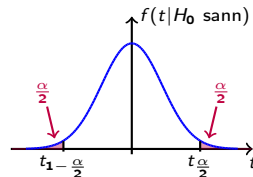
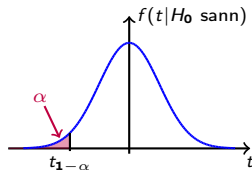
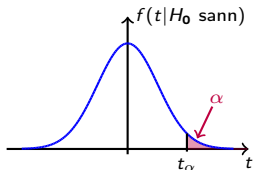


Plan for dagens undervisning

- ★ Regne på eksempler/situasjoner som inkluderer
 - hypotesetesting i binomisk fordeling: forkastningsregel og p-verdi, liten n og stor n
 - hypotesetesting når fordelingen til testobservator ikke er kjent for H_0 sann

Utleidning av hypotesetest

- ★ Formulere hypotesene H_0 og H_1
- ★ Velg/konstruer testobservator: T (kjent fordeling når H_0 er sann)
- ★ Bestem beslutningsregel:
 - forkast H_0 hvis $T \geq k$
 - forkast H_0 hvis $T \leq k$
 - forkast H_0 hvis $T \leq k_\ell$ eller $T \geq k_u$
- ★ Velge signifikansnivå α
- ★ Bestem kritisk verdi k (eventuelt k_ℓ og k_u) fra kravet
$$P(\text{Forkast } H_0 | H_0 \text{ er sann}) = \alpha$$
- ★ Skriv opp beslutningsregel: Forkast H_0 hvis ...
- ★ Regn ut observert verdi: t
- ★ Sammenlign observert verdi t med kritisk verdi og konkluder



p-verdi

Definisjon (p-verdi)

P-verdien er sannsynligheten, når H_0 er sann, for å observere en verdi for testobservatoren som er lik den observerte verdien eller en verdi som er mer ekstrem i retning av den alternative hypotesen.

Sentralgrenseteoremet og binomisk fordeling

Teorem (Sentralgrenseteoremet)

La $X_1, X_2, \dots, X_n$ være uavhengige og identisk fordelte variabler med forventningsverdi $E[X_i] = \mu$ og varians $\text{Var}[X_i] = \sigma^2$. La

$$Z = \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}},$$

der $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ er gjennomsnittet av X_i -ene. Da vil sannsynlighetsfordelingen til Z konvergere mot en standard normalfordeling når $n \rightarrow \infty$.

★ Praktisk konsekvens av teoremet: Når n er stor (nok) har vi

$$Z = \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \approx N(0, 1) \Leftrightarrow \sum_{i=1}^n X_i \approx N(n\mu, n\sigma^2)$$

★ La $Y \sim \text{binomisk}(n, p)$

- skriver $Y = \sum_{i=1}^n X_i$, der $X_i = \begin{cases} 1 & \text{hvis suksess i } i\text{-te forsøk,} \\ 0 & \text{ellers.} \end{cases}$
- $\mu = E[X_i] = 0 \cdot (1 - p) + 1 \cdot p = p$
- $\sigma^2 = \text{Var}[X_i] = E[X_i^2] - E[X_i]^2 = \dots = p(1 - p)$
- hvis n er stor (nok) gir sentralgrenseteoremet at

$$Y = \sum_{i=1}^n X_i \approx N(np, np(1 - p))$$

- tommelfingerregel: god approksimasjon hvis $np \geq 5$ og $n(1 - p) \geq 5$

Eksempel

- ★ Vi vet at systolisk blodtrykk i en populasjon av kvinner i alderen 20-29 år er normalfordelt med forventningsverdi $\mu_0 = 120$ mmHg og standardavvik $\sigma = 10$ mmHg.
- ★ Blant kvinner i denne alderen med en bestemt sykdom antar vi også at systolisk blodtrykk er normalfordelt med standardavvik $\sigma = 10$ mmHg, men forventningsverdien er kanskje annerledes enn blant kvinner uten denne sykdommen.
- ★ Ønsker nå å undersøke om forventet blodtrykk blant kvinner med denne sykdommen er høyere enn μ_0 .
- ★ Trekker ut tilfeldig $n = 100$ kvinner med sykdommen og måler systolisk blodtrykk

$$X_1, X_2, \dots, X_n$$