1. a)

X: Sammenfallende deler hos et barn og tilfeldig mann

X er binomisk fordelt fordi:

- Man kan se på forsøket som en Bernoullis forsøksrekke.
- Vi gjentar sammenligningen n ganger og får to mulige utfall.
- Hvert delforsøk har $p = 0{,}15$ for match

$n = 5$, $p = 0{,}15$

$$P(X=3) = \binom{5}{3} 0{,}15^3 (1-0{,}15)^{5-3} = \underline{\underline{0{,}024}}$$

$$P(X \geq 3) = 1 - P(X \leq 2) = 1 - 0{,}973 = \underline{\underline{0{,}027}}$$

$$P(X=3 \mid X \geq 2) = \frac{P(X=3)}{P(X \geq 2)} = \frac{0{,}024}{1-P(x \leq 1)}$$

$$= \frac{0{,}024}{1-0{,}83521} = \underline{\underline{0{,}148}}$$

b) $P(\text{Type 1-feil}) = P(\text{Forkast } H_0 \mid H_0 \text{ sann})$

$$= P(X=5 \mid p=0{,}15) = (0{,}15)^5 = \underline{\underline{7{,}594 \cdot 10^{-5}}}$$

$P(\text{Type 2-feil}) = P(\text{Behold } H_0 \mid H_1 \text{ sann}) = 0$

Hvis H_1 er sann, så blir $x = 5$ og forkastningsregelen sier at vi forkaster H_0.

2 a)

- Hvert forsøk er uavhengig
- Det er r/m sjanse for merket laks

- Vi kan trekke laks ∞ mange ganger

b) Punktsannsynlighet:

$$f(x;m) = \binom{x-1}{k-1}\left(\frac{r}{m}\right)^k\left(1-\frac{r}{m}\right)^{x-k} = L(m;x)$$

Finner m som maksimerer

$$\ln L(m;x) = \ln\binom{x-1}{k-1} + k\ln\left(\frac{r}{m}\right) + (x-k)\ln\left(1-\frac{r}{m}\right)$$

$$= \ln\binom{x-1}{k-1} + k(\ln r - \ln m) + (x-k)(\ln(m-r) - \ln m)$$

$$l'(m;x) = -\frac{k}{m} + \frac{(x-k)}{m-r} - \frac{(x-k)}{m} = 0$$

$$\frac{k}{m} = (x-k)\left(\frac{r}{m^2-rm}\right)$$

$$m = \frac{(x-k)r}{k} + r = \frac{xr}{k}$$

$$\underline{\underline{\hat{m} = \frac{Xr}{k}}}$$

$$E(\hat{m}) = \frac{r}{k}E(X) = \frac{r}{k}\cdot\frac{k}{p} = \frac{r}{\frac{r}{m}} = m$$

$$Var(\hat{m}) = \frac{r^2}{k^2}\cdot\frac{k\left(1-\frac{r}{m}\right)}{\frac{r^2}{m^2}} = \frac{m^2\left(1-\frac{r}{m}\right)}{k} = \frac{m}{k}(m-r)$$

c) $H_0: m = 50\,000 \qquad H_1: m < 50000$

$$X \sim N\left(\frac{k}{\left(\frac{r}{m}\right)}, k\,\frac{1-\left(\frac{r}{m}\right)}{\frac{r^2}{m^2}}\right)$$

$$X \sim N(1000, 49000)$$

Testobservator:

