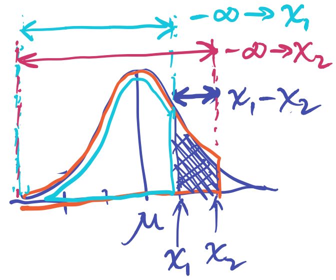
Ch 10: Normal Distribution (review)

10/15/2020

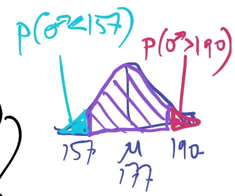
- Standard normal: $\mu=0, s.d.=1 \Rightarrow z=\frac{Y-\mu}{\sigma}$ {Z-distribution, Z-score}
- Probabilities for normal distribution
  $Pr[x_1 \le X \le x_2] = Pr[X \le x_2] - Pr[X \le x_1]$



Ex 10.4: Height of astronauts @ NASA
Must be b/w 157.5 – 190.5cm

Population: ♂ $\mu=177.6$, sd=9.7cm
♀ $\mu=163.2$, sd=10.1cm

Q: what is Prob. that ♂ or ♀ will NOT qualify?
$Pr[♂ < 157.5] \cup Pr[♂ > 190.5]$



(dqpr) ♂: pnorm(q_1, μ, sd) + pnorm(q_2, μ, sd, lower.tail=F) $\Rightarrow$ = 0.11 → 11.1%
$Pr[X \le q_1]$ $\quad$ $Pr[X > q_2]$

♀: 0.289 → 29%

CORRECTED → Equivalent: 1 – [pnorm(q_2, μ, sd) – pnorm(q_1, μ, sd)]
$P[X \le q_2]$ $\quad$ $Pr[X \le q_1]$

DEFAULT: lower.tail=F

- Distribution of sample means: If $Y \sim N$, then $\bar{Y} \sim N$
  Population SEM: $\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{n}} \Rightarrow z = \frac{\bar{Y}-\mu}{\sigma_{\bar{Y}}} \Rightarrow \bar{Y} = \mu + \sigma_{\bar{Y}} z$

Ex: Weight of babies at birth:
$n=80, \mu=3339, \sigma=573g$.

$\frac{\sigma}{\sqrt{n}} = \frac{573}{\sqrt{80}}$

① $Pr[\bar{Y} > 3370]$ = pnorm(q=3370, μ=3339, sd=SEM, lower.tail=F)

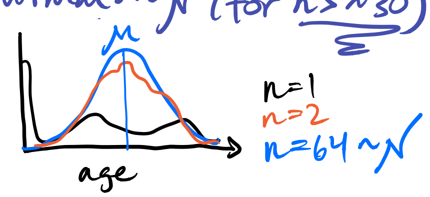
② $z = \frac{q-\mu}{SEM} = \frac{3370-3339}{SEM} \Rightarrow$ pnorm(z, 0, 1, lower.tail=F)

③ or equivalently $\Rightarrow$ pnorm($q-\mu$, 0, SEM, lower.tail=F)

- CLT: Σ or mean of random sample w/ large n from a non-normal population is approximately normal: $\sim N$ (for $n > \sim 30$)

⊛ Weak law of large numbers.

Ex 10.6 "Spanish" flu of 1918:



Ch 11: Inference for a Normal Population

- **Student's t-distribution**

$\Rightarrow$ sampling dist. for $\bar{Y} \sim N$ (CLT)

BUT don't know $\sigma_{\bar{Y}}$ = SEM. for population!

$\therefore$ Best estimate is: $SE_{\bar{Y}} = \frac{s}{\sqrt{n}}$

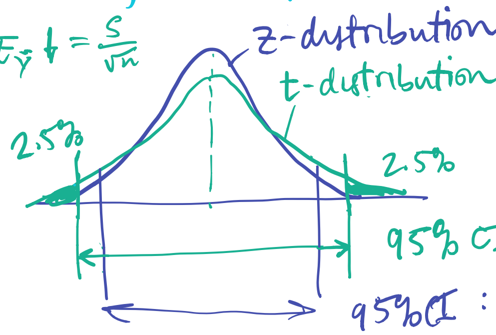
⊛ $t = \frac{\bar{Y} - \mu}{SE_{\bar{Y}}}$ "Student's t" w/ $n-1$ degrees of freedom

↳ Not a constant like $\sigma_{\bar{Y}}$ due to sampling error

$\therefore$ WIDER TAILS

why $n-1 = df$? Recall: $s^2 = \frac{\sum (Y_i - \bar{Y})^2}{n-1} \Rightarrow \frac{(\bar{Y}_i - \mu)^2}{n-1}$

As $n\uparrow$, $SE_{\bar{Y}}\downarrow = \frac{s}{\sqrt{n}}$



$\alpha = 0.05$

95% CI for $t_{2(\alpha),df} \to t_{2(0.05),df}$

95% CI : $z = \pm 1.96$.

e.g. If $df=5$, $t_{crit} \approx \pm 2.78$

vs. $z = 1.96$.

