

תרגיל בית 6

שאלה 1

חוקר מתעניין בהשפעה של הורמוני גדילה מסוגים שונים על קצב הגביהה של בני נוער בגילאי 14 עד 16. לצורך בחינת השפעה, החוקר אסף בני נוער מתנדבים, ומתוכם דגם באקראי 32 נערים בגיל 14 (כולם בנים). הנערים חולקו לארבע קבוצות בגודל שווה. הקבוצה הראשונה קיבלה למשך שנתיים הורמון מסוג x, הקבוצה השנייה קבלה למשך שנתיים הורמון מסוג y, הקבוצה השלישית קיבלה במשך שנתיים הורמון מסוג z, ואילו הקבוצה הרביעית לא קיבלה שום הורמון גדילה.

החוקר מדד את גובהו של כל נער פעמיים. בפעם הראשונה עם תחילת המחקר ובפעם השנייה כעבור שנתיים בסיום המחקר. הטבלה שלהלן מתארת את הפרש הגבהים בין שתי המדידות העוקבות על פני הנערים (כלומר את שיעור הגביהה בסנטימטרים של הנערים במשך השנתיים).

קבוצה 1 (הורמון x)	קבוצה 2 (הורמון y)	קבוצה 3 (הורמון z)	קבוצה 4 (שום הורמון)
8.6	11.5	14.2	6.7
14.1	12.7	13.9	9.1
9.4	16.3	12.5	4.8
12.0	14.4	9.6	5.6
7.9	15.1	11.2	7.2
13.3	9.3	13.1	5.1
11.4	8.9	18.6	8.8
11.7	14.5	16.2	7.4

החוקר השתמש בתוכנה סטטיסטית כדי לבחון את ההשערה כי להורמוני הגדילה אין השפעה על שיעור הגביהה.

א. תארו את המודל המתאים לבחינת ההשערה. מהן הנחות המודל ?

נתון להלן הפלט שהוציאה התוכנה :

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
treatment	3	222.053	74.018	12.942	1.738e-05
Residuals	28	160.136	5.719		

כתוצאה הסיק החוקר כי ההורמונים כן משפיעים על שיעור הגביהה. (בצעו את כל המבחנים ברמת מובהקות 0.05)

- בחנו את ההשערה כי קצב הגביהה של בני נוער שקיבלו הורמונים מאיזשהו סוג זהה לזה של בני נוער שלא קיבלו שום הורמון.
- בחנו את ההשערה כי קצב הגביהה של בני נוער שקיבלו את הורמון x זהה לקצב הגביהה של בני נוער שקיבלו את הורמון y.
- בחנו את ההשערה כי קצב הגביהה של בני נוער שקיבלו את הורמון x או y זהה לקצב הגביהה של בני נוער שקיבלו את הורמון z.
- מהי ההסתברות כי תוחלת הגביהה של נערים בגיל 14 שמקבלים את הורמון z גבוהה יותר מ 14.37 ס"מ לשנתיים?

- השוו בין שיעור הגביהה של נערים בכל זוגות הקבוצות שהשתתפו בניסוי. נסחו את המסקנות מהשוואה זו בשלושה או ארבעה משפטים קצרים.

שאלה 2

מתוך קונצרן הנעלה הכולל 30 מפעלי ייצור ברחבי העולם, נדגמו 4 מפעלים. בכל מפעל דגמו מספר זהה של נעליים ונתנו להם ציון על איכות הייצור בסקלה של 0 עד 100 (כאשר 100 הוא הציון עבור איכות מושלמת). להלן התוצאות:

מפעל בסין	מפעל בווייטנאם	מפעל במקסיקו	מפעל באנגליה
89	79	93	98
92	84	95	95
85	67	91	96
87	83	91	91
84	77	92	99

הערה: את שני הסעיפים הראשונים כבר פתרתם במסגרת תרגיל בית 2 (אין צורך לפתור מחדש).

- האם קיימת שונות מובהקת בין האיכות של נעליים מהמפעלים השונים?
- אם השונות מובהקת, אמדו את ערכה.
- העריכו את כל הפרמטרים של המודל ובנו עבורם רב"ס ברמת מובהקות של 0.05.
- העריכו את אחוז השונות המוסברת ע"י סוג המפעל מתוך סך השונות של תצפית בודדת.

