

# מה זה רב מימד?

The screenshot shows a software window titled "data" with a menu bar (PLOTS, VARIABLE, VIEW) and a toolbar with icons for opening, printing, inserting, deleting, and sorting. Below the toolbar, a "40x80 double" matrix is displayed. The matrix has 40 rows and 24 columns, with the first cell containing the value 73.0226. The interface also shows a status bar at the bottom with navigation arrows.

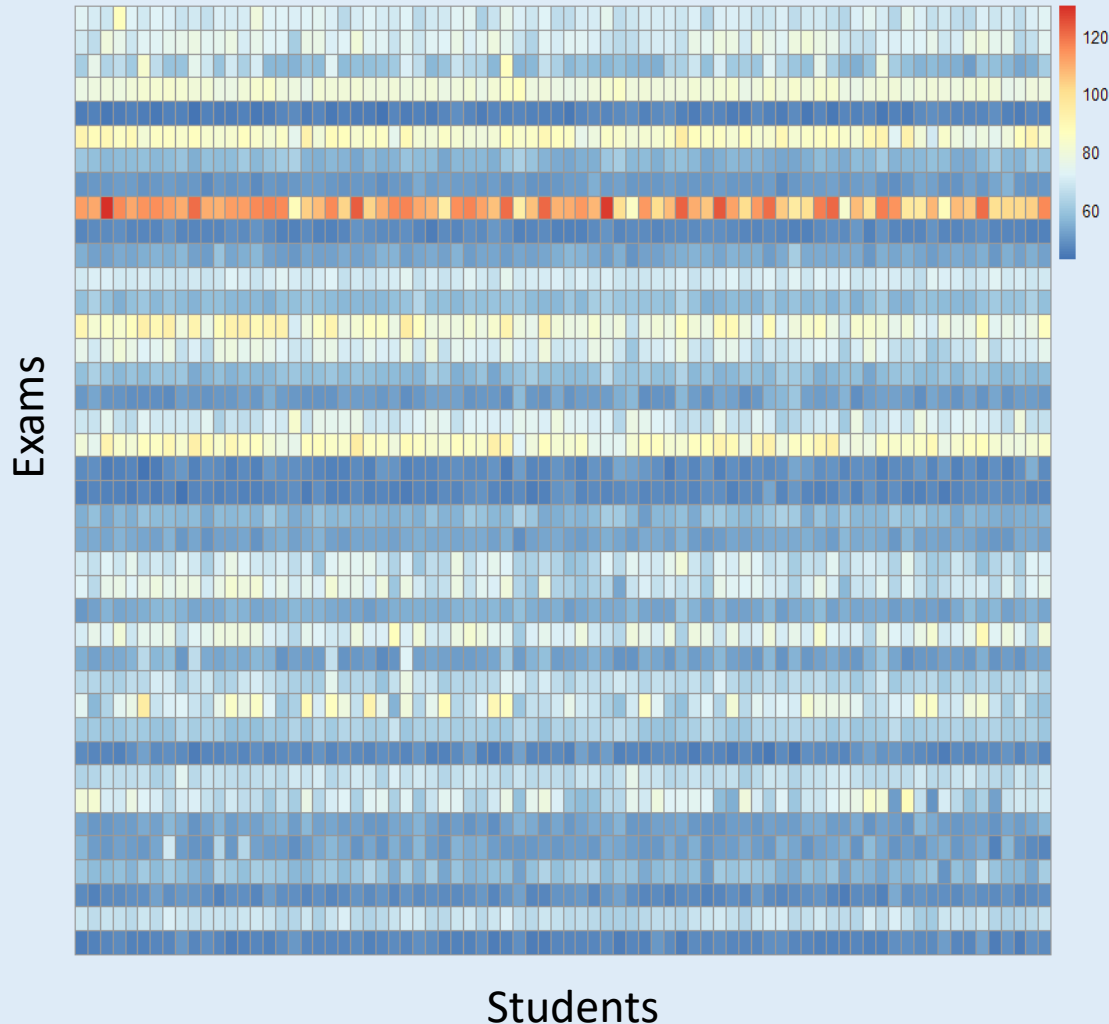
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	73.0226	70.2842	68.4022	86.7442	71.7994	69.1876	71.4856	71.0877	67.8353	69.5716	66.3824	71.1311	69.6583	70.1411	78.8956	71.8003	74.8049	71.4856	71.1788	71.3894	73.2055	71.0132	64.3886	69.9271
2	69.2283	65.5961	78.5237	75.6605	68.1950	72.6888	72.7764	74.5049	74.8569	77.2330	73.2592	74.5751	76.7224	70.1836	75.4056	72.1080	79.0490	71.4830	61.4092	76.1199	75.8806	71.1129	71.6269	79.5570
3	61.3745	75.3215	63.3232	65.7782	60.5553	82.6162	65.5606	58.6803	58.4965	74.3654	68.4490	58.8112	69.2075	62.3931	70.1377	56.2687	70	56.1040	55.5812	58.4956	57.2534	72.5545	56.7272	57.3731
4	76.6470	77.6595	78.4050	77.6031	79.6576	81.4451	81.1070	82.3275	81.1720	80.0954	80.1482	77.1688	77.2876	79.3802	82.0181	82.8702	70	78.9251	77.9733	75.6605	80.9579	80.0165	79.6229	80.8921
5	44.9873	45.1182	44.0467	44.6344	45.3357	44.9691	43.5275	45.9105	45.7614	42.4699	45.4632	44.3943	44.6587	45.3964	43.8751	45.0575	70	46.8424	47.2056	44.5252	46.3118	43.4330	45.2031	44.7811
6	90.7032	87.3822	88.8715	84.5719	88.5083	80.6831	85.0608	84.9099	82.8572	86.0802	82.7315	81.6713	82.3405	85.4925	82.7566	80.8105	70	84.1696	74.1218	91.8917	79.6637	87.5331	86.3498	82.6661
7	59.3746	57.8776	55.7390	57.0064	57.0957	57.7987	56.4611	57.2976	58.6837	54.6745	57.8169	56.9301	57.0957	60.0690	55.7303	58.7800	70	60.2380	62.4381	54.9294	54.4023	56.3501	55.3099	52.3101
8	49.2592	50.2179	49.9700	50.4251	51.1125	49.3381	50.1139	50.6453	49.5002	47.6000	49.5218	49.3866	49.7238	47.3694	50.2205	70	49.2834	50.1850	51.0631	50.7883	50.0810	48.5978	49.8201	
9	112.2899	110.2086	130	115.8311	110.2354	113.7610	114.7197	112.2899	109.6408	120.0501	109.6217	108.9698	112.2934	111.8201	115.7340	116.6563	70	115.2503	87.1144	103.7175	106.4897	115.3846	102.7206	122.3461
10	45.9842	47.1943	47.3937	47.8245	46.5103	46.8406	46.2173	48.8994	46.7184	46.1541	46.7652	46.7123	48.0698	47.0417	48.7911	46.9932	70	45.6062	45.4926	47.6572	45.2170	45.5750	46.4566	49.1661
11	54.1197	51.9508	51.7384	53.6897	52.6677	54.2168	54.8080	54.1249	55.6905	51.5910	51.1195	58.5095	52.6296	54.0391	55.7624	51.3674	70	50.5439	50.3410	52.8844	51.7861	51.3223	53.5242	51.2881
12	69.3089	70.5633	71.0305	67.1688	68.9188	69.6808	68.4871	68.5348	67.1686	69.4884	67.0768	68.1274	71.6113	70.4428	67.6810	70.4454	70	72.6143	70.2946	68.1941	69.7398	72.0690	66.7179	69.5501
