MATH 7343 Applied Statistics

Prof. (Aidong) Adam Ding

Northeastern University

Review

- Last time, we finished Module 3 (confidence intervals)

- We also started on the hypothesis testing (Chapter 10).

Two types of errors in hypothesis testing

Decision		True state: H_0 is true (innocent)	True state: H_A is true (murder)
	Fail to reject H_0 (Not guilty)	Correct $1-\alpha$	**Type II error** β
	Reject H_0 (Not guilty)	**Type I error** α	Correct $1-\beta$

α = P(reject H_0 | H_0 is true) is also called **significance level** (or size)

β = P(fail to reject H_0 | H_A is true). 1- β is called the **power**.

How to conduct hypothesis testing

At α level (type I error rate):

- To test **H$_0$**: $\mu \geq \mu_0$ versus **H$_A$** : $\mu < \mu_0$

  Reject **H$_0$** if $T_{obs} = \frac{\bar{X}_{obs} - \mu_0}{s/\sqrt{n}} < -t_{n-1,\alpha}$

  $\Leftrightarrow \bar{X}_{obs} < \mu_0 - t_{n-1,\alpha}\, s/\sqrt{n}$

- To test **H$_0$**: $\mu \leq \mu_0$ versus **H$_A$** : $\mu > \mu_0$

  Reject **H$_0$** if $T_{obs} = \frac{\bar{X}_{obs} - \mu_0}{s/\sqrt{n}} > t_{n-1,\alpha}$.

2-sided hypothesis testing

We may need to test 2-sided hypothesis (E.g., do the new drug perform different from the standard drug?)

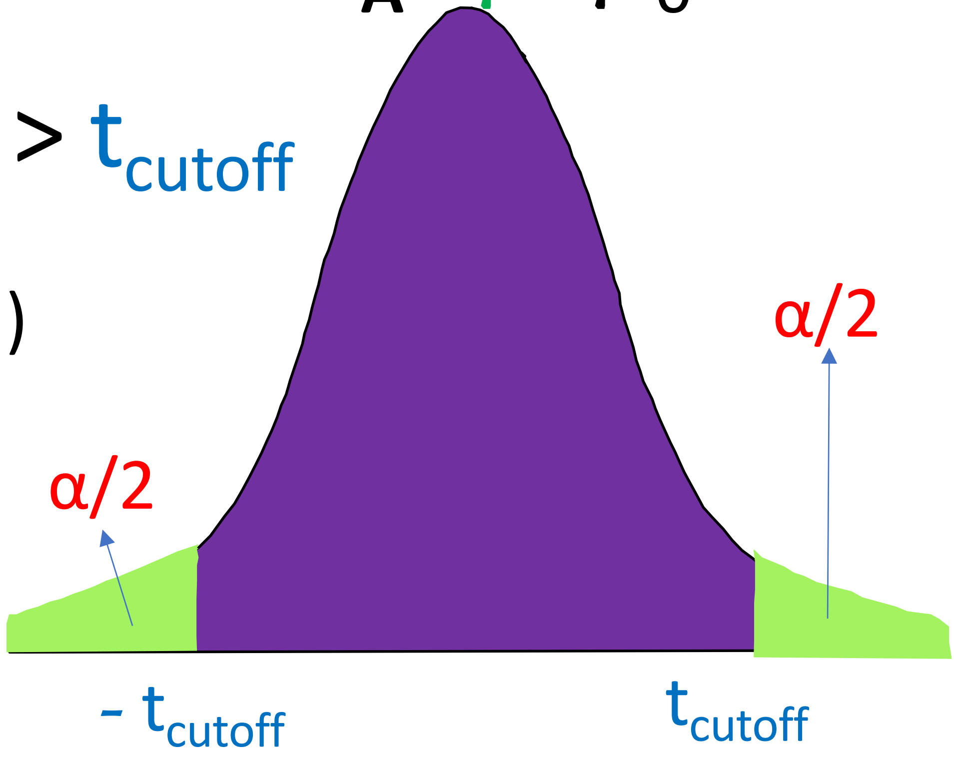
- To test, at α level, $\mathbf{H_0}$: $\mu = \mu_0$ versus $\mathbf{H_A}$: $\mu \neq \mu_0$

Reject $\mathbf{H_0}$ if $T_{obs} = \left|\frac{\bar{X}_{obs}-\mu_0}{s/\sqrt{n}}\right| > t_{cutoff}$

Since we want $\alpha = P\left(\left|\frac{\bar{X}_{obs}-\mu_0}{s/\sqrt{n}}\right| > t_{cutoff} \mid \mu=\mu_0\right)$

$= P(|t_{n-1}| > t_{cutoff})$

So $t_{cutoff} = t_{n-1,\alpha/2}$.



