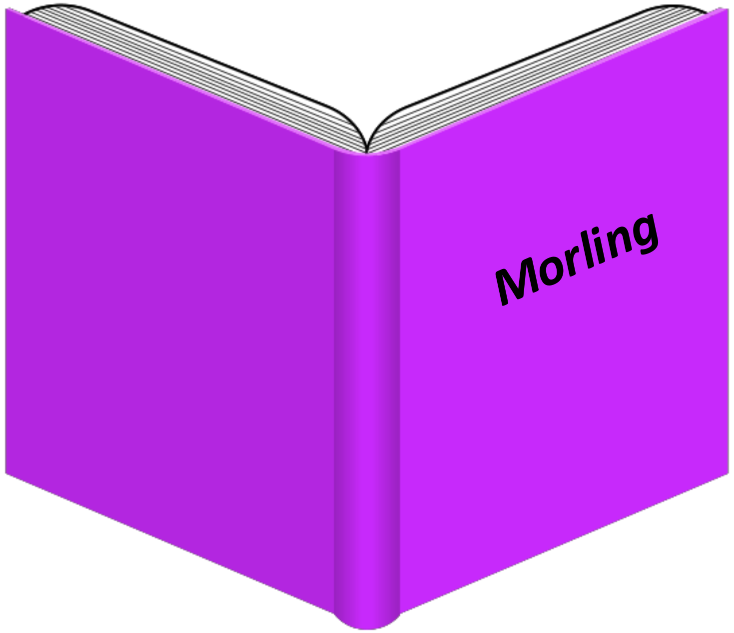
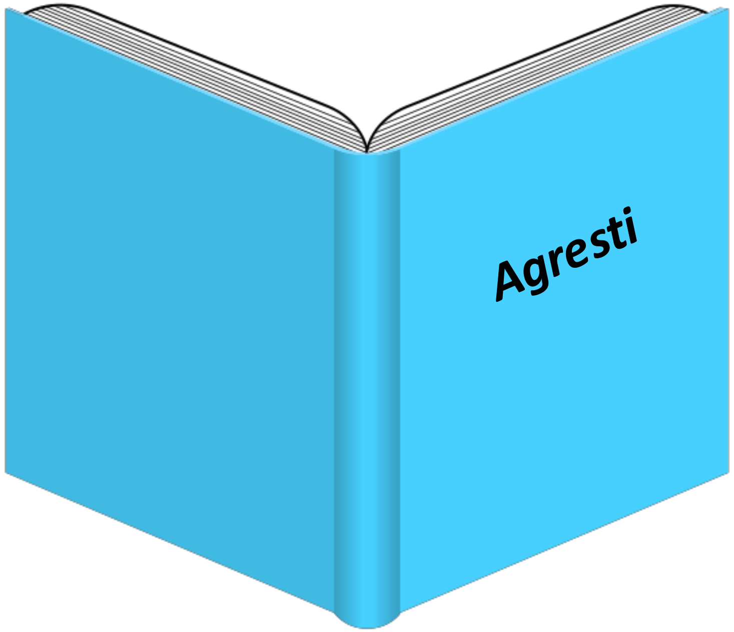




Research Methods and Statistics

Lecture 16: Comparing Two Groups

Johnny van Doorn

Pictures source: pixabay

Do “Smart Pills” Really Make You Smart?

Aa Aa Aa

*So-called "smart drugs" offer the promise of immediate and tangible cognitive benefits, but how effective are they really? In this guest post, **Camilla d'Angelo** takes a look at cognitive enhancers and asks whether, by focusing on quick fixes rather than adopting a healthy lifestyle, we are undermining our well being.*

Source: https://www.nature.com/scitable/blog/student-voices/do_8220smart_pills8221_really_make/

Overview of Today

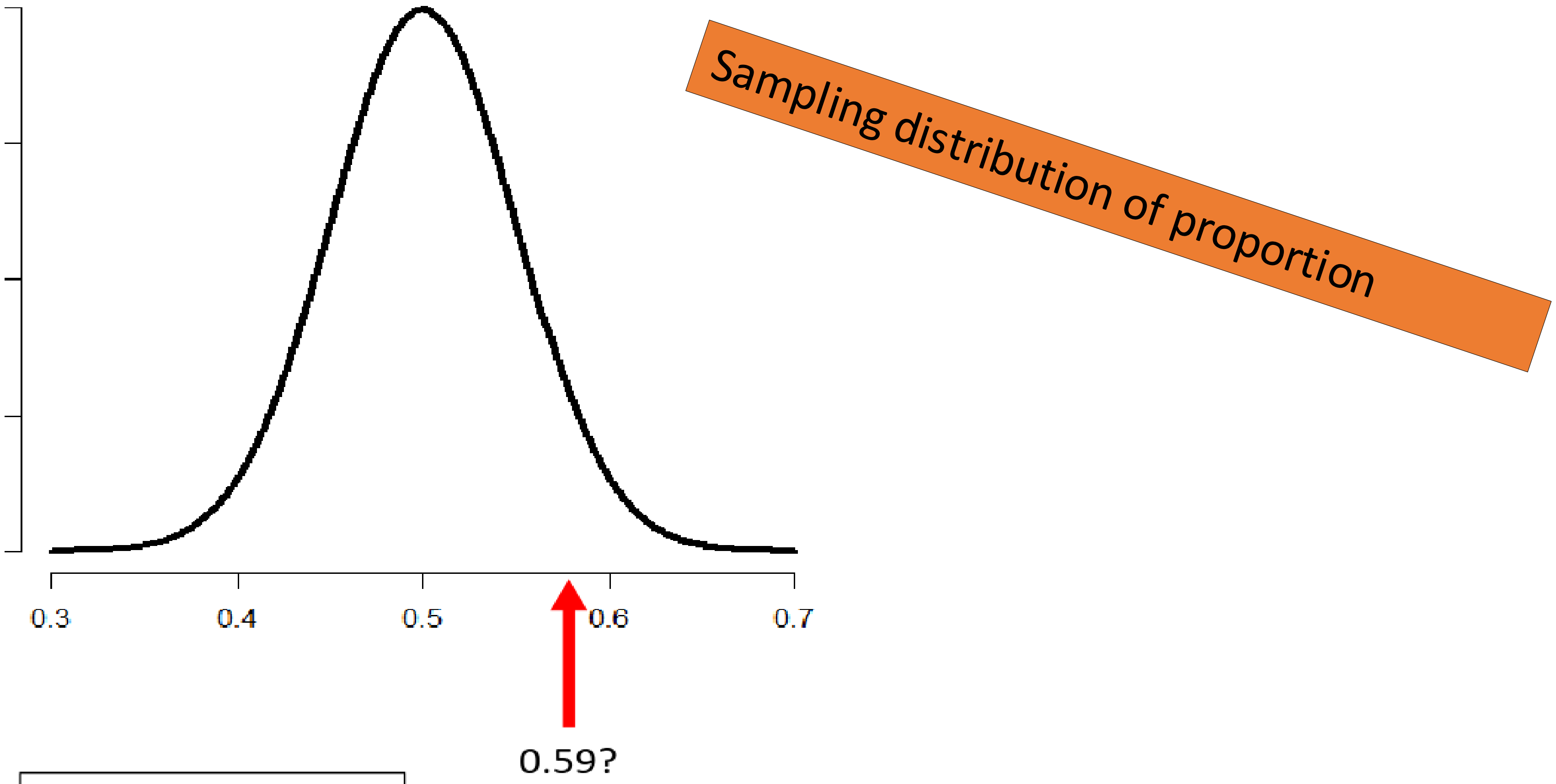
1. **Recap**
    1. Hypothesis testing
    2. Confidence interval

2. Comparing two groups
    1. Estimation: confidence interval
    2. Hypothesis Testing: *t*-test

3. Recap
    1. Next time
    2. Example exam question

How do we draw conclusions about the *difference* between two groups?

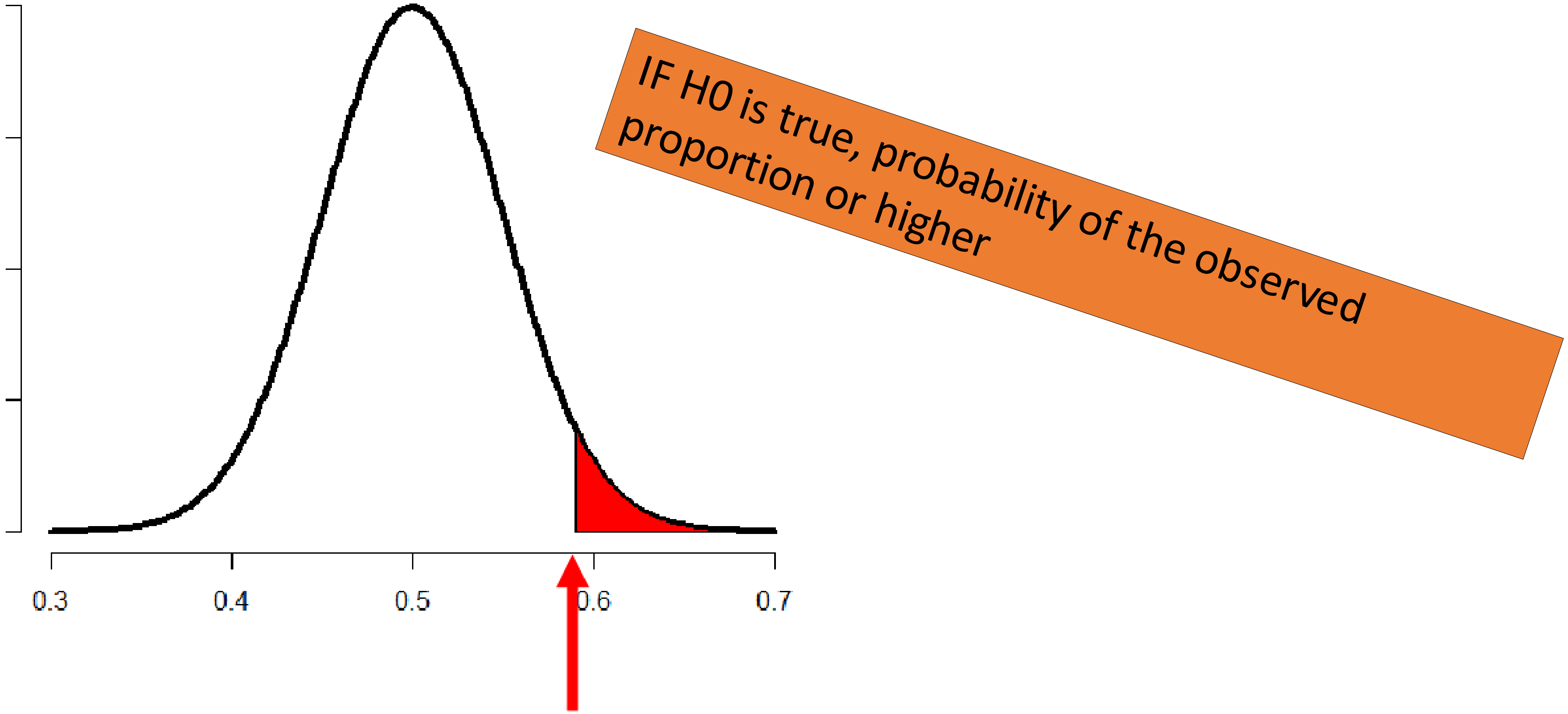
Hypothesis testing



We observe 0.59

Should we reject H_0?

Hypothesis testing



We observe 0.59

Should we reject H_0?

Hypothesis testing (conditional table)

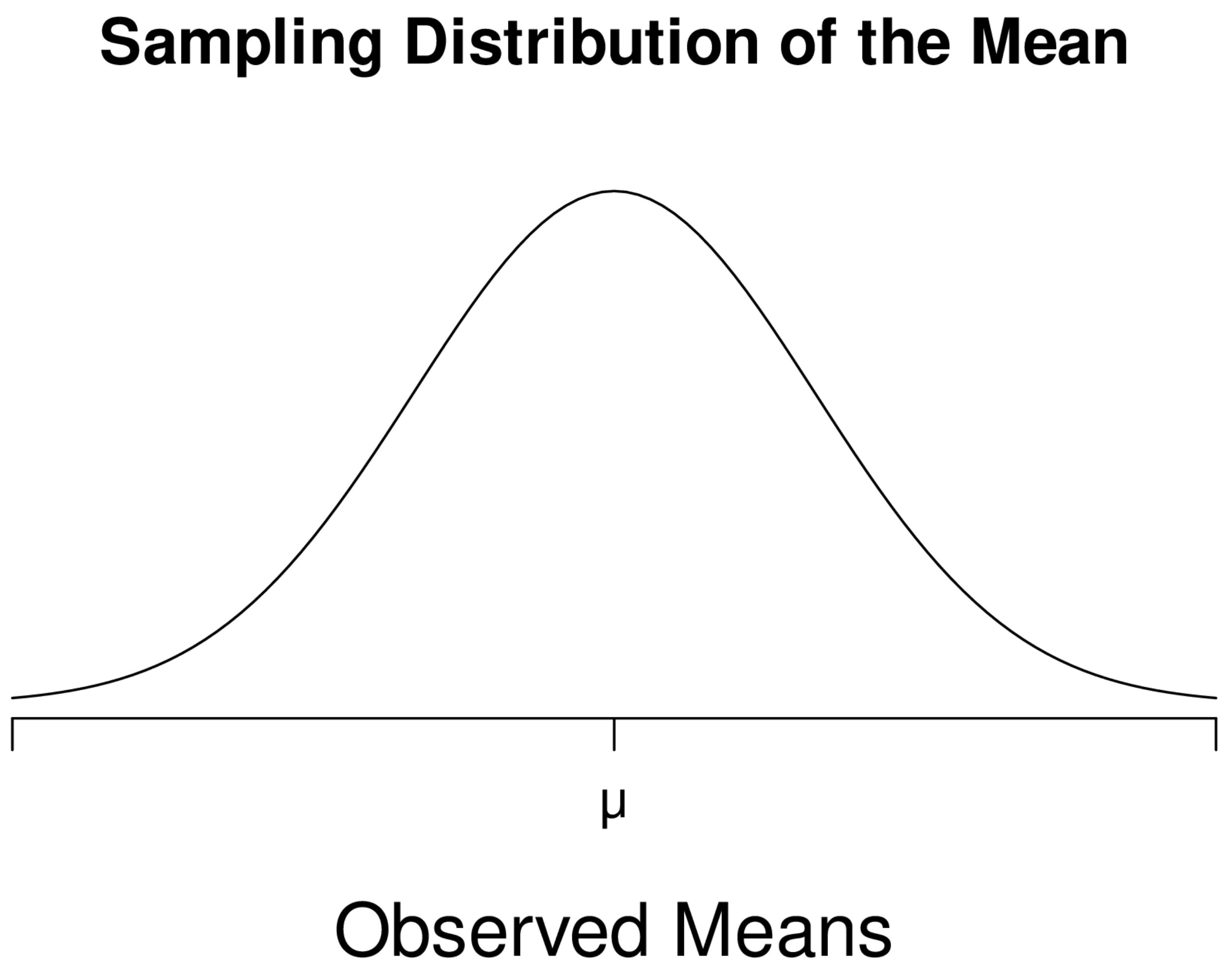
Statistical decision

True state of the world

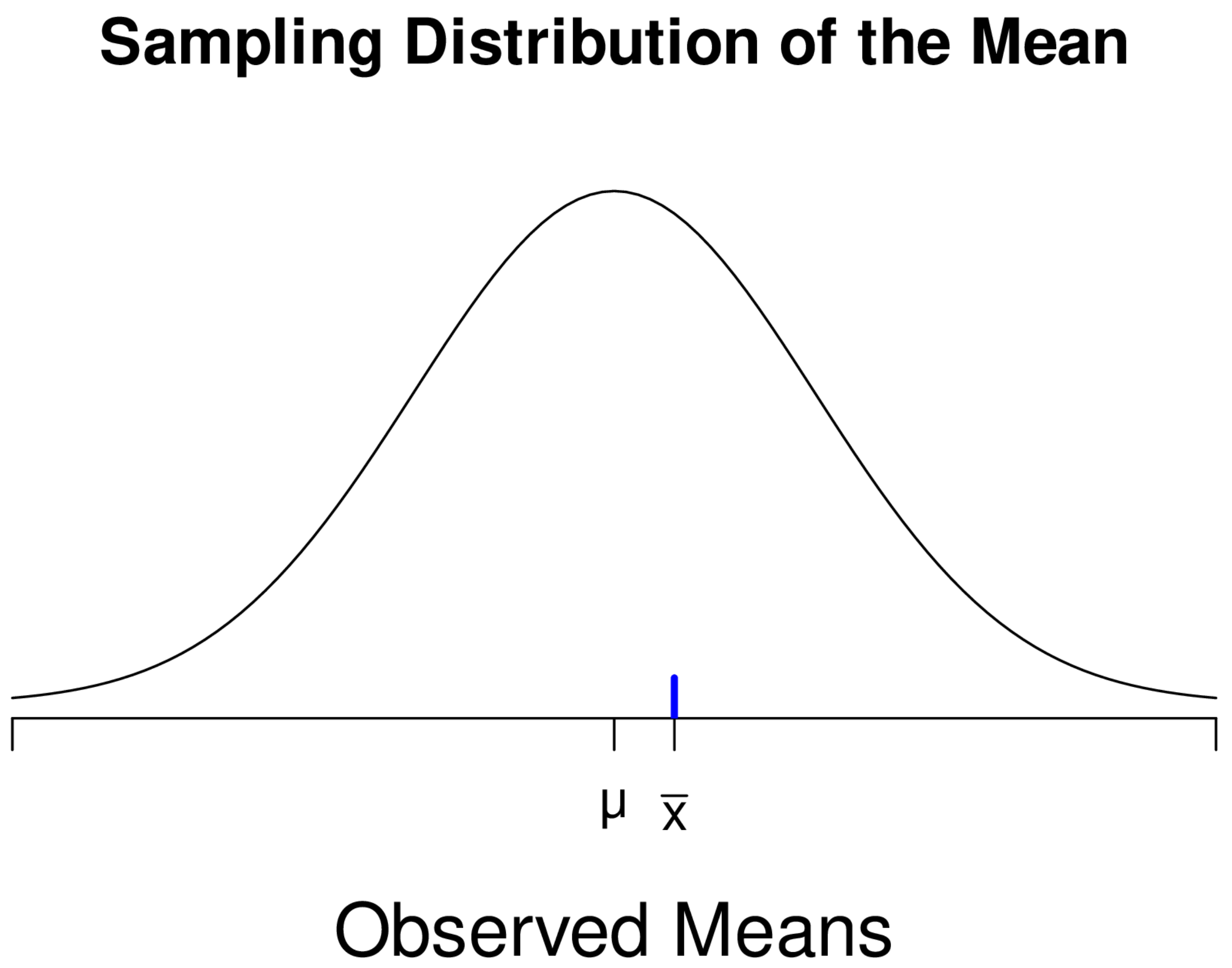
	H_0 is not rejected	H_0 rejected
H_0 is true	***1 – P(Type I error) =*** *P(not reject H_0 | H_0 is true)*	*Type I Error (α) =* *P(reject H_0 | H_0 is true)*
H_a is true	*Type II Error =* *P(not reject H_0 | H_A is true)*	***Power = 1 – P(Type II error) =*** *P(reject H_0 | H_A is true)*