TABLE XII Significant Ranges for Duncan's Multiple Range Test (continued)

f^*	$r_{\alpha}(p, f)$											
	p											
	2	3	4	5	6	7	8	9	10	20	50	100
1	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0	18.0
2	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09	6.09
3	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50	4.50
4	3.93	4.01	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02	4.02
5	3.64	3.74	3.79	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83	3.83
6	3.46	3.58	3.64	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68	3.68
7	3.35	3.47	3.54	3.58	3.60	3.61	3.61	3.61	3.61	3.61	3.61	3.61
8	3.26	3.39	3.47	3.52	3.55	3.56	3.56	3.56	3.56	3.56	3.56	3.56
9	3.20	3.34	3.41	3.47	3.50	3.52	3.52	3.52	3.52	3.52	3.52	3.52
10	3.15	3.30	3.37	3.43	3.46	3.47	3.47	3.47	3.47	3.48	3.48	3.48
11	3.11	3.27	3.35	3.39	3.43	3.44	3.45	3.46	3.46	3.48	3.48	3.48
12	3.08	3.23	3.33	3.36	3.40	3.42	3.44	3.46	3.46	3.48	3.48	3.48
13	3.06	3.21	3.30	3.35	3.38	3.41	3.42	3.44	3.45	3.47	3.47	3.47
14	3.03	3.18	3.27	3.33	3.37	3.39	3.41	3.42	3.44	3.47	3.47	3.47
15	3.01	3.16	3.25	3.31	3.36	3.38	3.40	3.42	3.43	3.47	3.47	3.47
16	3.00	3.15	3.23	3.30	3.34	3.37	3.39	3.41	3.43	3.47	3.47	3.47
17	2.98	3.13	3.22	3.28	3.33	3.36	3.38	3.40	3.42	3.47	3.47	3.47
18	2.97	3.12	3.21	3.27	3.32	3.35	3.37	3.39	3.41	3.47	3.47	3.47
19	2.96	3.11	3.19	3.26	3.31	3.35	3.37	3.39	3.41	3.47	3.47	3.47
20	2.95	3.10	3.18	3.25	3.30	3.34	3.36	3.38	3.40	3.47	3.47	3.47
30	2.89	3.04	3.12	3.20	3.25	3.29	3.32	3.35	3.37	3.47	3.47	3.47
40	2.86	3.01	3.10	3.17	3.22	3.27	3.30	3.33	3.35	3.47	3.47	3.47
60	2.83	2.98	3.08	3.14	3.20	3.24	3.28	3.31	3.33	3.47	3.48	3.48
100	2.80	2.95	3.05	3.12	3.18	3.22	3.26	3.29	3.32	3.47	3.53	3.53
∞	2.77	2.92	3.02	3.09	3.15	3.19	3.23	3.26	3.29	3.47	3.61	3.67

TABLE XII Significant Ranges for Duncan's Multiple Range Test

f^*	$r_{\alpha}(p, f)$											
	p											
	2	3	4	5	6	7	8	9	10	20	50	100
1	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0
2	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0
3	8.26	8.5	8.6	8.7	8.8	8.9	8.9	9.0	9.0	9.3	9.3	9.3
4	6.51	6.8	6.9	7.0	7.1	7.1	7.2	7.2	7.3	7.5	7.5	7.5
5	5.70	5.96	6.11	6.18	6.26	6.33	6.40	6.44	6.5	6.8	6.8	6.8
6	5.24	5.51	5.65	5.73	5.81	5.88	5.95	6.00	6.0	6.3	6.3	6.3
7	4.96	5.22	5.37	5.45	5.53	5.61	5.69	5.73	5.8	6.0	6.0	6.0
8	4.74	5.00	5.14	5.23	5.32	5.40	5.47	5.51	5.5	5.8	5.8	5.8
9	4.60	4.86	4.99	5.08	5.17	5.25	5.32	5.36	5.4	5.7	5.7	5.7
10	4.48	4.73	4.88	4.96	5.06	5.13	5.20	5.24	5.28	5.55	5.55	5.55
11	4.39	4.63	4.77	4.86	4.94	5.01	5.06	5.12	5.15	5.39	5.39	5.39
12	4.32	4.55	4.68	4.76	4.84	4.92	4.96	5.02	5.07	5.26	5.26	5.26
13	4.26	4.48	4.62	4.69	4.74	4.84	4.88	4.94	4.98	5.15	5.15	5.15
14	4.21	4.42	4.55	4.63	4.70	4.78	4.83	4.87	4.91	5.07	5.07	5.07
15	4.17	4.37	4.50	4.58	4.64	4.72	4.77	4.81	4.84	5.00	5.00	5.00
16	4.13	4.34	4.45	4.54	4.60	4.67	4.72	4.76	4.79	4.94	4.94	4.94
17	4.10	4.30	4.41	4.50	4.56	4.63	4.68	4.73	4.75	4.89	4.89	4.89
18	4.07	4.27	4.38	4.46	4.53	4.59	4.64	4.68	4.71	4.85	4.85	4.85
19	4.05	4.24	4.35	4.43	4.50	4.56	4.61	4.64	4.67	4.82	4.82	4.82
20	4.02	4.22	4.33	4.40	4.47	4.53	4.58	4.61	4.65	4.79	4.79	4.79
30	3.89	4.06	4.16	4.22	4.32	4.36	4.41	4.45	4.48	4.65	4.71	4.71
40	3.82	3.99	4.10	4.17	4.24	4.30	4.34	4.37	4.41	4.59	4.69	4.69
60	3.76	3.92	4.03	4.12	4.17	4.23	4.27	4.31	4.34	4.53	4.66	4.66
100	3.71	3.86	3.98	4.06	4.11	4.17	4.21	4.25	4.29	4.48	4.64	4.65
∞	3.64	3.80	3.90	3.98	4.04	4.09	4.14	4.17	4.20	4.41	4.60	4.68

