Attention !!! Please write or take a screenshot of all answers in the pdf file. You won't be graded if there is no pdf file in the submission.

Turn in Problem 1, 5, 7.

Extra credit (0.5 pionts) 4.

Problem 1: Moontaro prequel

You might wonder how Hamtaro came up with the mean for the growth rate of each coin in the previous homework. He estimated them using MLE!

To simplify the problem, consider a slightly different model for stock pricing. The price at the end of each day is the price of the previous day multiplied by a fixed, but unknown, rate of return, α, with some noise, w. For a two-day period, we can observe the following Markov process: $P(y_2, y_1, y_0|\alpha) = P(y_2|y_1)P(y_1|y_0)P(y_0|\alpha)$ where $y_2 \sim \mathcal{N}(\alpha y_1, \sigma^2), y_1 \sim \mathcal{N}(\alpha y_0, \sigma^2), y_0 \sim \mathcal{N}(0, \lambda)$

Find the MLE of the rate of return, α, given the observed price at the end of each day y_2, y_1, y_0. In other words, compute for the value of α that maximizes $P(y_2, y_1, y_0|\alpha)$.

Problem 1)

$y_0 \sim N(0, \lambda)$

$y_1 = \alpha y_0 + \omega_0 \longrightarrow \omega_0 = y_1 - \alpha y_0$

$y_2 = \alpha y_1 + \omega_1 \longrightarrow \omega_1 = y_2 - \alpha y_1$

$$P(y_i | y_{i-1}; \alpha) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - \alpha y_{i-1})^2}{2\sigma^2}\right)$$

$P(y_0 | \alpha) = N(0, \lambda)$ not depend on α

$P(y_2 y_1 y_0 | \alpha) = P(y_2 | y_1; \alpha) P(y_1 | y_0; \alpha) P(y_0 | \alpha)$

ln

$\ln[P(y_2 y_1 y_0 | \alpha)] = \ln[P(y_2 | y_1; \alpha)] + \ln[P(y_1 | y_0; \alpha)] + \ln[P(y_0 | \alpha)]$

$$\ln[P(y_2 y_1 y_0 | \alpha)] = \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{1}{2\sigma^2}(y_1 - \alpha y_0)^2 + \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{1}{2\sigma^2}(y_1 - \alpha y_0)^2 + \ln[N(0, \lambda)]$$

$$\frac{d}{d\alpha} \ln[P(y_2 y_1 y_0 | \alpha)] = -\cancel{\frac{1}{2\sigma^2}}\left[\cancel{2}(y_1 - \alpha y_0)(\cancel{-}y_0) + \cancel{2}(y_2 - \alpha y_1)(\cancel{-}y_1)\right] = 0$$

$$(y_1 - \alpha y_0) y_0 + (y_2 - \alpha y_1) y_1 = 0$$

$$y_1 y_0 - \alpha y_0^2 + y_2 y_1 - \alpha y_1^2 = 0$$

$$\alpha = \frac{y_2 y_1 + y_1 y_0}{y_0^2 + y_1^2}$$

Problem 2: Hamtaro and his entertainment

From the previous assignment, you might be wondered why Hamtaro is trying so hard is to build a cloud service. This is because he has a lifelong dream of opening his entertainment website, Hamhub, which needs a service from the cloud provider. After successfully building his own cloud service, he creates the website and monitors the number of visitors every day.

Recently, the most famous website in this field of entertainment was blocked by the government last Monday. From this news, he wants to know whether the blockade has a significant effect on the number of Hamhub's visitors?

- Before last Monday, the average number of visitors was $x_0 \sim \mathcal{N}(10000, \sigma^2)$.
- After last Monday, Hamtaro spent ten days collecting the number of users.

```
X1 = array([10190.25479236, 10082.65748517, 10161.37971691,
10042.27783459,
       10129.73858138,  9962.73586162, 10187.78833611,
10013.48007958,
       10372.98760763, 10238.55408072])
```

1. Can Hamtaro conclude that the blockade significantly increases the number of visitors with a significant level of 0.001?
2. If the sample mean and variance are held the same, what is the minimum number of samples Hamtaro need to reject the null hypothesis? For the same observation effect, larger sample size will result in a significant result.

