EXERCISE 1

(a) A histogram is a suitable representation for this variable. Boxplot is also good.

Mean and median are similar (7.34 and 7.42 respectively).

The dispersion can be described by the standard deviation of 1.498.

The distribution is pretty symmetric, as $Q_1 = 6.358$ and $Q_3 = 8.118$ have a similar distance from the median, the whiskers in the boxplot have similar length. We only note few lower outliers.

(b) quantile (SleepQuality, 0.95) = 9.668

quantile (SleepQuality, 0.90) = 9.26 (B)

(c) $Q_1 - 1.5(Q_3 - Q_1) = 3.267$ is the threshold

There are 3 outliers: 3.198778, 2.390676, 3.200246

(d) We investigate through a scatterplot and calculating the sample correlation $r = 0.3884$

Plausibly there a mild positive linear association

(e) Percentage of subjects who suffer from sleep disorder "insomnia" or "other"

among normal BMI: $(9+7)/(200+9+7) = 7.4\%$

among overweight BMI: $(64+65)/(64+65+19) = 87\%$

Clearly higher among overweight BMI!

EXERCISE 2

a) $W \sim N(250, 100^2)$

pnorm(0, 250, 100) = 0.0062 = p

b) qnorm(0.005, 250, 100) = −7.5829 qnorm(0.001, 250, 100) = −59.023

a loss of 7.5829 a loss of −59.023

c) X := number of investment resulting in a financial loss out of n=100

$X \sim$ Binomial(n, p)

$X \sim$ Poisson(np) approximately

i) 1 − pbinom(3, n, p) = 0.0036

1 − ppois(3, n·p) = 0.0038

both good

B

1 − pbinom(2, n, p) = 0.248

1 − ppois(2, n·p) = 0.252

both good

ii) $\Sigma W_i \sim N(250, 100)$

pnorm(0, 250, 10) ≈ 0

$\Sigma W_i \sim N(250, 100)$

pnorm(0, 250, 10) ≈ 0

B $P(T-20 > 40) \leq \dots \leq \frac{100}{40^2} = 0.0625$

EXERCISE 3

a $P(T>60) = P(T-15 > 45) \leq P(|T-15| > 45) \leq \frac{Var(T)}{45^2} = \frac{100}{45^2} = 4.94\%$

b $P(T \geq c) \leq \dots \leq \frac{Var(T)}{(c-15)^2} = 0.01$

$c = \left(\frac{100}{0.01}\right)^{1/2} + 15 = 115$

B $c = \left(\frac{100}{0.01}\right)^{1/2} + 20 = 120$

c $P(XY=0) = p \longrightarrow$ FALSE

$E[XY] = p \longrightarrow$ TRUE

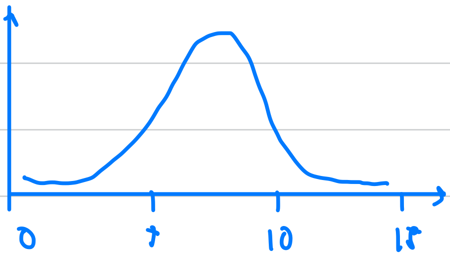
d As $X \sim \text{Bernoulli}(p)$ $Y \sim \text{Exp}(1)$, see notes:

$E[X] = p$ $\quad Var(X) = p(1-p)$ $\quad E[Y] = 1$ $\quad Var(Y) = 1$

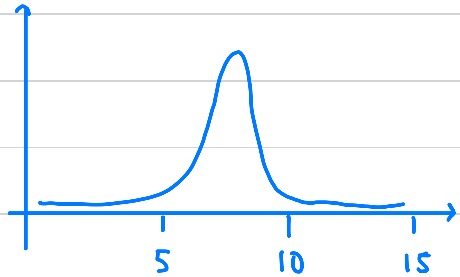
$$Var(XY) = E[X^2Y^2] - E^2[XY]$$
$$= E[X^2]E[Y^2] - E^2[X]E^2[Y]$$
$$= (Var(X) + E^2[X])(Var(Y) + E^2[Y]) - E^2[X]E^2[Y]$$
$$= (p(1-p) + p^2)(2) - p^2 = 2p - p^2$$

EXERCISE 4

ⓐ $K=5$



ⓑ $K=10$



By trying different values of K (or studying briefly the function):
- as K increases the mode slightly increases towards $\lambda = 8$
- as K increases the distribution becomes more concentrated around the mode.