Exam note: only need to know this table – no computations for power (Agresti Table 9.2)

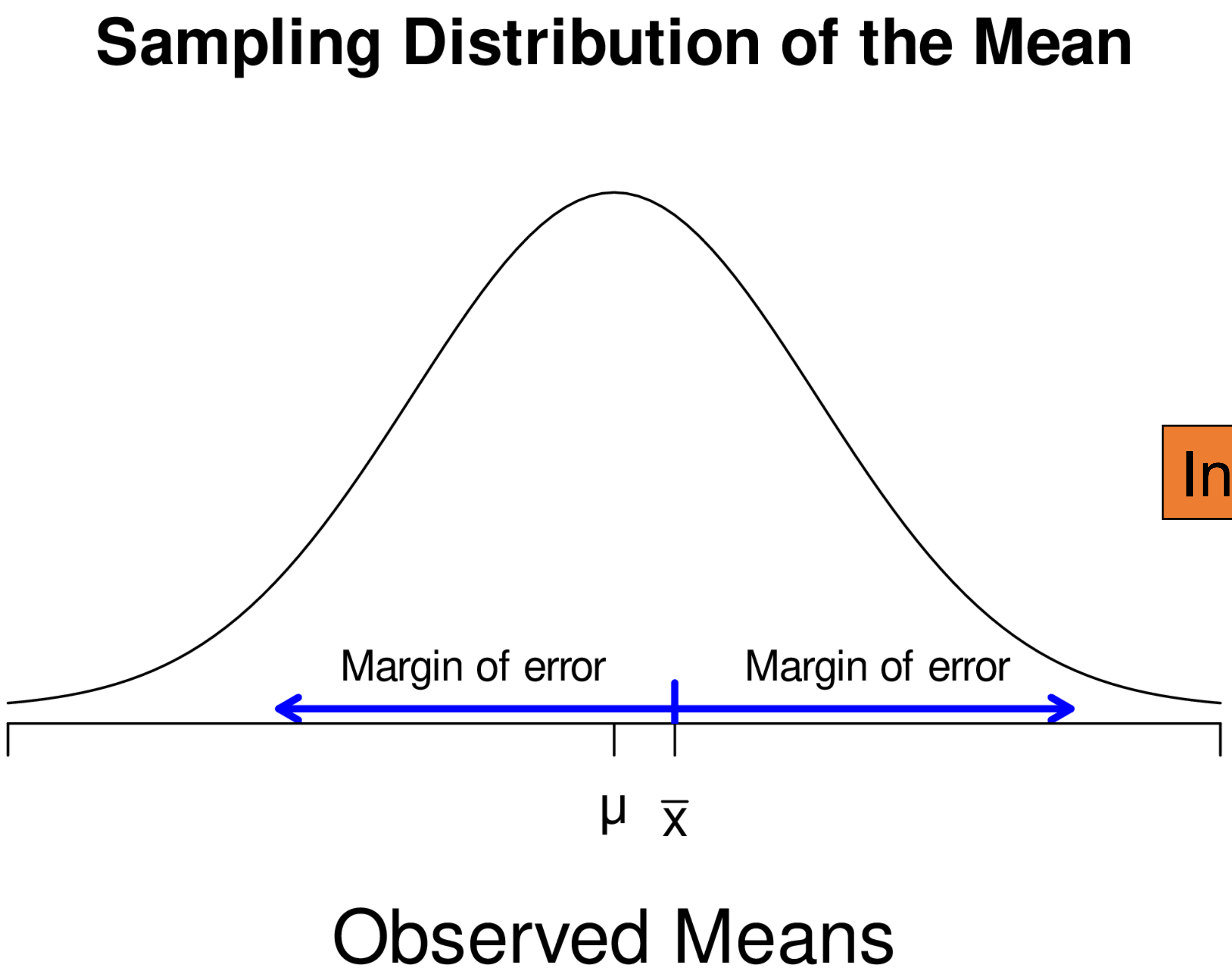
Confidence interval for the mean



Confidence interval for the mean

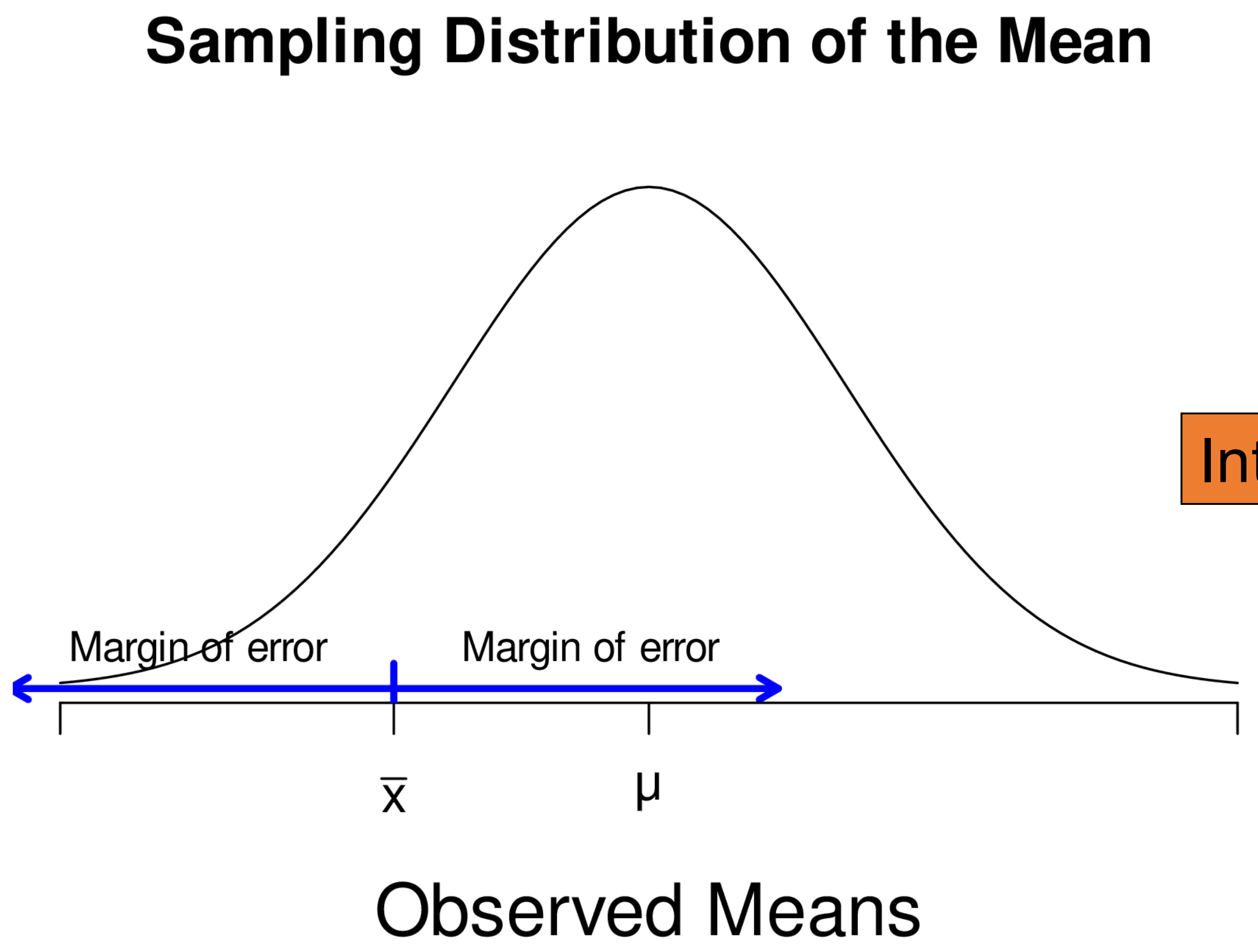


Confidence interval for the mean



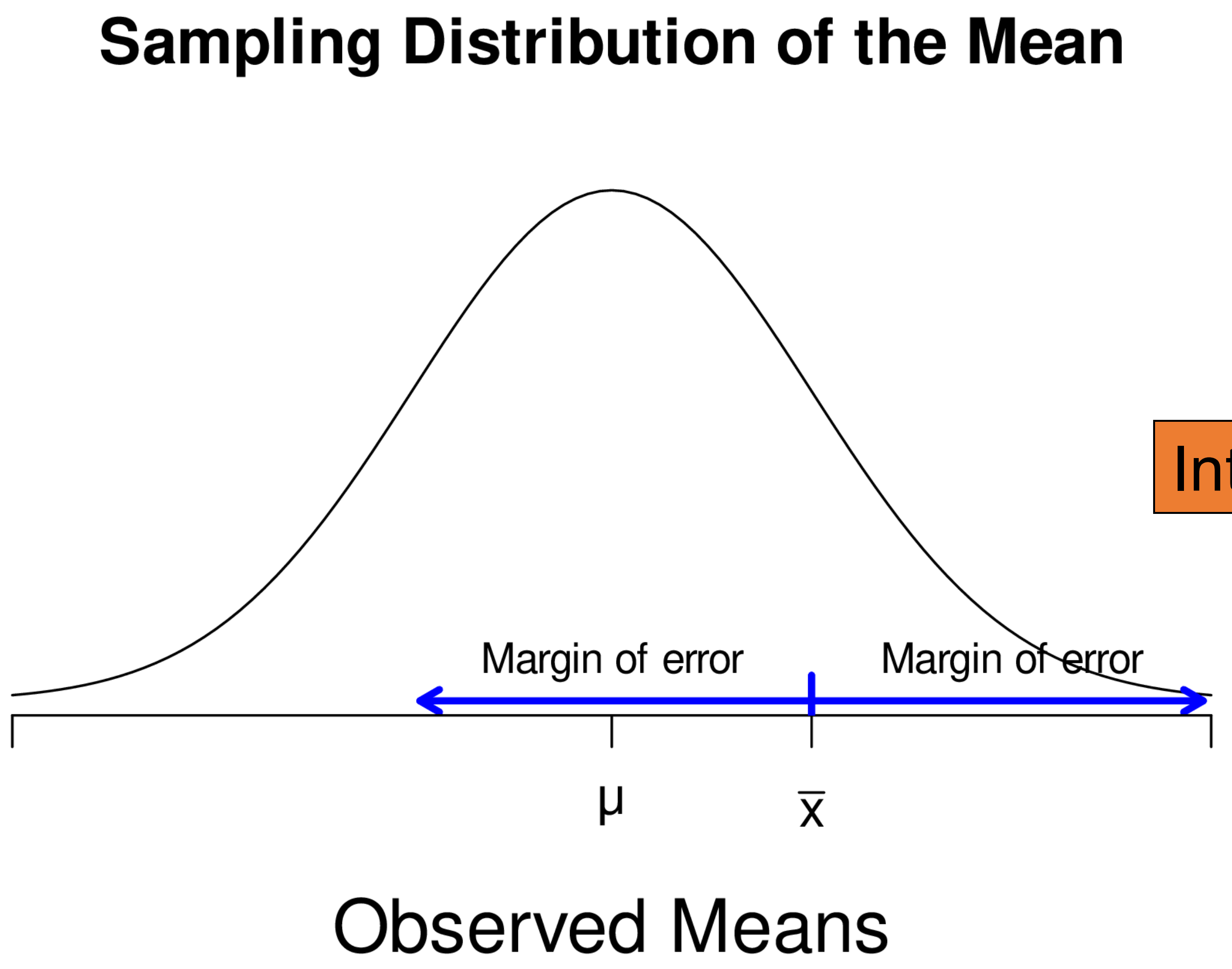
Interval contains μ

Confidence interval for the mean



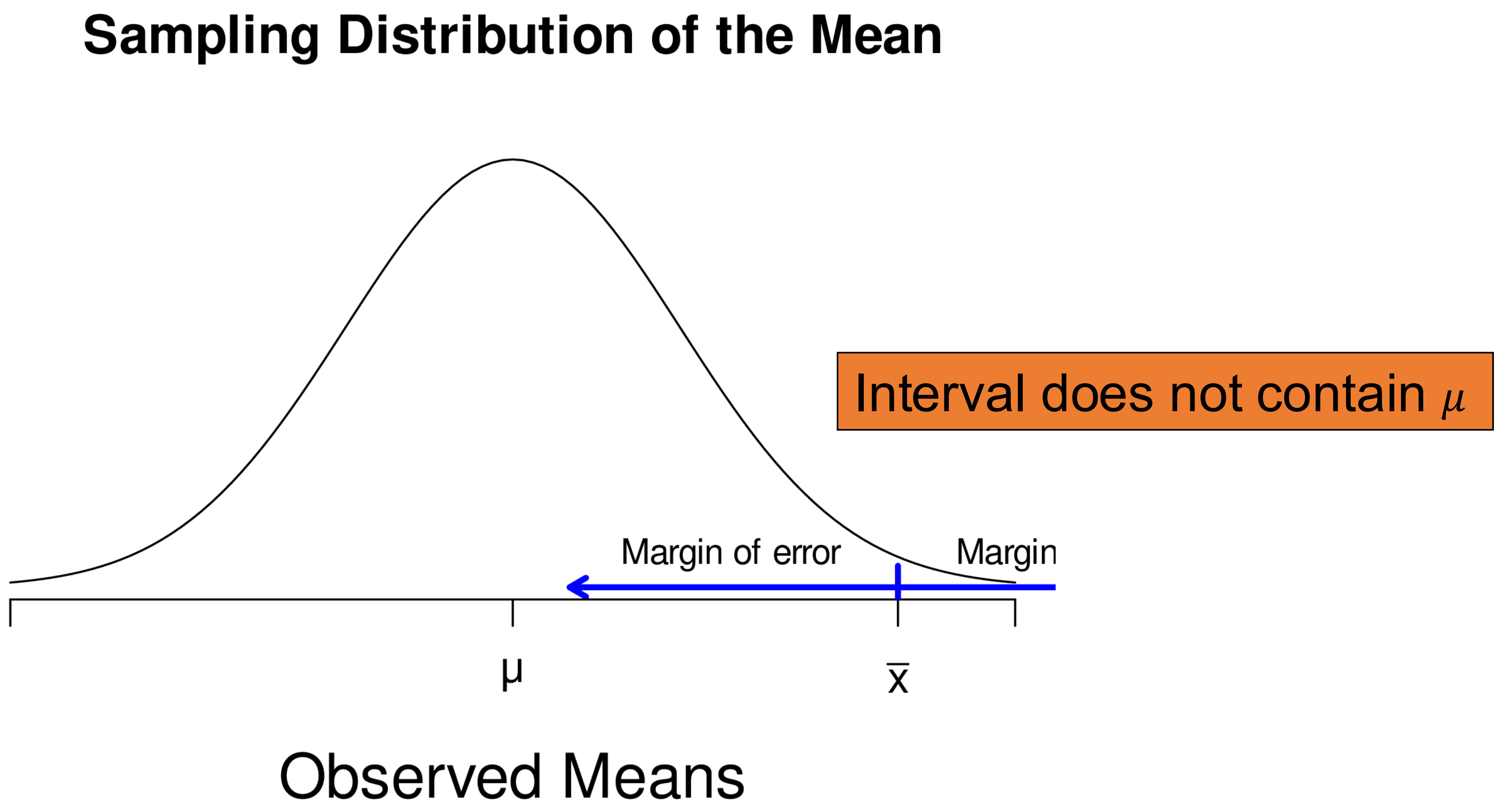
Interval contains μ

Confidence interval for the mean



Interval contains μ

Confidence interval for the mean



Confidence interval for the mean

Confidence interval for the mean: A 95% confidence interval for the population mean μ is

$$\bar{x} \pm margin\ of\ error$$

$$\bar{x} \pm t_{.025} \frac{s}{\sqrt{n}}$$

where $t_{.025}$ depends on the sample size *n*.