13	57.8611	61.7811	57.7077	54.0191	58.0249	59.3278	55.8560	58.3647	56.6232	55.5856	57.3687	57.8021	56.1196	55.8673	55.6593	54.5575	70	57.4164	57.3887	59.4180	57.4667	57.2864	56.2981	56.2111
14	90.3677	81.0524	84.1982	81.9106	86.9176	92.4595	89.5815	92.6589	78.5454	91.0829	75.9396	87.3770	91.5536	92.8296	88.8256	91.2389	70	92.2402	69.9157	78.1579	84.8796	91.4478	76.8125	81.7111
15	75.0753	68.8738	75.4368	79.2944	75.9483	71.9069	73.2887	77.2997	66.5098	71.3357	64.4380	76.9590	77.9351	73.9900	76.6426	75.3796	70	72.1722	68.0589	68.5036	75.6076	76.3002	69.9296	69.5911
16	57.7484	62.1391	58.2390	56.1638	58.6577	56.5036	59.4986	59.4986	59.3400	53.3863	56.3926	57.4069	55.2458	56.8841	56.2505	55.9358	70	59.0998	59.1337	57.6097	57.9067	55.9826	57.1390	58.2531
17	50.0818	52.4614	48.8908	49.1742	48.4582	48.5934	48.2545	47.3945	50.4555	50.0229	48.8275	50.2500	50.2578	51.7471	50.1408	53.9134	70	50.6982	49.8530	49.3216	48.6454	49.0147	51.6838	50.9661
18	68.0398	67.3403	74.8343	66.7283	64.6677	72.9168	69.5933	68.7559	69.4416	68.3259	73.7551	61.8747	67.0273	65.6065	66.3885	70.7462	70	68.7741	82.4142	67.7616	74.8890	74.0238	74.4928	76.4381
19	79.9150	74.0507	90.9711	83.2490	80.3988	85.9632	84.8675	89.1966	82.6387	92.3719	88.0246	83.5490	88.9166	84.6256	85.6398	80.5713	70	76.9391	75.4628	85.5254	86.5682	84.3915	83.3270	94.8051
20	47.4856	48.3490	44.5035	45.2976	46.7747	42.7403	44.5070	45.2395	46.7747	49.8010	45.2395	46.8944	45.6305	47.0036	45.1658	49.7351	70	48.2033	48.0178	46.9993	46.5173	45.8689	45.3635	47.5001
21	45.1277	46.3638	45.3956	45.4666	44.7861	46.5702	44.7506	46.6820	42.9969	46.0535	45.9998	45.7102	46.8155	45.5481	45.3357	46.3742	46.2390	47.5593	47.8340	44.8659	48.0005	45.7293	45.5421	46.9771
22	54.7126	57.4485	52.9156	56.1794	54.6814	58.8823	56.7402	57.3202	55.7295	56.9700	52.5775	56.5322	56.0658	56.4828	54.7733	60.3585	57.2343	56.4975	53.4401	55.2874	58.4748	56.0103	56.6596	55.5671
23	53.3057	53.0673	52.5143	54.7993	53.4956	52.6018	51.8572	55.5379	49.9795	52.5368	48.6784	51.4159	51.5416	52.9217	48.6758	53.0959	53.3985	54.3251	51.6578	53.0769	55.0238	51.9005	50.7008	52.8841
24	69.1988	66.9467	72.0092	67.9124	74.11573	66.4942	69.5967	66.5254	72.4817	69.5196	62.4329	69.7320	75.9673	71.6911	69.5681	67.8855	70.3440	70.3344	65.4991	70.4237	59.3668	72.8718	74.0854	73.7121
25	71.4449	65.0492	75.3562	76.2586	71.7266	75.1559	77.8458	74.5916	76.6686	77.4462	69.2682	76.5169	81.2674	77.6378	80.5245	72.2138	78.0764	75.2548	69.0940	74.4911	70.9204	80.6779	74.6948	75.2371
26	49.8694	51.7913	54.9718	52.7466	53.4132	54.3659	56.0762	57.0202	55.0004	52.9087	53.3699	53.0942	55.6896	55.5067	52.7761	53.7721	56.1256	53.9047	55.6237	55.5422	54.8219	53.5406	52.6313	52.8201