ⓑ $$\ell(K) = \prod_{i=1}^{n} P(x_i; K, \lambda) = \prod_{i=1}^{n} \frac{K}{\lambda}\left(\frac{x_i}{\lambda}\right)^{K-1} e^{-(x/\lambda)^K}$$

$$L(K) = -\log(\ell(K)) = -\sum\left[\log K - \log\lambda + (K-1)\log x_i - (K-1)\log\lambda - \left(\frac{x_i}{\lambda}\right)^K\right]$$

ⓒ $$L'(K) = -\sum\left[\frac{1}{K} + \log x_i - \log\lambda + \left(\frac{x_i}{\lambda}\right)^K \log\left(\frac{x_i}{\lambda}\right)\right]$$

to obtain the MLE $\hat{K}$ the condition is $L'(K) = 0$

ⓓ Using R we evaluate: ⓑ

$L(8) = 521.4$ $\quad$ $L(8.5) = 517.4$

$L(9) = 515.9$ $\quad$ $L(9.5) = 515.6$ ⟶ These are preferable, lowest value of L

$L(10) = 516.6$ $\quad$ $L(10.5) = 518.9$ hence highest likelihood!

(e) Through grid search in R, evaluate $L(k)$ for e.g

$k = 8, 8.01, 8.02, \ldots, 10.99, 11$

to find $\hat{k} = 9.36$

(f) Compute $\bar{x} = \text{mean}(\text{sleep\$sleep duration}) = 7.13$

Again we use grid search in R, for values of k and setting $\lambda = \dfrac{\bar{x}}{\Gamma(1 + \frac{1}{k})}$ every time to satisfy the constraint.

Obtain $\hat{k} = 10.15$ $\hat{\lambda} = 7.49$

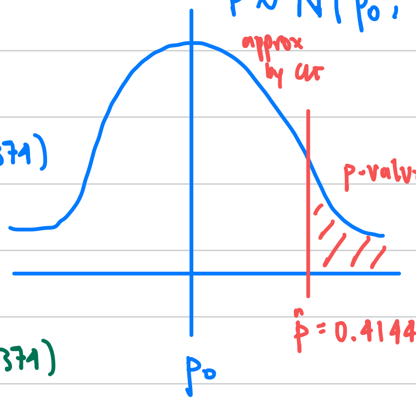
EXERCISE 5

(a) i) $H_0: p \leq 0.35 = p_0$

$H_1: p > 0.35$

ii) $\hat{p} = \frac{77+78}{374} = 0.4144$

$\hat{p} \overset{H_0}{\sim} N\left(p_0, \frac{p_0(1-p_0)}{n}\right)$ approx by CLT



p-value = 1- pnorm($\hat{p}$,0.35, sqrt (0.35·0.65 / 374))

= 0.0045 reject H_0 (iv)

(B) $H_0: p \leq 0.4 = p_0$

$H_1: p > 0.4$

p-value = 1- pnorm($\hat{p}$,0.4, sqrt (0.4·0.6 /374))

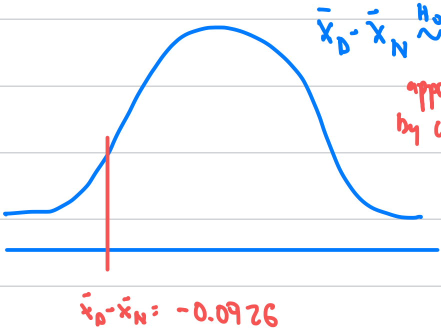
= 0.2843 do not reject H_0 (iv)

(iii) the p-value is the probability to obtain a $\hat{P}$ as extreme or more extreme than $\hat{p} = 0.4144$ if the null hypothesis H_0 is true.

(b) $H_0: \mu_D = \mu_N$

$H_1: \mu_D < \mu_N$

$\bar{x}_D - \bar{x}_N \overset{H_0}{\sim} N\left(0, \frac{\sigma_D^2}{n_D} + \frac{\sigma_N^2}{n_N}\right)$ approx by CLT



$\bar{x}_D = 6.97$ $\quad \bar{x}_N = 7.06$

$s_D^2 = 0.746$ $\quad s_N^2 = 0.98$

$n_D = 71$ $\quad n_N = 73$

p-value = pnorm ($\bar{x}_D - \bar{x}_N$, 0, sqrt ($\frac{s_D^2}{n_D} + \frac{s_N^2}{n_N}$)) = 0.2747

↓

do not reject H_0 !