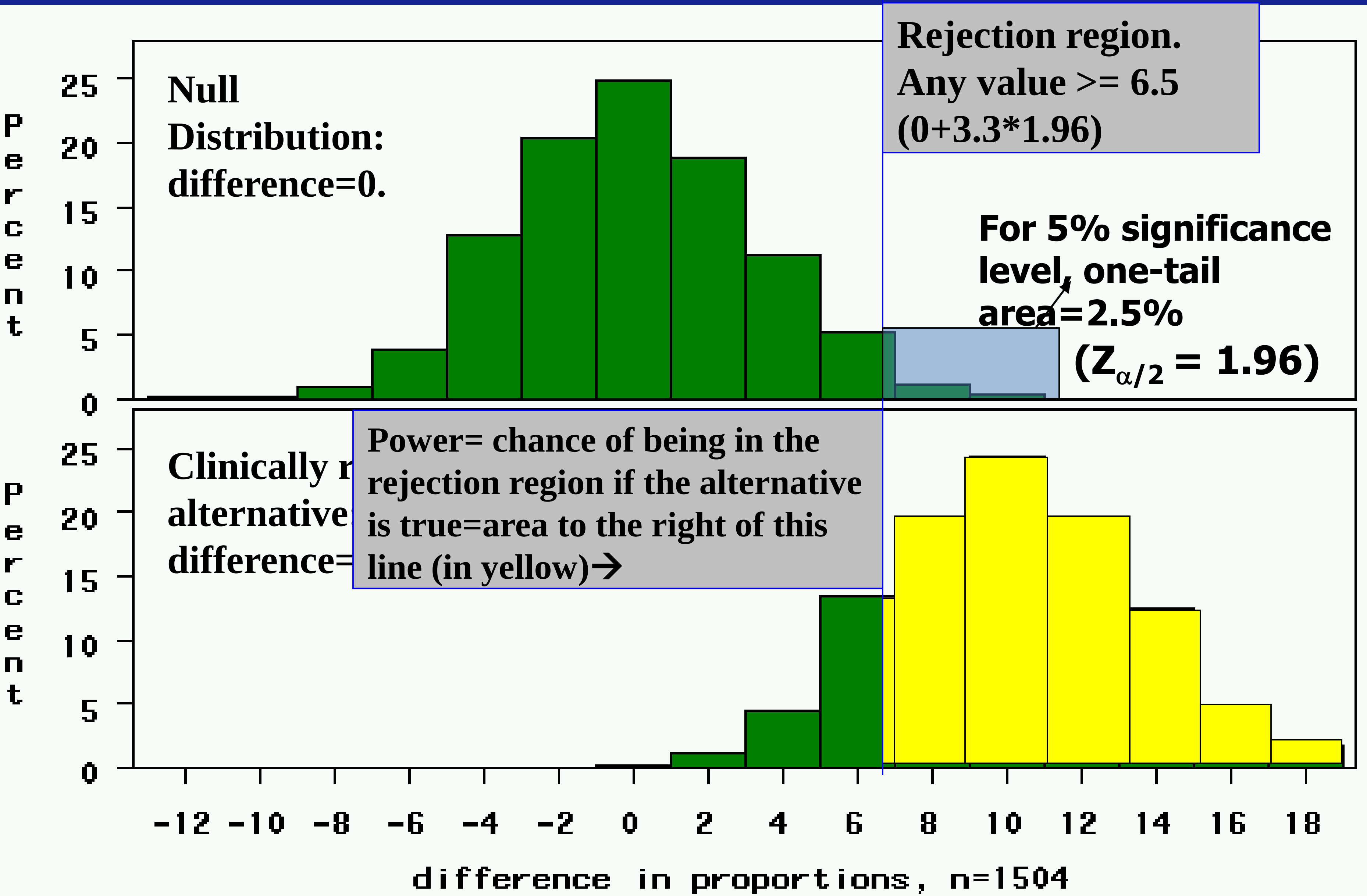
Study 1: 263 cases, 1241 controls



Study 1: 263 cases, 1241 controls



Study 1: 263 cases, 1241 controls

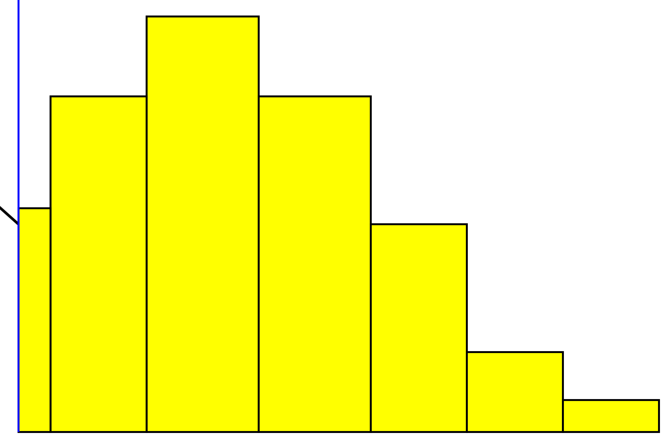


Study 1: 263 cases, 1241 controls

Rejection region. Any value >= 6.5 (0+3.3*1.96)

Power= chance of being in the rejection region if the alternative is true=area to the right of this line (in yellow)

Power here = >80%



study 1: 263 cases, 1241 controls

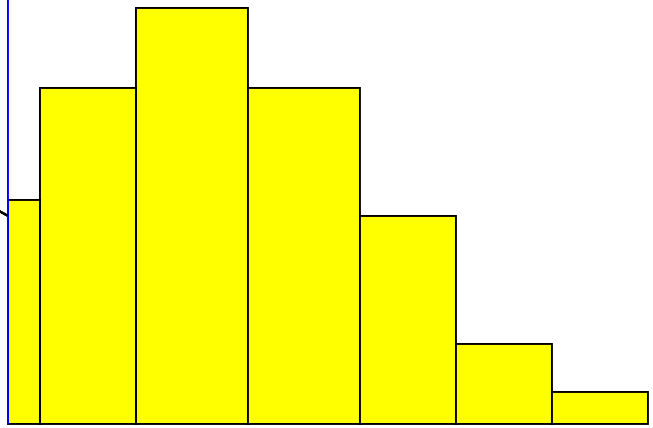
Rejection region. Any value >= 6.5 (0+3.3*1.96)

Power= chance of being in the rejection region if the alternative is true=area to the right of this line (in yellow)

Power here:

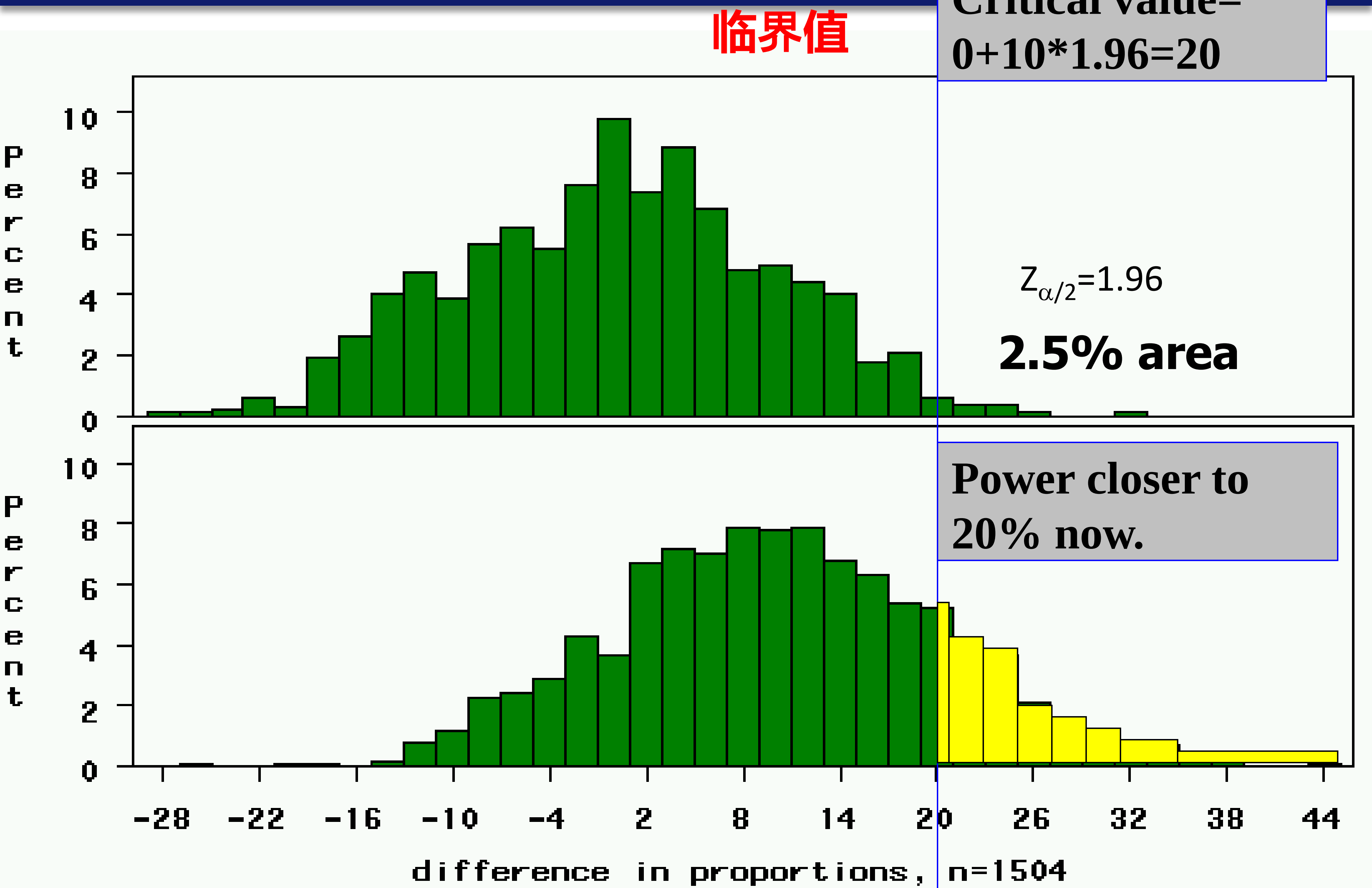
$$\mathbf{P(Z > \frac{6.5 - 10}{3.3}) =}$$

