

ch 10: Normal Distribution (review)

10/15/2020

- Standard normal: $\mu=0, s.d.=1 \Rightarrow z = \frac{y-\mu}{\sigma}$ $\left\{ \begin{array}{l} z\text{-distribution} \\ z\text{-score} \end{array} \right.$

- Probabilities for normal distribution

$$\Pr[X_1 \leq X \leq X_2] = \Pr[X \leq X_2] - \Pr[X \leq X_1]$$

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

Population: σ^2 $\mu=177.6, sd=9.7\text{cm}$
 φ $\mu=163.2, sd=10.1\text{cm}$

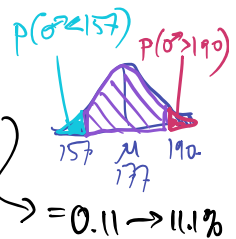
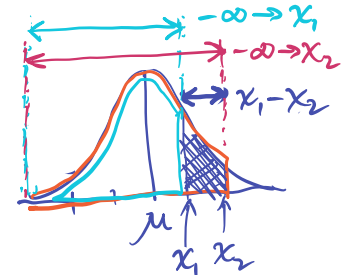
Q: what is Prob. that σ^2 or φ will NOT qualify?

$$\Pr[\sigma^2 < 157.5] \cup \Pr[\sigma^2 > 190.5]$$

(dqqpr) σ^2 : $\text{pnorm}(q_1, \mu, sd) + \text{pnorm}(q_2, \mu, sd, \text{lower.tail}=F)$
 $\Pr[X \leq q_1]$ $\Pr[X > q_2]$

φ : $0.289 \rightarrow 29\%$

CORRECTED \Rightarrow Equivalent: $1 - [\text{pnorm}(q_2, \mu, sd) - \text{pnorm}(q_1, \mu, sd)]$ DEFAULT: lower.tail=F
 $\Pr[X \leq q_2]$ $\Pr[X \leq q_1]$



- 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:
 $n=80, \mu=3339, \sigma=573\text{g}.$

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

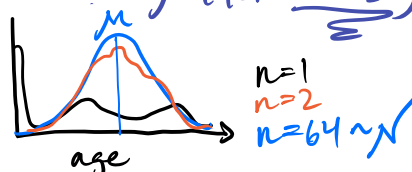
② $z = \frac{q - \mu}{\text{SEM}} = \frac{3370 - 3339}{\text{SEM}} \Rightarrow \text{pnorm}(z, 0, 1, \text{lower.tail}=F)$

③ Or equivalently $\Rightarrow \text{pnorm}(q - \mu, 0, \text{SEM}, \text{lower.tail}=F)$

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

⊗ Weak law of large numbers.

Ex 10.6 "Spanish" flu of 1918:



Ch 11: Inference for a Normal Population

• Student's t-distribution

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

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

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

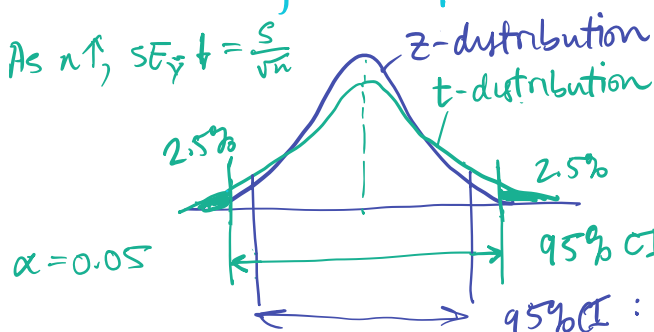
$$\otimes t = \frac{\bar{Y} - \mu}{SE_{\bar{Y}}} \quad \text{"Student's t" w/ } n-1 \text{ degrees of freedom}$$

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

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



e.g. If $df=5$, $t_{crit} \approx \pm 2.78$
vs. $z = 1.96$.

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

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

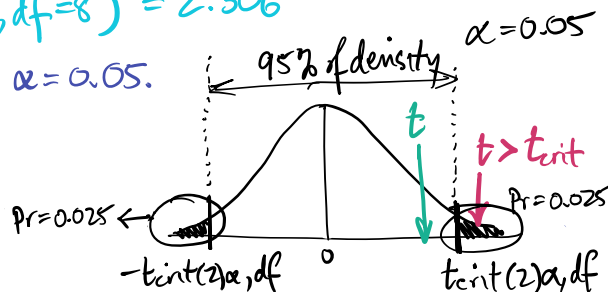
$$-t_{crit(2)\alpha, df} < \frac{\bar{Y} - \mu}{SE_{\bar{Y}}} < t_{crit(2)\alpha, df} \Rightarrow \bar{Y} = \pm t_{crit} \cdot SE_{\bar{Y}} + \mu$$

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

e.g. $t_{crit} = qt(0.025, df=8, \text{lower.tail}=F)$
 $= qt(0.975, df=8) = 2.306$

The significance cutoff is $\alpha = 0.05$.

\otimes 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 = μ_0

H_A : " " $\neq \mu_0$

Assumptions:

- Random sample
- " variable $n \sim N$

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

⊗ sampling distribution of t gives $Pr[\text{obs. data}]$ under H_0 .

Ex 11.3 Human body temp. $37^\circ\text{C} \approx 98.6^\circ\text{F}$
(36.8)

$n=25$ vs $n=130$
 $p > 0.05$ $p < 0.05$

REJECT H_0

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

ONE-SAMPLE ttest

- compare to Exp. value of Pop. param

⊗ t.test (body temp, $\mu = 98.6$)

\hookrightarrow data for your sample = VECTOR.

H_0 : Mean body temp = 98.6

H_A : " " " $\neq 98.6$

} TWO-SIDED TEST

\rightarrow { computes d.f. from the sample data
can also directly access p-value & 95% CI } { my.ttest & p-value
my.ttest & conf.int

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

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

DO NOT REJECT H_0

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

(NOT EXACTLY
TO SCALE!)