$$Z = \frac{X-1000}{\sqrt{49000}} \sim N(0,1) \text{ når } H_0 \text{ sann}$$

$$P(Z < z) = 0{,}05 \Rightarrow z_\alpha = -1{,}645$$

Forkast H_0 hvis $Z < -1{,}645$

$$z_{obs} = \frac{728-1000}{\sqrt{49000}} = -1{,}228$$

Vi beholder H_0 fordi $-1{,}228 \not< -1{,}645$

3a) $\hat{r} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

$$E(\hat{r}) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = r$$

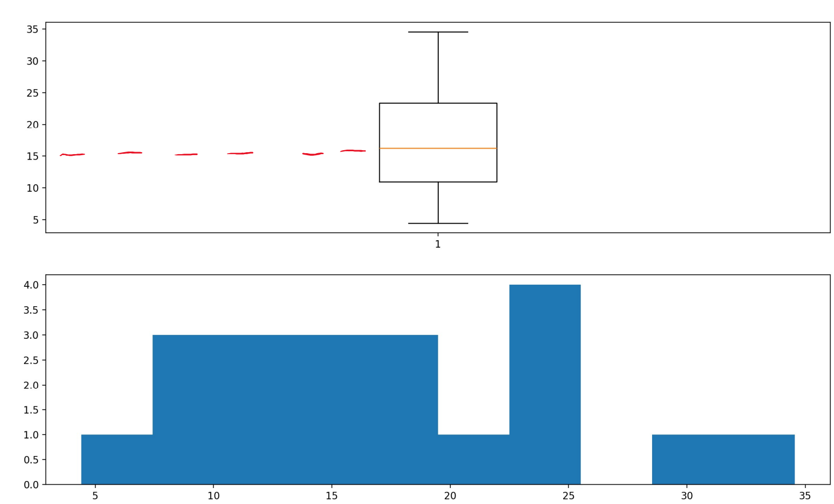
$$Var(\hat{r}) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{r^2}{an}$$

Sentralgrenseteoremet sier at

$$Z = \frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1) \text{ når } n \to \infty$$

$$\frac{\bar{X} - r}{\frac{r}{\sqrt{an}}} = \frac{\hat{r} - r}{r}\sqrt{na} \sim N(0,1) \text{ når } n \to \infty$$

b)




```python
import numpy as np
import matplotlib.pyplot as plt
data = np.array([7.98,
                 10.82,
                 15.88,
                 17.00,
                 24.22,
                 12.20,
                 8.17,
                 16.53,
                 7.46,
                 14.31,
                 34.55,
                 19.46,
                 20.21,
                 13.58,
                 10.98,
                 4.42,
                 24.92,
                 30.29,
                 23.45,
                 23.36])

fig , axs = plt.subplots(2)
axs[0].boxplot(data)
axs[1].hist(data)
plt.show()
```

Vi ser at medianen er over 12,5.

Vi kan påstå at r_0 har økt.

c) $H_0 : r_0 = 12,5 \qquad H_1 : r_0 > 12,5$

`print(data.mean())` $\rightarrow 16,9895$
$\hat{r}$

$$Z = \frac{\hat{r} - r}{r}\sqrt{na} \sim N(0,1)$$

Forkaster hvis

$$\frac{\hat{r} - r}{r}\sqrt{na} > z_{0,10}$$

$$\frac{16,9895 - 12,5}{12,5}\sqrt{20,5} > 1,282$$

$$3,5916 > 1,282$$

Forkast H_0

d) $P\left(\frac{\hat{r} - r_0}{r_0}\sqrt{na} < z_a \,\middle|\, r\right) \leq \beta_0, \quad r = r_0 + \delta$

$\alpha = 0{,}1 \quad \beta = 0{,}2 \quad \delta = 2{,}5 \quad r_0 = 12{,}5$

$$P\left(\hat{r} < \frac{z_\alpha r_0}{\sqrt{na}} + r_0 \,\middle|\, r\right) \leq \beta_0$$

$$P\left(\frac{\hat{r} - r}{r}\sqrt{na} < \frac{\left(\frac{z_\alpha r_0}{\sqrt{na}} + r_0\right) - r}{r}\sqrt{na} \,\middle|\, r\right) \leq \beta_0$$

$$P\left(\frac{\hat{r} - r}{r}\sqrt{na} < \frac{\frac{z_\alpha r_0}{\sqrt{na}} - \delta}{r}\sqrt{na}\right) \leq \beta_0$$

$$\frac{\frac{z_\alpha r_0}{\sqrt{n\alpha}} - \delta}{r} \leq -z_{\beta_0}$$

$$z_\alpha r_0 - \delta\sqrt{na} \leq -r z_{\beta_0}$$

$$z_\alpha r_0 + r z_{\beta_0} \leq \delta\sqrt{na}$$

$$\frac{\left(\frac{z_\alpha r_0 + r z_{\beta_0}}{\delta}\right)^2}{a} \leq n \Rightarrow n \geq 26{,}3$$