$$\mathbf{P(Z > -1.06) = 85\%}$$

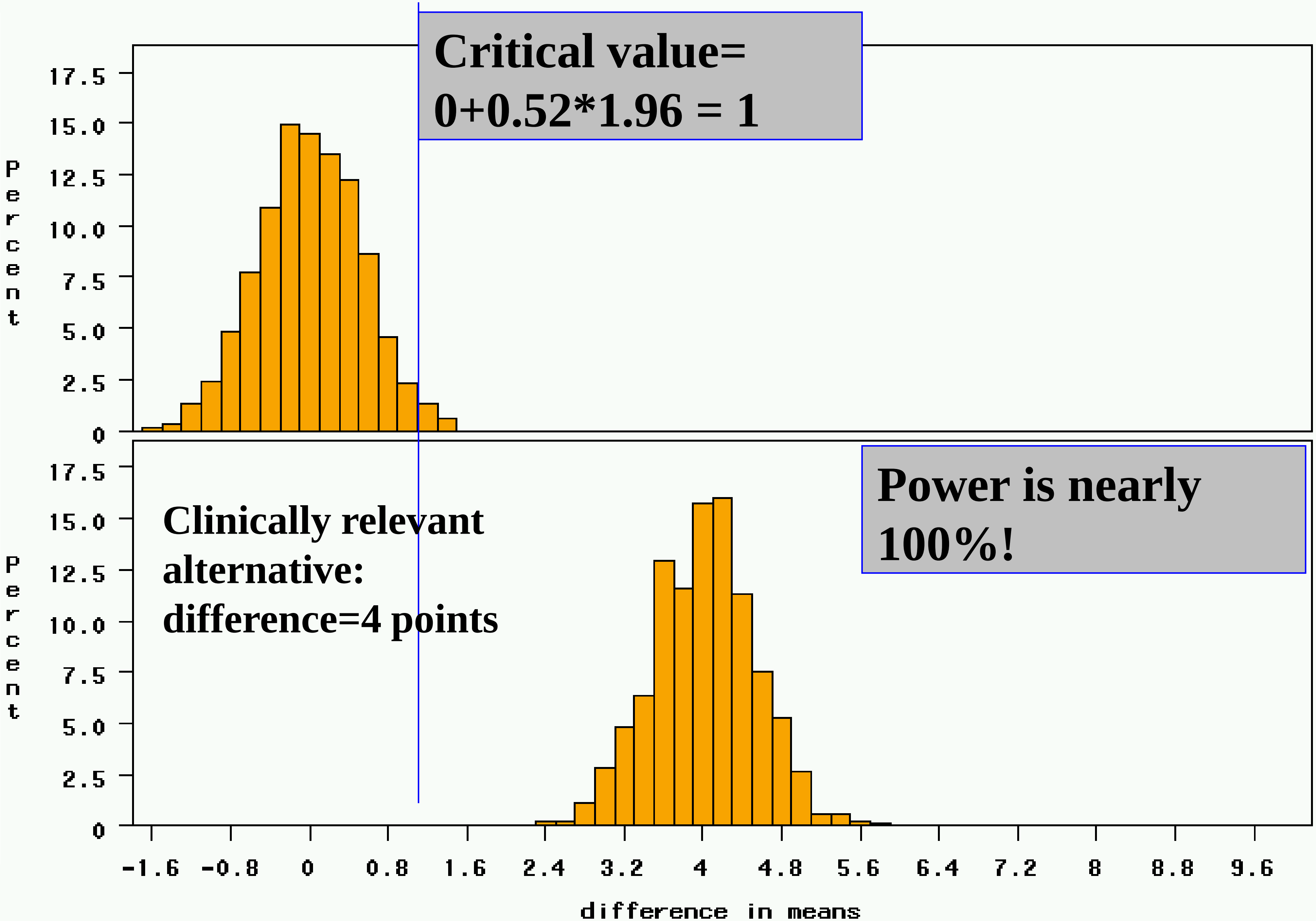


Study 1: 50 cases, 50 controls

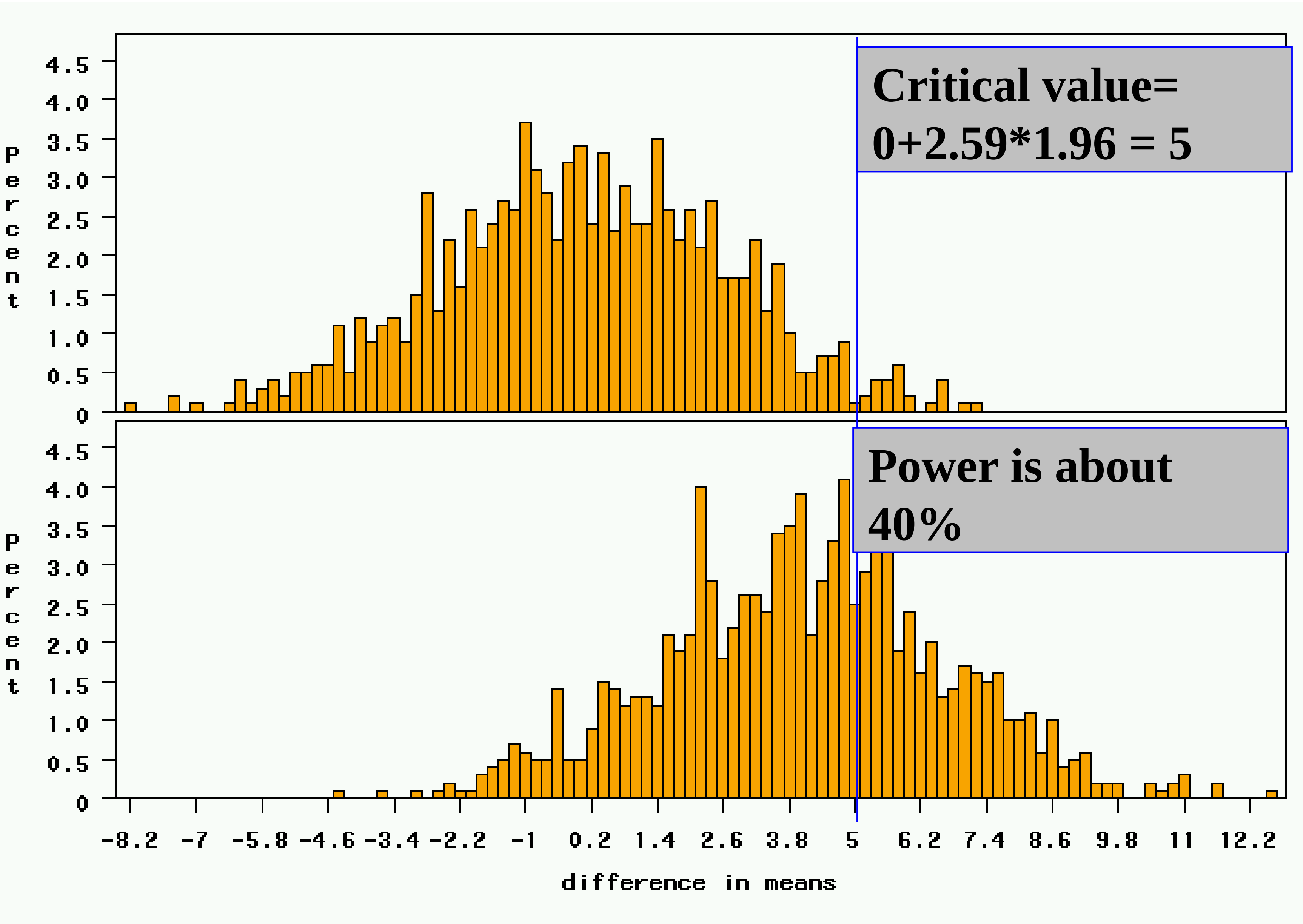
Critical value= 0+10*1.96=20



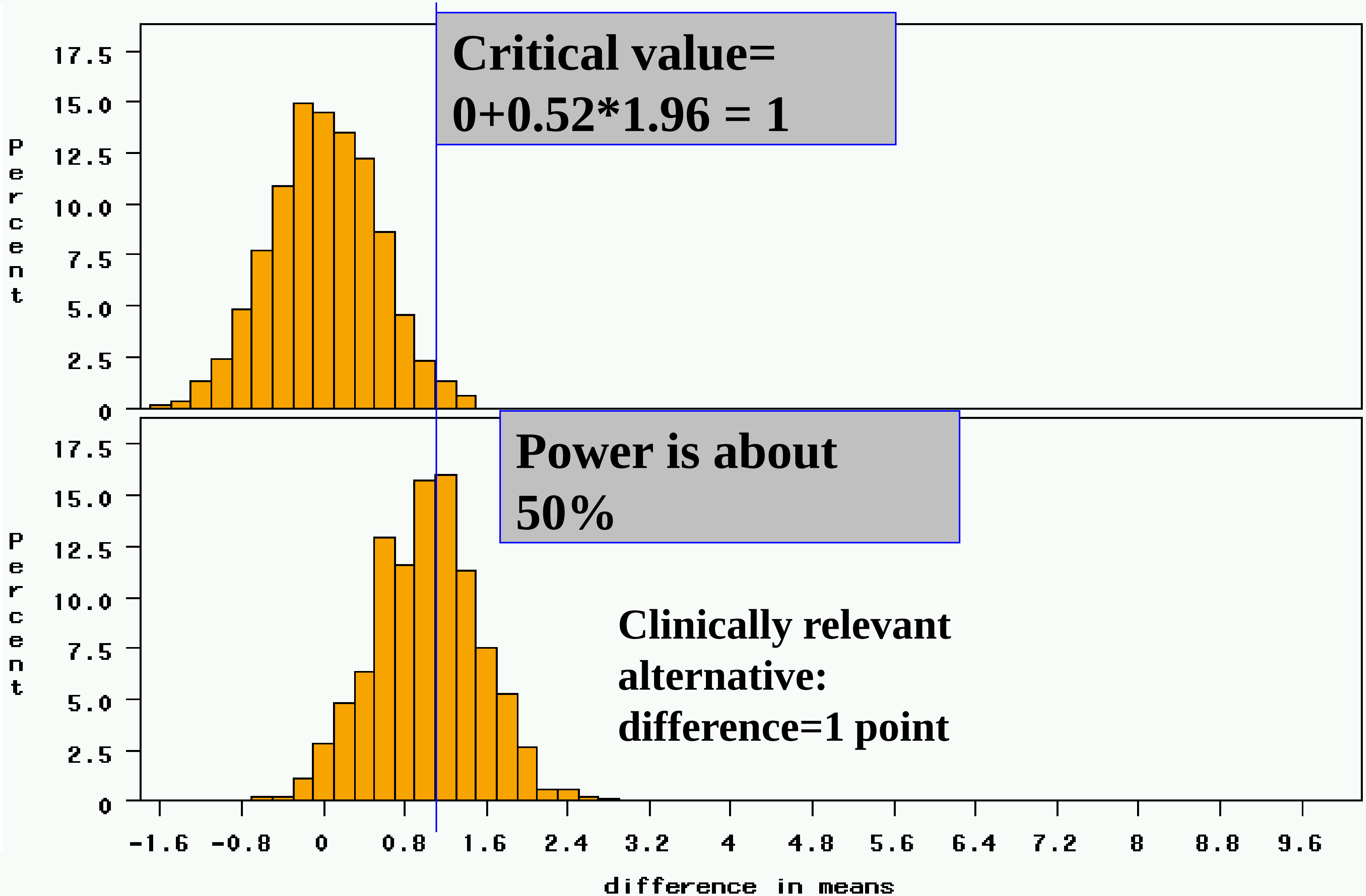