① נראה שבנתונים הנכון ביותר למקרה זה הינו מבחן ANOVA
סט קואלסטים. הקואלסטים מראה שמשנו בהם הונס (3,-1,-1,-1)

עם תוצאות מבחן נראה שכן יש הבדל בין קבוצות הורמון /
גלול קבוצות הורמון. (בכמה מאבקים שהא שאלתם).

② עם תוצאות קלטו נראה שיש הבדל א שאלו בהורמון.
בהם זמי שאלו בסוף הקומ.

②

```
# Perform ANOVA test between Col2 and Col3
anova_result = f_oneway(df['x'], df['y'])

# Display the ANOVA test results
anova_result
```

✓ 0.0s

F_onewayResult(statistic=2.0661816712134993, pvalue=0.17257203321426834)

Generate + Code + Markdown

```
print(f'The p-value is {anova_result.pvalue:.4f} which is less than 0.05, so w
```

✓ 0.0s

The p-value is 0.1726 which is less than 0.05, so we reject the null hypothesis.

③

```

avg_z = df['z'].mean()
std_z = df['z'].std()
✓ 0.0s

avg_z, std_z
✓ 0.0s
(13.6625, 2.813202293269566)

# assume avg_z is the mean of the z column and std_z is the standard deviation
# calculate the z-score for each value in the z column
z_scores = (14.374 - avg_z) / std_z
p_less_avg = norm.cdf(z_scores)
p_greater_avg = 1 - p_less_avg
print(f'The probability of a value in the z column being more then the 14.37 is
✓ 0.0s

The probability of a value in the z column being more then the 14.37 is 0.4002

```

```

# Perform Tukey's HSD Test
comp = mc.MultiComparison(values, labels)
tukey_result = comp.tukeyhsd() "tukeyhsd":

# Print the test results
print(tukey_result.summary())

```

```

[38] ✓ 0.4s
... Multiple Comparison of Means - Tukey HSD, FWER=0.05

```

group1	group2	meandiff	p-adj	lower	upper	reject
control	x	4.2125	0.0076	0.9478	7.4772	True
control	y	6.0	0.0001	2.7353	9.2647	True
control	z	6.825	0.0	3.5603	10.0897	True
x	y	1.7875	0.4539	-1.4772	5.0522	False
x	z	2.6125	0.1522	-0.6522	5.8772	False
y	z	0.825	0.9001	-2.4397	4.0897	False

```
# Perform Tukey's HSD Test as a substitute for Duncan's Test
comp = mc.MultiComparison(df_melt['score'], df_melt['plant'])
tukey_result = comp.tukeyhsd() "tukeyhsd": Unknown word.
```

```
# Print the results of Tukey's test
print(tukey_result.summary())
```

✓ 0.4s

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
China_Plant	English_Plant	8.4	0.0257	0.8956	15.9044	True
China_Plant	Mexico_Plant	5.0	0.2643	-2.5044	12.5044	False
China_Plant	Vietnam_Plant	-9.4	0.0119	-16.9044	-1.8956	True
English_Plant	Mexico_Plant	-3.4	0.5782	-10.9044	4.1044	False
English_Plant	Vietnam_Plant	-17.8	0.0	-25.3044	-10.2956	True
Mexico_Plant	Vietnam_Plant	-14.4	0.0003	-21.9044	-6.8956	True