27	70.1013	75.6527	70.9343	79.4461	67.9393	74.1851	74.5994	75.9049	65.8727	72.4045	74.8265	78.2793	77.5849	78.0010	77.0873	70.8355	89.5035	69.5603	63.9760	74.5882	71.9485	69.0133	68.0537	77.5461
28	53.8761	51.9491	53.6776	53.1176	55.2917	64.4753	55.7416	57.9339	49.5201	65.3933	53.8137	52.2074	53.9957	53.0309	56.2487	54.8496	54.8678	48.9272	51.9647	50.8941	50.6964	66.5618	49.3675	50.1381
29	65.1341	63.6396	62.9981	62.5075	66.0365	64.3817	69.6470	68.7836	62.4069	66.6277	68.7247	61.9241	63.3328	67.9783	70.2608	66.2584	60.3083	64.2430	62.2258	62.3341	61.5513	74.4573	63.0762	63.6431
30	73.6857	56.3519	63.4715	73.5566	76.5282	94.4949	67.2129	70.8312	72.0066	67.0646	64.9685	75.2660	83.6755	77.0925	83.8289	71.7630	84.0535	59.5142	62.0299	87.9864	70.1585	85.4387	78.3053	66.6221
31	60.4244	60.1687	59.6061	59.6971	61.6909	60.1410	58.5685	58.9083	63.7671	61.2948	62.5846	61.6658	57.7007	58.3959	58.0891	60.4470	59.8245	59.5584	62.8846	59.2966	64.3045	60.3811	63.5001	62.8081
32	47.3997	46.5702	46.7652	45.4658	48.1375	45.17757	47.5931	48.3117	46.9724	44.3267	46.5875	46.0050	47.5792	46.7444	48.2831	46.0284	46.2711	47.2723	47.8332	46.1983	46.1983	46.8398	47.8418	46.9561
33	63.4628	63.9699	66.6867	65.5623	65.6958	66.3009	62.4590	64.2976	73.1301	64.4753	67.0646	66.1093	65.2979	65.0292	65.7201	66.9242	66.3330	64.8012	69.4728	67.0230	69.6201	64.2716	66.2688	69.5101
34	79.7486	81.2648	70.2053	69.3679	76.7293	71.7474	69.6279	64.7570	70.4038	71.0253	66.5211	62.7892	69.4771	66.0911	66.0495	60.8318	61.2245	66.7898	61.6996	76.4432	63.7142	69.8403	71.8740	62.4021
35	52.9590	49.7351	49.9258	50.5101	50.5274	53.0205	51.7063	56.0910	55.1253	48.6151	51.9655	55.1253	50.7172	52.8142	51.5399	53.9108	49.3901	54.7412	51.1819	56.1317	51.8199	52.7474	51.0553	51.5911
36	55.7667	50.5829	49.8747	52.0583	50.6765	50.0619	52.5359	69.3219	50.7311	49.3684	49.1023	63.7541	49.9076	63.5373	51.2365	50.9253	50.6132	48.2866	48.2866	50.6132	53.5363	56.4308	53.2260	49.3901
37	59.1952	56.6579	56.8079	57.5273	55.1261	60.2710	56.8538	60.6784	57.6886	55.0863	53.6819	63.0198	60.4262	57.5950	61.4820	60.5805	60.3481	59.3833	55.7650	55.6627	59.8696	60.1583	57.2404	59.5701
38	47.0383	45.9555	47.6199	47.6121	47.9233	48.3698	51.5598	48.4963	50.3185	49.3268	48.4504	45.8914	48.1210	48.1912	46.7747	51.0917	49.0867	47.8141	48.2553	46.2546	48.4357	46.5468	47.2116	49.2971
39	68.2522	65.3994	67.4053	66.1752	64.5698	67.8864	66.0374	71.0349	67.9193	66.8167	67.1201	67.6576	66.9363	68.6319	69.2240	67.3576	66.0296	63.8685	64.0306	68.1092	61.7663	67.1461	70.5780	65.1081
40	45.0705	45.5455	45.9998	46.8918	45.7510	45.7215	46.5225	46.5025	50.3454	46.8380	47.9424	46.4410	44.8321	45.1910	46.3075	45.2837	47.9485	47.1345	49.7524	46.6490	46.2815	46.2390	46.5953	47.6361