Pain reliever example: 2-sided test

Do the new drug and standard drug perform differently?

- Test **H$_0$**: μ=3.5 versus **H$_A$** : $\mu \neq$ 3.5
- At α=0.05 level, from Table A.4, $t_{119,0.025} \approx t_{120,0.025}$ = 1.98

(If use R, qt(1-0.025, df=119) gives a more accurate number)

In fact, we observe T_{obs}= $\left|\frac{3.2-3.5}{1.14/\sqrt{120}}\right|$ = 2.88 indeed >1.98

Hence, we do <u>reject</u> **H$_0$** at α=0.05 level

- Conclusion: Two drugs do perform differently.

Significance test (p-value)

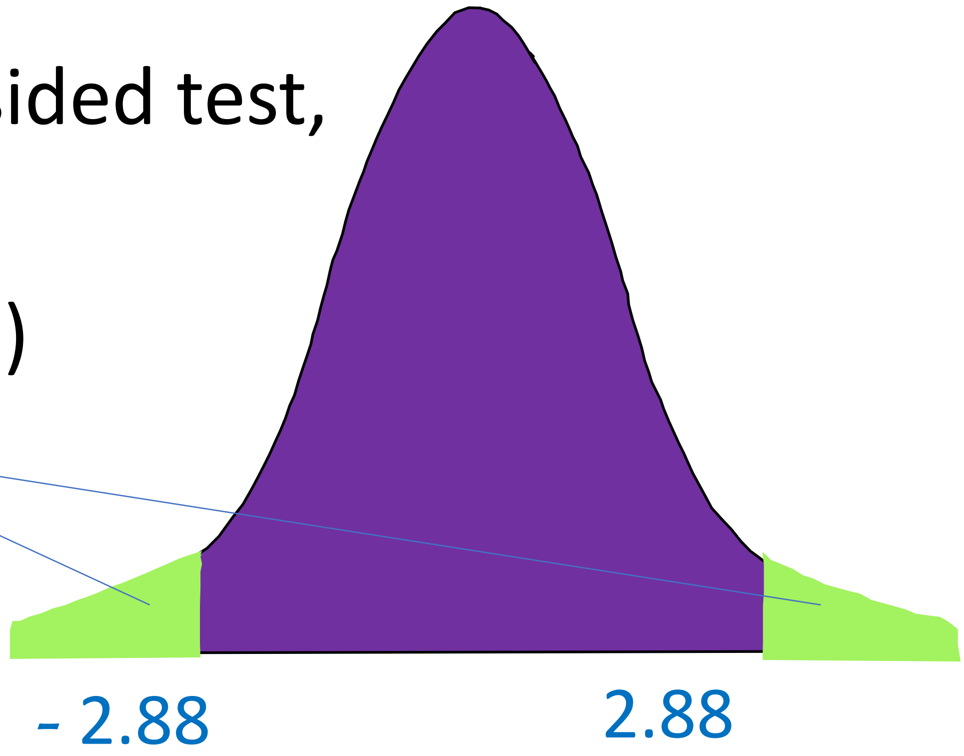
P-value

= P(obtaining data as extreme or more extreme than observed | H_0 is true)

- In the pain reliever example, for 2-sided test,
  T_{obs}=2.88.