Overview of Today

1. Recap
    1. Hypothesis testing
    2. Confidence interval

2. **Comparing two groups**
    1. Estimation: confidence interval
    2. Hypothesis Testing: *t*-test

3. Recap
    1. Next time
    2. Example exam question

Comparing two groups

- To understand whether a ***dependent variable*** differs between two groups of people (= populations)
- Estimate two population parameters
    - + uncertainty (e.g., using confidence interval)
- Estimate the uncertainty around the ***difference*** between the parameters
    - How certain can we be that the ***difference*** deviates from zero?

Example: Do Nootropics make you smarter?

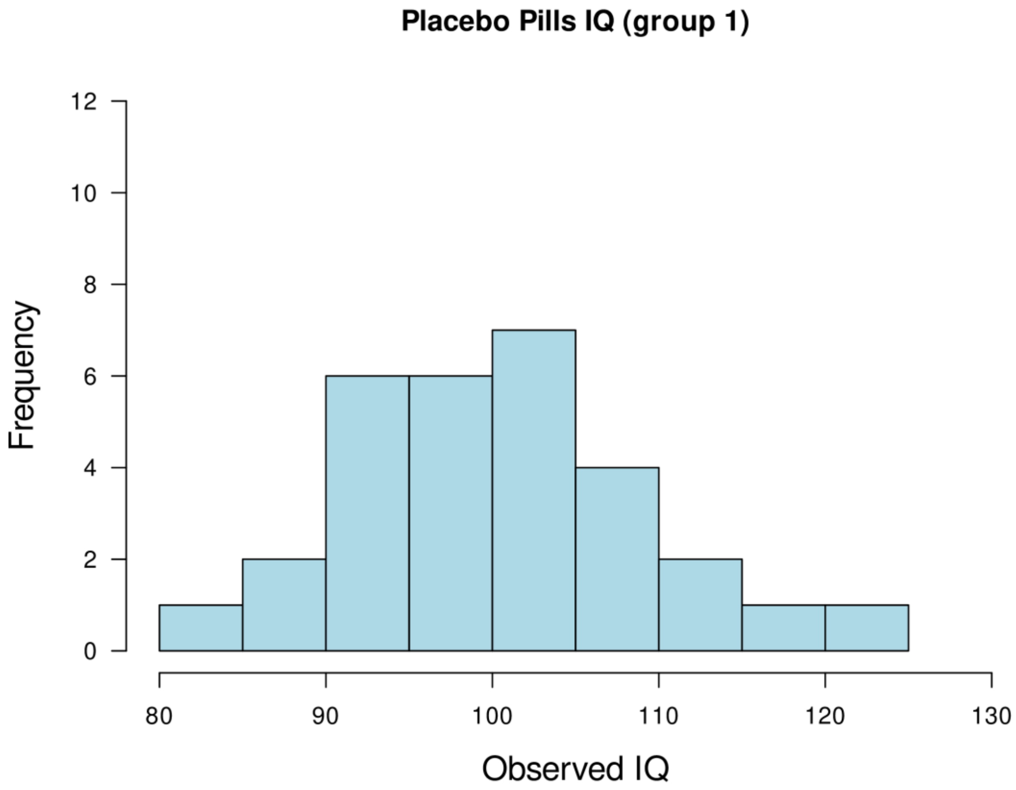
- Hypothesis:
  - One-sided: Consumption of nootropics **increases** intelligence
  - Two-sided: Consumption of nootropics **changes** intelligence

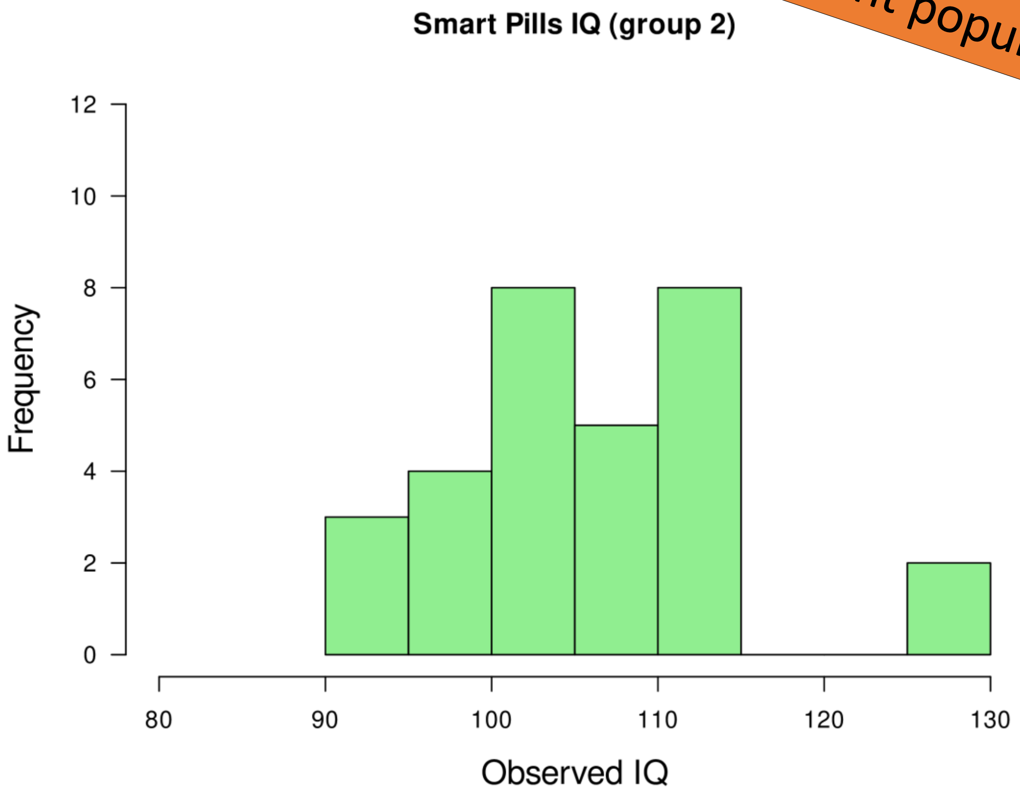
- Operationalization:
  - IV: Participants receive either a placebo or a nootropic pill
  - DV: Score on IQ test

The data

- 60 participants (30 per condition/factor level)

Different populations?

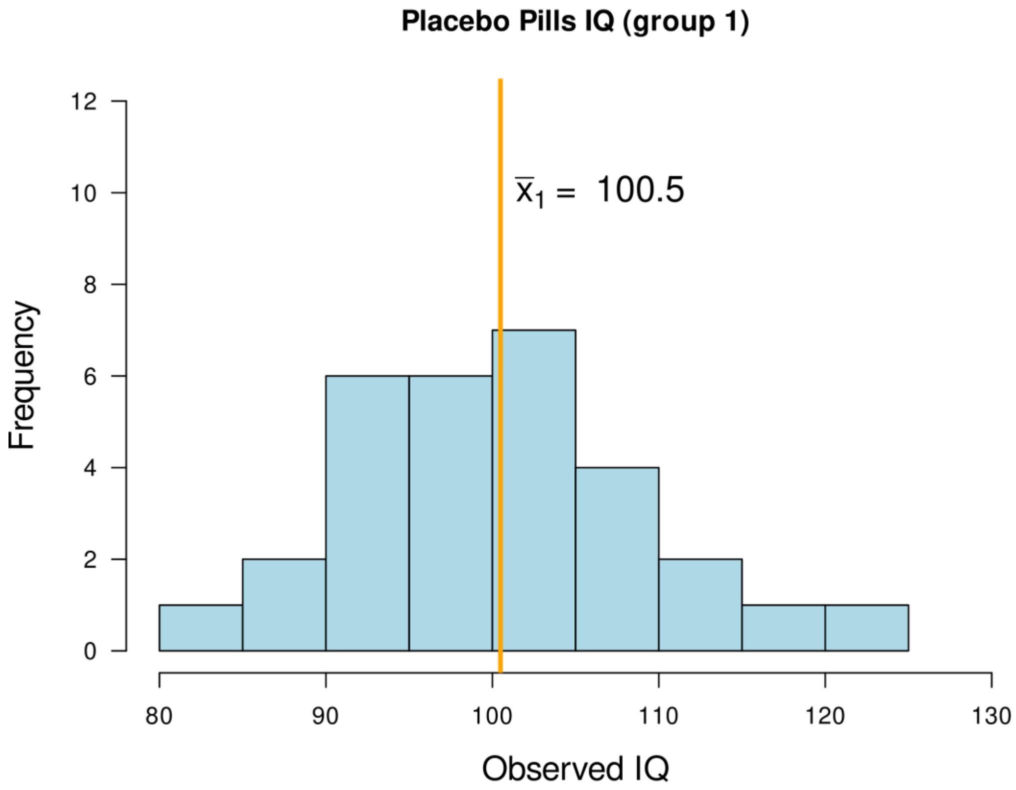


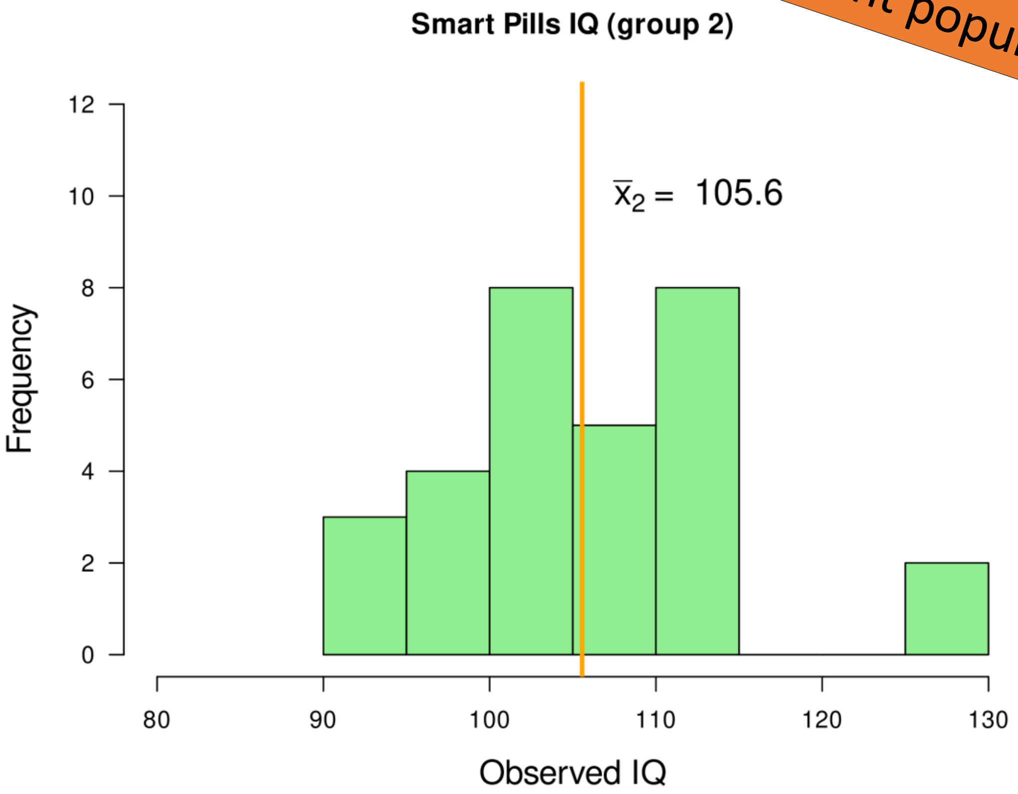


The data

- 60 participants (30 per condition/factor level)

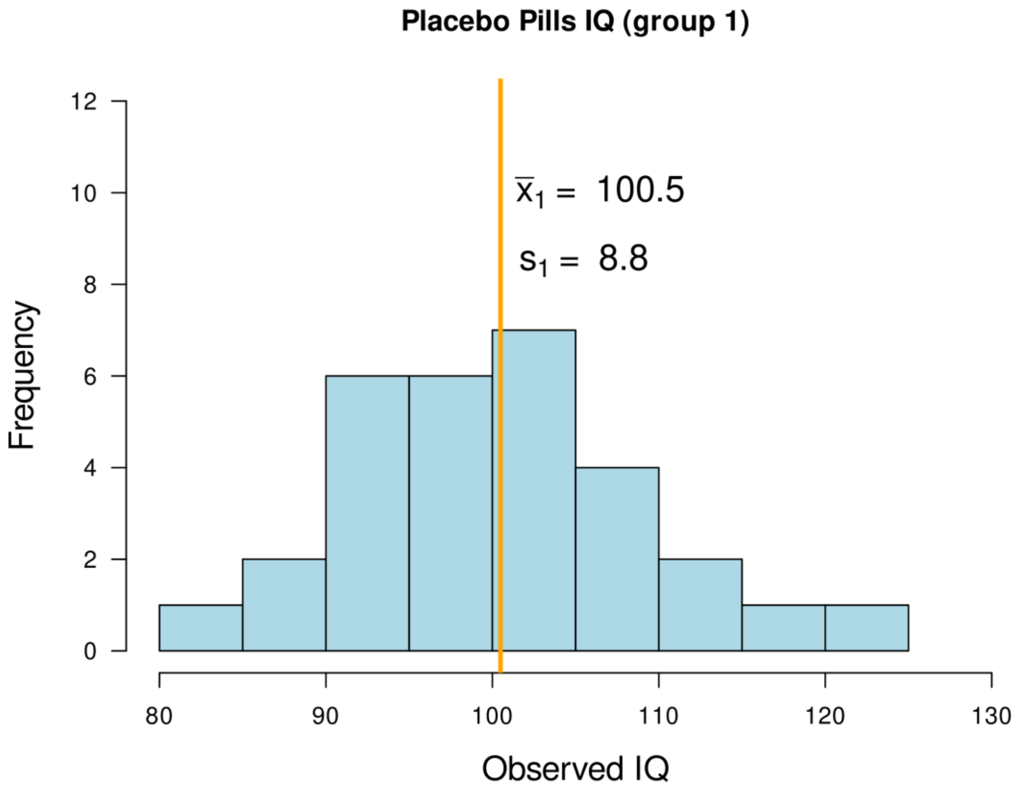
Different populations?