# מה נעשה בשיעור

1. נגדיר מהן השערות המחקר
2. נתאר את סף הטעות שמקובל עלינו עבור מבחן יחיד ( $alpha$ )
3. נראה שככל שמספר ההשוואות (או המבחנים) גדל – כך הסיכוי לטעות מסוג 1 גדל
  - Family-wise error rate
4. נרצה לקבוע  $alpha$  חדש עבורו הסיכוי לטעות קטן (ולחשב  $p value$ )
  - תיקון בונפרוני
5. חישוב  $p value$  בעזרת פרמוטציות
6. נתקן עם False Discovery Rate (שבוע הבא)

# איך אנחנו אמורים לנתח את זה??

נרצה לדעת באילו בחינות יש הבדל בציון בין סטודנטים הנוטלים ריטלין לבין אלו שלא



40 בחינות

78 סטודנטים – חצי מטופלים בריטלין וחצי לא

שתי קבוצות שנרצה לבחון את ההבדלים בינהן ←

איך ניתן למצוא את התשובה?

נבצע מבחן t על מנת לבדוק את ההבדל בין הקבוצות.

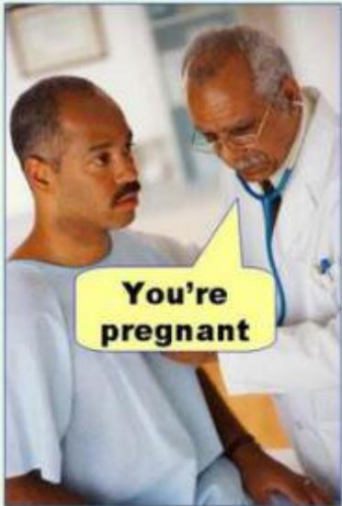
## מה למדתם על t-test ?

# נגדיר את השערות מחקר

השערת האפס ( $H_0$  - null hypothesis) - המניחה שממוצעי שתי הקבוצות שווים (כלומר אין הבדל בין הקבוצות).

השערה אלטרנטיבית ( $H_1$  - alternative hypothesis) - המניחה שממוצעי שתי הקבוצות לא שווים (כלומר יש הבדל בין הקבוצות).

**Type I error**  
(false positive)



**Type II error**  
(false negative)



השערת האפס היא		מציאות החלטה
לא נכונה	נכונה	
החלטה נכונה (true positive) $1 - \beta$	טעות מסוג 1 (false positive) $\alpha$	דחיית $H_0$
טעות מסוג 2 (false negative) $\beta$	החלטה נכונה (true negative) $1 - \alpha$	קבלת $H_0$

$\alpha, \beta$  - הסתברויות

# דוגמא – בדיקת קורונה

## Sensitivity

All samples are COVID-19 **positive**

Days Post-Symptom Onset	n	Positive	Negative
< 3	4	0	4
3 - 7	8	2	6
8 - 13	22	19	3
≥ 14	88 <sup>a</sup>	88	0
<b>Total</b>	<b>122</b>	<b>109</b>	<b>13</b>

## Specificity

All samples are COVID-19 **negative**

Category	n	Positive	Negative
Pre-COVID-19 Outbreak	997	4	993
Other Respiratory Illness	73	0	73
<b>Total</b>	<b>1070</b>	<b>4</b>	<b>1066</b>

המטופל הוא		מציאות
חולה	בריא	
		החלטה
		חולה
החלטה נכונה (true positive) $1 - \beta$	טעות מסוג 1 (false positive) $\alpha$	
		בריא
טעות מסוג 2 (false negative) $\beta$	החלטה נכונה (true negative) $1 - \alpha$	

נרצה לדעת:

1. כמה דוגמאות מחולים נמצאו חיוביות (true positive)
2. כמה דוגמאות מחולים נמצאו שליליות (false positive)
3. כמה דוגמאות מבריאים נמצאו חיוביות (true negative)
4. כמה דוגמאות מבריאים נמצאו שליליות (false negative)

מהי הטעות החמורה יותר במקרה הזה?  $\alpha$  או  $\beta$ ?

# אבל נניח שמישהו עומד למשפט על רצח...

מהי הטעות החמורה יותר במקרה הזה?  $\alpha$  או  $\beta$ ?

רצח או לא?		מציאות
לא רצח	רצח	החלטה
החלטה נכונה (true positive) $1 - \beta$	טעות מסוג 1 (false positive) $\alpha$	לא רצח
טעות מסוג 2 (false negative) $\beta$	החלטה נכונה (true negative) $1 - \alpha$	רצח

רוצח יוצא לחופשי

אדם יושב בכלא לשווא

# נחזור לדוגמת הסטודנטים והריטלין

נרצה לדעת באילו בחינות יש הבדל בציון בין סטודנטים הנוטלים ריטלין לבין אלו שלא

השערת האפס ( $H_0$  – null hypothesis) – מניחה שאין הבדל בין סטודנטים הנוטלים ריטלין לאלו שלא

השערה אלטרנטיבית ( $H_1$  – alternative hypothesis) – מניחה שיש הבדל בין סטודנטים הנוטלים ריטלין לאלו שלא

השערת האפס היא		מציאות החלטה
לא נכונה	נכונה	
החלטה נכונה (true positive) $1 - \beta$	טעות מסוג 1 (false positive) $\alpha$	דחיית $H_0$
טעות מסוג 2 (false negative) $\beta$	החלטה נכונה (true negative) $1 - \alpha$	קבלת $H_0$

לפני שנבצע את המבחן נקבע  $\alpha$  – רמת המובהקות הסטטיסטית – ההסתברות לטעות טעות מסוג ראשון.