So the p-value = P(T≥2.88 | H_0 is true)

= P(| t_{119} |≥2.88)



Pain reliever example: 2-sided p-value

Do the new drug and standard drug perform differently?

p-value = $P(|\, t_{119} \,| \geq 2.88)$

- From Table A.4,

$t_{120,0.005}$ =2.617, $t_{120,0.0005}$ =3.373

TABLE A.4
Percentiles of the *t* distribution

	Area in Upper Tail					
df	0.10	0.05	0.025	0.01	0.005	0.0005
1	3.078	6.314	12.706	31.821	63.657	636.619
2	1.886	2.920	4.303	6.965	9.925	31.599
3	1.638	2.353	3.182	4.541	5.841	12.924
4	1.533	2.132	2.776	3.747	4.604	8.610
...						
110	1.289	1.659	1.982	2.361	2.621	3.381
120	1.289	1.658	1.980	2.358	2.617	3.373
∞	1.282	1.645	1.960	2.327	2.576	3.291

So 2(0.0005)<p-value<2(0.005)

Or 0.001< p-value< 0.01

So reject $\boldsymbol{H_0}$ *at 0.01 level but not at 0.001 level.*

Using R, we get a more accurate number:

2(1-pt(2.88, df=119)) = 0.004717498

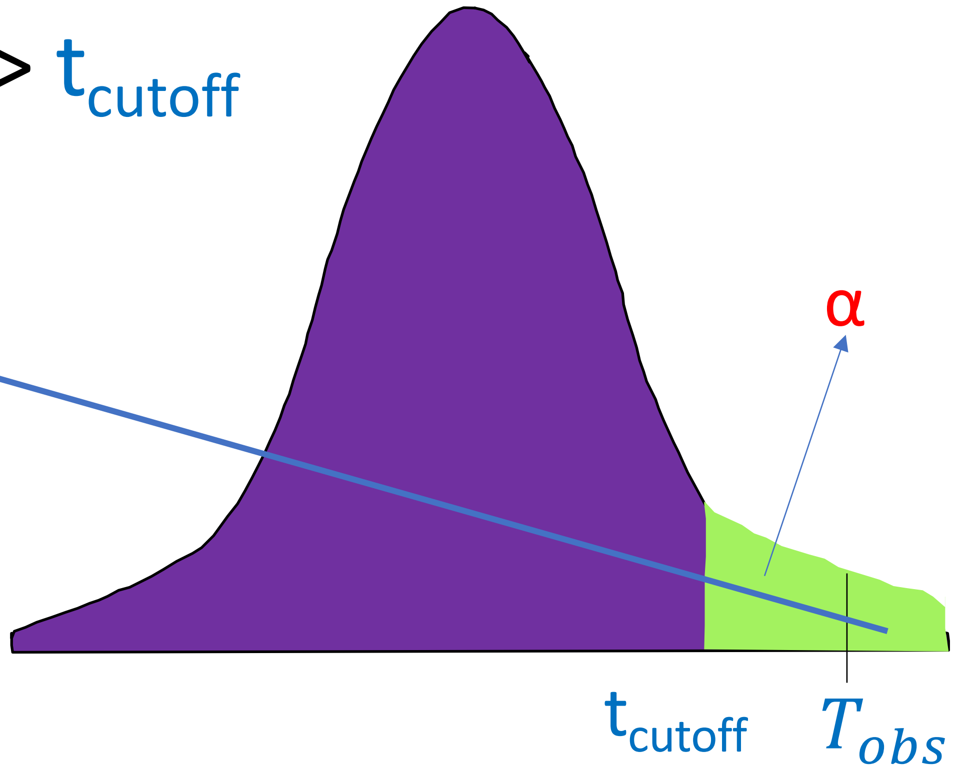
Significance test versus hypothesis test

- A hypothesis test rejects $\mathbf{H_0}$ if and only if p-value of the corresponding significance test is less than α.
- For example, to test $\mathbf{H_0}$: $\mu \leq \mu_0$ versus $\mathbf{H_A}$: $\mu > \mu_0$

1-sided test rejects $\mathbf{H_0}$ when $T_{obs} > t_{cutoff}$

where $P(T \geq t_{cutoff} \mid H_0) = \alpha$.

Recall p-value=$P(T \geq T_{obs} \mid H_0)$



So p-value< $\alpha \Leftrightarrow T_{obs} > t_{cutoff}$

Confidence interval versus hypothesis test

- A (1- α) 1-sided upper CI for μ is $(-\infty, \bar{X} + t_{n-1,\,\alpha}\frac{s}{\sqrt{n}})$.
- The 1-sided test for $\mathbf{H_0}$: $\mu = \mu_0$ versus $\mathbf{H_A}$: $\mu < \mu_0$ rejects $\mathbf{H_0}$ at α level when $T_{obs} < -t_{n-1,\,\alpha}$

$$\Leftrightarrow \frac{\bar{X}_{obs} - \mu_0}{s/\sqrt{n}} < -t_{n-1,\alpha}$$

$$\Leftrightarrow \bar{X}_{obs} - \mu_0 < -\, t_{n-1,\alpha}\, s/\sqrt{n}$$

$$\Leftrightarrow \mu_0 > \bar{X}_{obs} + t_{n-1,\alpha}\, s/\sqrt{n}$$

So <u>rejects</u> $\mathbf{H_0}$ at α level if and only if true parameter value μ_0 is <u>outside</u> the (1- α) confidence interval.