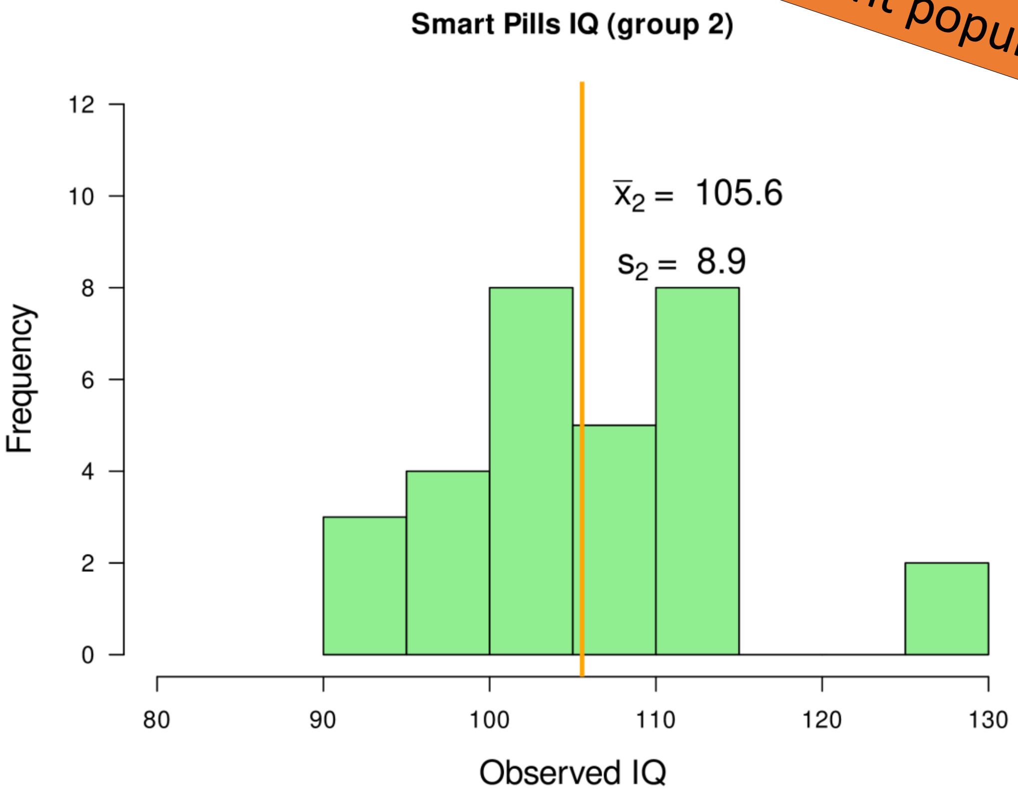




The data

- 60 participants (30 per condition/factor level)

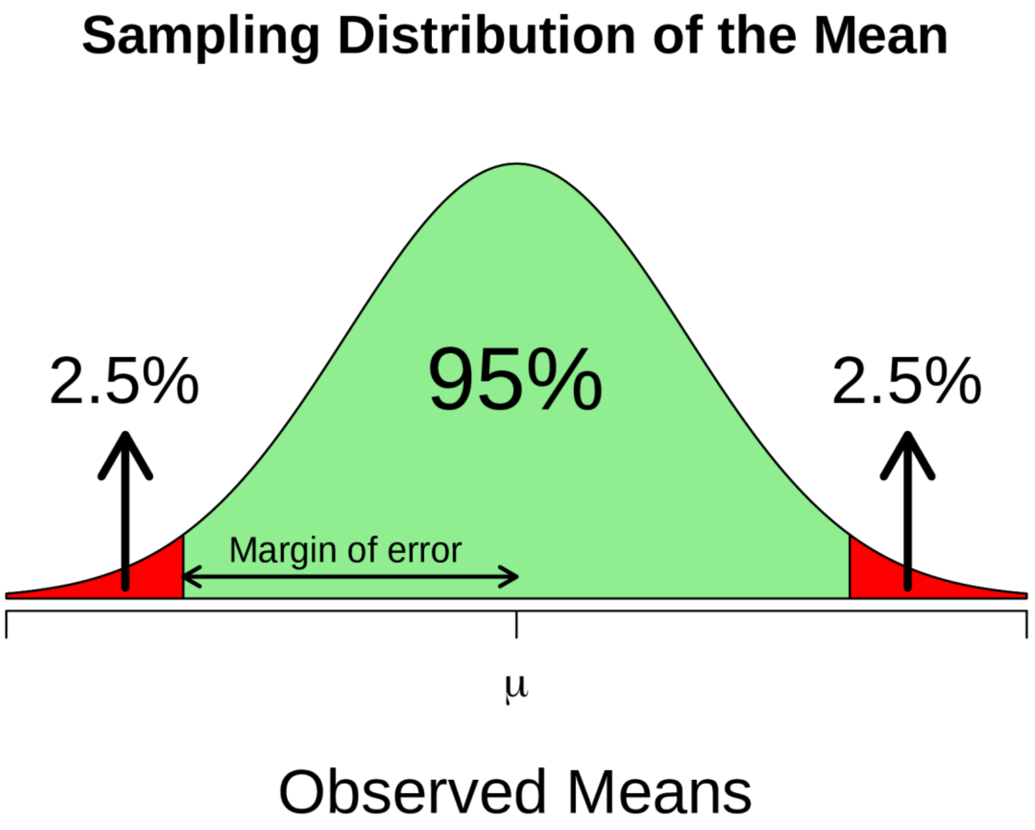




Different populations?

Reasoning

- We can estimate μ_1 and μ_2



Reasoning

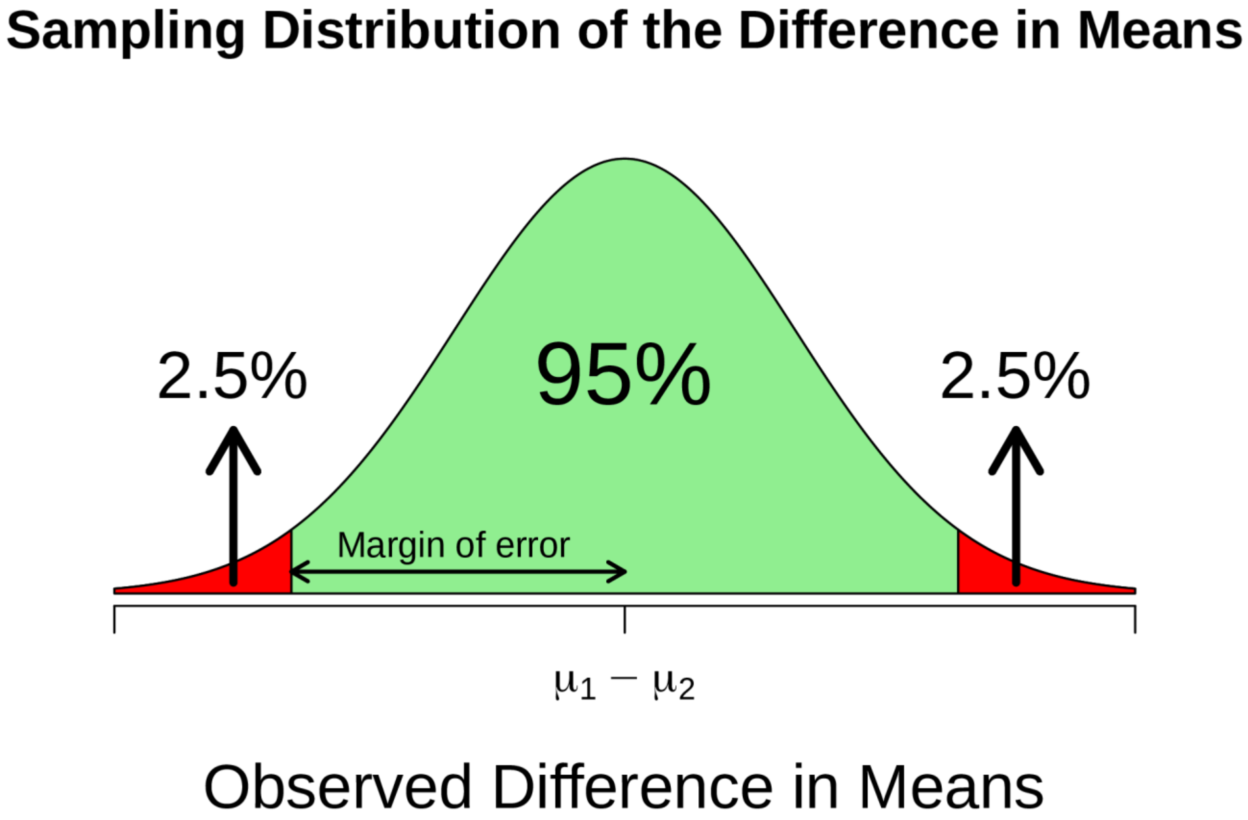
- But we are really interested in $\mu_1 - \mu_2$, since
  - $H_0 : \mu_1 - \mu_2 = 0$
  - $H_A : \mu_1 - \mu_2 \neq 0$

Consumption of nootropics **does not affect** intelligence

Consumption of nootropics **does affect** intelligence

Reasoning

- So we look at the sampling distribution of the ***difference*** in means



Overview of Today

1. Recap
    1. Hypothesis testing
    2. Confidence interval

2. Comparing two groups
    1. **Estimation: confidence interval**
    2. Hypothesis Testing: *t*-test

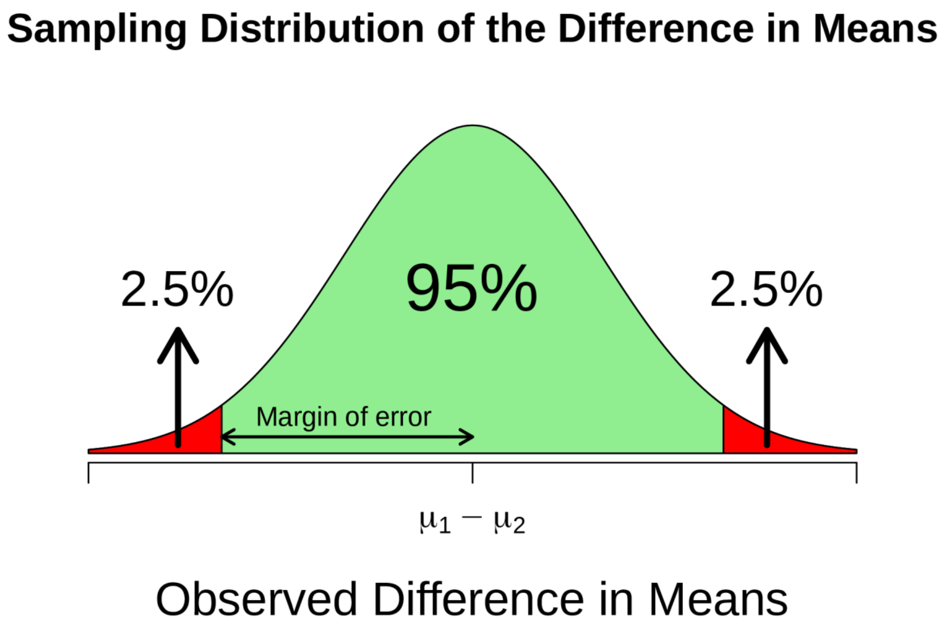
3. Recap
    1. Next time
    2. Example exam question

Reasoning

- So we look at the sampling distribution of the ***difference*** in means

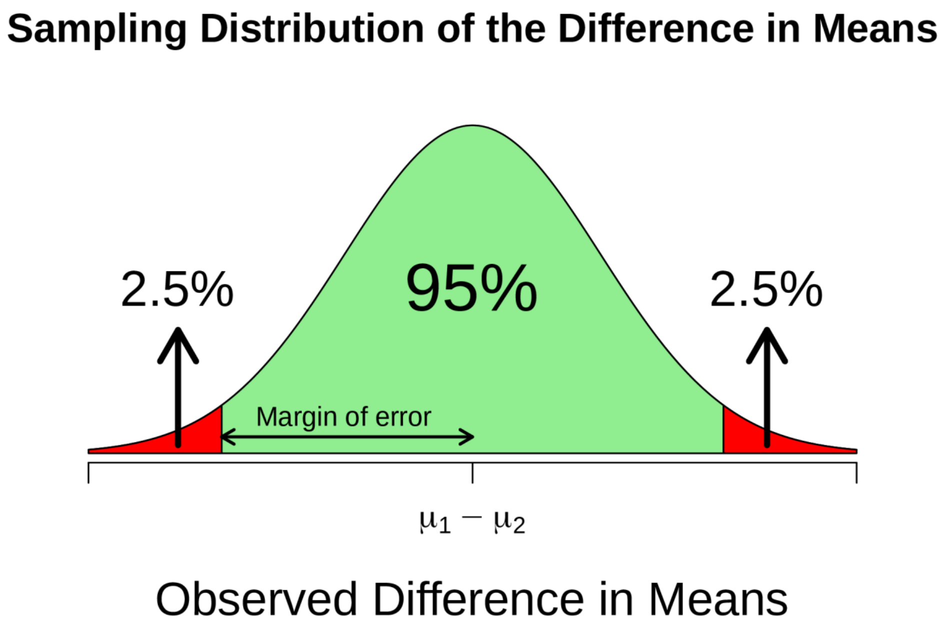
- Central questions:
  - What is the sampling distribution of $\bar{x}_1 - \bar{x}_2$?
  - How do compute the margin of error?

Sampling distribution of the difference



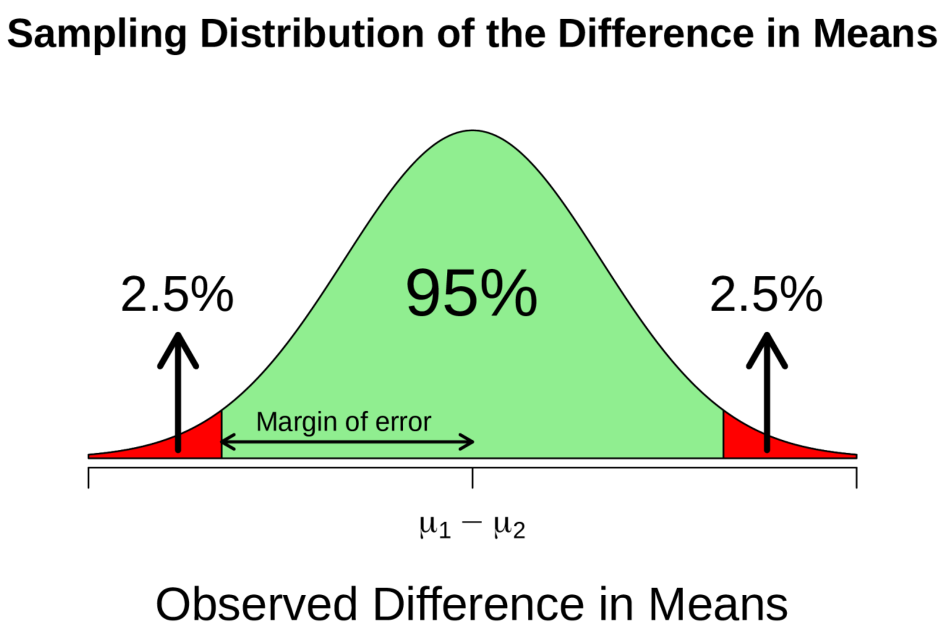
- Normally distributed if sample is large enough (Central Limit Theorem)
- Mean: $\mu_1 - \mu_2$

Sampling distribution of the difference



- Normally distributed if sample is large enough (Central Limit Theorem)
- Mean: $\mu_1 - \mu_2$
- Standard deviation: $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

Sampling distribution of the difference

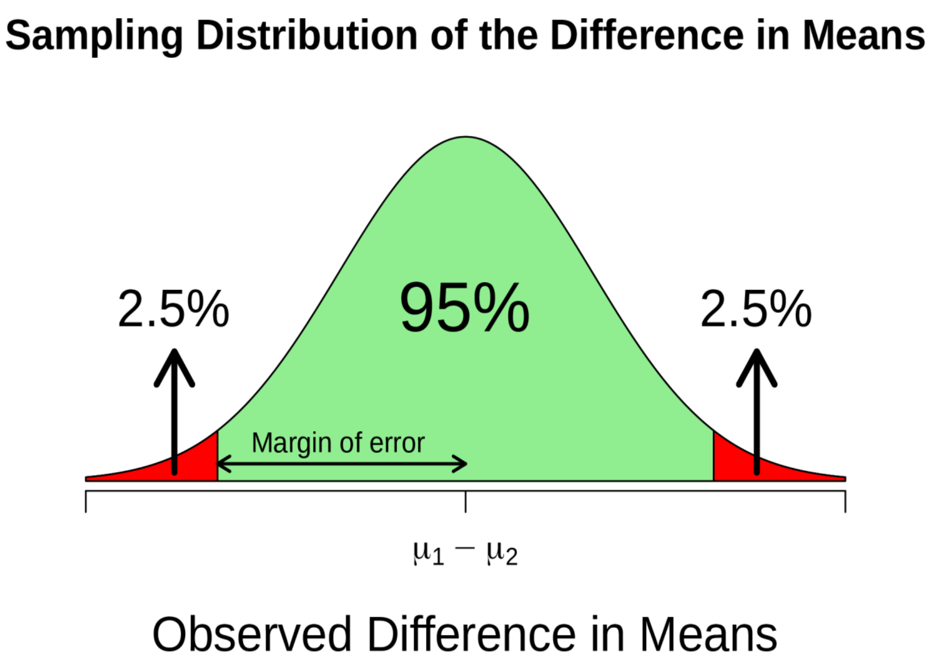


- Normally distributed if sample is large enough (Central Limit Theorem)
- Mean: $\mu_1 - \mu_2$
- Standard deviation: $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
- Standard error: $\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

Standard error: The ***estimated*** standard deviation of a sampling distribution.

