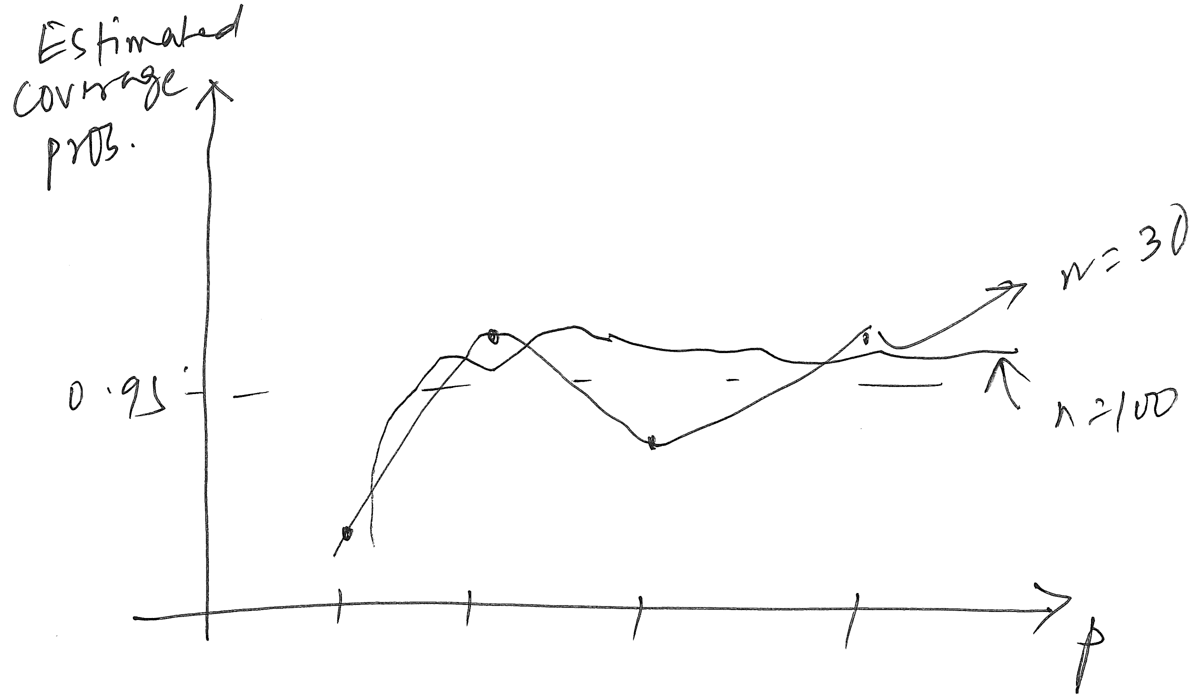
Project #2

#2]



In R:

```
plot(            , xlim = c(-,-),
                 , ylim = c(-,-), type = "l")

lines(           , lty = 2)

lines(           , lty = 3)

legend(locator(1), legend = c("n=30", - -),
                   lty = 1:3)
```

Summary of steps in a hypothesis test:

- Formulate H_0 and H_1
- Find a test statistic T and get its null distribution
- Compute T_{obs}
- Use the null distribution to compute either the critical point or the p-value for the test.
- State your conclusion. (in lay man terms also).

don't just say H_0 is accepted or rejected.

Recall: · p-value = $P[T$ is as extreme or more extreme than $T_{obs} \mid H_0$ is true$]$

· Level-α test:

If p-value $\le \alpha$, reject H_0, but accept it

if p-value $> \alpha$.

Some specific tests

One-sample tests for μ where $X \sim N(\mu, \sigma^2)$

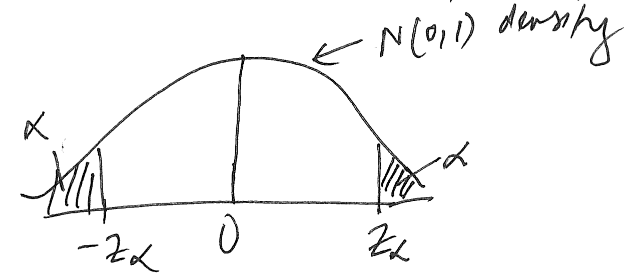
NOTE: when H_0 is true: $\bar{X} \sim N\left[\mu = \mu_0, \frac{\sigma^2}{n}\right]$.

Case 1: z**-test (known σ^2):** $H_0 : \mu = \mu_0$

Test statistic: General: $T = \frac{\hat{\theta} - \theta_0}{se(\hat{\theta}_0)}$ Here: $z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \sim N(0,1)$ when H_0 is true.