Confidence interval versus hypothesis test

- For one-sided test, rejects **H_0** at α level if and only if true parameter value μ_0 is outside the (1- α) one-sided confidence interval.

- Similarly, for two-sided test, rejects **H_0** at α level if and only if true parameter value μ_0 is outside the (1- α) two-sided confidence interval.

- Due to this equivalence, t.test() gave CI.

R commands for hypothesis testing

Recall the grocery example (In CalculateCI.pdf)

- Test **H_0**: μ=22 versus **H_A** : $\mu \neq$ 22

```
> t.test(x, mu=22)
        One Sample t-test
data:  x
t = 1.6794, df = 49, p-value = 0.09945
alternative hypothesis: true mean is not equal to 22
95 percent confidence interval:
 21.24567 30.42713
sample estimates:
mean of x
  25.8364
```

- $T_{obs} = \left|\frac{\bar{X}_{obs}-22}{s/\sqrt{n}}\right| = 1.6794$, df=n-1=49, p-value=P($|t_{49}|\geq 1.6794$)=0.09945 >0.05
- So fail to reject **H_0: μ=22** at α=0.05 level
- Corresponding, the 95% 2-sided CI (21.25, 30.43) contains 22.

R commands for hypothesis testing

Recall the grocery example (In CalculateCI.pdf)

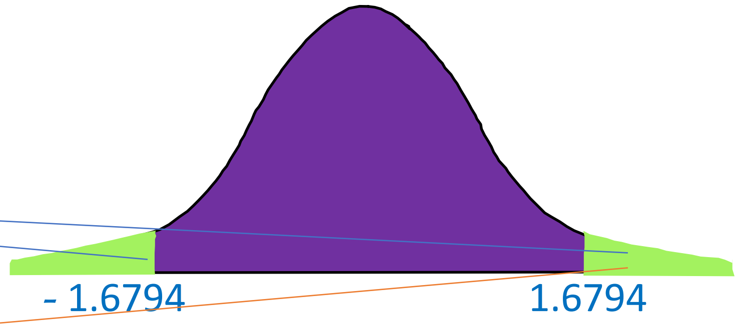
- How about 1-sided test **H$_0$**: $\mu \le 22$ versus **H$_A$** : $\mu > 22$

From R outputs above, we already get

$T_{obs} = \left|\frac{\bar{X}_{obs} - 22}{s/\sqrt{n}}\right|$ = 1.6794, df=n-1=49,

2-sided p-value=P(|t_{49}|≥1.6794)=0.09945

1-sided p-value=?



$$=P(t_{49} \ge 1.6794) = \frac{1}{2}P(|t_{49}| \ge 1.6794) = \frac{1}{2}(0.09945) = 0.0497$$

- So reject **H$_0$: $\mu \le 22$** at α=0.05 level.

- Notice that we do the test through p-value in R outputs always.
- For 1-sided test, we can let R do it directly by

t.test(x, mu=22, alternative="greater")

Statistical significance vs. practical significance

- Example: Does Coca-Cola 2-liter bottle contains 2 liter coca on average?
- Statistical Test **H$_0$**: μ= 2 versus **H$_A$** : $\mu \neq 2$

  Assume that σ=0.01 liter. Test at α=0.05 level

- Sample 10 bottles, reject H$_0$ if Z = $\left|\frac{\bar{X}-2}{0.01/\sqrt{10}}\right|$ >$z_{\alpha/2}$ =1.96

⇔ $\bar{X}$> 2+ (1.96)0.01/$\sqrt{10}$= 2.0062 or $\bar{X}$< 2 - (1.96)0.01/$\sqrt{10}$= 1.9938