Study 2: 18 treated, 72 controls, STD DEV = 2



Study 2: 18 treated, 72 controls, STD DEV=10



Study 2: 18 treated, 72 controls, effect size=1.0



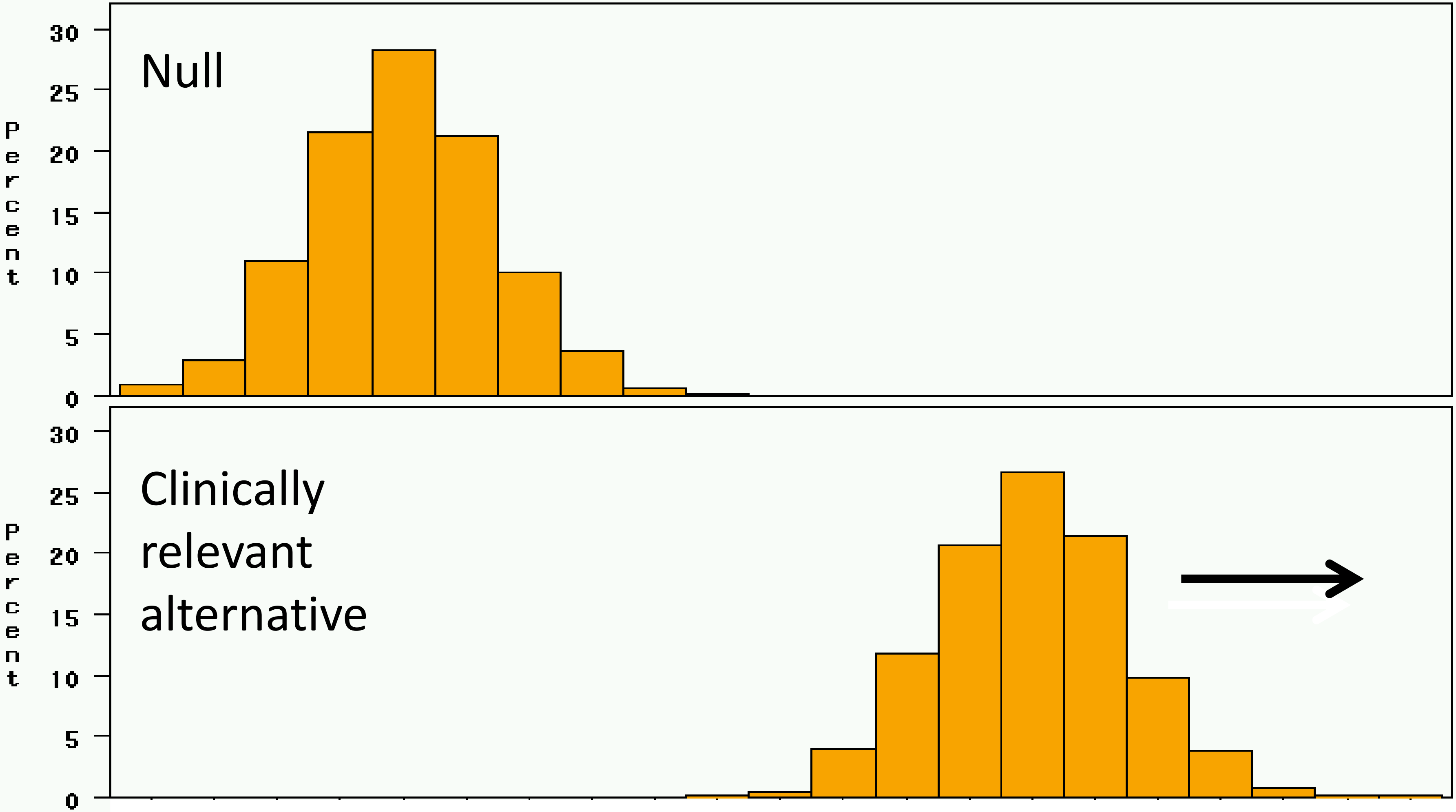
Group discussion (5 min)

What's possible factor affecting statistics power ?