Problem 3: T-Test

Hamtaro performs a t-test for the null hypothesis $H_0 : \mu = 10$ at significance level $\alpha = 0.05$ from a dataset consisting of $n = 16$ elements with sample mean 11 and sample variance 4.

1. Should we reject the null hypothesis in favor of $H_a : \mu \neq 10$
2. What if we test against $H_{a'} : \mu > 10$?

Problem 4: Hamtaro and his entertainment - 2

The story in this problem is a parallel universe of problem 2.

Last Monday, Hamtaro added the new channel to the website, and he wanted to know its effects on the number of visitors. However, the most famous website in this field of entertainment was also blocked by the government on the same day. Since there was no sign of unblocking from the government, Hamtaro could not perform a hypothesis testing on only the factor of adding the new channel. How could Hamtaro know that the changes from adding the new channel are significant?

There are four scenarios in this problem,

1. Before the last Monday, the average number of visitors was $x_0 \sim \mathcal{N}(\mu_0, \sigma^2)$ (no block + no new channel).
2. After the last Monday, the average number of visitors are $x_1 \sim \mathcal{N}(\mu_1, \sigma^2)$ (block + new channel).

3. Days after removing the channel, the average number of visitors are $x_2 \sim \mathcal{N}(\mu_2, \sigma^2)$ (block + no new channel).
4. In an imaginative scenario that the new channel is added but the most famous website haven't been blocked, the average number of visitors is $x_3 \sim \mathcal{N}(\mu_3, \sigma_3^2)$ (no block + new channel).

Assuming that a user decides to visit the website because of the blockade, a new channel, or none of the two (independent).

1. Hamtaro found the p-value of 0.03 from doing a t-test on $H_a : x_1 > x_0$. Can he conclude that adding the new channel significantly increases the number of visitors? Justify your answer.
2. Hamtaro did another t-test and found the p-value of 0.1 from testing $H_a : x_1 > x_2$. Does he now have enough information to conclude anything about x_3 ?
3. Does the current setups, 1. and 2., lead to the final question about the significance of adding the new channel?

- If yes, what should you do next to get the final answer?
- If no, Can we use the hypothesis testing answer to solve this problem?
  - If yes, design your testing, describe assumptions you made.
  - If no, explain why.

Problem 4

4.1) $H_a: x_1 > x_0 \longrightarrow H_0: x_1 \leq x_0$

reject H_0 if p-value $< \alpha$ } ไม่รู้ α Let $\alpha = 0.05$
can't reject H_0 if p-value $\geq \alpha$ }

$\therefore$ p-value < 0.05 reject H_0 *

4.2) Let $\alpha = 0.5$ $H_a: x_1 > x_2 \longrightarrow H_0: x_1 \leq x_2$

p-value $= 0.1 \geq 0.05$

can't reject H_0

information now : 1) $x_0 > x_1$ ✓ } can't conclude anything about x_3
2) $x_1 > x_2$? }

4.3) No, it doesn't.

Yes, we can.

1) test $H_a: x_3 > x_0$ same "no block" $\longleftarrow$ fix control variable (ตัวแปรควบคุม)

2) test $H_a: x_3 > x_1$ same "new channel"

assumption

1) $x_i \sim N(\mu_i, \sigma_i^2)$ # visitor follow Normal Dist. } given
2) (block/noblock) and (new ch/no new ch.) are independent }

3) fix another control variable such as normal day/holiday time season festival etc.

Problem 5: Hamtaro and his casino

After opening Hamhub for a short while, the website was also banned by the government since it contains some 'immoral' videos. Hamtaro then moves on and follows his other passionate dream of creating a gambling empire. Therefore, he hones his skills on public gambling websites which can be easily found even if they are illegal.

After playing for a while, he notices that the online gambling business has great business potential since the risk of gambling websites being banned is much lower than his previous entertainment business. Thus, he decides to open his own online casino.