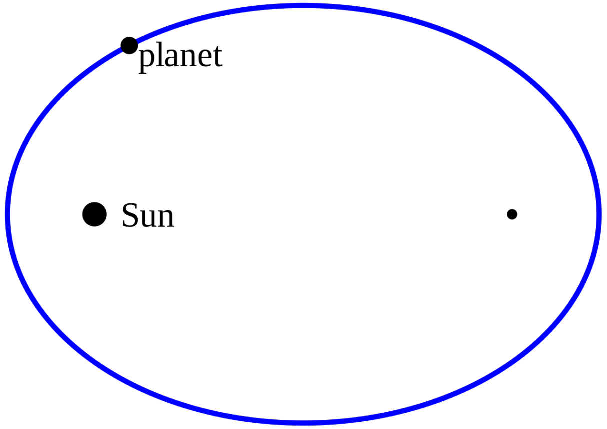
- Sample 1 million bottles, reject H$_0$ if Z = $\left|\frac{\bar{X}-2}{0.01/\sqrt{1{,}000{,}000}}\right|$ > 1.96

⇔ $\bar{X}$> 2+ (1.96)0.00001= 2.000022 or $\bar{X}$< 1.99998

- Do we really care if μ= 2.00001 or 1.999?

Statistical significance vs. practical significance

- Example: Kepler's Law says that the orbit of a planet is an ellipse with the Sun at one of the two foci.
- Statistical Test formulation



  $\mathbf{H_0}$: The orbit is an ellipse.

  $\mathbf{H_A}$: The orbit is NOT an ellipse.

- Using Kepler's measurements, we can not reject $\mathbf{H_0}$ (so accept $\mathbf{H_0}$).
- If we use today's measurements, we will reject $\mathbf{H_0}$
- More accurate measurements bring punishment.

Statistical significance vs. practical significance

How to deal with this issue?

- One way: define practical importance first then test.

Cola example:

We are concerned only if the Cola 2-liter bottle μ<1.99

Then we test **H_0**: $\mu \geq 1.99$ (instead of $\mu \geq 2$) versus **H_A** : $\mu < 1.99$

- Second way: always report the confidence interval.

For example, in two cases the 95% CIs are respectively (1.85,1.95) and (1.9997,1.99998). While both case reject $\mu \geq 2$, you can make judgement on the CI on whether practically the results are significant.

Power calculation

Recall that **power** = 1- β = P(Reject H_0 | H_A is true).

We can calculate this if we know the truth under H_A.

		True state	
		H_0 is true (innocent)	H_A is true (murder)
Decision	Fail to reject H_0 (Not guilty)	Correct $1-\alpha$	**Type II error** β
	Reject H_0 (Not guilty)	**Type I error** α	Correct $1-\beta$

Power calculation

- Example: 1-sided Test $\mathbf{H_0}$: μ= 2 versus $\mathbf{H_A}$: μ<2

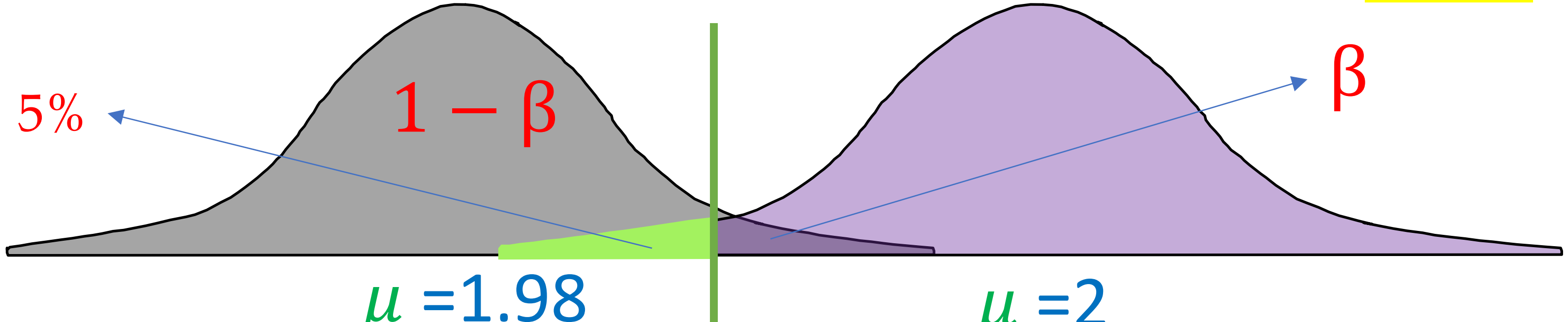
Assume that σ=0.1. Reject H_0 if $\frac{\bar{X}-2}{0.1/\sqrt{n}} < -z_{\alpha}$.