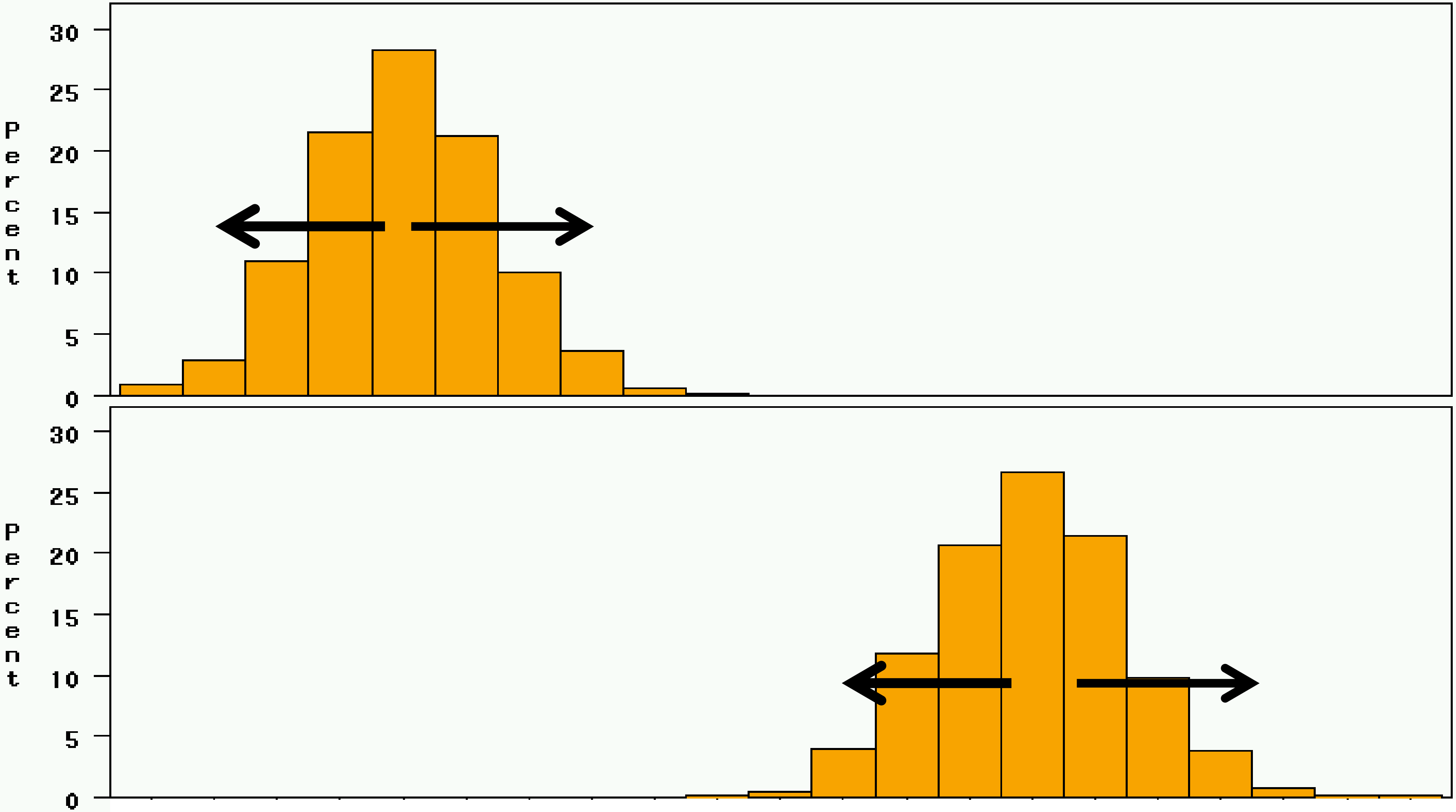
Factors Affecting Power

1. Size of the effect $\uparrow$
2. Standard deviation of the characteristic $\downarrow$
3. Bigger sample size $\uparrow$
4. Significance level desired $\downarrow$