Critical point for the level α test:

One-sided alternative: z_α (right-tailed) or $-z_\alpha$ (left-tailed)

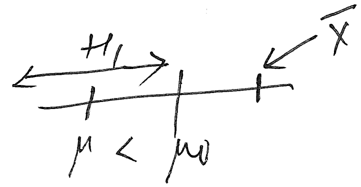


Two-sided alternative: $z_{\alpha/2}$

NOTE: These critical points guarantee that

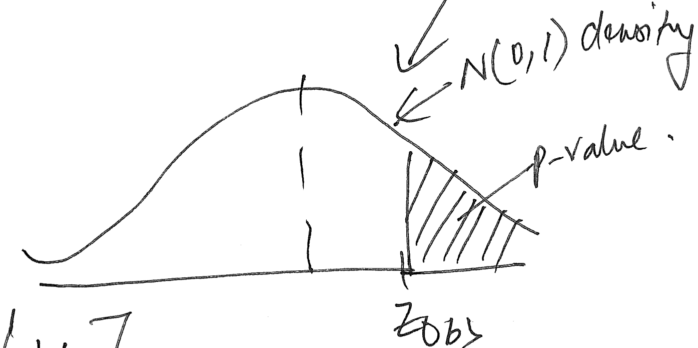
$$\underbrace{P[\text{Reject } H_0 \mid H_0]}_{P[\text{Type I error}]} = \alpha.$$

p-value:



$F(x) = \text{CDF of } N(0,1)$.

H_1	reject when	p-value	computing p-value
$\mu \neq \mu_0$	$\lvert z_{obs} \rvert > z_{\alpha/2}$	$P[\lvert Z \rvert \geq \lvert z_{obs} \rvert \mid H_0]$	$2(1 - F(\lvert z_{obs} \rvert))$.
$\mu > \mu_0$	$z_{obs} > z_\alpha$	$P[Z \geq z_{obs} \mid H_0]$	$1 - F(z_{obs})$, [1-pnorm(z_{obs})]
$\mu < \mu_0$	$z_{obs} < -z_\alpha$	$P[Z \leq z_{obs} \mid H_0]$	$F(z_{obs})$

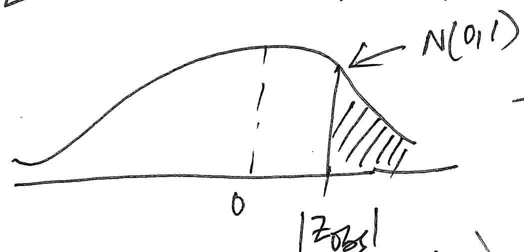


Computing: $P[\lvert Z \rvert \geq \lvert z_{obs} \rvert \mid H_0]$

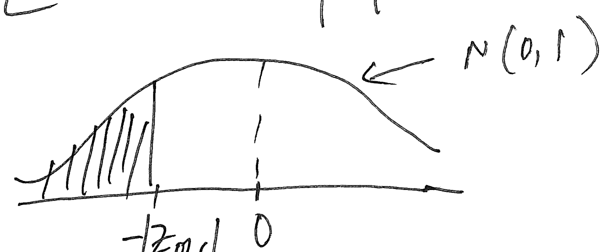
$= P[Z \geq \lvert z_{obs} \rvert \vee Z \leq -\lvert z_{obs} \rvert \mid H_0]$

$= P[Z \geq \lvert z_{obs} \rvert \mid H_0] + P[Z \leq -\lvert z_{obs} \rvert \mid H_0]$.

$=$



$+$



$= 2(1 - F(\lvert z_{obs} \rvert))$.

Case 2: t-**test (unknown σ^2):** $H_0 : \mu = \mu_0$

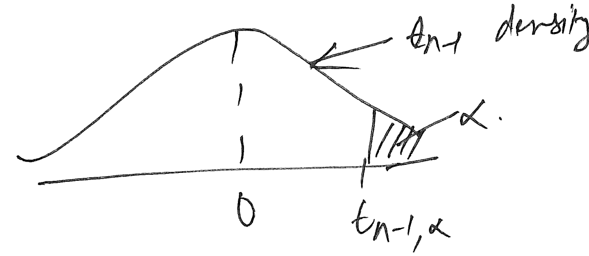
$\hat{s}^2 =$ sample variance

Test statistic: $t = \dfrac{\bar{X} - \mu_0}{\widehat{se}(\bar{X})} = \dfrac{\bar{X} - \mu_0}{s/\sqrt{n}}$, $\sim t_{n-1}$ when H_0 is true.