Computing a 95%CI for $\mu_1 - \mu_2$

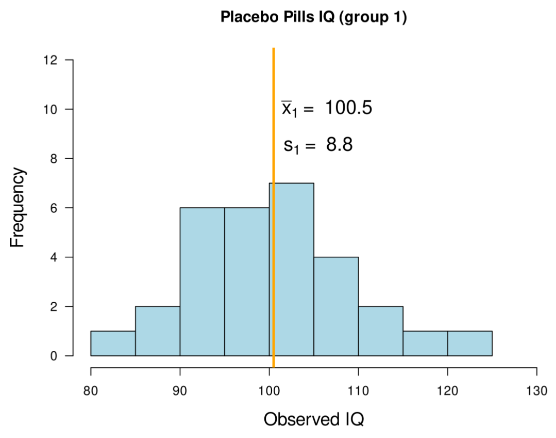
- $\bar{x}_1 - \bar{x}_2 \pm$ margin or error
- Because we do not know σ_1 and σ_2:
  - $\bar{x}_1 - \bar{x}_2 \pm t_{.025} se$

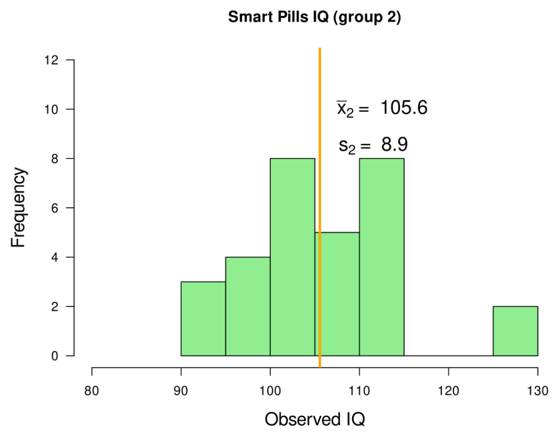


Computing a 95%CI for $\mu_1 - \mu_2$

- $\bar{x}_1 - \bar{x}_2 \ \pm$ margin or error
- Because we do not know σ_1 and σ_2:
  - $\bar{x}_1 - \bar{x}_2 \ \pm \boldsymbol{t}_{.025} se$
- What *t*-distribution? (degrees of freedom?)
  - Depends on s_1, s_2, n_1, n_2
  - IF we can assume that $s_1 \approx \ s_2$, THEN $df \approx n_1 + n_2 - 2$

Computing a 95%CI for $\mu_1 - \mu_2$





- $\bar{x}_1 = 100.5$ $\bar{x}_2 = 105.6$
- $s_1 = 8.8$ $s_2 = 8.9$

Standard deviations are close enough

- $n_1 = n_2 = 30$

Df = $n_1 + n_2 - 2 = 58$

- $t_{.025}$ for 58 df is:
  - MS Excel: `=T.INV(0.975,58)`: 2.001717484

Type this in any cell

Computing a 95%CI for $\mu_1 - \mu_2$

- $\bar{x}_1$ = 100.5 $\bar{x}_2$ =105.6
- s_1 = 8.8 s_2 = 8.9
- n_1 = n_2 = 30
- $t_{.025}$= 2.002

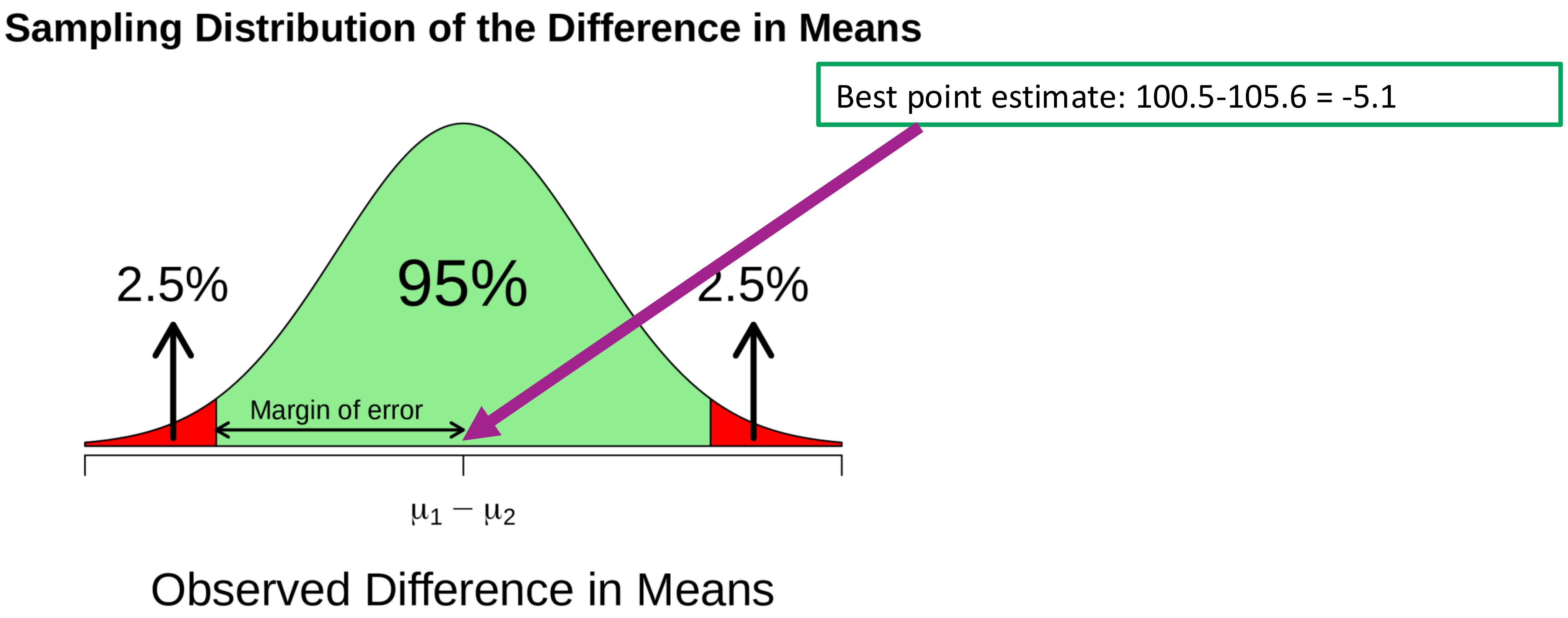
SE = $\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ = 2.285

95% CI = $\bar{x}_1 - \bar{x}_2 \pm t_{.025} se$

95% CI = 100.5 – 105.6 + (2.002 * 2.285) = -0.525

95% CI = 100.5 – 105.6 – (2.002 * 2.285) = -9.675

What do we know now?

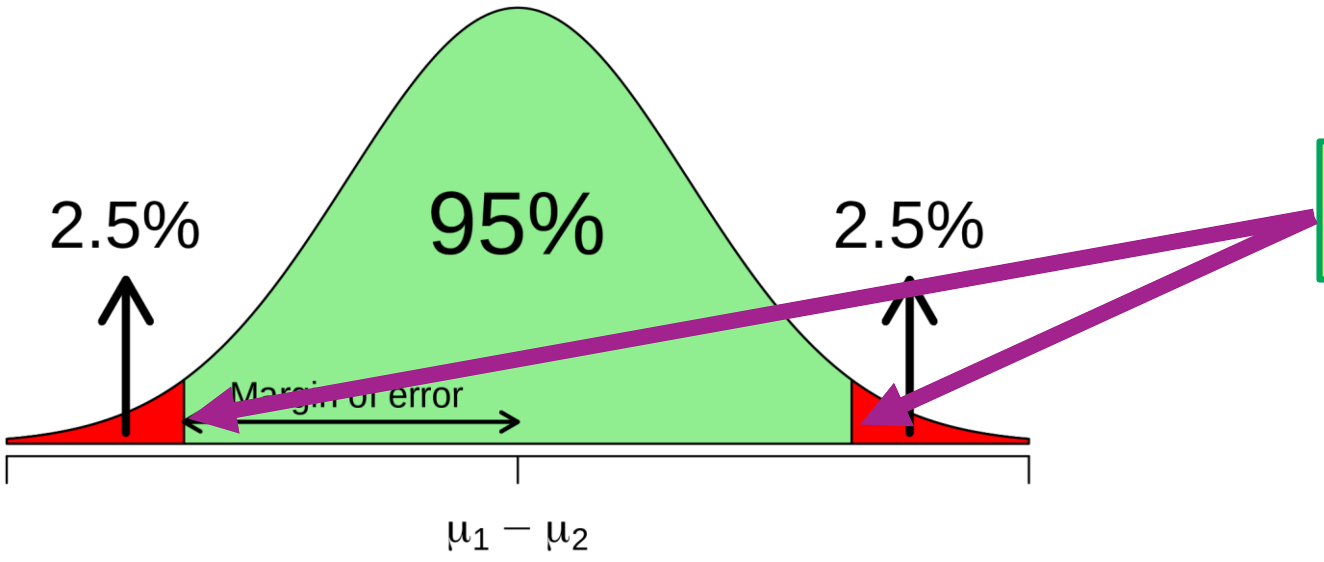


What do we know now?

95% CI = 100.5 – 105.6 + (2.002 * 2.285) = -0.525

95% CI = 100.5 – 105.6 – (2.002 * 2.285) = -9.675

Sampling Distribution of the Difference in Means



Observed Difference in Means

95% CI for this sample:
(-9.675; -0.525)

What do we know now?

A 95% confidence intervals means that *in the long run,*
95% of your confidence intervals will include the true parameter value

↓

95% CI for this sample:
(-9.675; -0.525)

↓

It is likely that the true difference is between these bounds!

Overview of Today

1. Recap
    1. Hypothesis testing
    2. Confidence interval

2. Comparing two groups
    1. Estimation: confidence interval
    2. **Hypothesis Testing: *t*-test**

3. Recap
    1. Next time
    2. Example exam question

Hypothesis testing

- Another way of drawing inference (i.e., not using a confidence interval)

- The five steps of a significance test
  1. Assumptions
  2. Hypothesis
  3. Test statistic
  4. P-value
  5. Conclusion

Step 0: specify your alpha level!

1. Assumptions of two sample *t*-test

- Quantitative variable
- Random sample from the population
- Normally distributed populations (two of them) OR large samples (because of CLT)
    - Because if this is the case, we know the shape of the sampling distribution (→ t-distribution)

Hypothesis testing

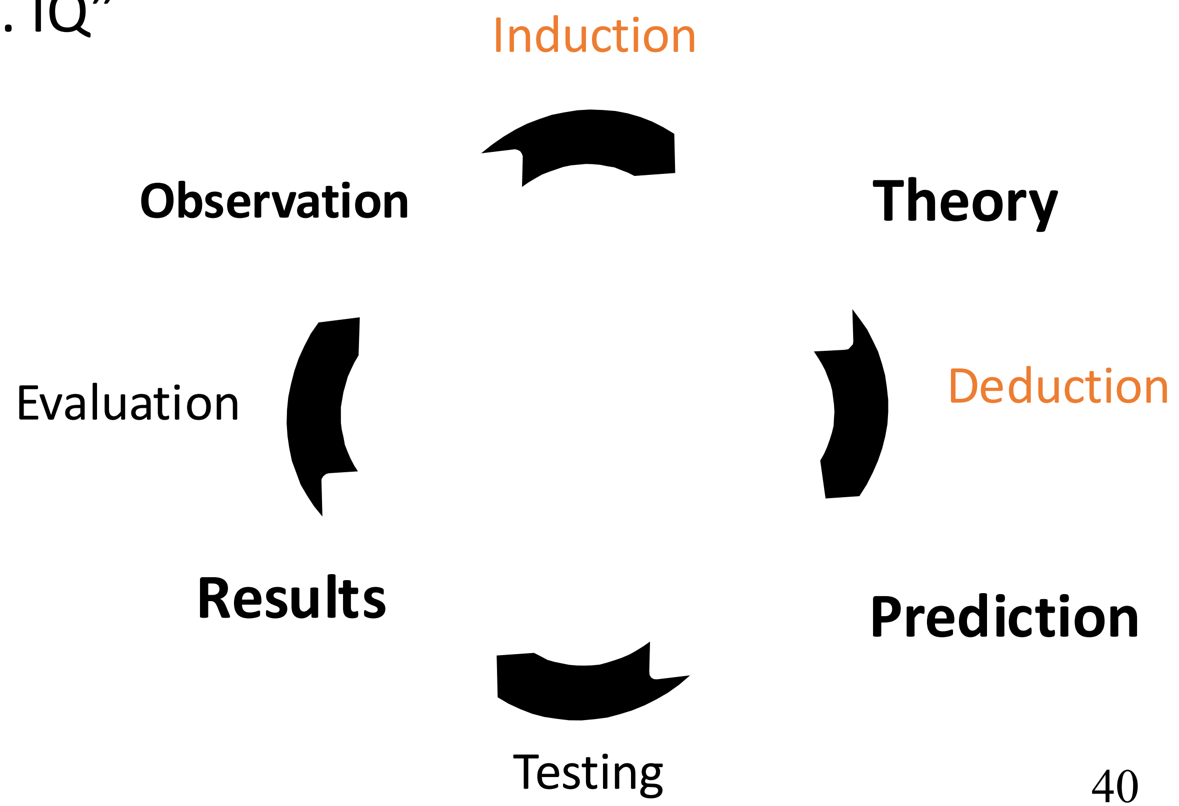
- Another way of drawing inference (ie, not using a confidence interval)

- The five steps of a significance test
    1. Assumptions
    2. **Hypothesis**
    3. Test statistic
    4. P-value
    5. Conclusion

2. Hypothesis

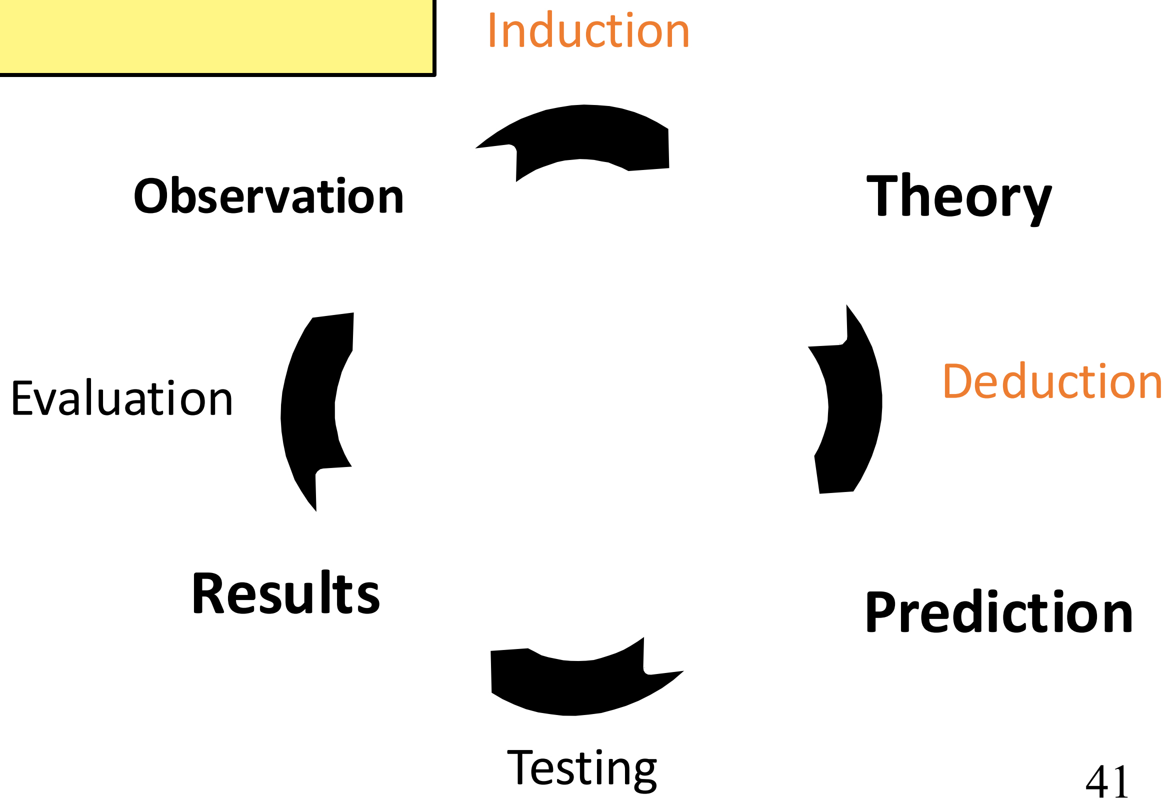
- H_0: “The populations do not differ w.r.t. IQ”
- H_A: “The populations do differ w.r.t. IQ”
  -

