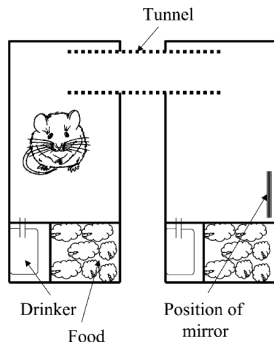


Прикладная статистика. Занятие 3. Непараметрическая проверка гипотез.

22 февраля 2011 г.

Зеркала в клетках мышей

Shervin, Mirrors as potential environmental enrichment for individually housed laboratory mice (2004): 16 лабораторных мышей были помещены в двухкомнатные клетки, в одной из комнат висело зеркало. Измерялось доля времени, которое каждая мышь проводила в каждой из своих двух клеток.



Общая постановка:

H_0 : мышам всё равно, висит в клетке зеркало или нет.

H_1 : у мышей есть какие-то предпочтения насчёт зеркала.

Зеркала в клетках мышей

Уточнённая постановка-1:

H_0 : вероятность того, что мышь предпочтёт комнату с зеркалом, равна $\frac{1}{2}$.

H_1 : вероятность того, что мышь предпочтёт комнату с зеркалом, не равна $\frac{1}{2}$.

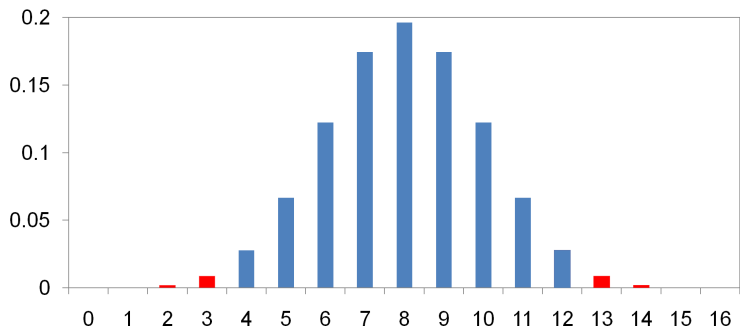
Редуцированные данные: 0 — мышь провела больше времени в комнате с зеркалом, 1 — в комнате без зеркала.

0000000000000000	0100000000000000	1000000000000000	1100000000000000
0000000000000001	0100000000000001	1000000000000001	1100000000000001
0000000000000010	0100000000000010	1000000000000010	1100000000000010
0000000000000011	0100000000000011	1000000000000011	1100000000000011
0000000000000100	0100000000000100	1000000000000100	1100000000000100
0000000000000101	0100000000000101	1000000000000101	1100000000000101
0000000000000110	0100000000000110	1000000000000110	1100000000000110
...

Используем критерий знаков.

Зеркала в клетках мышей

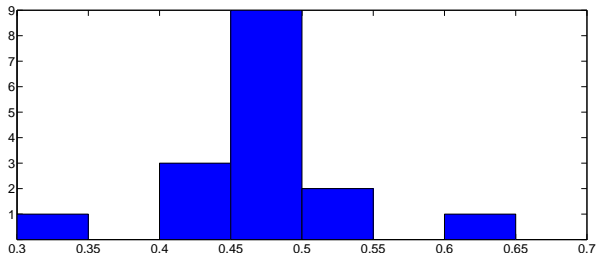
Статистика: t — число единиц в выборке.



13 из 16 мышей провели больше времени в комнате без зеркала.

Критерий знаков: $p = 0.0213$.

Зеркала в клетках мышей



Гистограмма распределения доли времени, проводимого в клетке с зеркалом.

Средняя доля времени, проводимого в клетке с зеркалом — $47.6 \pm 4.7\%$.

Зеркала в клетках мышей

Действительно ли мышей волнуют зеркала?



Как проверить гипотезу, что средняя доля времени, проводимого в клетке с зеркалом, равна 0.5?

Причины использовать критерий знаков

- разности $y_i - x_i$ при H_1 могут быть небольшими по модулю, хотя и имеют систематический характер по знаку (пример с мышами);
- разности $y_i - x_i$ при H_0 могут быть большими по модулю, хотя и случайны по знаку (влияние меди на число личинок комаров);
- точные разности $y_i - x_i$ неизвестны, известны только их знаки (сравнение агрессивности комаров).



Плотность населения охотников-собирателей

Binford, The role of category definition in habitat models (2002): измерена плотность населения (в людях на 100 кв. км) 86 племён охотников-собирателей по всему миру. Медиана распределения плотности населения — 7.38.

13 из 86 племён обитают в лесах Северной Австралии. Можно ли распределение плотности населения отдельно по этим 13 племенам описать той же медианой?



Плотность населения охотников-собирателей

H_0 : распределение плотности населения охотников-собирателей Австралии имеет медиану 7.38.

H_1 : медиана распределения плотности населения охотников-собирателей Австралии не равна 7.38.