[9] ✓ 0.0s

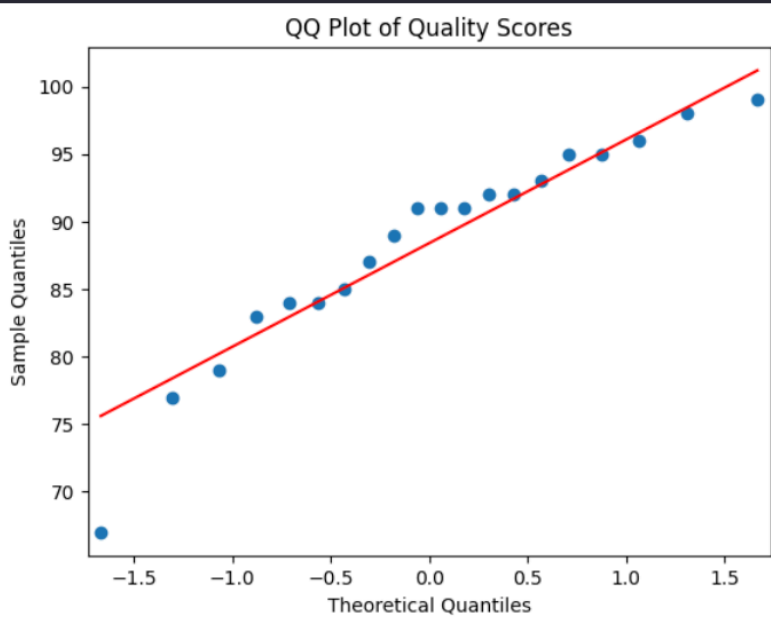
Deviations:

	English_Plant	Mexico_Plant	Vietnam_Plant	China_Plant
0	2.2	0.6	1.0	1.6
1	-0.8	2.6	6.0	4.6
2	0.2	-1.4	-11.0	-2.4
3	-4.8	-1.4	5.0	-0.4
4	3.2	-0.4	-1.0	-3.4

Z-scores (normalized deviations):

	English_Plant	Mexico_Plant	Vietnam_Plant	China_Plant
0	0.706377	0.358569	0.147442	0.498542
1	-0.256865	1.553797	0.884652	1.433307
2	0.064216	-0.836660	-1.621862	-0.747812
3	-1.541187	-0.836660	0.737210	-0.124635
4	1.027458	-0.239046	-0.147442	-1.059401

✓ 0.1s



[9] ✓ 0.0s

```
Overall Mean (ȳ): 88.4
Means per Plant (ȳi): {'English_Plant': 95.8, 'Mexico_Plant': 92.4, 'Vietnam_Plant': 78.0, 'China_Plant': 87.4}
MSE: 17.2
Confidence Interval for μ: (86.43408870084448, 90.36591929915554)
Confidence Intervals for τi: {'English_Plant': (3.0040708088094767, 11.795929191190506), 'Mexico_Plant': (-0.39592919119051473, 8.395929191190515),
Confidence Interval for σ2: (9.540532290242052, 39.8398048771261)
```

[9] ✓ 0.0s

```
'Vietnam_Plant': (-14.79592919119052, -6.004070808809491), 'China_Plant': (-5.395929191190515, 3.395929191190515)
```

```
# Mean Square for Treatments (MStr)
treatment_mean_diffs = means_per_plant - overall_mean
SSTr = n * (treatment_mean_diffs**2).sum()
MS_tr = SSTr / (a - 1)

# Ratio of variances (στ2 / σ2)
variance_ratio = (MS_tr / MSE) - 1
std_ratio = np.sqrt(variance_ratio / n) # √(στ2 / σ2)

# Displaying the results
ratio_results = {
    "Mean Square for Treatments (MStr)": MS_tr,
    "Variance Ratio (στ2 / σ2)": variance_ratio,
    "Standard Deviation Ratio (στ / σ)": std_ratio,
}

ratio_results
```

[8] ✓ 0.0s

```
{'Mean Square for Treatments (MStr)': 299.86666666666673,
'Variance Ratio (στ2 / σ2)': 16.434108527131787,
'Standard Deviation Ratio (στ / σ)': 1.8129593777650832}
```