w05 sem II

$\boxed{\text{One simple problem} \quad y_i \in \{0,1\}}$

out : 20 years. 2 bad years

exp: +80 years : a) point prediction (easy)
b) interval prediction (hard !!)
95%

conserv & prediction

a) Assume $y_1, \dots y_{20}, y_{21}, \dots \dots y_{100}$

$y_i \sim \text{Bernoulli}(p)$ indep

$$\hat{p} = \frac{2}{20}$$

$$\mu_F = E(y_{21} + y_{22} + \dots + y_{100}) = 80(p)$$

$$\hat{\mu}_F = 80 \cdot \hat{p} = 80 \cdot \frac{2}{20} = 8$$

(point est)

Recap. interval estimation ?

desired

α = prob-ty of error = 0.05

$1-\alpha$ = prob-ty of coverage = 0.95

* exact interval $P(\text{quantity} \in \text{Interval}) = 1-\alpha$

dream lucky !!

unicorn

* asymptotic interval $\lim_{n\to\infty} P(\text{quant} \in \text{Int}) = 1-\alpha$

* conservative interval $P(\text{quant} \in \text{Int}) \geq 1-\alpha$

exact.

classical inference (not Bayesian)

$\boxed{\text{confidence}}$ interval = CI

θ - some parameter. (constant)

$$P(\theta \in CI) = 1-\alpha.$$

unknown constant quantity

random interval

$CI(y_1, y_2 \dots y_n)$

!! $\boxed{\text{prediction}}$ interval = PI

n - sample size

$$P(Y_{n+1} \in PI) = 1-\alpha.$$

$$\boxed{P(Y_{n+1}, Y_{n+2} \dots Y_{n+10} \in PI) = 1-\alpha}$$

$$P(\text{at least 3 future obser-ns out of 10 } : (Y_{n+1}, Y_{n+2} \dots Y_{n+10}) \in PI) = 1-\alpha$$

$$P\left(\bar{Y}_F = \frac{Y_{n+1} + \dots + Y_{n+10}}{10} \in PI\right) = 1-\alpha$$

$\boxed{\text{Tolerance}}$ interval

(limiting case of prediction interval)

$$P(80\% \text{ of all future obs-ns } Y_{n+1}, Y_{n+2}, Y_{n+3} \dots \in TI) = 1-\alpha$$

---

goal $\quad P(Y_{21} + Y_{22} + \dots Y_{100} \in PI) \geq 1-\alpha$

asymptotic PI (not very appropriate because $n = 20$)

$$Y_{21} + Y_{22} + \dots + Y_{100} \sim Bin(p ; 80)$$

CLT: $$Y_{21} + \dots + Y_{100} \overset{\text{approx}}{\sim} N(80p ; 80 \cdot p(1-p))$$

$$E(Y_{21} + \dots + Y_{100}) = 80p$$

$$Var(Y_{21} + \dots + Y_{100}) = 80p(1-p)$$

number of future bad years.

$$Y = Y_{21} + \dots + Y_{100}$$

$$P\left(-1.96 \le \frac{Y - 80p}{\sqrt{80p(1-p)}} \le 1.96\right) \simeq 0.95$$

1 ← Approx 1 due to CLT
80 obs.

$$P\left(80p - 1.96\sqrt{80p(1-p)} \le Y \le 80p + 1.96\sqrt{80p(1-p)}\right) \simeq 0.95$$

LLN: $$\operatorname*{plim}_{n \to \infty} \hat{p} = p \qquad \hat{p} \simeq p$$

2 ← Approx 2 due to LLN.
20 obs.

asy PI $$\left[80 \cdot 0.1 - 1.96\sqrt{80 \cdot 0.1 \cdot 0.9} ; 80 \cdot 0.1 + \underbrace{1.96\sqrt{80 \cdot 0.1 \cdot 0.9}}_{5.3}\right]$$

$$\hat{p} = \frac{2}{20}$$

$$[2.3 ; 13.3]$$

conservative PI

$$P(Y \in PI) \geq \underline{0.95}$$

trivial one !! $\quad P(Y \in [0;80]) = 1$

Hypergeometric distribution.

$n = 20$ | $m = 80$

bad years : X | Y

(def. details)

$N = n + m = 100$

$D = X + Y$

$\boxed{N, n, m, D \text{ - constants (known).}}$

a) $P(X = x \mid n, D, N) \overset{?}{=}$

sample size — total number of Def. — total number of obs.-s

number of bad years in a sample

number of good years in observed sample

total number of good years in life

$$= \frac{C_D^x \cdot C_{N-D}^{n-x}}{C_N^n}$$

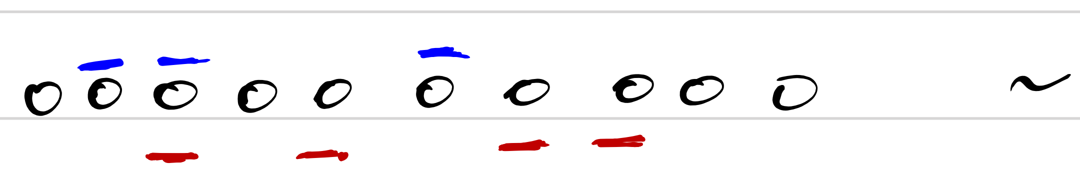
I am choosing n objects out of N.

total number of bad years in life

b) $P(X = x \mid D, n, N) \overset{!!}{=} P(X = x \mid n, D, N)$

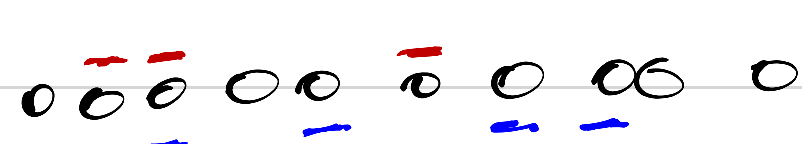
$n = 3$

$D = 4$



$n = 4$

$D = 3$



one - tail

$\underbrace{n}_{X} + \underbrace{m}_{Y} = k$

$X + Y = D$

$$\boxed{P(Y \in [0; Y_R] \mid X = x) \geq 0.95}$$

cdf cumul. distr function for x

cdf $\boxed{P(X \leq x \mid n, D, N)}$

$x = 2$, $n = 20$, $N = 100$

$D = x + Y$, $x = 2$

Y	cdf	
1		
2		
3	0.946	
4	0.907	
5		
10	0.55	
15		
20	0.12	
24	0.055	
25	0.044	

$P(Y \in [0 : 25]) \geq 0.95$

$P(X \leq 2 \mid n, D, N) =$
$= P(X=0 \mid n, D, N) +$
$+ P(X=1 \mid n, D, N) +$
$+ P(X=2 \mid n, D, N)$

my	Soft "like"
n	$\to k = 20$
D	$\to m = 2 + (1 \ldots)$
N	$m + n = 100$
x	$= x$

Meeker, Hahn „Statistical intervals"

study design.