בדוגמא שלנו, זו ההסתברות להסיק שיש הבדל בין סטודנטים הנוטלים ריטלין לאלו שלא – למרות שאין כזה הבדל.

דוחים את השערת האפס אם  $p \leq \alpha$

# נחזור לדוגמת הסטודנטים והריטלין

נרצה לדעת באילו בחינות יש הבדל בציון בין סטודנטים הנוטלים ריטלין לבין אלו שלא

```
> num_sig  
[1] 11
```

11 מבחנים בהם ההבדלים בין הקבוצות מובהקים ( $p < 0.05$ )

האם זה אומר שיש סיכוי של 0.05 לטעות מסוג 1 ב-11 המבחנים האלו?



# השערות מרובות

כאשר עושים השערות מרובות יש סיכוי לקבל תוצאות חיוביות שגויות (false positive או false discovery). למה?

אם עשינו למשל 200 מבחנים סטטיסטיים בהם השערת האפס נכונה, ובכל אחד רצינו  $p < 0.05$

בעצם רצינו שהסיכוי שלנו לפחות לטעות אחת ( $\alpha$ ) יהיה 5%, כלומר:

$$200 * 0.05 = 10$$

נצפה ש-10 מתוך 200 המבחנים שלנו יהיו מובהקים רק במקרה (false positive).

ככל שכמות ההשוואות גדלה, כך הסיכוי שטעינו לפחות פעם אחת גדל, על אף ששמרנו על סיכוי נמוך לטעות בכל מבחן בפני עצמו.

# השערות מרובות

## שאלה

נניח ועשינו 40 מבחני t עם  $\alpha = 0.05$ , מהי ההסתברות לעשות לפחות טעות אחת מסוג 1?

## פתרון

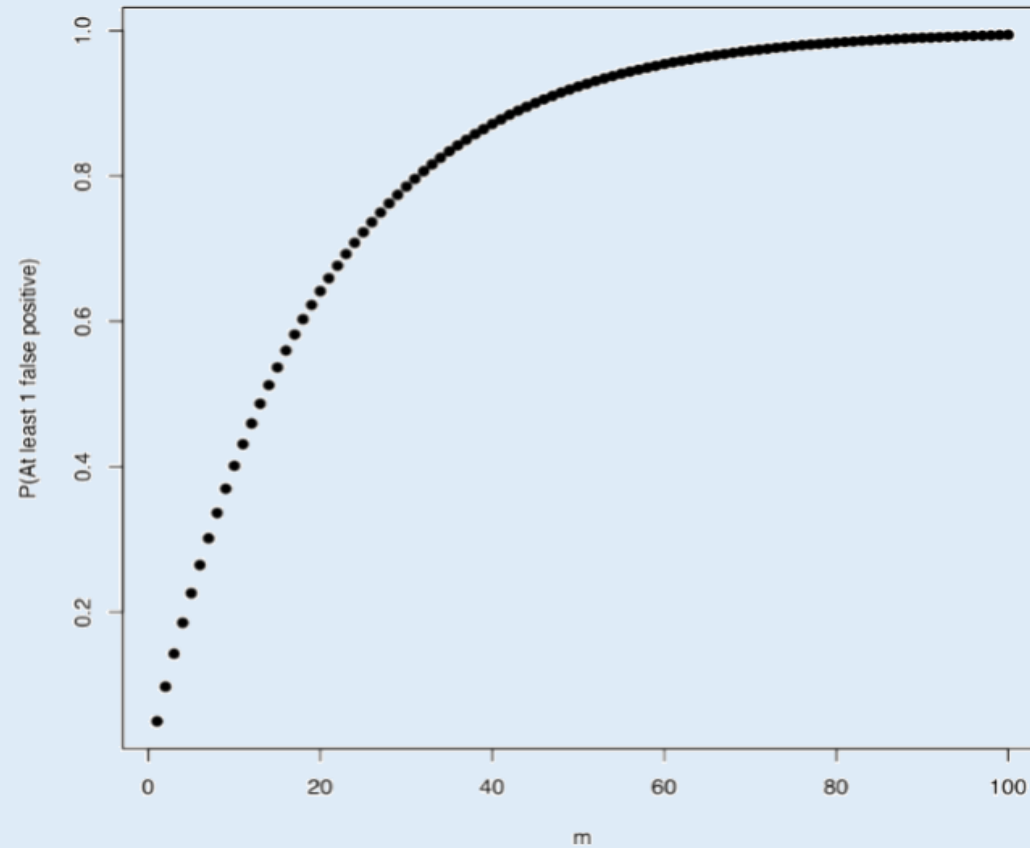
אנחנו בעצם מחפשים מה הסיכוי לעשות טעות מסוג 1 אחת או יותר בכל הניסוי במקום בכל מבחן בנפרד.

לשם כך נשתמש בהגדרה אחרת לשגיאה, (FWER) Family-wise error rate, על מנת לתאר את ה- $\alpha$  הזה:

$$FWER = \alpha = 1 - \underbrace{(1 - \alpha_{comparison})}^{number\ of\ comparisons}$$

החלטה נכונה (true negative)

# Family-Wise Error Rate (FWER)



ככל שעושים יותר מבחנים הסיכוי לטעות טעות מסוג 1 שואף ל-100%

# תיקון בונפרוני

על מנת לפתור את בעיית ההשוואות המרובות (FWER גבוה), נצטרך לשלם "קנס" על כל מבחן שעשינו.

