t distribution

Dr. Mine Çetinkaya-Rundel
Duke University

review:
what purpose does a large sample serve?

As long as observations are independent, and the population distribution is not extremely skewed, a large sample would ensure that...

▸ the sampling distribution of the mean is nearly normal

▸ the estimate of the standard error is reliable: $\frac{s}{\sqrt{n}}$

t distribution

▸ when **σ** unknown (almost always), use the t distribution to address the uncertainty of the standard error estimate
▸ bell shaped but thicker tails than the normal
  ▸ observations more likely to fall beyond 2 SDs from the mean
  ▸ extra thick tails helpful for mitigating the effect of a less reliable estimate for the standard error of the sampling distribution
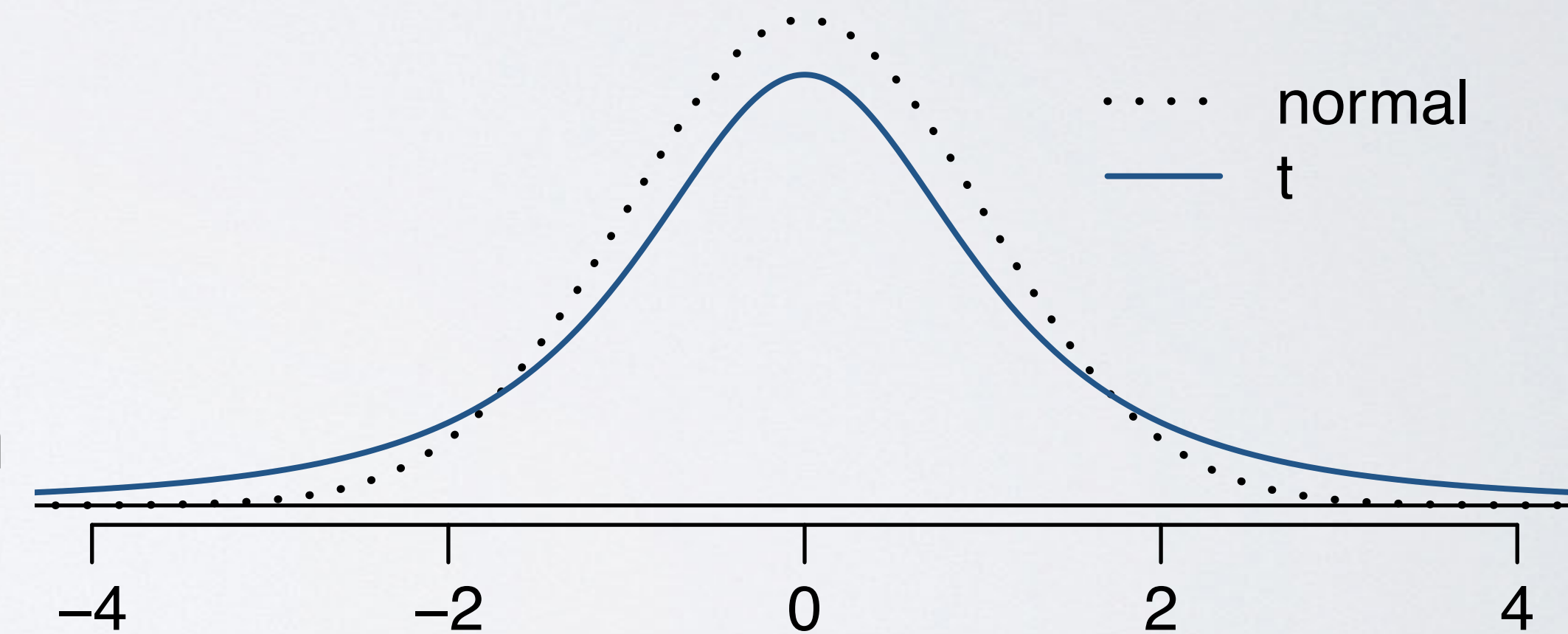
t distribution