At the opening date, he offers only a dice game. The rule is simple, the player selects a number and rolls a die. The player will receive a reward if the rolled number is the same as the one he chooses. Hamtaro wants to maximize his profit by cheating using a biased die. Since it is an online casino, he could easily change the biasness of the die after the player selects a number. However, the player is not a fool and would notice if it is too biased.

As a player,

1. Formulate the null hypothesis H_0 and alternative hypothesis H_a to investigate the biasness of the dice.
2. Should the H_a be one-sided or two-sided? What are the differences and benefits over another in this problem?
3. The player found the selected number is rolled out 3 out of 30 attempts. If he wants no more than 10% of type-I error, can he reject the H_0? Justify your answer.
4. If the player plays 200 games, what is the rejection region if he wants no more than 10% type-I error?
5. What would be the result in 4. if the true distribution is approximated by the Normal distribution?

As Hamtaro,

6. The mastermind Hamtaro observes that players will play no more than 200 games a day. He knows that some players studied Com Eng Math 2 and might perform hypothesis testing to check whether Hamtaro cheats. Hamtaro assumes that the players will use a significant level of 0.01. He thinks that it is safe enough if the probability of being caught by a player is less than 0.05. What should be the lowest probability of rolling the selected number? (How much bias can he put in the dice) Answer in floating number with a precision of 3.
7. What if Hamtaro accepts the probability of being caught $= 0.01$ instead? Answer in floating number with the precision of 5.

(Hint Problem 6 and 7 are related to test power)

...

Problem 5

5.1) $H_0 : p = \frac{1}{6}$, $H_a : p \neq \frac{1}{6}$

5.2) 2 sides Hypothesis. We are testing if the dice is fair p exactly $\frac{1}{6}$ so we can't use 1 side Hypothesis (1 direction)
2 directions

5.3) $x \sim binomial(30, 1/6)$

$10\% \longrightarrow \alpha' = 0.1$

P-value $= 0.46308 \geq 0.1$

can't reject H_0

5.4) $x \sim Binomial(200, 1/6) \longrightarrow \mu = \frac{200}{6} = 33.33$ $\quad \sigma = \sqrt{200 \cdot \frac{1}{6} \cdot \frac{5}{6}} = 5.27$

$\alpha = 0.1 \rightarrow \frac{\alpha}{2} = 0.05$

$Z_{\alpha/2} = \pm 1.645$

$x = \mu \pm Z_{\alpha/2}\, \sigma = 24.64 - 42.02$

Rejection Region $x \leq 24 \cup x \geq 42$

5.5) $x \sim N(33.33, 5.27^2)$

$\left.\begin{array}{l} \text{Lower} = 24.66 \\ \text{Upper} = 42.00 \end{array}\right\}$ Reject Region $x \leq 24 \cup x \geq 42$

5.6) $\mu = \frac{200}{6}$ $\quad \sigma = \sqrt{np(1-p)}$ $\quad Z_{\alpha/2} = \pm 2.576$

$\alpha = 0.01$

$x_{lower} = 33.33 - 2.576(5.27) = 19.76$

$x_{higher} = 33.33 + 2.576(5.27) = 46.9$

Reject Region $x \leq 19 \cup x \geq 47$

5.7) 0.136292

5.3

In [279...

```python
from scipy.stats import binomtest

# Parameters
k = 3  # Observed successes
n = 30  # Number of trials
p_null = 1 / 6  # Null hypothesis proportion

# Perform the binomial test
result = binomtest(k, n, p_null, alternative='two-sided')

# Output the results
print("P-Value:", result.pvalue)
```

```
P-Value: 0.46308240295825426
```

5.5

In [280…

```
from scipy.stats import norm

n = 200  # number of trials
p_null = 1 / 6  # null hypothesis probability
alpha = 0.10  # significance level

# Mean and standard deviation under H0 for the Normal approximation
mean = n * p_null
std_dev = (n * p_null * (1 - p_null)) ** 0.5

# Two-tailed test: split alpha equally between the two tails
alpha_half = alpha / 2

# Critical z-scores for two-tailed test
z_critical_low = norm.ppf(alpha_half)
z_critical_high = norm.ppf(1 - alpha_half)

# Convert z-scores to X values (Normal approximation to Binomial)
x_low = mean + z_critical_low * std_dev
x_high = mean + z_critical_high * std_dev

# Output rejection region
print("x_lower",x_low)
print("x_high",x_high)
```

```
x_lower 24.66419353540737
x_high 42.002473131259286
```