**תיקון בונפרוני** הוא תיקון מאוד מחמיר שמנסה לשלוט ב-FWER – רמת המובהקות של הניסוי מותאמת לכמות ההשוואות שעשינו.

עבור כל השוואה הקבוצות יוכרזו כשונות רק אם:

$$p \text{ value} \leq \frac{\alpha_{total}}{m}$$

כאשר  $m$  הוא מספר ההשוואות.

נניח ועשינו 100 מבחני  $t$  עם  $\alpha = 0.05$ , כל אחד מהמבחנים צריך:  $p \text{ value} \leq 0.0005$  →  $p \text{ value} \leq \frac{0.05}{100}$

על מנת להיות מובהק.

אבל כך יכול להיות שנפספס הבדלים אמיתיים! בונפרוני מעלה את הסיכוי לטעות מסוג 2.

# נחזור לדוגמת הסטודנטים והריטלין

נרצה לדעת באילו בחינות יש הבדל בציון בין סטודנטים הנוטלים ריטלין לבין אלו שלא

11 מבחנים בהם ההבדלים בין הקבוצות מובהקים ( $p < 0.05$ )

```
> sig_tets  
[1] 1 3 5 9 14 15 17 20 21 25 28  
> num_sig  
[1] 11
```

$$p \text{ value} \leq \frac{\alpha_{total}}{m} \quad \text{בונפרוני}$$

4 מבחנים בהם ההבדלים בין הקבוצות מובהקים ( $p < 0.05/40$ )

```
> alpha/40  
[1] 0.00125  
> sig_tets<-which(data_p < alpha/40)  
> sig_tets  
[1] 14 15 21 25
```

# פרמוטציות

נניח ויש לנו קבוצה של אנשים עם סוכרת וקבוצה לא, ואנו רואים הבדל של  $20 \frac{mg}{dL}$  ברמות הכולסטרול הממוצעות בין שתי הקבוצות.

האם זהו הבדל משמעותי?

כדי לבדוק זאת אפשר לערבב באופן אקראי את ה"תוויות" עבור כל אחד מהאנשים, לחלק לשתי קבוצות בגודלן המקורי ולבדוק שוב. האם יש הבדל (מבחן  $t$ ,  $p$ -value...) בממוצע הכולסטרול בין קבוצת ה"סכרתיים" לקבוצת ה"לא סכרתיים".

$$p = \frac{\# \text{ times passed real threshold}}{\# \text{ of permutations}}$$

Original Data

Patient	Age	Gender	Diabetes	Cholesterol
1001	54	M	Yes	170
1002	45	M	No	135
1003	34	F	Yes	185
1004	61	M	Yes	200
1005	58	F	No	140

1 permutation →

Permuted Data

Patient	Age	Gender	Diabetes	Cholesterol
1001	54	M	No	170
1002	45	M	Yes	135
1003	34	F	No	185
1004	61	M	Yes	200
1005	58	F	Yes	140

# פרמוטציות - דוגמא

Perm. #1 3547 3529 3588 3571 3525 3498 $t = +0.7361$	Perm. #2 3547 3529 3588 3571 3525 3498 $t = -0.4094$	Perm. #3 3547 3529 3588 3571 3525 3498 $t = -1.7448$	Perm. #4 3547 3529 3588 3571 3525 3498 $t = +0.0892$
Perm. #5 3547 3529 3588 3571 3525 3498 $t = +1.5510$	Perm. #6 3547 3529 3588 3571 3525 3498 $t = +0.3158$	Perm. #7 3547 3529 3588 3571 3525 3498 $t = +3.3399$	Perm. #8 3547 3529 3588 3571 3525 3498 $t = -0.8470$
Perm. #9 3547 3529 3588 3571 3525 3498 $t = +0.6567$	Perm. #10 3547 3529 3588 3571 3525 3498 $t = -0.2928$	Perm. #11 3547 3529 3588 3571 3525 3498 $t = +0.2928$	Perm. #12 3547 3529 3588 3571 3525 3498 $t = -0.6567$
Perm. #13 3547 3529 3588 3571 3525 3498 $t = +0.8470$	Perm. #14 3547 3529 3588 3571 3525 3498 $t = -3.3399$	Perm. #15 3547 3529 3588 3571 3525 3498 $t = -0.3158$	Perm. #16 3547 3529 3588 3571 3525 3498 $t = -1.5510$
Perm. #17 3547 3529 3588 3571 3525 3498 $t = -0.0892$	Perm. #18 3547 3529 3588 3571 3525 3498 $t = +1.7448$	Perm. #19 3547 3529 3588 3571 3525 3498 $t = +0.4094$	Perm. #20 3547 3529 3588 3571 3525 3498 $t = -0.7361$

**שאלה:** בדוגמא הבאה מוצג משקל ההיפוקמפוס אצל 6 אנשים. 3 עם מחלת אלצהיימר (ירוק) ו-3 בריאים (כחול).

חישבנו את  $t$  וקיבלנו הבדל בין הקבוצות עם  $t_{real} = 0.7361$ . על מנת לבדוק את ההשערה, ערכנו 20 פרמוטציות שתוצאותיהן מוצגות בתמונה. חשבו את ה- $p$ -value עבור המבחן האמיתי.