- always centered at 0 (like the standard normal)
- has one parameter: degrees of freedom (df) - determines thickness of tails
  - remember, the normal distribution has two parameters: mean and SD

normal
t, df=10
t, df=5
t, df=2
t, df=1
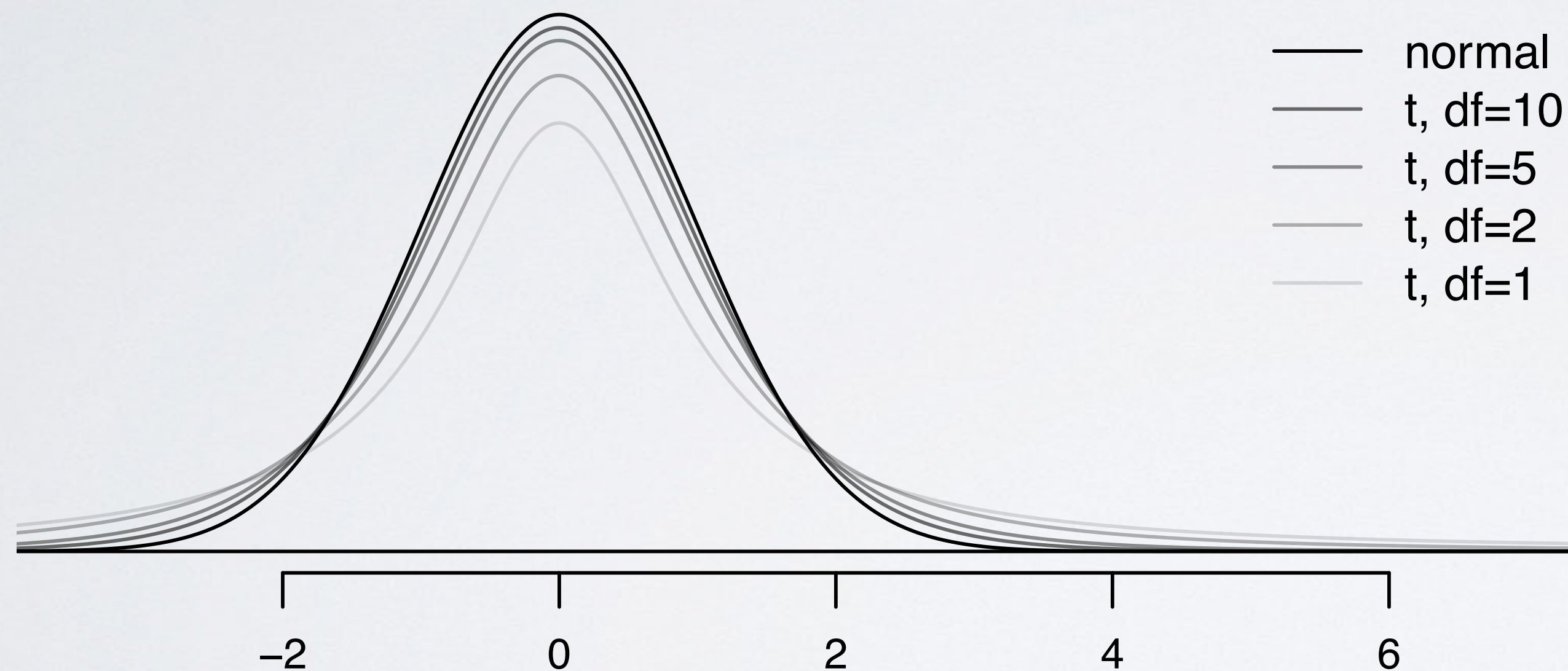
What happens to the shape of the t-distribution as degrees of freedom increases?

approaches the normal dist.

t statistic

▸ for inference on a mean where

    ▸ σ unknown, which is almost always

▸ calculated the same way

$$T = \frac{obs - null}{SE}$$

▸ p-value (same definition)