5.7

In [289…

```
from scipy.stats import binom
import numpy as np
k_low = 19
k_high = 47
safe_probability = 0.05
for p in np.arange(0.000001,1.000001,0.000001):
    prob_caught = binom.cdf(k_low, n, p) + (1 - binom.cdf(k_high, n, p))
    if prob_caught < safe_probability:
        print(round(p, 6))
        break
```

```
0.136292
```

Problem 6: Hamtaro and the new AC company

From the previous assignment, Hamtaro tried to control the temperature of the cloud storage room so that the failure rate of storage disks is as low as possible. Later, a new company came to him and offered a new AC system that, in theory, could provide a more stable room temperature. To strengthen their claim, the company sends the historical room temperature of the new AC, of which the target temperature is set 15, to Hamtaro.

Given that Hamtaro's existing AC provides the room temperature of $t \sim \mathcal{N}(15, 0.5^2)$:

TODO:6

1. Formulate the null and alternative hypotheses for determining whether the new AC is better than the existing one or not. List your assumptions that are required to make this experiment possible.
2. Can you decide which AC system is better? Justify your answer.

In [282…

```
import numpy as np
temp_log = np.array([14.66017243, 14.82134507, 14.75354867, 15.02847413, 15.9633
       15.46598137, 15.35605532, 14.91048177, 15.13237189, 14.38789873,
       15.76833691, 14.85383663, 15.28335022, 15.06718901, 15.44364169,
       14.29511914, 15.13458572, 14.57428013, 15.14885716, 14.08580661,
       15.6006654 , 14.98109974, 14.95059512, 14.91460432, 14.68809902,
       15.4988617 , 14.99646465, 15.00654947, 14.65024467, 15.20684546,
       15.540787  , 15.39207656, 14.53129171, 14.27527689, 14.37856735,
       15.4685476 , 14.94268835, 15.28311368, 14.8878152 , 15.52350034,
       14.35791689, 15.11741279, 15.41721681, 15.56690632, 15.30108101,
       14.7138976 , 15.39536719, 15.02994055, 14.74887633, 14.81419334,
       15.36735467, 14.89706838, 14.89134826, 15.19781408, 15.3273354 ,
       15.16729623, 14.82748547, 15.59488402, 15.49763473, 15.12876929,
       14.11446324, 14.61298282, 14.57006854, 15.13227246, 14.68369474,
       14.96443757, 15.73872741, 15.48498884, 15.35770021, 15.13471147,
       14.94871779, 13.91322937, 14.84786617, 14.42086587, 15.26216287,
       14.33225067, 14.94179209, 14.57095395, 15.1261513 , 14.93201265,
       14.82252959, 15.19061294, 15.33257912, 14.72448901, 15.54406202,
       14.72704346, 14.9902773 , 14.71477903, 14.90866689, 14.28862563,
       15.04302902, 15.06973955, 14.51951387, 14.61413562, 14.58725869,
       14.41407727, 15.05585075, 14.69229146, 14.30425173, 14.76913898,
       14.27819269, 14.93917912, 14.22675051, 15.20964   , 14.96122782,
       14.05371218, 15.10273752, 15.50886439, 15.43965366, 14.98863063,
       15.34326459, 15.23694786, 13.90170147, 15.29660252, 15.26635161,
       15.34710713, 14.34928594, 15.61509746, 15.80476574, 15.36769161,
       14.52027993, 14.80624255, 14.58269606, 15.58830065, 14.25665696,
       14.86914893, 15.40500584, 15.28855103, 15.43907472, 15.18196326,
       15.47088551, 15.06327054, 15.01022434, 14.43508736, 15.3791887 ,
       14.86202479, 15.1697766 , 14.6434633 , 15.7263277 , 14.31813452,
       15.30657752, 14.91471004, 15.1456617 , 14.93856484, 15.14098396,
       14.76996958, 15.3890821 , 15.53549397, 15.28528007, 15.61416247,
       14.4514347 , 14.75105769, 14.22367585, 14.93898327, 14.61033024,
       14.96348807, 15.24771829, 14.84653005, 15.36780845, 14.96846837,
       14.66094081, 14.75905691, 14.96864336, 15.55687252, 14.62138304,
       16.02201637, 14.95786084, 14.98549356, 15.18029872, 14.82305383,
       15.093562  , 15.98065684, 15.27950419, 15.42169411, 15.66950953,
       14.90725077, 13.69523862, 15.7470953 , 14.93824139, 15.65590845,
       14.69911713, 14.63306529, 15.09566097, 15.00531748, 15.0664824 ,
       15.00496274, 15.1577527 , 15.26365236, 14.98708579, 14.43256043,
       15.5816707 , 14.69227952, 15.22774367, 15.01510129, 15.03105086,
       15.07222669, 15.22579141, 15.34835664, 15.14017702, 15.12604511])
```