$\Rightarrow$ CI expresses precision of an estimate, follows t-dist. for $n<30$

"Critical value" of t-statistic for $n-1$ d.f. at significance level α:

$$-t_{crit(2)\alpha,df} < \frac{\bar{Y}-\mu}{SE_{\bar{Y}}} < t_{crit(2)\alpha,df} \Rightarrow \bar{Y} = \pm t_{crit} \cdot SE_{\bar{Y}} + \mu$$

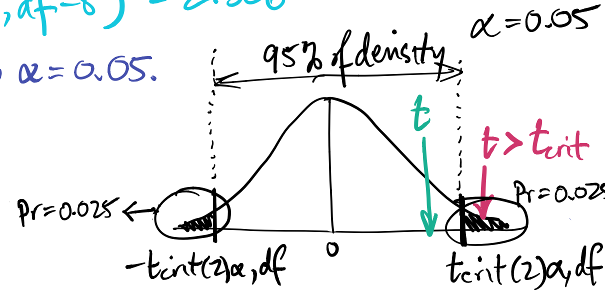
$$\left.\bar{Y} - t_{crit} \cdot SE_{\bar{Y}} < \mu < \bar{Y} + t_{crit} \cdot SE_{\bar{Y}} \right\} \text{ CI for } \mu$$

e.g. $t_{crit} = qt(0.025, df=8, lower.tail=F)$

$= qt(0.975, df=8) = 2.306$

The significance cutoff is $\alpha = 0.05$.

⊛ If the sample statistic $t_{obs} > t_{crit}$, then $p < \alpha$ and we reject H_0.



Ch 11.3 One-sample t-test.

H_0: True mean $= \mu_0$

H_A: " " $\neq \mu_0$

Assumptions:
- Random sample
- " variable $\sim N$

$\Rightarrow$ compare sample to pop $\quad t = \frac{\bar{Y} - \mu}{SE_{\bar{Y}}}$ THIS IS THE t-statistic

⊗ sampling distribution of t gives Pr[obs. data] under H_0.

Ex 11.3 Human body temp. 37°C ≈ 98.6°F (36.8)

$\underline{n = 25}$ vs $\underline{n = 130}$

$p > 0.05 \qquad p < 0.05$ REJECT H_0

since $\uparrow n \Rightarrow \downarrow SE_{\bar{Y}} = \frac{s}{\sqrt{n}}$

ONE-SAMPLE t-test
- compare to Exp. value of pop. param

⊗ t.test(body_temp, mu=98.6)

↳ data for your sample = VECTOR.

H_0: Mean body temp $= 98.6$

H_A: " " " $\neq 98.6$

} TWO-SIDED TEST

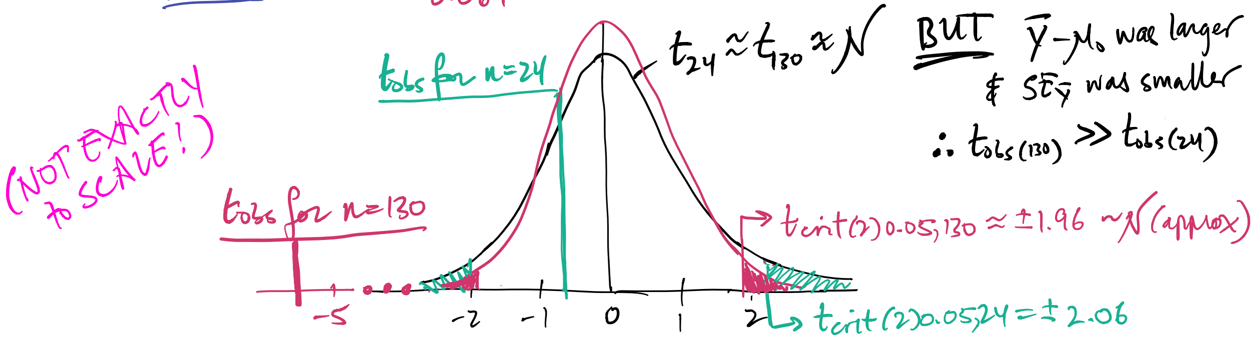
{ computes d.f. from the sample data
{ can also directly access p-value & 95% CI { my_ttest$p.value
{ my_ttest$conf.int

$n = 25$: $t_{obs} = \frac{\bar{Y} - \mu_0}{SE_{\bar{Y}}} = \frac{98.52 - \mu_0}{0.136} = -0.56$

$$\text{p-value} = \underbrace{Pr[t < -0.56]}_{\text{left tail}} + \underbrace{Pr[t > 0.56]}_{\text{right tail}} = 2 * Pr[t > 0.56] = 0.58$$

DO NOT REJECT H_0

$n = 130$: $t_{obs} = \frac{98.25 - \mu_0}{0.064} = -5.44 \Rightarrow \text{p-value} = 1.6 \cdot 10^{-5}$ REJECT H_0



BUT $\bar{Y} - \mu_0$ was larger & $SE_{\bar{Y}}$ was smaller

$\therefore t_{obs(130)} \gg t_{obs(24)}$