→ **Two-sided**



2. Hypothesis

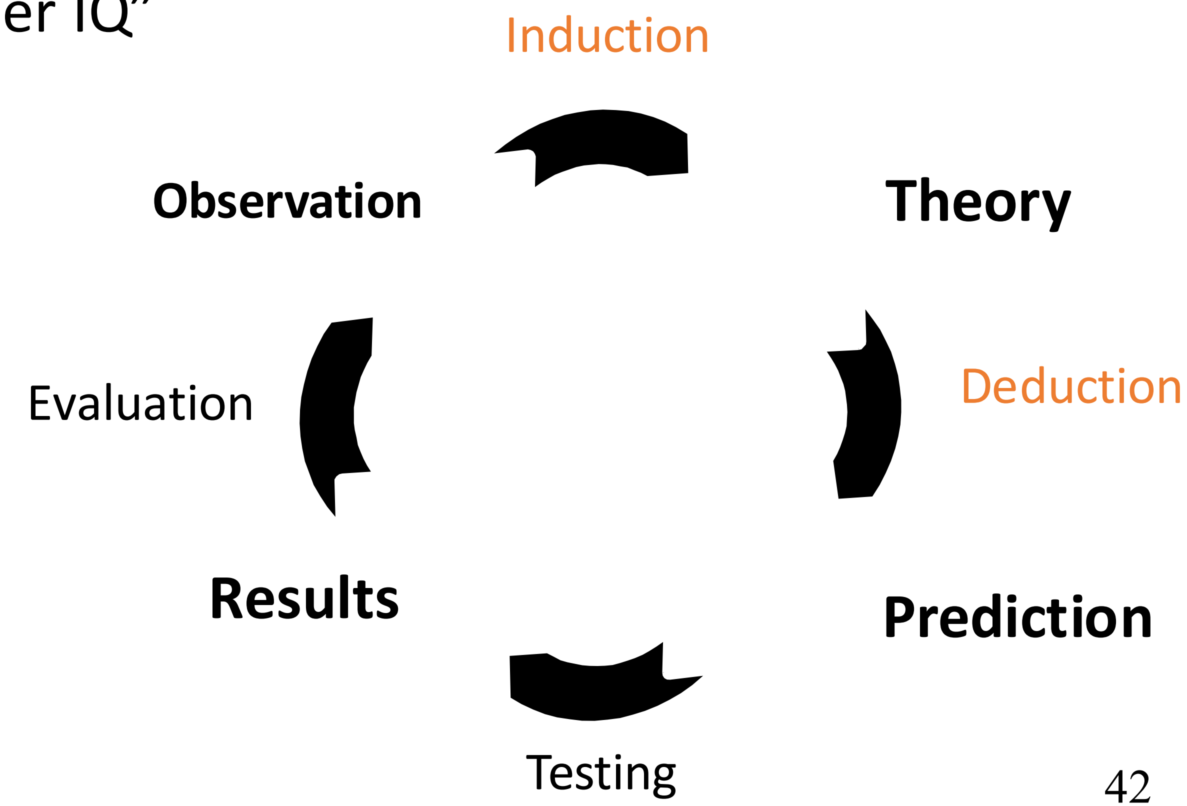
- H_0: $\mu_1 - \mu_2 = 0$
- H_A: $\mu_1 - \mu_2 \neq 0$
  - → **Two-sided**



2. Hypothesis

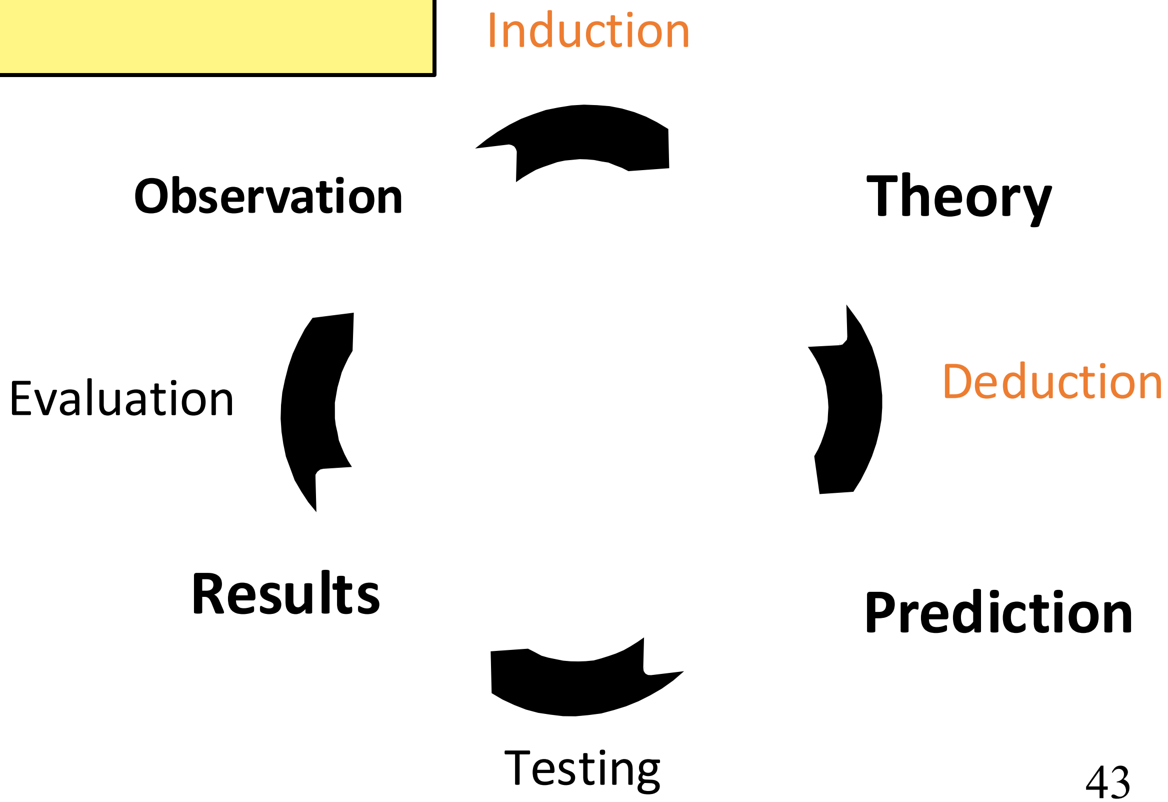
- H_0: "Population 2 will not have a higher IQ"
- H_A: "Population 2 will have a higher IQ"
  -

→ **One-sided**



2. Hypothesis

- H_0: $\mu_1 - \mu_2 = 0$
- H_A: $\mu_1 - \mu_2 < 0$
  - → **One-sided**

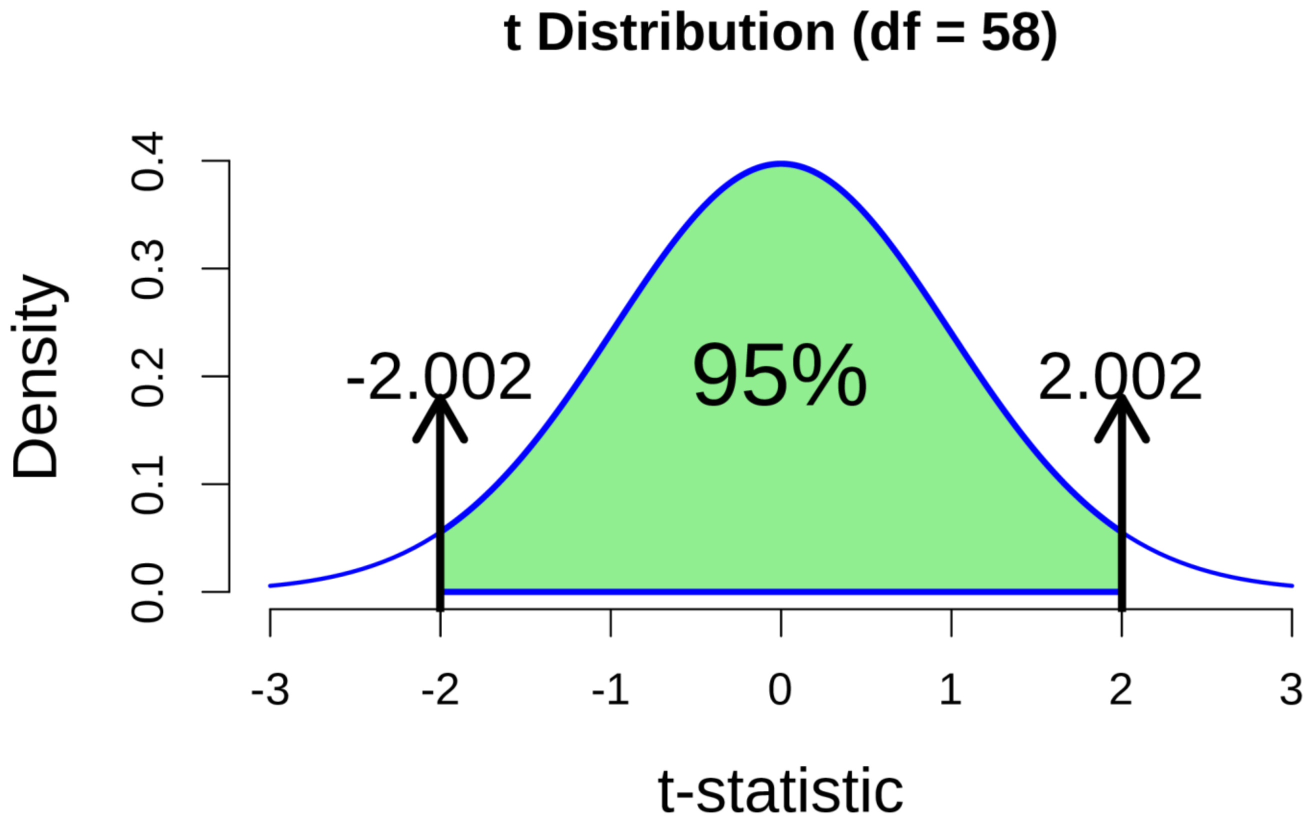


Hypothesis testing

- The five steps of a significance test
  1. Assumptions
  2. Hypothesis
  3. **Test statistic**
  4. P-value
  5. Conclusion

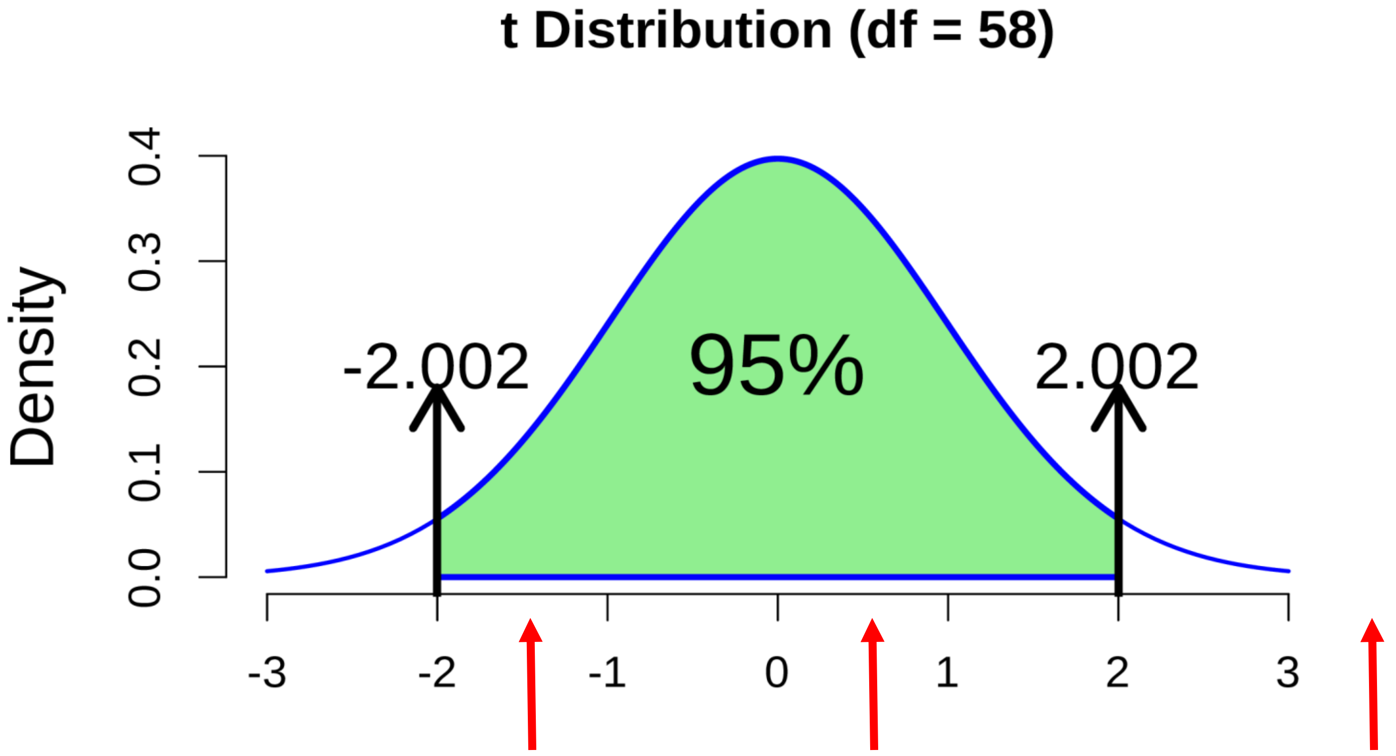
3. Test statistic

Test statistic: A measure of how far the point estimate falls from the parameter value, **if** the null hypothesis is true



3. Test statistic

Test statistic: A measure of how far the point estimate falls from the parameter value, **if** the null hypothesis is true



t = -1.5
Not unlikely

t = 0.5
Likely

t = 3.23
Very unlikely

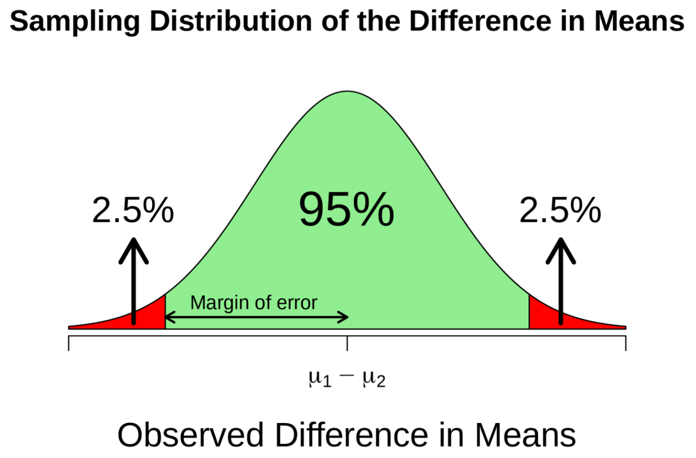
3. Test statistic

This means "Assuming H_0 is true"

- If the assumptions are met, then ***under*** H_0, $\bar{x}_1 - \bar{x}_2$ is distributed with:
- Mean: $\mu_1 - \mu_2 = 0$

This is 0 because we assume H_0 is true when doing the test

$H_0: \mu_1 - \mu_2 = 0$

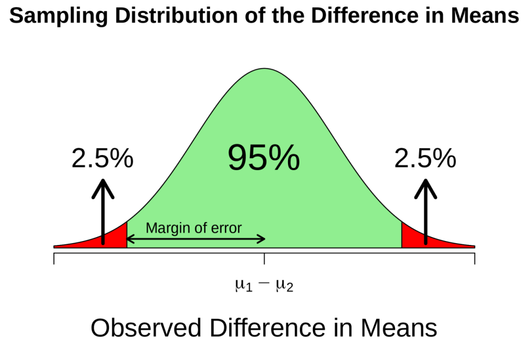


This means "Assuming H_0 is true"

3. Test statistic

- If the assumptions are met, then ***under*** H_0, $\bar{x}_1$ - $\bar{x}_2$ is distributed with:

- Mean: μ_1 - μ_2 = 0
- Standard deviation: $\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_1^2}{n_2}}$ (can be estimated using $\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$)



H_0 : μ_1 - μ_2 = 0

$H_0: \mu_1 - \mu_2 = 0$

3. Test statistic: *t*

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{se} =$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{se}$$

- Two questions:
  - What is *se*? (needed to compute *t*-statistic)
  - What is *df*? (needed to compare the found *t*-statistic to its sampling distribution)

3. Test statistic: *t*

- In general $se = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
  - Exact df can only be determined by computer (eg MS Excel)

3. Test statistic: t

- In general $se = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
  - Exact df can only be determined by computer (eg MS Excel)
- IF we can assume that $s_1 \approx s_2$, THEN $df \approx n_1 + n_2 - 2$
- IF we *additionally* assume that $\sigma_1 = \sigma_2$, THEN $df = n_1 + n_2 - 2$

Note on standard error in independent samples

- We discuss general case and simplify it
- Always use "unpooled" SE:

$$SE = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

- Always use, as df, $n_1 + n_2 - 2$
- For example, if $n_1 = 20$ and $n_2 = 10$, df = 28