$p$  value =

# שאלה 1

במטריצה Diabetes מתוארות 10 קבוצות של חולים/בריאים בסכרת, ורמות הסוכר שלהם בדם (הקבוצות אחידות-או חולים או בריאים בתוך אותה קבוצה). כל שורה מתארת קבוצת חולים, וכל עמודה היא מטופל.

נרצה לדעת אילו מן הקבוצות הן קבוצות של חולים ואלו של בריאים על ידי עריכה של סדרת מבחני t ברמת מובהקות 0.01. אם ידוע כי ישנן 3 קבוצות חולים, כיצד יתבצעו מבחני ה-t?

יש לבחור תשובה אחת:

תלוי במה היא השערת האפס

השוואת כל קבוצה לשאר הקבוצות

השוואת כל 3 קבוצות לשאר הקבוצות

ביצוע מבחן t במקרה זה הוא חסר משמעות

מהו ה  $H_0$  ? איך נוכל להשוות אליו ?

כמה מבחני t נצטרך לבצע?

תשובה:



# שאלה 1

מה בעצם ה  $H_0$  שלנו ?

עבור כל קבוצה, ה  $H_0$  הוא שזו קבוצה של בריאים.

איך נדע איך  $H_0$  מתפלג ?

מכוון שרוב הקבוצות מורכבות מבריאים, נניח שכל שאר הקבוצות הן  $H_0$

אז איך נשווה ל  $H_0$  ?

נשווה כל קבוצה ל  $\theta$  הקבוצות האחרות. מכוון שה "קבוצות האחרות" מתפלגות כמו  $H_0$  נבדוק עד כמה אנחנו שונים מהממוצע שלהם.

# שאלה 1

במטריצה Diabetes מתוארות 10 קבוצות של חולים/בריאים בסכרת, ורמות הסוכר שלהם בדם (הקבוצות אחידות-או חולים או בריאים בתוך אותה קבוצה). כל שורה מתארת קבוצת חולים, וכל עמודה היא מטופל.

נרצה לדעת אילו מן הקבוצות הן קבוצות של חולים ואלו של בריאים על ידי עריכה של סדרת מבחני t ברמת מובהקות 0.01. אם ידוע כי ישנן 3 קבוצות חולים, כיצד יתבצעו מבחני ה-t?

יש לבחור תשובה אחת:

תלוי במה היא השערת האפס

השוואת כל קבוצה לשאר הקבוצות

השוואת כל 3 קבוצות לשאר הקבוצות

ביצוע מבחן t במקרה זה הוא חסר משמעות

מהו ה  $H_0$  ? איך נוכל להשוות אליו ?

כמה מבחני t נצטרך לבצע?

תשובה: 10

# שאלה 3

במקרה של המטריצה הנתונה, מהו הסיכוי לעשות לפחות טעות מסוג ראשון (false positive) עבור  $\alpha=0.01$ ?  
אנא כתבו את תשובתכם עם דיוק של שתי ספרות אחרי הנקודה

תשובה:

ועבור  $\alpha=0.05$ ?

תשובה:

# שאלה 3

הקבוצה היא		מציאות החלטה
חולה	בריאה	
החלטה נכונה (true positive) $1 - \beta$	טעות מסוג 1 (false positive) $\alpha$	חולים
טעות מסוג 2 (false negative) $\beta$	החלטה נכונה (true negative) $1 - \alpha$	בריאים

$$P(\text{at least 1 FP}) = 1 - P(\text{no FP}) = 1 - (1 - \alpha)^7 = 1 - (1 - 0.01)^7 = 0.067$$

$$P(\text{at least 1 FP}) = 1 - P(\text{no FP}) = 1 - (1 - \alpha)^7 = 1 - (1 - 0.01)^7 = 0.31$$

## שאלה 5

על מנת להוריד את הסיכוי לטעויות מסוג 1, החלטנו לעשות תיקון בונפרוני על מנת לתקן את בעיית ההשוואות המרובות. מהו ה- $\alpha$  החדש שנדרוש מכל מבחן כתוצאה מהתיקון? הניחו  $\alpha=0.01$  (דיוק של 5 ספרות אחרי הנקודה).

תשובה:

# שאלה 5

על מנת להוריד את הסיכוי לטעויות מסוג 1, החלטנו לעשות תיקון בונפרוני על מנת לתקן את בעיית השוואות המרובות. מהו ה- $\alpha$  החדש שנדרוש מכל מבחן כתוצאה מהתיקון? הניחו  $\alpha=0.01$  (דיוק של 5 ספרות אחרי הנקודה).

**זוהי כמובן חלוקה פשוטה במספר ההיפותזות.**

0.001

תשובה:

# שאלה 6

```
1 load("ritalin_data.Rdata")
2 ritalin_heatmap <- heatmap(Data, Rowv=NA, Colv=NA,scale="none",xlab="Students",ylab="Exams")
3 alpha<-0.05
4
5 #define the groups to be compared
6 group1 <- 1:39
7 group2 <- 40:78
8
9 #find p-values for each comparison
10 data_p <- numeric(dim(Data)[1])
11 for (i in 1:40) {
12   data_p[i]<-t.test(Data[i,group1],Data[i,group2])$p.value
13 }
14
15 #check for significant comparisons
16 sig_tets<-which(data_p < alpha)
17
18 #number of significant tests
19 num_sig<-sum(data_p <= alpha)
```