Say n=100, α=0.05 and true μ=1.98.

Then power =P(Reject H_0 | μ=1.98) =P($\frac{\bar{X}-2}{0.1/\sqrt{100}}$ < -1.645 | μ=1.98)

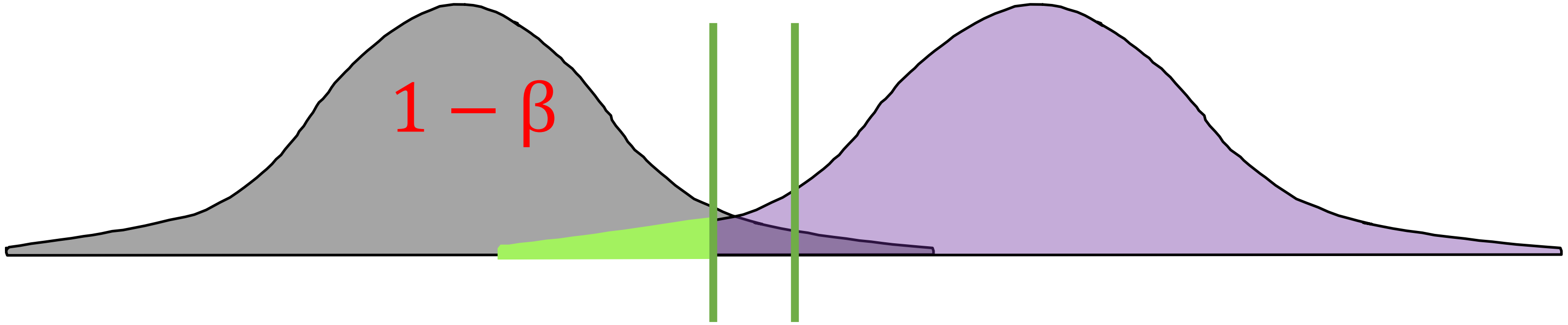
=P($\frac{\bar{X}-1.98-0.02}{0.01}$ < -1.645 | μ=1.98) = P($Z - \frac{0.02}{0.01}$ < -1.645)

=P(Z < 0.355) = 1- P(Z ≥ 0.355) = 1-0.361 = 0.639 (TableA.3)



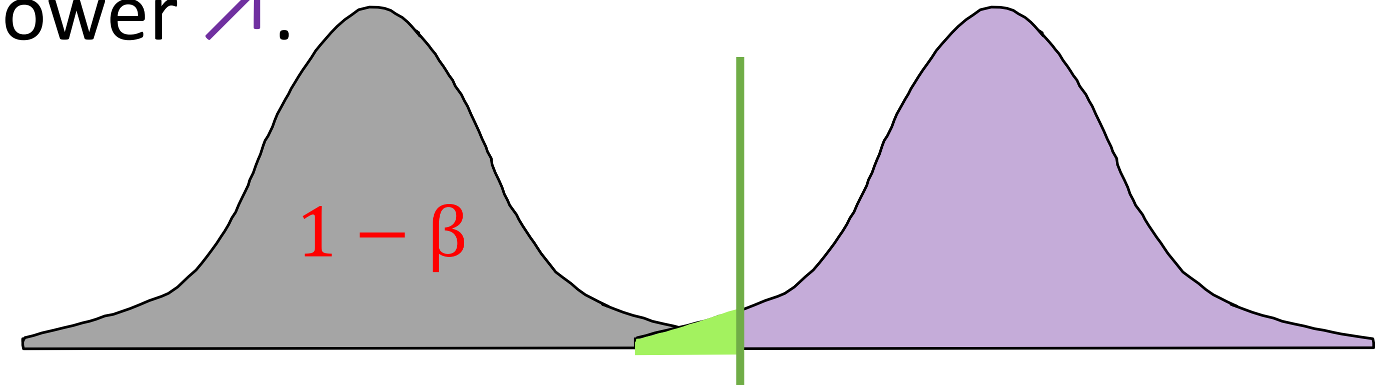
Power calculation

- Notice that the power changes with α and n.
- As α $\nearrow$, the cutoff line $\nearrow$, thus the power $\nearrow$.