Problem 7: Hamtaro Empire Part 3

After Hamtaro has successfully established his factories (in Problem 4.2 HW 3), he further boosts the factory productivity by replacing the old machines with a new type-II variant.

However, there is a concern from the local factory managers that Hamtaro might get bamboozled, since they do not observe an increase in productivity compared to the previous one. Therefore, to ease their concern, he decided to conduct a z-testing.

Given that the number of goods produced each day by the old machines was $x \sim \mathcal{N}(5000, 20^2)$:

1. Formulate the null and alternative hypothesis for determining whether the new machine is better than the previous one at a significant level = 0.05.
2. From the testing, can Hamtaro conclude that factory productivity increased as a whole?
3. Can Hamtaro say the same for each individual factory?
4. Repeat 1-3 again but with a t-test. Is there any difference from the z-test? What, and why does it happen?

In [283…

```
from scipy.stats import norm,t
import numpy as np

# 30 days of product quantity in 4 factories

fac_0 = np.array([4993.89323126, 5021.67118211, 5023.54710937, 4999.11746331,
       5001.53450095, 4986.27990953, 4987.12362188, 5004.91535087,
       4999.97591193, 5038.09176163, 4993.94184053, 5026.5264468 ,
       5040.62862593, 4979.91124088, 5008.59143715, 5016.45331659,
       5013.63203948, 5010.84253735, 5014.99772195, 5002.39462129,
       5047.80507624, 5007.23005532, 5019.87205007, 5005.76363012,
       4997.09106036, 4982.80291132, 5037.18158407, 4996.54197735,
       5007.57964251, 4971.18880247])

fac_1 = np.array([5036.80041897, 4989.33779117, 4971.68709581, 5041.92493487,
       5041.64823146, 5026.33602398, 5009.58334612, 4989.05382998,
       5031.17423169, 4992.20198911, 4970.63425555, 5007.17615704,
       4993.84416738, 5028.59671588, 5009.91388156, 5049.64187466,
       5015.12711371, 5032.2900513 , 5013.66869347, 4988.21257317,
       5020.44276181, 4985.62886808, 5022.46800468, 5042.35501669,
       5001.6153908 , 5012.14209858, 5006.14666402, 4999.93219541,
       5002.77927647, 5002.20750425])

fac_2 = np.array([5029.95293241, 5019.47959949, 4976.8427836 , 4985.22792264,
       4994.97618684, 5026.75059569, 5015.71350753, 5008.46632673,
       5037.96915682, 4990.38948551, 4988.7082206 , 5032.42440206,
       5036.41040953, 5003.75236158, 5002.62361815, 4998.8932057 ,
       5000.51153033, 5002.19196574, 5023.74534474, 5032.03601587,
       5000.10614764, 4989.74566985, 4985.97436664, 4973.63380449,
       5028.58100504, 4997.8426781 , 5011.4202198 , 5018.71432385,
       4969.03296199, 5009.23456565])
fac_3 = np.array([4962.36508403, 5015.91734917, 5030.86885403, 5012.74787091,
       5036.94455211, 4995.2103757 , 5029.84241184, 5015.68062582,
       4996.43546786, 4999.57614716, 5006.88735305, 5035.10432486,
       5017.33437936, 5006.70625696, 5007.97827037, 4981.80482708,
       5020.78603239, 4993.12742287, 4996.10718141, 4988.00315629,
       5003.00004152, 4949.54117305, 5008.6250048 , 5004.09075453,
       5026.56246304, 5011.02296759, 5010.67413795, 4990.58062539,
       5009.64435703, 5001.9413428 ])
```