$$\Rightarrow \underline{\underline{n = 27}}$$

e)

```
import numpy as np
a = 5
r = 12.5
n = 20
z_0_1 = 1.2816
def typeI_error():
    data = [np.random.gamma(a,r/a) for i in range(n)]
    z = (np.mean(data)-r)/r*np.sqrt(a*n)
    if (z > z_0_1):
        return 1
    else:
        return 0

def simulate(m):
    total_typeI_errors = 0
    for i in range(m):
        total_typeI_errors += typeI_error()

    return total_typeI_errors/m
alpha = simulate(100000)
print(alpha)
# alpha = 0.10301
```

Estimat 0,103

$P(\text{Type } 1\text{-feil})$

$$= P\left(\frac{\hat{r} - r_0}{r_0}\sqrt{na} > z_\alpha \,\middle|\, r_0\right)$$

$$P\left(\hat{r} > \frac{z_\alpha}{\sqrt{na}} r_0 + r_0 \,\middle|\, r_0\right)$$

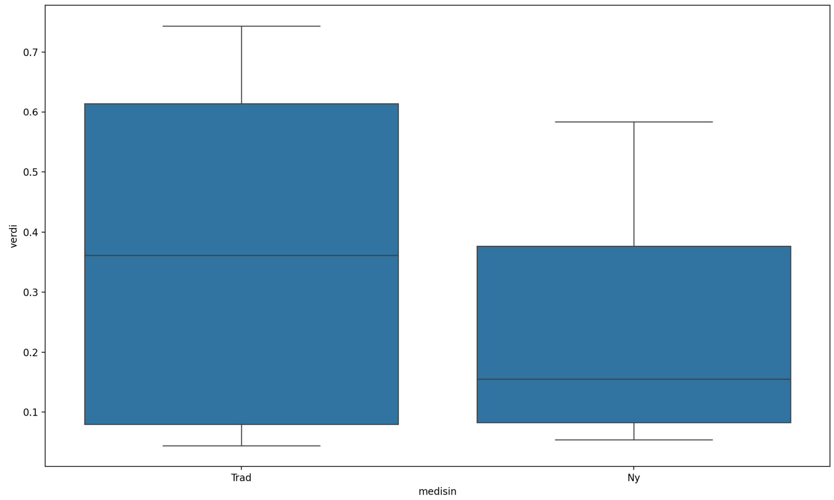
$$\hat{r} \sim Y, \quad Y \sim \text{Gamma}\left(na, \frac{r}{na}\right)$$

$$P(Y > 14{,}1025)$$

$$= 1 - \int_0^{14{,}025} \frac{1}{\left(\frac{r}{na}\right)^{na}\Gamma(na)} t^{na-1} e^{-\frac{tna}{r}}\, dt$$

$$= 1{,}103119$$

4.a)



Virker som at Ny er bedre.
Medianen ligger langt under.
IQR er mindre og forskjellig fra Trad.

H_0: Behandlingen er like effektiv H_1: Ny bedre enn Trad

μ: Gjennomsnittlig blodtrykk

Testobservator:

$$U = \underbrace{\frac{1}{8}\sum_{i=1}^{8} X_i}_{\text{Trad}} - \underbrace{\frac{1}{7}\sum_{i=9}^{15} X_i}_{\text{Ny}}$$

Mer effektiv medisin => U er større

Forkast H_0 hvis U > kritisk verdi

Observert verdi: −0,11723214285714284

Estimert p-verdi: 0,15109

```python
# regner ut observert verdi av testobservatoren:
statisticObserved = testStatistic(x = x,nTrad = 8)
print('Observert verdi: ',statisticObserved)
m = 10000
p_counter = 0
for i in range(m):
    # genererer tilfeldig en permutasjon av (alle) elementene i lista x:
    xPermuted = sample(x,len(x))
    # regner ut simulert verdi av testobservatoren
    statisticSimulated = testStatistic(xPermuted,8)
    # print('Simulert verdi: ',statisticSimulated)
    if statisticSimulated < statisticObserved:
        p_counter += 1
print("estimert p-verdi: ", p_counter/m)
```