- As n $\nearrow$, the two distributions are more separated.

So the power $\nearrow$.



Sample size calculation

- For cola bottles, we want to test $\mathbf{H_A}$: $\mu<2$ at level α=0.05 (assume that we know σ=0.1).

We also want the power to be 90% when μ =1.99.

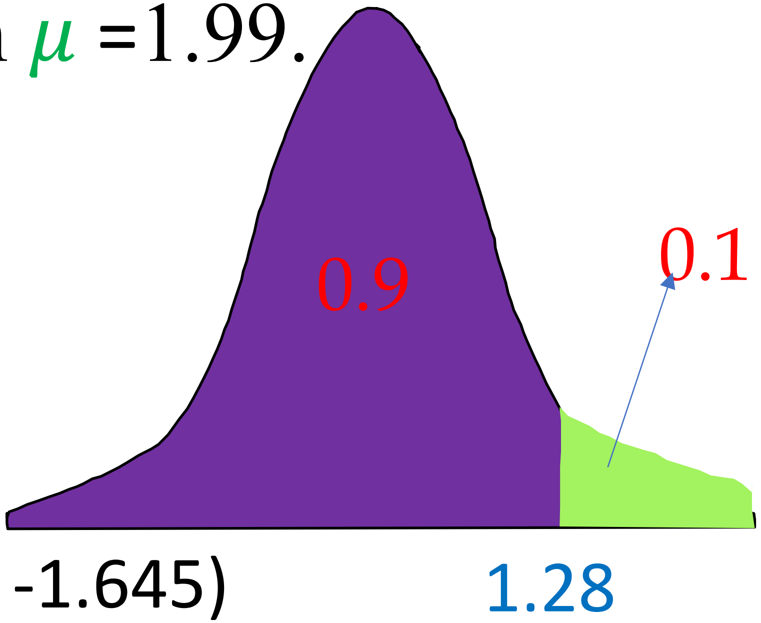
How many bottles are needed?

- Solution: Want

$$0.90 = \text{power} = P\left(\frac{\bar{X}-2}{0.1/\sqrt{n}} < -1.645 \mid \mu=1.99\right)$$

$$= P\left(\frac{\bar{X}-1.99}{0.1/\sqrt{n}} - \frac{0.01}{0.1/\sqrt{n}} < -1.645 \mid \mu=1.99\right)$$

$$= P(Z - 0.1\sqrt{n} < -1.645) = P(Z < 0.1\sqrt{n} - 1.645)$$



So 1.28 = $0.1\sqrt{n}$ -1.645 (TableA.3) $\Rightarrow \sqrt{n}$ =29.25,

so n=$(29.25)^2$ =855.6 = 856 (Round up!)

At least 856 bottles are needed for the test.

Sample size calculation

- Notice that for sample size calculation, we have used z-interval and z-test. Why?

- Generally for inference, we use t-interval and t-test. For sample size calculation in experimental design, we used z-interval and z-test. (Actually, more precise calculation should also use t-interval and t-test. Special software is needed for this.)

Inferences about the population mean μ

- Point estimator $\overline{X}$: $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ ~ N(0,1), $\frac{\overline{X}-\mu}{s/\sqrt{n}} \sim t_{n-1}$
- Confidence intervals: σ known or σ unknown

2-sided: $(\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}})$ or $(\overline{X} - t_{n-1,\,\alpha/2}\frac{s}{\sqrt{n}}, \overline{X} + t_{n-1,\,\alpha/2}\frac{s}{\sqrt{n}})$

1-sided: $(\overline{X} - z_{\alpha}\frac{\sigma}{\sqrt{n}}, \infty)$ / $(-\infty, \overline{X} + z_{\alpha}\frac{\sigma}{\sqrt{n}})$ or $(\overline{X} - t_{n-1,\,\alpha}\frac{s}{\sqrt{n}}, \infty)$ / $(-\infty, \overline{X} + t_{n-1,\,\alpha}\frac{s}{\sqrt{n}})$

- Test **H_0**: $\mu = \mu_0$ versus **H_A** : $\mu \neq \mu_0$

Reject **H_0** if $T_{obs} = \left|\frac{\bar{X}_{obs}-\mu_0}{s/\sqrt{n}}\right| > t_{n-1,\alpha/2}$

p-value=$P(T \geq T_{obs} \mid H_0)$ = max level to reject **H_0** for observed data.