Problem 7

7.1) $H_0 : \mu_{new} \leq 5000$ $\alpha = 0.05$ $Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

$H_A : \mu_{new} > 5000$

7.2) $\bar{x}_{all} = 5007.734666198916 \rightarrow Z_{all} = 4.23641152 \rightarrow P(z > z_{overall}) = 1.135401 \cdot 10^{-5} < 0.05$ reject H_0

$\therefore$ new machine significantly increased productivity

7.3 $\bar{x}_{fac0} = 5007.904222020666 \rightarrow Z_{fac0} = 2.16466035 \rightarrow P(z > z_{fac0}) = 0.0152069 < 0.05$ reject H_0

$\bar{x}_{fac1} = 5011.152371922332 \rightarrow Z_{fac1} = 3.05420283 \rightarrow P(z > z_{fac1}) = 0.0011283 < 0.05$ reject H_0

$\bar{x}_{fac2} = 5006.378377153666 \rightarrow Z_{fac2} = 1.74679052 \rightarrow P(z > z_{fac2}) = 0.0403368 < 0.05$ reject H_0

$\bar{x}_{fac3} = 5005.503693698999 \rightarrow Z_{fac3} = 1.50724859 \rightarrow P(z > z_{fac3}) = 0.0658735 \geq 0.05$ not reject H_0

$\therefore$ factory 0,1,2 significantly increased productivity

but factory 3 cannot (not have enough evidence to conclude that improved)

7.4) $t_{all} = 4.340952 \rightarrow P(t > t_{all}) = 1.498021 \cdot 10^{-5} < 0.05$

$t_{fac0} = 2.326243 \rightarrow P(t > t_{fac0}) = 0.013600 < 0.05$ reject H_0

$t_{fac1} = 2.882684 \rightarrow P(t > t_{fac1}) = 0.003676 < 0.05$ reject H_0

$t_{fac2} = 1.804536 \rightarrow P(t > t_{fac2}) = 0.040771 < 0.05$ reject H_0

$t_{fac3} = 1.558858 \rightarrow P(t > t_{fac3}) = 0.064939 \geq 0.05$ not reject H_0

$\therefore$ Over all : new machine significantly increased productivity

Each : fac 0,1,2 significantly increased productivity

but fac 3 does not

In [284…

```
fac_all = np.concatenate([
    fac_0, fac_1, fac_2, fac_3
])
```

7.1)Ans H0 : mu_new <= 5000, Ha : mu_nex > 5000

In [285…

```
mean_fac0 = np.mean(fac_0)
mean_fac1 = np.mean(fac_1)
mean_fac2 = np.mean(fac_2)
mean_fac3 = np.mean(fac_3)
mean_overAll = np.mean(fac_all)

print("mu_all :",mean_overAll)
print("mu_fac0 :",mean_fac0)
```

```
print("mu_fac1 :",mean_fac1)
print("mu_fac2 :",mean_fac2)
print("mu_fac3 :",mean_fac3)
```

```
mu_all : 5007.734666198917
mu_fac0 : 5007.904222020666
mu_fac1 : 5011.152371922332
mu_fac2 : 5006.378377153666
mu_fac3 : 5005.503693698999
```