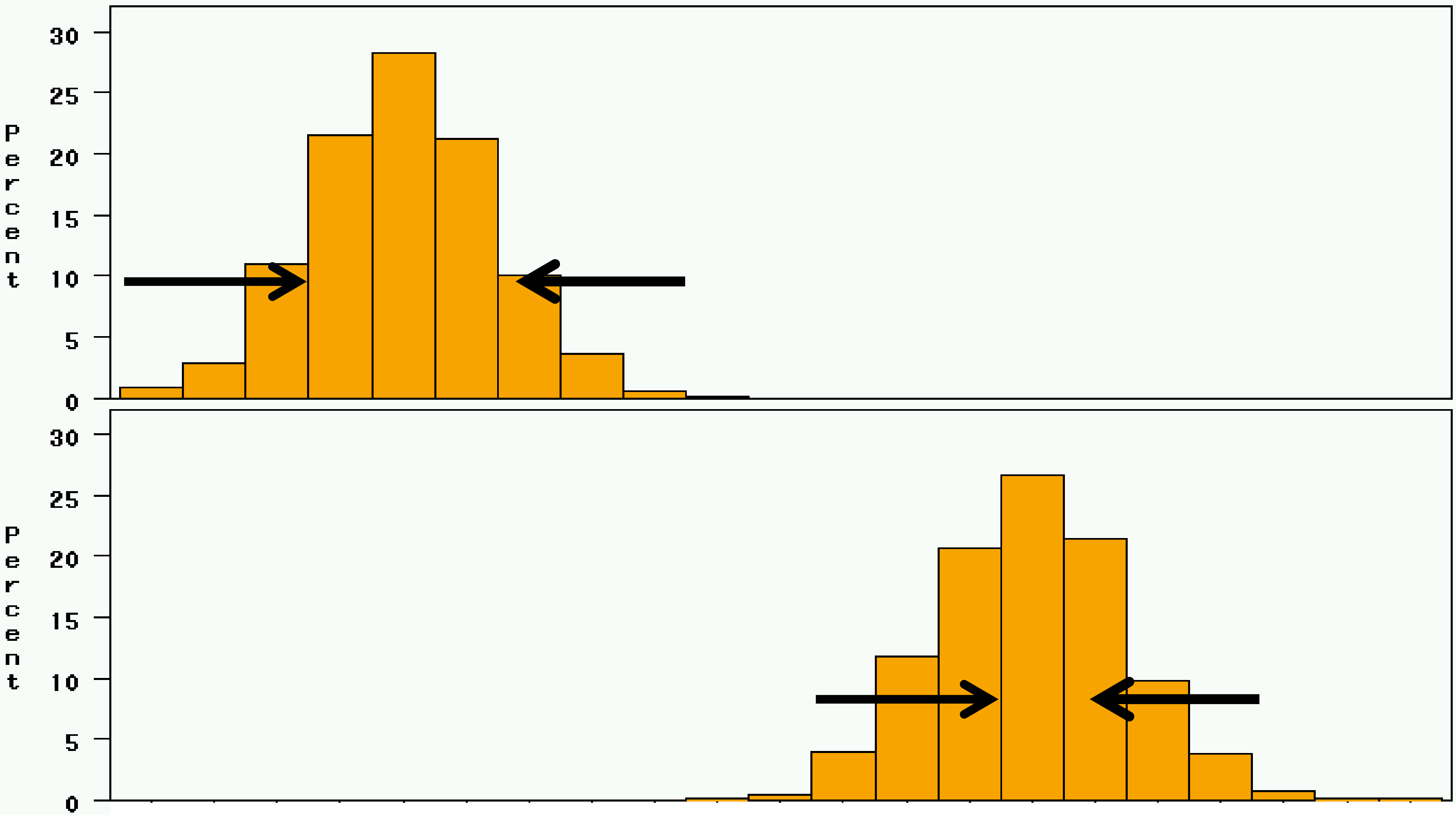
1. Bigger difference from the null mean



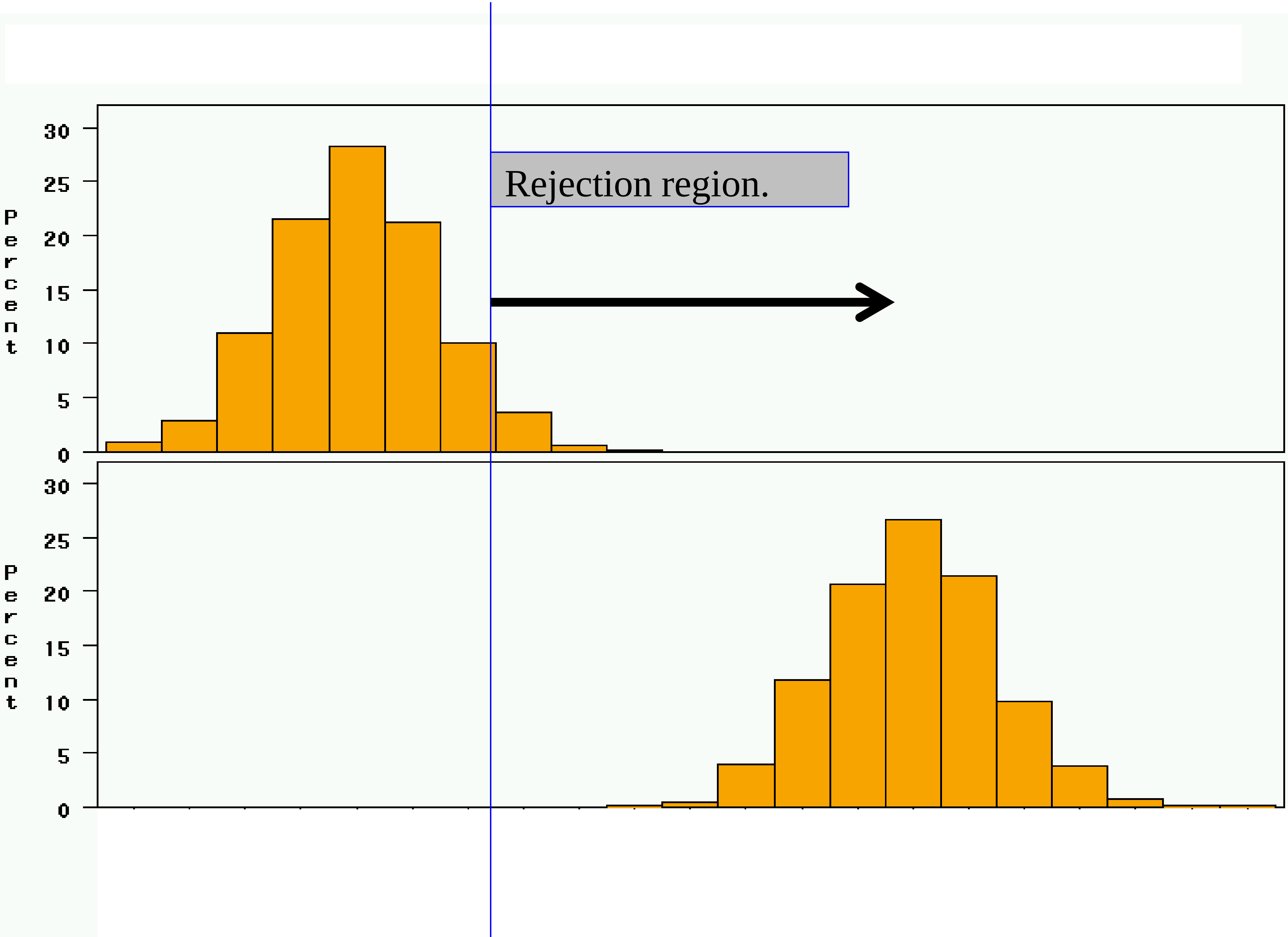
2. Bigger standard deviation



3. Bigger Sample Size



4. Higher significance level



Sample size calculations

- Based on these elements, you can write a formal mathematical equation that relates power, sample size, effect size, standard deviation, and significance level…

Simple formula for difference in proportions

Represents the desired power (typically .84 for 80% power).

Sample size in each group (assumes equal sized groups)

$$\boxed{n} = 2 \times \frac{\boxed{(\bar{p})(1-\bar{p})}\boxed{(Z_{\beta}} + \boxed{Z_{\alpha/2}})^2}{\boxed{(p_1 - p_2)^2}}$$

A measure of variability (similar to standard deviation)

Effect Size (the difference in proportions)

Represents the desired level of statistical significance (typically 1.96).