<u>Reject H_0 if p-value<α</u>. (equivalent to μ_0 outside 1-α CI).

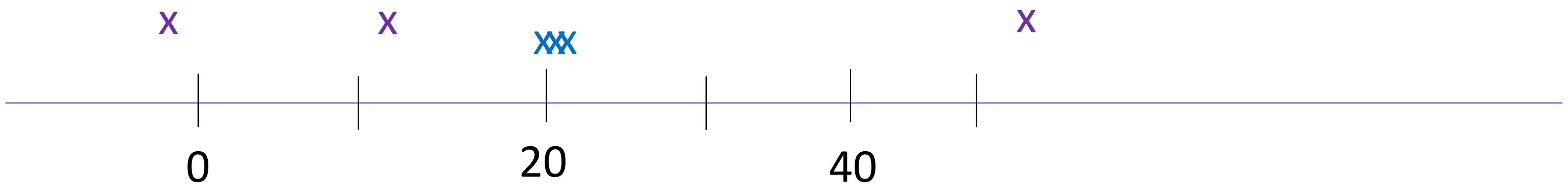
Inferences about the population mean μ

How much of these inferences we can do without probability theory?

- Point estimator is the best guess. So obviously we will use the sample mean $\bar{X}$.
- Interval estimate gives an assessment of uncertainty. We may not get the interval, but can have an intuitive comparison.
- Data set one: 11, 52, -3. Data set two: 20,21,20.5

$$\bar{X}_1 = \frac{11+52+(-3)}{3} = 20, \qquad \bar{X}_2 = \frac{20+21+20.5}{3} = 20.5$$

Obviously $\bar{X}_2$ is more accurate and data set two should have a shorter confidence interval.



Inferences about the population mean μ

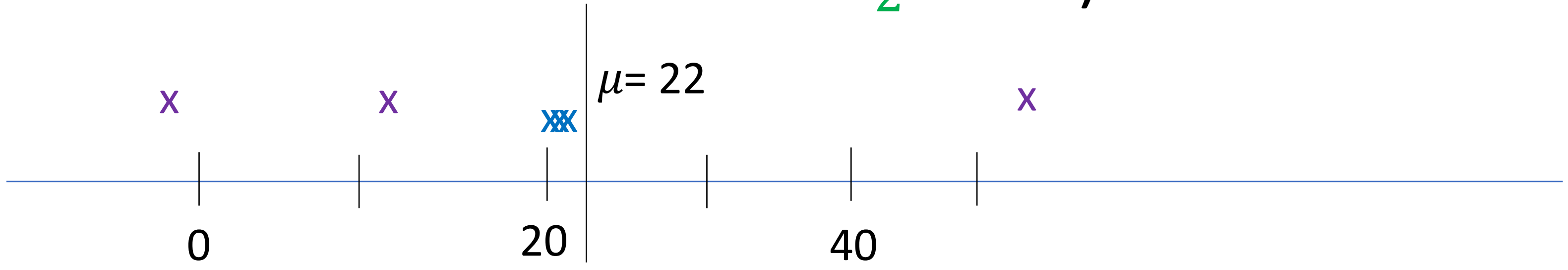
How much of these inferences we can do without probability theory?

- Test **H$_0$**: μ= 22 versus **H$_A$** : μ≠22

Without theory, can not say if reject or not. But Date Two is more likely to reject by intuition.

- Test **H$_0$**: μ= 20.5 versus **H$_A$** : μ≠20.5

Data set one more likely to reject. (That is also very unlikely, but more likely than data set two which has $\bar{X}_2$=20.5)



- Theory enable us to quantify the inference.

Summary

Module 4 done. You should know:

- How to set up your study as a hypothesis test
- Which test to use? t or z, 1-sided vs. 2-sided
- Sample size calculation for achieving power.
- Concepts related to hypothesis testing: Type I/II error rates, practical significance, p-value, etc.

- Homework 4 will be on this module and next together.
- Next lecture we cover the next module on two sample comparison (chapter 11).