7.2&7.3

In [286…

```
z_all=(mean_overAll-5000)/(20/np.sqrt(120))
z0=(mean_fac0-5000)/(20/np.sqrt(30))
z1=(mean_fac1-5000)/(20/np.sqrt(30))
z2=(mean_fac2-5000)/(20/np.sqrt(30))
z3=(mean_fac3-5000)/(20/np.sqrt(30))

print("Z Stat all",z_all)
print("Z Stat fac0",z0)
print("Z Stat fac1",z1)
print("Z Stat fac2",z2)
print("Z Stat fac3",z3)
print("---------------------------")


p_value_fac_all = 1 - norm.cdf(z_all)
p_value_fac0 = 1 - norm.cdf(z0)
p_value_fac1 = 1 - norm.cdf(z1)
p_value_fac2 = 1 - norm.cdf(z2)
p_value_fac3 = 1 - norm.cdf(z3)


print("p_value all",p_value_fac_all)
print("p_value fac0",p_value_fac0)
print("p_value fac1",p_value_fac1)
print("p_value fac2",p_value_fac2)
print("p_value fac3",p_value_fac3)
```

```
Z Stat all 4.2364511519194625
Z Stat fac0 2.1646603501238224
Z Stat fac1 3.05420283577434
Z Stat fac2 1.7467905236693428
Z Stat fac3 1.5072485942704237
---------------------------
p_value all 1.1354014685771574e-05
p_value fac0 0.015206852813733351
p_value fac1 0.0011282972610209274
p_value fac2 0.040336840487064096
p_value fac3 0.0658734743204481
```

7.4

In [287…

```
sd_fac_all = np.std(fac_all)
sd_fac_0 = np.std(fac_0)
sd_fac_1 = np.std(fac_1)
sd_fac_2 = np.std(fac_2)
sd_fac_3 = np.std(fac_3)

print("mu_all, sd_fac_all :",mean_overAll,sd_fac_all)
```

```python
print("mu_fac0, std_fac0 :",mean_fac0,sd_fac_0)
print("mu_fac1, std_fac1 :",mean_fac1,sd_fac_1)
print("mu_fac2, std_fac2 :",mean_fac2,sd_fac_2)
print("mu_fac3, std_fac3 :",mean_fac3,sd_fac_3)
```

```
mu_all, sd_fac_all : 5007.734666198917 19.437039601366116
mu_fac0, std_fac0 : 5007.904222020666 18.297977890357476
mu_fac1, std_fac1 : 5011.152371922332 20.83383275578031
mu_fac2, std_fac2 : 5006.378377153666 19.034592878918822
mu_fac3, std_fac3 : 5005.503693698999 19.01282477077151
```

In [288…

```python
t_overAll=(mean_overAll-5000)/(sd_fac_all/np.sqrt(120-1))
t0=(mean_fac0-5000)/(sd_fac_0/np.sqrt(30-1))
t1=(mean_fac1-5000)/(sd_fac_1/np.sqrt(30-1))
t2=(mean_fac2-5000)/(sd_fac_2/np.sqrt(30-1))
t3=(mean_fac3-5000)/(sd_fac_3/np.sqrt(30-1))

print("t Stat all",t_overAll)
print("t Stat fac0",t0)
print("t Stat fac1",t1)
print("t Stat fac2",t2)
print("t Stat fac3",t3)
print("---------------------------")


p_value_all = t.sf(t_overAll, 120-1)
p_value_fac0 = t.sf(t0,30-1)
p_value_fac1 = t.sf(t1,30-1)
p_value_fac2 = t.sf(t2,30-1)
p_value_fac3 = t.sf(t3,30-1)

print("p_value_all",p_value_all)
print("p_value_fac0",p_value_fac0)
print("p_value_fac1",p_value_fac1)
print("p_value_fac2",p_value_fac2)
print("p_value_fac3",p_value_fac3)
```

```
t Stat all 4.340951533630494
t Stat fac0 2.3262427416035476
t Stat fac1 2.8826842135207373
t Stat fac2 1.8045362143046846
t Stat fac3 1.5588581904284349
---------------------------
p_value_all 1.4980212512074299e-05
p_value_fac0 0.013599797606286 93
p_value_fac1 0.0036764678223551303
p_value_fac2 0.04077133002863415
p_value_fac3 0.06493877130251395
```