Simple formula for difference in means

Sample size in each group (assumes equal sized groups)

Represents the desired power (typically .84 for 80% power).

$$n = 2 \times \frac{s^2 (Z_b + Z_{a/2})^2}{\text{difference}^2}$$

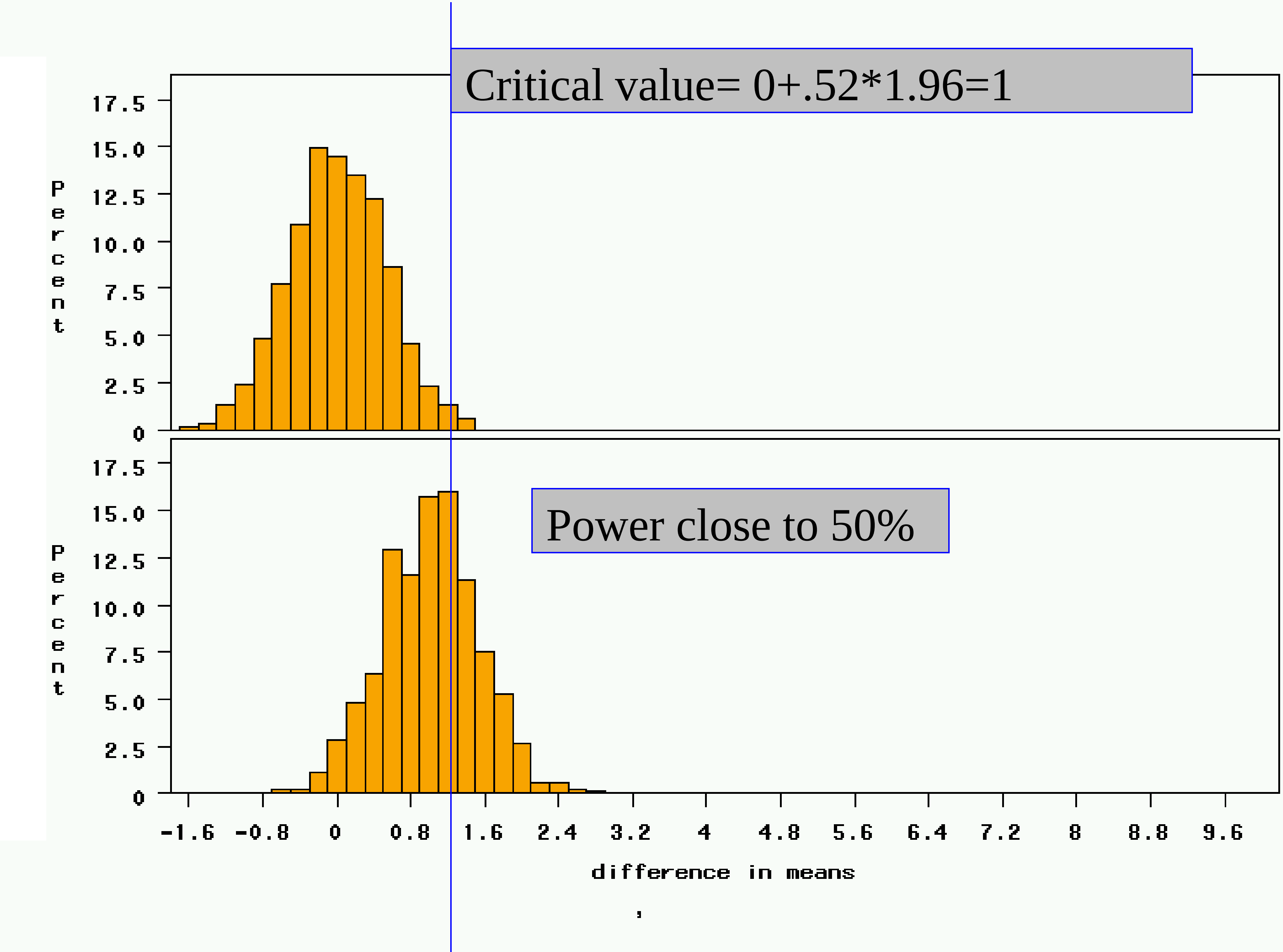
Standard deviation of the outcome variable

Effect Size (the difference in means)

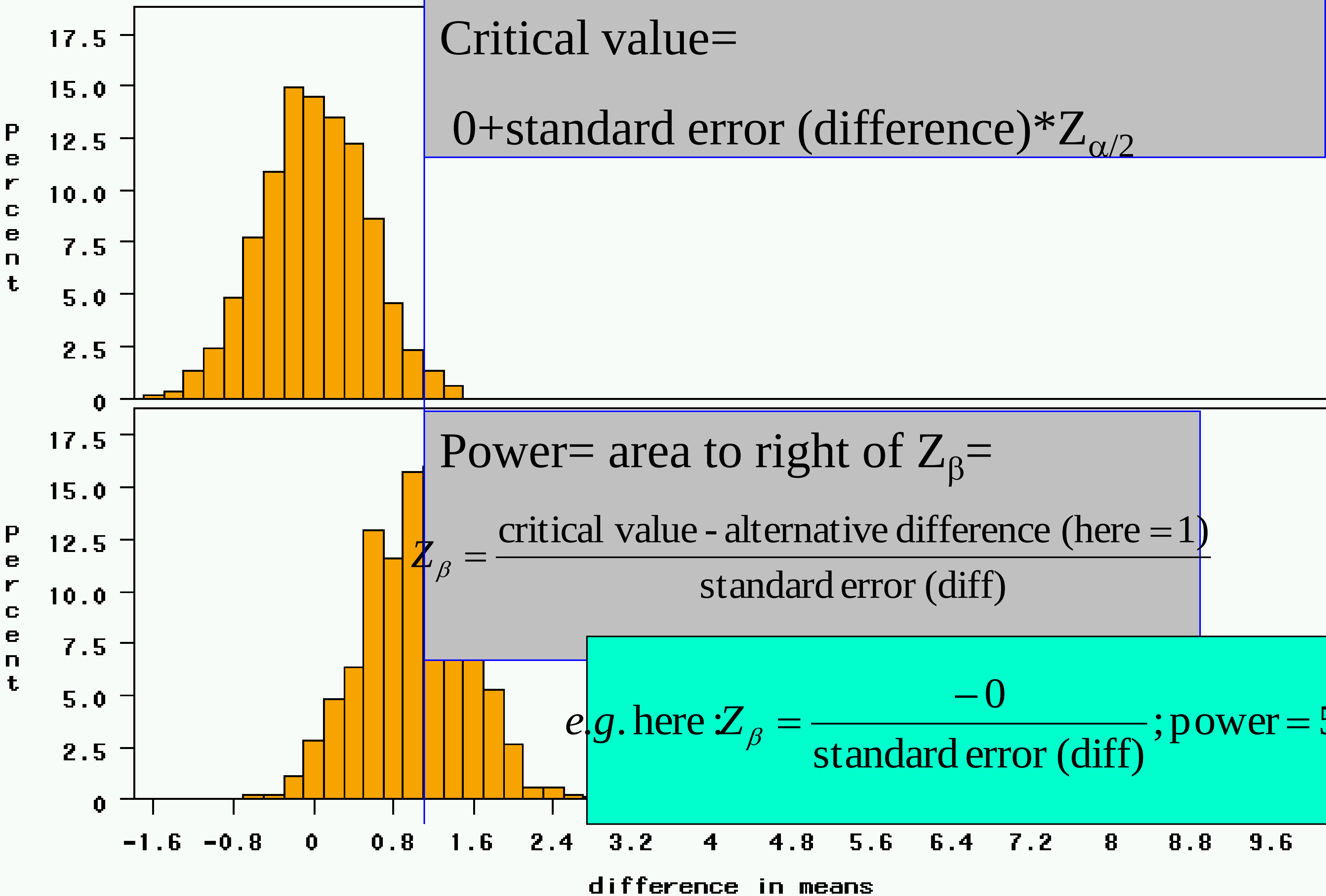
Represents the desired level of statistical significance (typically 1.96).

