Elementaire Statistiek: Werkcollege 1

Axel De Leeuw

February 2024

1 Opgave

De opgave is bijgevoegd in het ZIP-bestand.

2 Oefening 1

Gegeven zijn het gemiddelde en de standaardafwijking per dag, maar we willen een kans berekenen van het aantal geboortes per jaar. We moeten dus met een 365 keer zo groot gemiddelde en grote variantie werken:

$\mu = 365 * 228 = 83220$
$\sigma = \sqrt{365} * 20 \approx 382$

Belangrijk: Het is de variantie σ^2 die vermenigvuldigd wordt met 365, daarom dat bij σ de wortel van 365 genomen moet worden.

De tabel geeft Z-scores weer voor een standaard normale verdeling met $\mu = 0$ en $\sigma = 1$, de nodige Z-waarde kan als volgt berekend worden:

$$Z = \frac{X - \mu}{\sigma} = \frac{82000 - 83220}{382} \approx -3,19$$

Nu moet dus de p-waarde voor $X > -3,19$ gezocht worden. Omdat de tabel enkel waarden bevat voor positieve waarden van X, en ook uitgaat van $X \leq B$, is nog een eigenschap nodig om de p-waarde voor $X > -3,19$ te kunnen bepalen.

$P(X \leq B) = P(X > -B)$

Als we de p-waarde voor 3,19 nemen verkrijgen we de correcte p-waarde:

$P(X > -3,19) = P(X \leq 3,19) = 0.999 \approx 1$

De kans op meer dan 82000 baby's is dus zo goed als gegarandeerd.

3 Oefening 2

Deze vraag gebruikt de Centrale Limietstelling die stelt dat het steekproefgemiddelde normaal verdeeld is, als er genoeg grote steekproeven genomen worden.

De normaalverdeling dat deze zal volgen wordt gegeven door:

$$\bar{X} = N(\mu, \frac{\sigma^2}{n})$$

De standaardafwijking van deze verdeling is:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{0.3}{\sqrt{125}} \approx 0.03$$

De tabel van de standaarnormaalverdeling geeft aan dat voor $X \leq 2.31$ de kans al 0.99 is, wat overeenkomt met een kans van 0.01 voor $X > 2.31$.

Het is nu dus een kwestie van uitzoeken voor welke waarde L de volgende vergelijking resulteert in 2.31:

$$Z = \frac{L - 1.4}{\frac{0.3}{\sqrt{125}}} = 2.31$$

Verplaatsen van termen leidt tot:

$$L = \frac{2.31 * 0.3}{\sqrt{125}} + 1.4 \approx 1.46$$

4 Oefening 3(a)

Volgens de definitie van schatters is een schatter onvertekend als haar bias gelijk is aan 0. De bias wordt gegeven via de volgende formule:

$b(W) = E[W] - \mu$

Invullen van W geeft:

$b(W) = E[a * X + b * Y] - \mu$

Voor stochastische variabelen geldt:

$E[X + Y] = E[X] + E[Y]$

En dus:

$b(w) = E[a * X] + E[b * y] - \mu$

Ook geldt er dat constanten uit de verwachtingswaarde mogen:

$E[cX] = c * E[X]$

Wat leidt tot:

$b(W) = a * E[X] + b * E[Y] - \mu$

Het gegeven geeft dat $E[X] = E[Y] = \mu$, de bias is dus ten slotte:

$b(W) = a * \mu + b * \mu - \mu = (a + b - 1) * \mu$

Het is makkelijk in te zien dat de bias 0 wordt als $a + b = 1$.

5 Oefening 3(b)

Naast dat we willen dat de schatter onvertekend is, hebben we ook liefst dat deze zo nauwkeurig mogelijk is. Dit betekent dat we een minimale MSE verkiezen. De variantie wordt gegeven door volgende formule:

$Var(w) = Var(aX + bY)$

Aangezien X en Y onafhankelijk:

$Var(w) = Var(aX) + Var(bY)$

Constanten mogen buitengezet worden maar moeten gekwadrateerd worden:

$Var(w) = a^2Var(X) + b^2Var(Y)$

De variantie van X en Y zijn gekend:

$Var(w) = a^2\sigma^2 + 4b^2\sigma^2$

Eerder werd gesteld dat $a + b = 1$, we hebben dus de keuze om a in te vullen als $1 - b$, of om b in te vullen als $1 - a$:

$Var(w) = (1 - b)^2\sigma^2 + 4b^2\sigma^2$
$Var(w) = \sigma^2 - 2b\sigma^2 + b^2\sigma^2 + 4b^2\sigma^2$
$Var(w) = \sigma^2 - 2b\sigma^2 + 5b^2\sigma^2$
$Var(w) = (1 - 2b + 5b^2)\sigma^2$

Het minimum kan bepaald worden via de afgeleide in b voor $f(b) = 1-2b+5b^2$:

$f'(b) = 10b - 2 = 0$

Duidelijk is $b = \dfrac{1}{5}$ en dus $a = \dfrac{5-1}{5} = \dfrac{4}{5}$.

6 Oefening 4(a)

Deze vraag is zeer gelijkaardig aan 3(a), waarbij werd gezien dat de schatter zuiver was indien $a + b = 1$, of nog: $b = 1 - a$, maar nu is nog een extra eigenschap uit de theorie nodig: $E\left[\overline{X}_m\right] = E\left[X\right]$.