- 10.3 is not part of exam

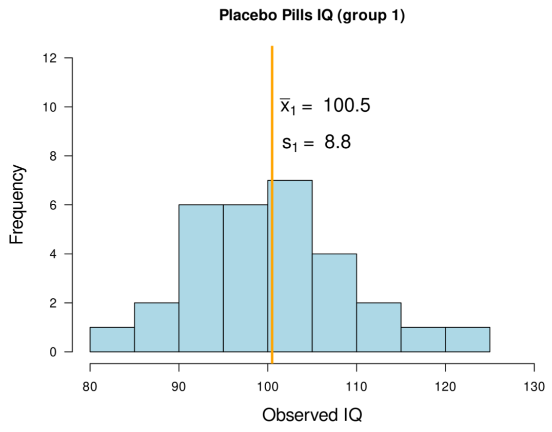
3. Test statistic: t

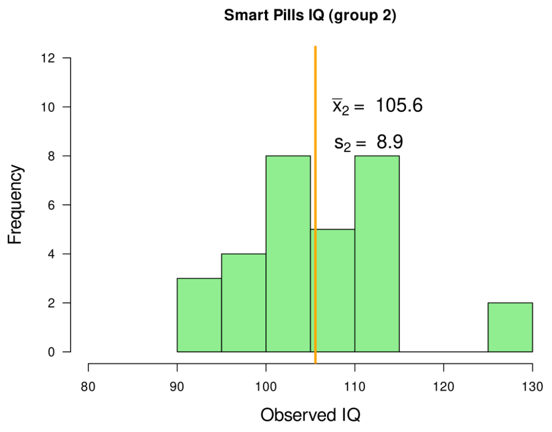
$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{se}$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{se}$$

$$SE = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} = \frac{8.8^2}{30} + \frac{8.9^2}{30} = 2.285$$

$$t = \frac{(100.5 - 105.6) - 0}{2.285} = -2.23$$

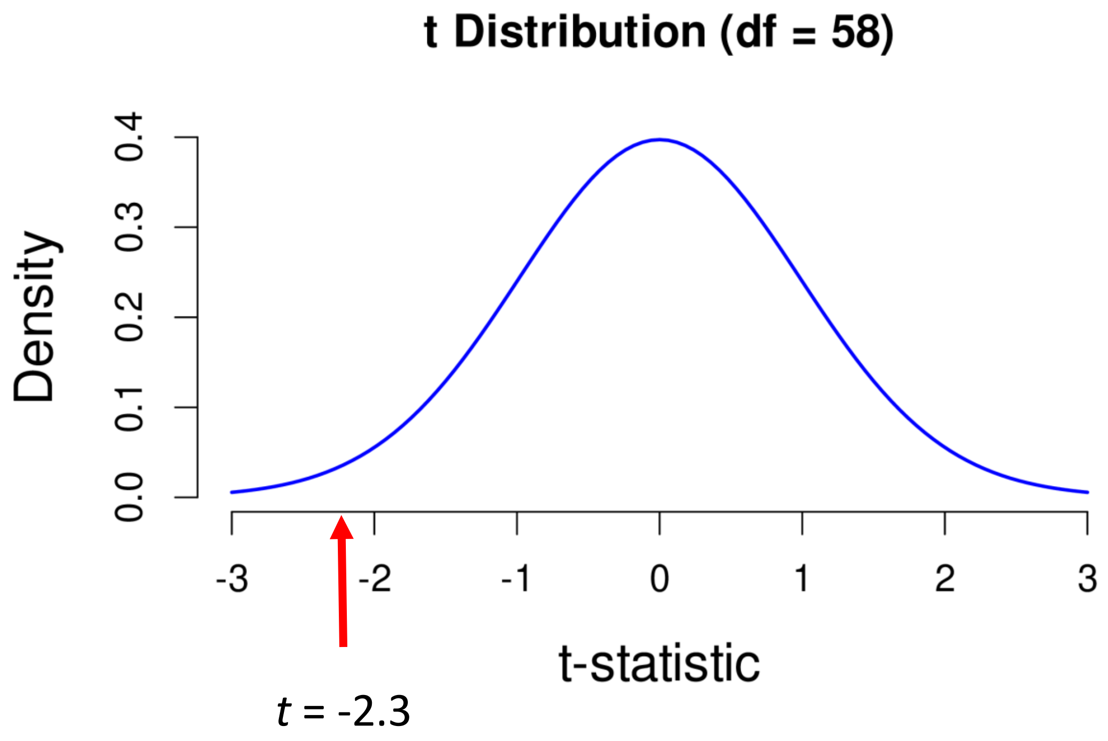




Hypothesis testing

- The five steps of a significance test
  1. Assumptions
  2. Hypothesis
  3. Test statistic
  4. **P-value**
  5. Conclusion

4. p-value



- How likely is our observed result, ***if*** the null hypothesis were true?

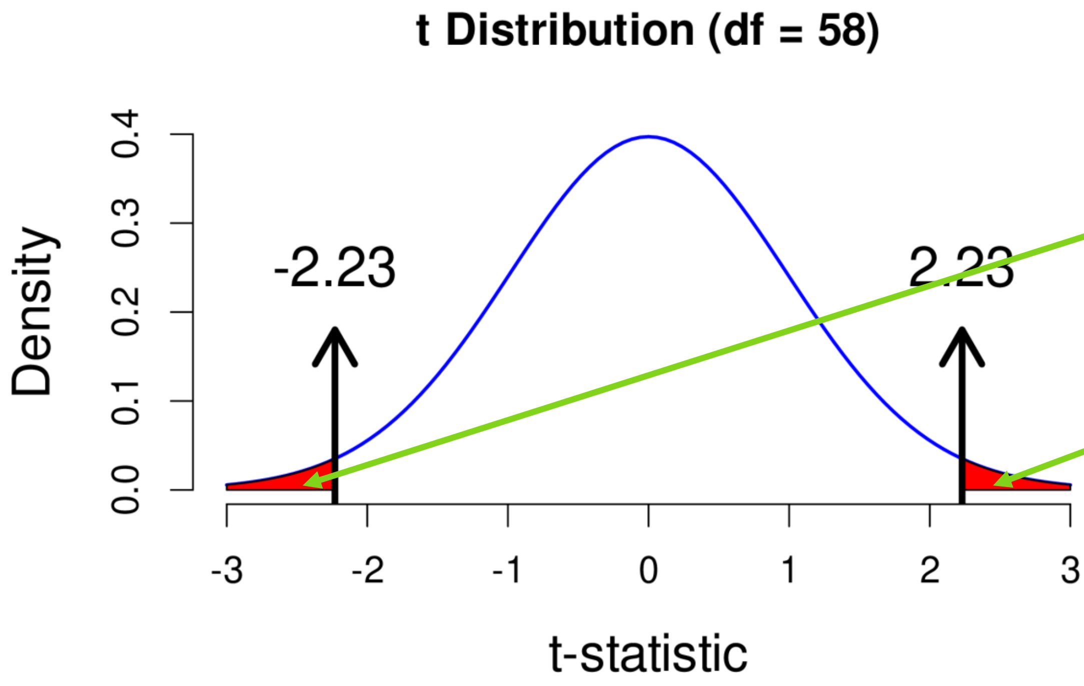
p-value: Probability that you observe the observed test statistic *or more extreme*, ***if*** H_0 is true.

Exam Tip:
Make a drawing of the sampling distribution and indicate the observed statistic, and which tails are indicated by the alternative hypothesis.

4. p-value for the two-sided test

- Because H_A: $\mu_1 - \mu_2 \neq 0$ we perform a two-sided test

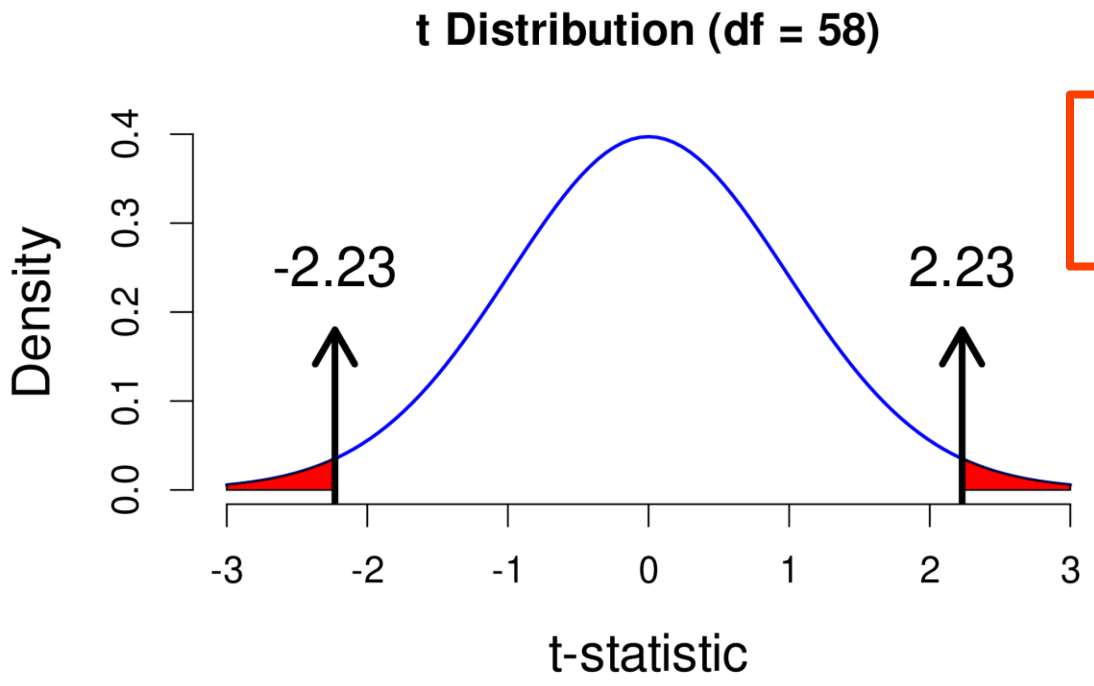
p-value is $P(t \leq -2.23) + P(t \geq 2.23)$



p-value: Probability that you observe the observed test statistic *or more extreme*, ***if*** H_0 is true.

4. p-value for the two-sided test

- Because H_A: $\mu_1 - \mu_2 \neq 0$ we perform a **two-sided test**



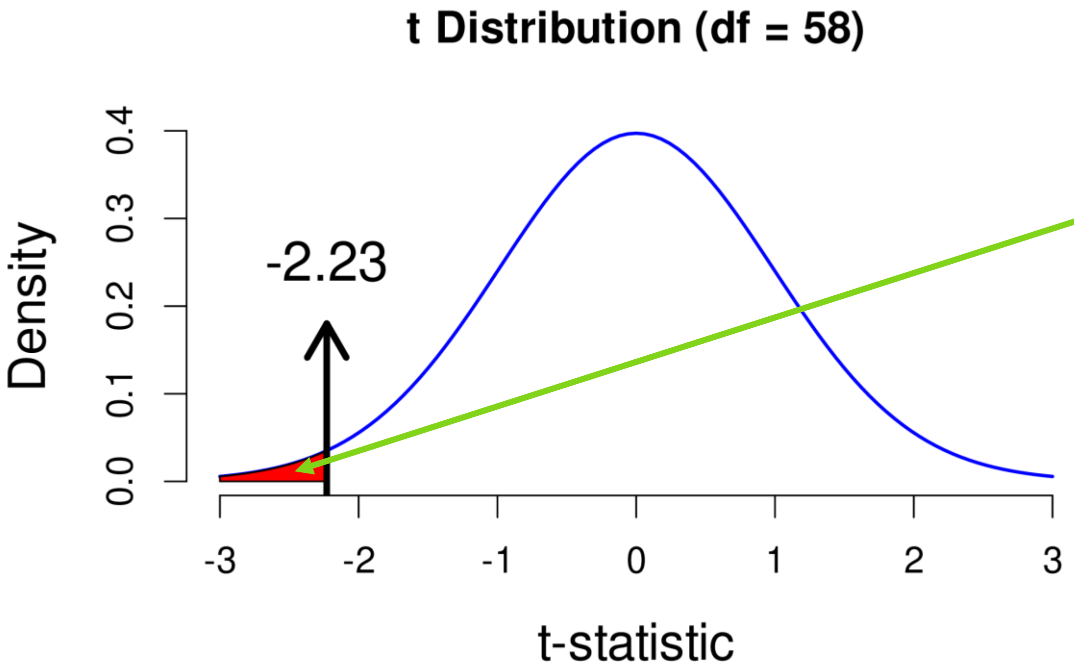
p-value is $P(t \leq -2.23) + P(t \geq 2.23)$

MS Excel: = `2*T.DIST(-2.23; 58; TRUE)`
→ p = 0.03

p-value: Probability that you observe the observed test statistic *or more extreme*, ***if*** H_0 is true.

4. p-value for the **one-sided test**

- Because H_A: $\mu_1 - \mu_2 < 0$ we perform a *negative* **one-sided test**



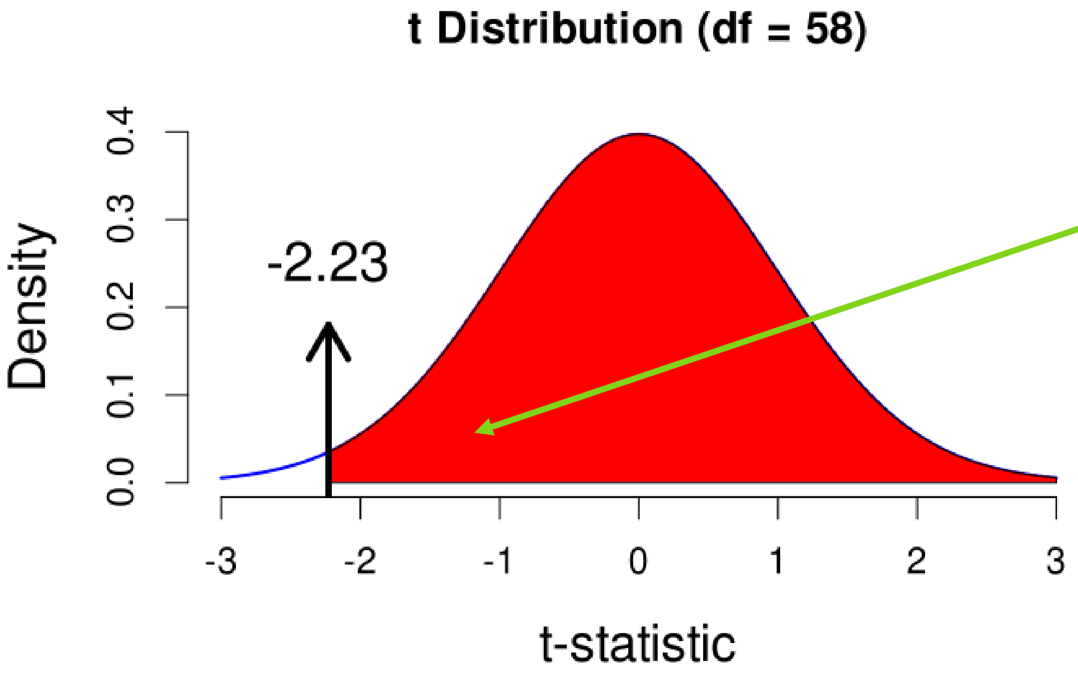
p-value is $P(t \leq -2.23)$

MS Excel: = `T.DIST(-2.23; 58; TRUE)`
→ p = 0.015

p-value: Probability that you observe the observed test statistic *or more extreme*, ***if*** H_0 is true.

4. p-value for the **other one-sided test**

- In case H_A: $\mu_1 - \mu_2 > 0$ we perform a *positive* **one-sided test**



p-value is $P(t \geq -2.23)$

MS Excel: = `1-T.DIST(-2.23; 58; TRUE)`
→ p = 0.985

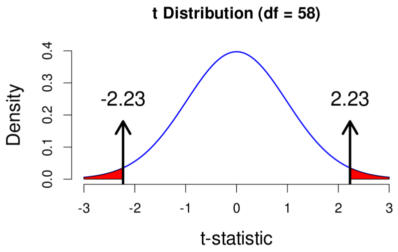
p-value: Probability that you observe the observed test statistic *or more extreme*, ***if*** H_0 is true.

What does it mean to be extreme?

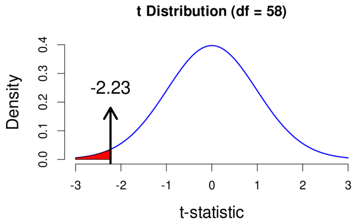
p-value: Probability that you observe the observed test statistic *or more extreme*, ***if*** H_0 is true.