    ▸ one or two tail area, based on H_A

    ▸ using R, applet, or table

```R
R
> pnorm(2, lower.tail = FALSE) * 2
[1] 0.0455
> pt(2, df = 50, lower.tail = FALSE) * 2
[1] 0.0509
```
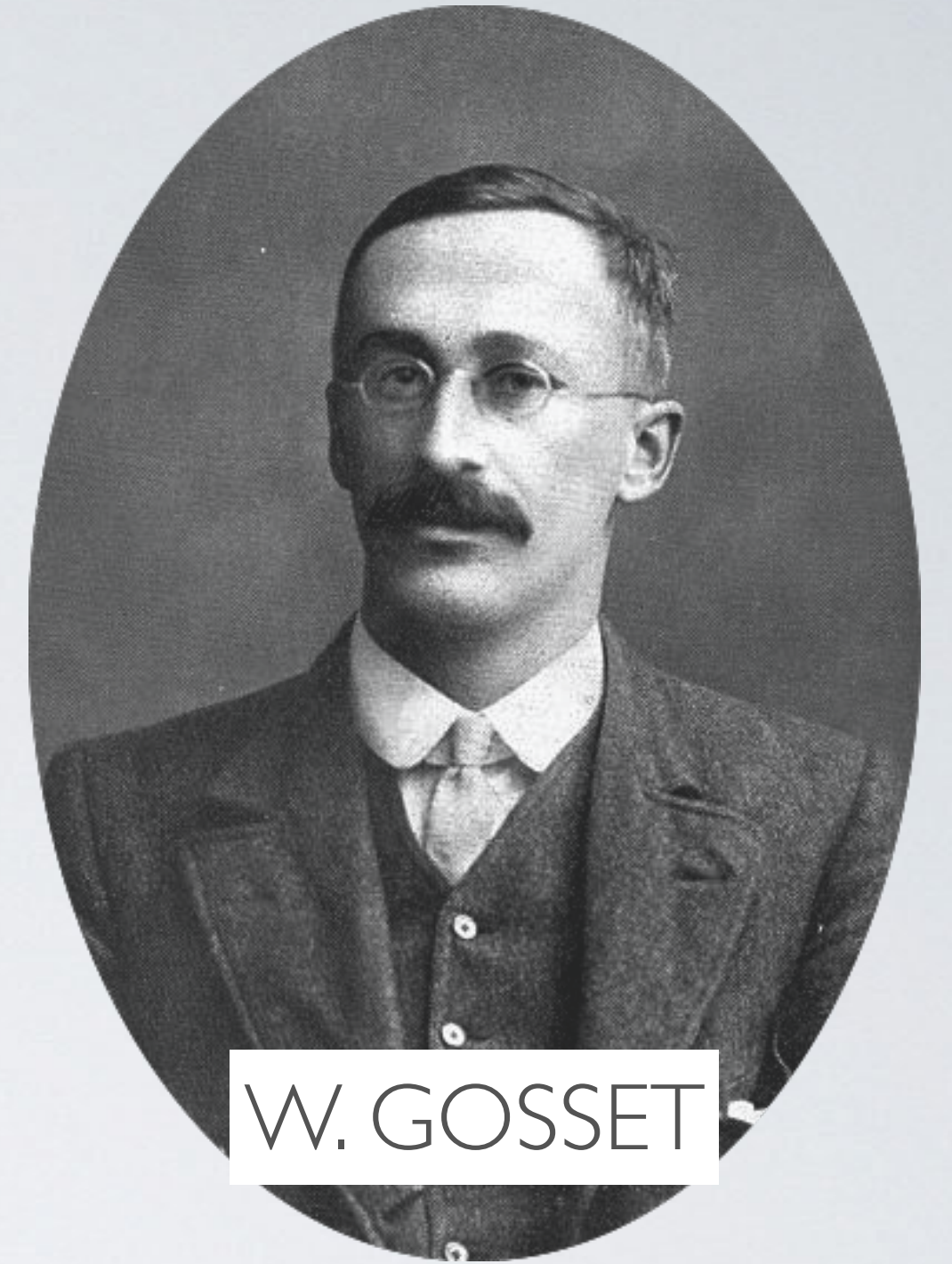
Suppose you have a two sided hypothesis test, and your test statistic is 2. Under which of these scenarios would you be able to reject the null hypothesis at the 5% sig. level?

Find the following probabilities.

a. $P(|Z| > 2)$ $0.0455 \longrightarrow$ *reject*

b. $P(|t_{df = 50}| > 2)$ $0.0509 \longrightarrow$ *fail to reject?*

c. $P(|t_{df = 10}| > 2)$ $0.0734 \longrightarrow$ *fail to reject*

origins of the t distribution

W. GOSSET

‣ Student's t

‣ William Gosset (1876 - 1937)

‣ "Head Experimental Brewer" at the Guinness brewing company