$E\left[\hat{\mu}\right] = E\left[a\overline{X}_m + (1-a)\overline{Y}_n\right] = aE\left[\overline{X}_m\right] + (1-a)E\left[\overline{Y}_n\right] =$
$aE\left[X\right] + (1-a)E\left[Y\right] = a\mu + \mu - a\mu = \mu$

7 Oefening 4(b)

Opnieuw is de minimale MSE gezocht, we weten al dat de bias gelijk is aan 0:

$$Var(\hat{\mu}) = a^2 Var(\overline{X}_m) + (1-a)^2 Var(\overline{Y}_n)$$

De varianties zijn nu echter niet simpelweg σ^2 en $4\sigma^2$ omdat we te maken hebben met steekproefgemiddelden, er geldt echter wel:

$$Var(\overline{X}_m) = \frac{1}{m} Var(X)$$

En dus:

$$Var(\hat{\mu}) = \frac{a^2\sigma^2}{m} + \frac{4(1-a)^2\sigma^2}{n} = \left(\frac{a^2}{m} + \frac{4 - 8a + 4a^2}{n}\right)\sigma^2$$

We leiden af in a:

$$\left(\frac{2a}{m} + \frac{8a - 8}{n}\right)\sigma^2 = 0$$

$$\left(\frac{2an + 8am - 8m}{mn}\right)\sigma^2 = 0$$

Het is duidelijk dat de teller 0 moet worden (de noemer zal ook nooit 0 zijn):

$$2an + 8am - 8m = 0$$

De ideale waarde voor a kan bepaald worden door verplaatsing van termen:

$$a = \frac{8m}{2n + 8m} = \frac{4m}{n + 4m}$$

8 Oefening 5(a)

Voor de schatter $X - \dfrac{X^2}{n}$ verwachten we dat:

$$E\left[X - \frac{X^2}{n}\right] = Var(X)$$

Dankzij rekenregels voor de verwachtingswaarde weten we:

$$E\left[X - \frac{X^2}{n}\right] = E\left[X\right] - E\left[\frac{X^2}{n}\right] = E\left[X\right] - \frac{E\left[X^2\right]}{n}$$

Er geldt bovendien ook:

$Var(X) = E\left[X^2\right] - E\left[X\right]^2$ en dus $E\left[X^2\right] = Var(X) + E\left[X\right]^2$

Waardoor volgt:

$$E\left[X - \frac{X^2}{n}\right] = E\left[X\right] - \frac{Var(X) + E\left[X\right]^2}{n}$$

Voor de binomiaalverdeling zijn:

$E\left[X\right] = np$ en $Var(X) = np(1-p)$

En dus:

$$E\left[X - \frac{X^2}{n}\right] = np - \frac{np(1-p) + n^2p^2}{n}$$

$$E\left[X - \frac{X^2}{n}\right] = np - p(1-p) - np^2 = np(1-p) - p(1-p) = (n-1)p(1-p)$$

Het gevonden resultaat lijkt grotendeels overeen te komen met de variantie van X, enkel hebben we $(n-1)$ in plaats van n, en is de schatter dus niet zuiver.

9 Oefening 5(b)

Een idee voor een schatter is:

$$\left(\frac{n}{n-1}\right)\left(X - \frac{X^2}{n}\right)$$

Deze is onvertekend:

$$E\left[\left(\frac{n}{n-1}\right)\left(X - \frac{X^2}{n}\right)\right] = \left(\frac{n}{n-1}\right)E\left[X - \frac{X^2}{n}\right] = \left(\frac{n}{n-1}\right)(n-1)p(1-p)$$

$$= np(1-p) = Var(X)$$

10 Oefening 6

Gegeven is $T = c\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2$, voor de verwachtingswaarde geldt:

$$E\left[T\right] = E\left[c\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2\right] = cE\left[\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2\right]$$

De schatter S^2 voor de variantie is gedefinieerd als:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2 \text{ en } E\left[S^2\right] = \sigma^2$$

Deze schatter wordt deels teruggevonden in de verwachtingswaarde van T:

$$E\left[T\right] = cE\left[(n-1)\,S^2\right] = c(n-1)E\left[S^2\right] = c(n-1)\sigma^2$$

Voor de variantie vinden we:

$$Var(T) = c^2 Var\left(\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2\right) = c^2(n-1)^2 Var\left(S^2\right)$$

De stelling van Helmert vertelt ons dat:

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

Bovendien weten we ook dat:

$$Var\left(\frac{(n-1)S^2}{\sigma^2}\right) = Var(\chi^2_{n-1}) = 2(n-1)$$

Uiteindelijk vinden we voor $Var(S^2)$:

$$Var(S^2) = \frac{2\sigma^4}{n-1}$$

$Var(T)$ is dus uiteindelijk:

$Var(T) = 2c^2(n-1)\sigma^4$

Indien T als schatter gebruikt wordt voor de variantie σ^2, geldt voor de bias:

$b(T) = E[T] - \sigma^2 = c(n-1)\sigma^2 - \sigma^2 = (c(n-1)-1)\sigma^2$

De MSE wordt gegeven door:

$MSE(T) = (b(T))^2 + Var(T) = (c(n-1)-1)^2\sigma^4 + 2c^2(n-1)\sigma^4$

Na vereenvoudiging:

$MSE(T) = \left[c^2(n-1)^2 - 2c(n-1) + 1 + 2c^2(n-1)\right]\sigma^4$

Afleiden in c geeft:

$\left[2c(n-1)^2 - 2(n-1) + 4c(n-1)\right]\sigma^4 = 0$

$2c(n-1)^2 - 2(n-1) + 4c(n-1) = 0$

$2c(n-1) - 2 + 4c = 0$

$c(n-1) + 2c - 1 = 0$

$c = \frac{1}{n+1}$