The alternative hypothesis determines what is "more extreme"
It's all the values that would provide more evidence in favor of H1 than the observed test statistic

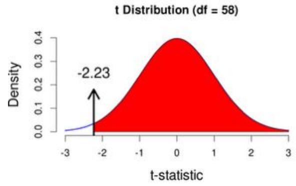
T = -2.3



If H1 is two-sided, an observed t-value of -2.5 would provide more evidence than -2.3



If H1 is one-sided **negative**, an observed t-value of -2.5 would provide more evidence than -2.3

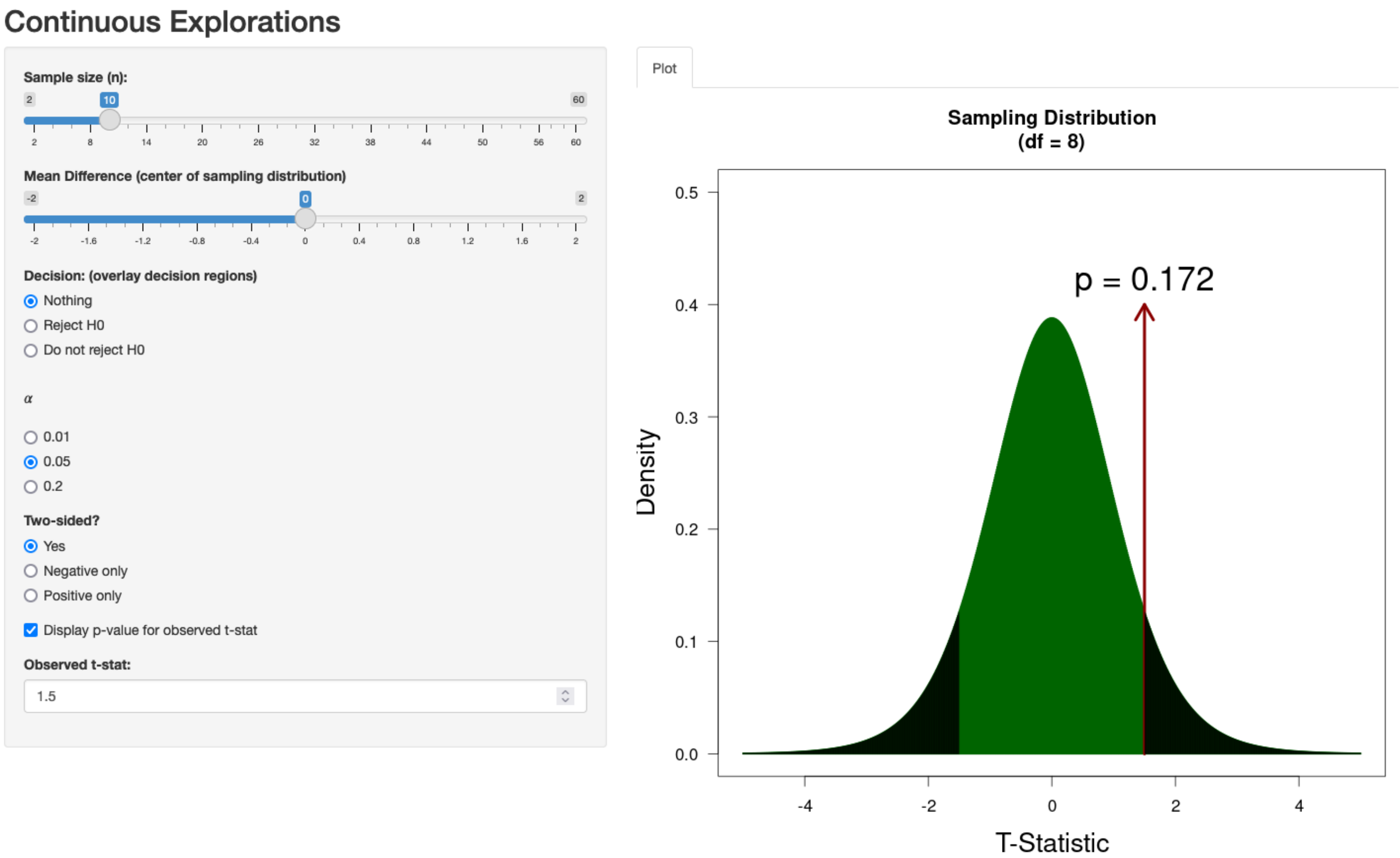


If H1 is one-sided **positive**, an observed t-value of -2.5 would not provide more evidence than -2.3, but 0.5 would, for instance

Source: pixabay.com

Stuff for understanding these concepts

Link to the app, link to flowchart, link to the extra explanation



Hypothesis testing

- The five steps of a significance test
  1. Assumptions
  2. Hypothesis
  3. Test statistic
  4. P-value
  5. **Conclusion**

5. Conclusion

- If we chose α = 0.05, then we should reject H_0
- Because $p = 0.03 < 0.05$

Under the null hypothesis, there is a 3% probability of observing the observed test statistic, or more extreme

We have some evidence that the pills change your IQ (two-sided test) and increase your IQ (one-sided test)

But remember…

Statistical decision

True state of the world

	H_0 is not rejected	H_0 rejected
H_0 is true	***1 – P(Type I error)***	*Type I Error (α)*
H_a is true	*Type II Error*	***Power = 1 – P(Type II error)***

Connection between p and CI

When $p < \alpha$, 0 will be outside a $(1-\alpha)\%$ CI

For example, we collect some data and…

- if 0 is not in the 95% confidence interval, the corresponding t-test/*p*-value will be significant for $\alpha = 0.05$
- if 0 is not in the 99% confidence interval, the corresponding t-test/*p*-value will be significant for $\alpha = 0.01$
- if 0 is not in the 90% confidence interval, the corresponding t-test/*p*-value will be significant for $\alpha = 0.1$

Overview of Today

1. Recap
    1. Hypothesis testing
    2. Confidence interval

2. Comparing two groups
    1. Estimation: confidence interval
    2. Hypothesis Testing: *t*-test

3. **Recap**
    1. Next time
    2. Example exam question

Recap

1. In order to analyze a difference in the means of two *populations*, we look at the means of two *samples*
2. To do so, we can use confidence intervals (estimation), or the *t*-test (hypothesis testing)
3. When we do hypothesis testing, we look at how likely our observed difference is, if the null hypothesis were to be true
4. If the observed difference is very unlikely under the null hypothesis, the null hypothesis is probably not true and we reject it

Recap – basic flow of conducting *t*-test:

Step 0: determine your alpha and alternative hypothesis

1. Compute the sample statistics (mean and sd), get df
2. Calculate se
3. Calculate *t*
4. Look up the probability of t, using Excel
5. If the test is two-sided (i.e., if the H_a is two-sided), multiply by 2 → this is your p-value
6. Compare to the pre-specified alpha level

Next time

- Today I introduced the basic method to compare two groups
  - ❼ Really: two uncertain estimates of two population parameters

- Next week: Variants of this
  - Does *t*-test change for a within-groups design? Yes
  - Different IV (categorical)
  - Different DV *and* IV (yet another distribution “Chisquare”)

Example exam question

Assume we conduct an experiment with two groups (treatment and control group), with 70 participants randomly assigned to one of each group (i.e., 35 per group). The data can be found in ExampleExamQ.xls (on Canvas).

What is the value of the *t* statistic if you want to test whether the difference in means of these groups deviates from 0?

a) 1.96

b) 4.03

c) 2.69

Solution

- $t = \frac{(\bar{x}_{treatment} - \bar{x}_{control}) - 0}{se}$
- Check the Excel file
- $\bar{x}_{treatment} = 11.06$
- $\bar{x}_{control} = 8.37$

- $se = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

Solution

- $se = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
- $s_1 = 2.92$
- $s_2 = 2.66$
- $n_1 = n_2 = 35$

- $se = \sqrt{\frac{2.92^2}{35} + \frac{2.66^2}{35}} = 0.67$

$$t = \frac{(\bar{x}_{treatment} - \bar{x}_{control}) - 0}{se}$$

$$t = \frac{(11.06 - 8.37) - 0}{0.67} = 4.03$$

Example exam question

Assume we conduct an experiment with two groups (treatment and control group), with 70 participants randomly assigned to one of each group (ie 35 per group). The data can be found in ExampleExamQ.xls (on Canvas).

What is the value of the *t* statistic if you want to test whether the difference in means of these groups deviates from 0?

a) ~~1.96~~ → critical z-value for 95%CI

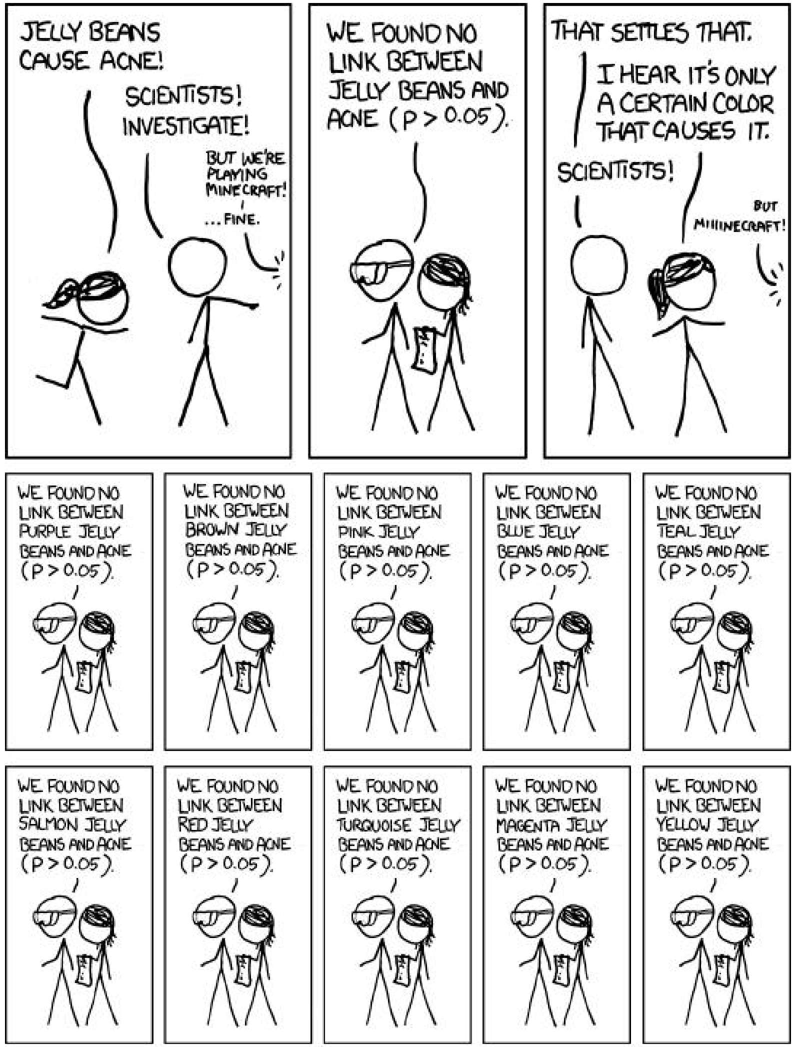
b) 4.03

c) ~~2.69~~ → unstandardized mean difference

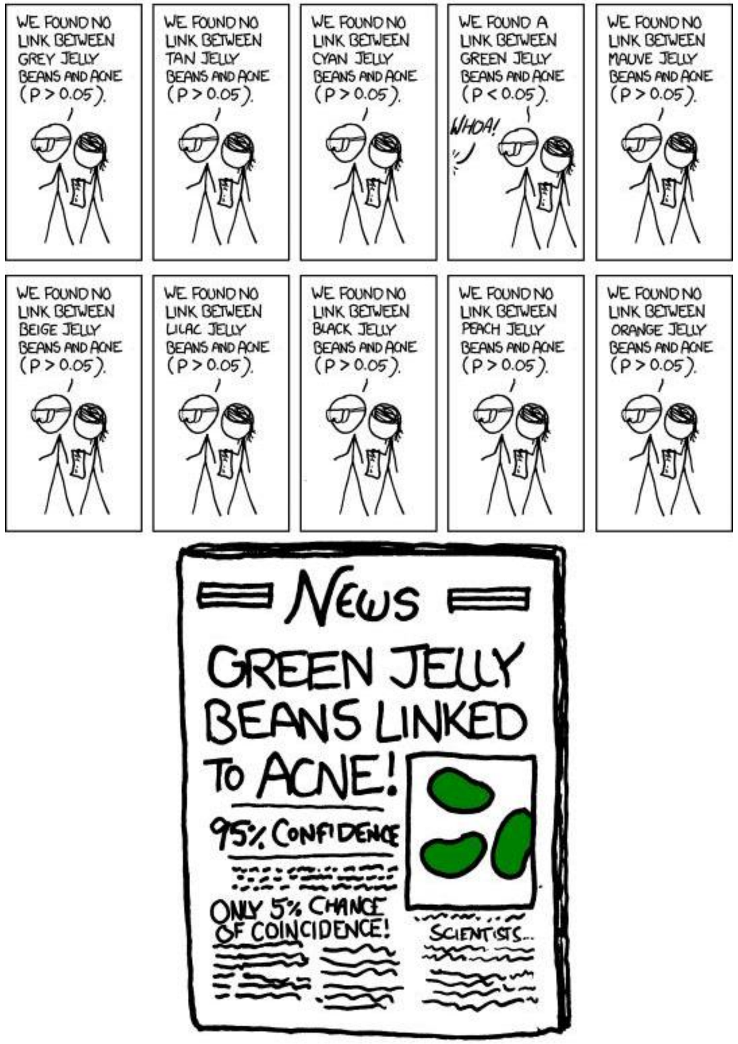
Note that this answer is slightly different from the Excel answer on Canvas (which is 4.024) due to rounding – on the exam we advise to always round to 4 decimals. Also, the options will not be so close together.

Questions?

Thank you for your attention



https://xkcd.com/882/



Excel commands for calculations:

Regular numeric operations:
= 2.3 + 3 - 7 * SQRT(10) / 4^2

For distributions → surface area:
= NORM.DIST(2, 0, 1, TRUE)
= T.DIST(2, 14, TRUE)
= CHISQ.DIST(8, 1, TRUE)

You start with inputting the observed test statistic. Then you give the df (for X^2 and t distributions), or the mean and sd (for the normal/z distribution). Then you add TRUE to get the area under the curve, to the LEFT of the test statistic.

For distributions-→ critical value:
= NORM.INV(0.025, 0, 1)
= T.INV(0.025, 20)
= CHISQ.INV(0.95, 1)

You start with inputting the desired probability. For instance, 0.025 gives you the value of the distribution, where to the left of that value, 2.5% of the distribution is situated. This is useful when constructing the 95% CI (not relevant for X^2).
Then you give the df (for X^2 and t distributions), or the mean and sd (for the normal/z distribution).

→ Paste these commands into any cell and press enter!

Excel commands for calculations:

Regular numeric operations:
= 2.3 + 3 - 7 * SQRT(10) / 4^2

If you do not want to pay for Microsoft office and do not like pirates, take a look at the open-source (/free) LibreOffice!
https://www.libreoffice.org/

Bonus Video

Going from a number (i.e., p-value) to a verdict (i.e., how much evidence do I have for my hypothesis) can be a pretty tricky thing. Over the past 50 years, researchers have been using certain thresholds for the p-value that dictate how happy you should be when observing a certain p-value. This has been dubbed "The Dance of the p-values" by Geoff Cumming, and his YouTube video on the topic provides a good disclaimer for working with p-values (or any arbitrary statistical decision making).

p value scale					
***	.001	Very highly significant!!!	There IS an effect. Definitely, for sure!	Elation!! Exuberance!! Smugness?	Nobel Prize, Tenure, Research grant
**	.01	Highly significant!!	There is an effect.	Great pleasure, Dancing, Drinking	PhD, Prize, Top publication
*	.05	Significant (phew!)	Most likely, there is an effect.	Relief, Cheerfulness	Consolation prize, Fair publication
?	.10	Approaching significance	Almost. Probably an effect, but low power?	Frustration, 'if only'	Counselling, stress leave
p>.10		Nonsignificant	No effect (effect is zero?)	Despair, depression	Medication, Reconsider life goals

Source: https://www.youtube.com/watch?v=5OL1RqHrZQ8

Highlighted exercises from the book

→ try yourself first, then check next slide for answer

- 10.22
  - "Chelation is an alternative therapy…"
- 10.28 (skip plot; see Canvas for excel file – link below lecture slides)
  - "refer to the FL Student Survey data file on the book's website… "

Highlighted excercises from the book

10.22:

A) The 95% confidence interval for the difference in population means (chelation − placebo) is from −53 to 36 seconds. This means we are 95% confident that the true average difference in treadmill time between the chelation and placebo groups lies somewhere between 53 seconds less and 36 seconds more for chelation. Because the interval includes 0, it suggests that there may be no true difference between the two treatments.

B) Hypotheses for the two-sided test
   H_0: μ_chelation = μ_placebo
   H_a: μ_chelation ≠ μ_placebo

C) The p-value = 0.69 is much greater than 0.05, so we fail to reject H_0.
   This indicates no statistically significant difference between the groups.
   This conclusion is consistent with the confidence interval, which includes 0 and thus also suggests no evidence of a real effect of chelation therapy compared with placebo.

Highlighted excercises from the book

10.28:
Means: 3.7742 & 4.4138
Sd's: 2.9293 & 3.0997
N: 31 & 29

SE: 0.7798

CI: (-2.201, 0.921)
0 is included in the interval, so we cannot conclude that newspaper reads differ between the groups

Continued on next slide...

Highlighted exercises from the book

10.28:

1)Assumptions – enough observations for CLT (30), sd's are similar

2)Hypothesis – two-sided hypothesis test

3)Test statistic – (3.7742 - 4.4138) / 0.7798 = -0.8202

4)P-value - T.DIST(-0.8202, 58, TRUE) * 2 = 0.415

5)Conclusion – not significant against alpha 0.05