שורה	מה עושים בשורה?
1	טוענים את מטריצת הנתונים בעזרת פונקציית load לתוך אובייקט שקוראים לו Data
2	מייצרים מפת חום (heatmap) בעזרת פונקציית heatmap. בחרנו לא לעשות אשכול לשורות והעמודות (Rowv,Colv) ולא לעשות נרמול ומרכז לנתונים (scale="none"). xlab ו-ylab נותנים שמות לצירים.
3	קביעת ה-alpha הרצוי לנו
6-7	הגדרת קבוצות הסטודנטים – קבוצה 1 (מטופלים בריטלין) וקבוצה 2 (לא מטופלים בריטלין)
10	יצירת וקטור מספרי (בפונקצייה numeric) באורך של מספר השורות של Data (כלומר – מספר הבחינות)
11	יצירת לולאת for שנותנת למשתנה i ערך עולה בין 1 ל-20 בכל איטרציה
12	הכנסה לתוך וקטור במיקום i את ה-p value המתקבל מה-t test בין שתי הקבוצות. הוקטור הסופי נותן לנו את ה-p value המתקבל לכל בחינה בנפרד.
16	בדיקה אילו מהבחינות עומדות בתנאי הסף של alpha
19	ספירה של כמות הבחינות שעומדות בתנאי הסף של alpha

# שאלה 7

בשאלה הקודמת ראינו שביצוע השוואות לאחר תיקון בונפרוני צמצם את מספר המובהקיות.

בחר את המשפט הנכון:

יש לבחור תשובה אחת:

- אין סיכוי שהמובהקיות שנפלו הן מובהקויות אמיתיות מאחר ותיקון בונפרוני מוצא את האמיתיות.
- המובהקויות שנשארו לאחר תיקון בונפרוני אינן שגיאה.
- המובהקויות שנפלו היו נכונות, אבל מטרת מבחן  $t$  היא למצוא את המובהקות ביותר.
- יש סיכוי שהמובהקויות שנפלו אחרי תיקון בונפרוני הן טעויות מסוג 1 שכן תיקון בונפרוני מעלה את הסיכוי לאלה.
- יש סיכוי שהמובהקויות שנפלו אחרי תיקון בונפרוני הן טעויות מסוג 2 שכן תיקון בונפרוני מעלה את הסיכוי לאלה.



# שאלה 7

בשאלה הקודמת ראינו שביצוע השוואות לאחר תיקון בונפרוני צמצם את מספר המובהקיות.

בחר את המשפט הנכון:

יש לבחור תשובה אחת:

- אין סיכוי שהמובהקיות שנפלו הן מובהקיות אמיתיות מאחר ותיקון בונפרוני מוצא את האמיתיות.
- המובהקיות שנשארו לאחר תיקון בונפרוני אינן שגיאה.
- המובהקיות שנפלו היו נכונות, אבל מטרת מבחן  $t$  היא למצוא את המובהקות ביותר.
- יש סיכוי שהמובהקיות שנפלו אחרי תיקון בונפרוני הן טעויות מסוג 1 שכן תיקון בונפרוני מעלה את הסיכוי לאלה.
- יש סיכוי שהמובהקיות שנפלו אחרי תיקון בונפרוני הן טעויות מסוג 2 שכן תיקון בונפרוני מעלה את הסיכוי לאלה.

# שאלה 7

נתון כי רק 27 פעמים ההבדל שהתקבל היה קטן או שווה להבדל בנתונים הגולמיים.  
מהו ה-p-value?

תשובה:

## שאלה 8

נתון כי רק 27 פעמים ההבדל שהתקבל היה קטן או שווה להבדל בנתונים הגולמיים.  
מהו ה-p-value?

שימו לב שהמבחן חד צדדי !

$$\frac{27}{1000} = 0.0027$$

תשובה:

# שאלה 10

חוקר ביצע מבחן סטטיסטי המשווה את רמת הביטוי של 2,000 גנים בין שתי קבוצות. מאחר וביצע השערות מרובות, חפץ החוקר ליישם את שיטת תיקון בונפרוני על רמת המובהקות שהתקבלה. עבור ארבעה גנים התקבלו רמות המובהקות הבאות:

רמת מובהקות	גן
0.1	ABI2
0.01	GSX1
0.001	TBXA2R
0.0001	COX3

עבור  $\alpha=0.05$ , ולאחר תיקון בונפרוני, מה מספר הגנים (מתוך ארבעת הגנים הנ"ל) שיהיו מובהקים סטטיסטית?

יש לבחור תשובה אחת:

- 2
- 0
- 1
- 3
- לא ניתן לדעת ללא מידע נוסף על יתר הגנים
- 4

# שאלה 10

חוקר ביצע מבחן סטטיסטי המשווה את רמת הביטוי של 2,000 גנים בין שתי קבוצות. מאחר וביצע השערות מרובות, חפץ החוקר ליישם את שיטת תיקון בונפרוני על רמת המובהקות שהתקבלה. עבור ארבעה גנים התקבלו רמות המובהקות הבאות:

רמת מובהקות	גן
0.1	ABI2
0.01	GSX1
0.001	TBXA2R
0.0001	COX3

$$\alpha_{corrected} = \frac{0.05}{2000} < 0.0001$$

עבור  $\alpha=0.05$ , ולאחר תיקון בונפרוני, מה מספר הגנים (מתוך ארבעת הגנים הנ"ל) שיהיו מובהקים סטטיסטית?

יש לבחור תשובה אחת:

- 2
- 0
- 1
- 3
- לא ניתן לדעת ללא מידע נוסף על יתר הגנים
- 4