Derivation of sample size formula….

Study 2: 18 treated, 72 controls, effect size=1.0



SAMPLE SIZE AND POWER FORMULAS



Critical value=

0+standard error (difference)*$Z_{\alpha/2}$

Power= area to right of Z_β=

$$Z_\beta = \frac{\text{critical value - alternative difference (here = 1)}}{\text{standard error (diff)}}$$

$$e.g.\ \text{here}: Z_\beta = \frac{-0}{\text{standard error (diff)}};\ \text{power} = 50\%$$

Power= area to right of Z_β=

$$Z_\beta = \frac{\text{critical value - alternative difference}}{\text{standard error (diff)}}$$

$$Z_\beta = \frac{Z_{\alpha/2} * \text{standard error (diff) - difference}}{\text{standard error(diff)}}$$

$$Z_\beta = Z_{\alpha/2} - \frac{\text{difference}}{\text{standard error(diff)}}$$

$$-Z_\beta = \frac{\text{difference}}{\text{standard error(diff)}} - Z_{\alpha/2}$$

$$Z_{power} = -Z_\beta \longrightarrow$$

$$\text{the area to the left of } Z_{power} = \text{the area to the right of } Z_\beta$$

Power is the area to the *right* of Z_β. OR power is the area to the *left* of - Z_β.
Since normal charts give us the area to the left by convention, we need to use
- Z_β to get the correct value. Most textbooks just call this "Z_β"; I'll use the term Z_{power} to avoid confusion.

All-purpose power formula…

$$Z_{power} = \frac{\text{difference}}{\text{standard error(difference)}} - Z_{\alpha/2}$$

Derivation of a sample size formula…

$$s.e.(diff) = \sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}$$

Sample size is embedded in the standard error….

$$\text{if ratio } r \text{ of group 2 to group 1}: s.e.(diff) = \sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{rn_1}}$$

Algebra…

$$\therefore Z_{power} = \frac{\text{difference}}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{rn_1}}} - Z_{\alpha/2}$$

$$Z_{power} = \frac{\text{difference}}{\sqrt{\frac{(r+1)\sigma^2}{rn_1}}} - Z_{\alpha/2}$$

$$(Z_{power} + Z_{\alpha/2})^2 = \left(\frac{\text{difference}}{\sqrt{\frac{(r+1)\sigma^2}{rn_1}}}\right)^2$$

$$(r+1)\sigma^2(Z_{power}+Z_{\alpha/2})^2 = rn_1\text{difference}^2$$

$$rn_1\text{difference}^2 = (r+1)\sigma^2(Z_{power}+Z_{\alpha/2})^2$$

$$n_1 = \frac{(r+1)\sigma^2(Z_{power}+Z_{\alpha/2})^2}{r\text{difference}^2}$$

$$\boxed{n_1 = \frac{(r+1)}{r}\frac{\sigma^2(Z_{power}+Z_{\alpha/2})^2}{\text{difference}^2}}$$

$$\text{If r} = 1\text{(equal groups), then } n_1 = \frac{2\sigma^2(Z_{power}+Z_{\alpha/2})^2}{\text{difference}^2}$$

Sample size formula for difference in means

$$n_1 = \frac{(r+1)}{r} \frac{\sigma^2 (Z_{power} + Z_{\alpha/2})^2}{\text{difference}^2}$$

where:

n_1 = size of smaller group

r = ratio of larger group to smaller group

σ = standard deviation of the characteristic

diffference = clinically meaningful difference in means of the outcome

Z_{power} = corresponds to power (.84 = 80% power)

$Z_{\alpha/2}$ = corresponds to two - tailed significance level (1.96 for $\alpha = .05$)

Sample size needed for comparing two proportions

Example: I am going to run a case-control study to determine if pancreatic (胰腺) cancer is linked to drinking coffee. If I want **80% power** to detect a **10% difference** in the proportion of coffee drinkers among cases vs. controls (if coffee drinking and pancreatic cancer are linked, we would expect that a higher proportion of cases would be coffee drinkers than controls, **α=0.05**), how many cases and controls should I sample? About **half the population drinks coffee**.

Sample size

$$n = 2 \times \frac{(\overline{p})(1-\overline{p})(Z_{\beta} + Z_{\alpha/2})^2}{(p_1 - p_2)^2}$$

For 80% power…

$$n = \frac{.5(.84 + 1.96)^2}{.10^2} = 392$$

There is 80% area to the left of a Z-score of .84 on a standard normal curve; therefore, there is 80% area to the right of -.84.

Would take 392 cases and 392 controls to have 80% power!
Total=784

For 90%, 95%, 99% power...

Sample size calculators on the web…

- http://biostat.mc.vanderbilt.edu/twiki/bin/view/Main/PowerSampleSize
- http://calculators.stat.ucla.edu
- http://hedwig.mgh.harvard.edu/sample_size/size.html

A group of investigators are studying a treatment that can reduce LDL Cholesterol level. the standard deviation of the reduction in LDL for this population from this treatment is around 20. Find the sample size so that one can have a 90% power to detect a 10 units average reduction in LDL (i.e., effect size of 10 units) at 5% level of significance for one-sided test.

Two-sided test

$$n = 2 \times \frac{\sigma^2 (Z_\beta + Z_{\alpha/2})^2}{\text{difference}^2} = 2\text{x} \ \frac{20^2 \times (1.28 + 1.96)^2}{10^2} = 83.98$$

One-sided test ?