Critical point for the level α test:

One-sided alternative: $t_{n-1,\alpha}$, or $-t_{n-1,\alpha}$

Two-sided alternative: $t_{n-1,\alpha/2}$

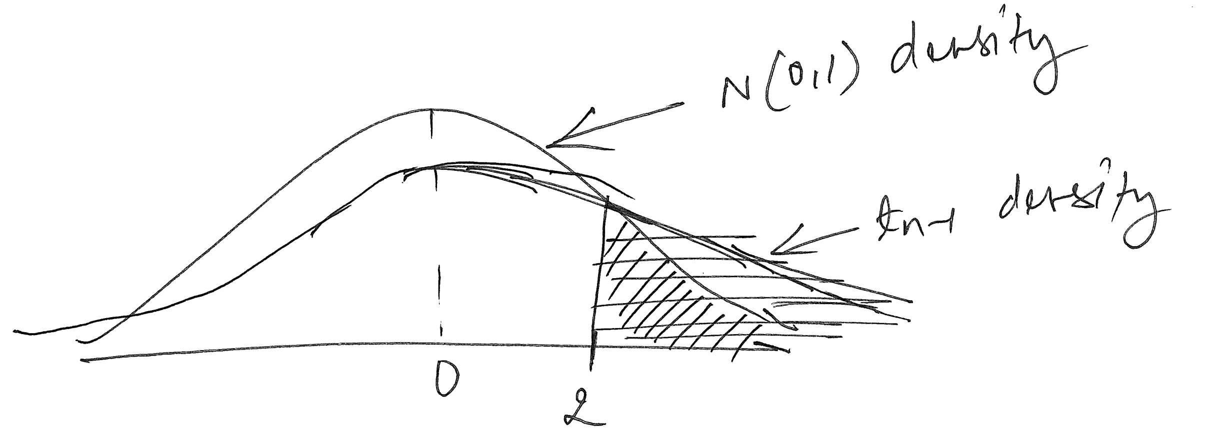


p-**value:**

Here: $F(x) =$ CDF of t_{n-1} dist.

H_1	reject when	p-value	computing p-value
$\mu \neq \mu_0$	$\lvert t_{obs}\rvert \geq t_{n-1,\alpha/2}$	$P(\lvert t\rvert \geq \lvert t_{obs}\rvert \,\vert H_0)$	$2(1 - F(\lvert t_{obs}\rvert))$
$\mu > \mu_0$	$t_{obs} \geq t_{n-1,\alpha}$	$P(t \geq t_{obs} \,\vert H_0)$	$1 - F(t_{obs})$
$\mu < \mu_0$	$t_{obs} \leq -t_{n-1,\alpha}$	$P(t \leq t_{obs} \,\vert H_0)$	$F(t_{obs})$

Recall:



~~p-value assuming t_{n-1} > p-value assuming~~

$1-F(2)$ assuming $N(0,1)$ $\leq$ $1-F(2)$ assuming t_{n-1}.

One-sample test for μ when X is nonnormal

Large-sample z-test: $H_0 : \mu = \mu_0$

- Need large n but works for mean of any (non-normal) population
- Use the z-test with test statistic

$$z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \quad \sim \quad t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \stackrel{\cdot}{\sim} N(0,1) \text{ when } n \text{ is large.}$$

$\stackrel{\cdot}{\sim} N(0,1)$ when n is large

- When n is large, the null distribution is approximately $N(0, 1)$ due to central limit theorem.
- This test has approximate level α.

One-sample test for population prop p

- The large-sample z-test works because in this case $X \sim$ Bernoulli(p) and $E(X) = p$. $H_0: p = p_0$.
- Use the z-test with test statistic

$$z = \left\{ \underbrace{\frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}}_{se(\hat{p}) \text{ when } H_0 \text{ is true.}} \right\} \overset{\cdot}{\sim} N(0,1) \text{ when } n \text{ is large.} \longrightarrow \text{book uses.}$$

$$\text{or } \frac{\hat{p} - p_0}{\sqrt{\hat{p}(1-\hat{p})/n}} \longrightarrow \text{ok to use.}$$

- This test has approximate level α.

Ex 1: A long-time authorized user of a computer account takes 0.2 seconds on average between keystrokes. One day, when a user typed in the username and password, 15 times between keystrokes were recorded. These data had mean of 0.3 $\bar{x}$ seconds and standard deviation of 0.12 s seconds. Do these data given evidence of an unauthorized login attempt? Assume normality for the time b/w keystrokes, and 5% level of significance.

Recall: X = time b/w key strokes for the person who is trying to log in

$\theta = E[X]$

$H_0: \underbrace{\theta = 0.2\ s}_{\text{authorized attempt}}$ vs $H_1: \underbrace{\theta \neq 0.2\ s}_{\text{unauthorized attempt}}$

t-test because population is normal and SD of pop. is unknown.

$t_{obs} = \dfrac{0.3 - 0.2}{0.12/\sqrt{15}} = \underline{3.227}$

p-value: $2\left[1 - pt(|3.227|, df = 14)\right] = 0.006.$

compute: $t_{n-1,\alpha/2} = qt(0.975, df = 14) = 2.145$

Level: 5%. test: since p-value $< 0.05 \Rightarrow$ reject H_0.

$\Rightarrow$ attempt is ~~unathor~~ unauthorized