Племя	Плотность X_i	$X_i - 7.38$	Знак	$ X_i - 7.38 $	Ранг	Знак*ранг
jeidji	17	9.62	+	9.62	7	7
kuku	50	42.62	+	42.62	12	12
mamu	45	37.62	+	37.62	11	11
ngatjan	59.8	52.42	+	52.42	13	13
undanbi	21.74	14.36	+	14.36	8	8
jiniarra	16	8.62	+	8.62	6	6
ualaria	9	1.62	+	1.62	1	1
barkindji	15.43	8.05	+	8.05	5	5
wongaibon	5.12	-2.26	-	2.26	2	-2
jaralde	40	32.62	+	32.62	10	10
tjapwurong	35	27.62	+	27.62	9	9
tasmanians	13.35	5.97	+	5.97	3	3
badjalang	13.4	6.02	+	6.02	4	4

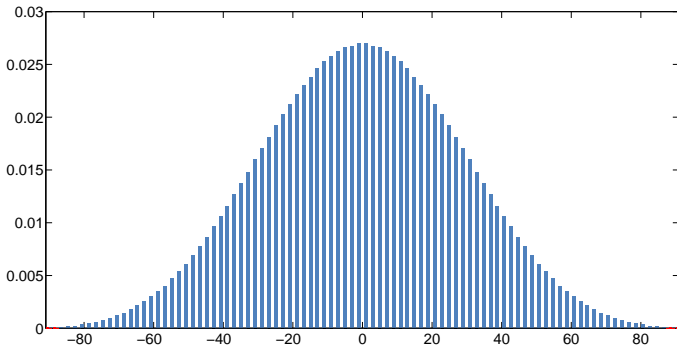
Статистика критерия W — сумма знаковых рангов.
 $W = 87$.

Плотность населения охотников-собирателей

1	2	3	4	5	6	7	8	9	10	11	12	13	W
-	-	-	-	-	-	-	-	-	-	-	-	-	-91
+	-	-	-	-	-	-	-	-	-	-	-	-	-89
-	+	-	-	-	-	-	-	-	-	-	-	-	-87
+	+	-	-	-	-	-	-	-	-	-	-	-	-85
-	-	+	-	-	-	-	-	-	-	-	-	-	-85
+	-	+	-	-	-	-	-	-	-	-	-	-	-83
+	-	-	+	-	-	-	-	-	-	-	-	-	-81
-	+	-	+	-	-	-	-	-	-	-	-	-	-79
+	+	-	+	-	-	-	-	-	-	-	-	-	-77
-	-	+	+	-	-	-	-	-	-	-	-	-	-77
...	
+	+	-	-	+	+	+	+	+	+	+	+	+	77
-	-	+	-	+	+	+	+	+	+	+	+	+	77
+	-	+	-	+	+	+	+	+	+	+	+	+	79
-	+	+	-	+	+	+	+	+	+	+	+	+	81
-	+	-	+	+	+	+	+	+	+	+	+	+	83
+	+	-	+	+	+	+	+	+	+	+	+	+	85
-	-	+	+	+	+	+	+	+	+	+	+	+	85
+	-	+	+	+	+	+	+	+	+	+	+	+	87
-	+	+	+	+	+	+	+	+	+	+	+	+	89
+	+	+	+	+	+	+	+	+	+	+	+	+	91

Плотность населения охотников-собирателей

Распределение возможных значений статистики:



Критерий знаковых рангов: $p = \frac{6}{8192} \approx 0.0007$.

Зеркала в клетках мышей

Уточнённая постановка-2:

H_0 : распределение доли времени, проведённого в клетке с зеркалом имеет медиану 0.5.

H_1 : медиана распределения доли времени, проведённого в клетке с зеркалом, не равна 0.5.

Критерий знаковых рангов: $p = 0.0703$.

Кофеин и респираторный обмен

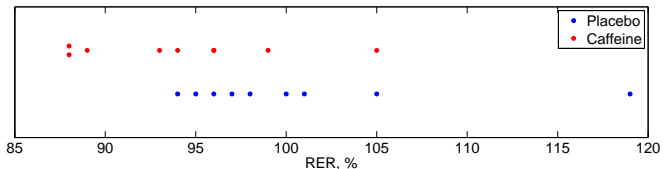
RER (респираторный обмен)— соотношение числа молекул CO_2 и O_2 в выдыхаемом воздухе. Является косвенным признаком того, из жиров или углеводов вырабатывается энергия в момент измерения.

Изучалось влияние кофеина на мышечный метаболизм. В эксперименте принимало участие 18 испытуемых, респираторный обмен которых измерялся в процессе физических упражнений. За час до этого 9 из них получили таблетку кофеина, оставшиеся 9 — плацебо.

Повлиял ли кофеин на значение показателя респираторного обмена?



Кофеин и респираторный обмен



Значение показателя респираторного обмена в двух группах.

H_0 : среднее значение показателя респираторного обмена не отличается в двух группах.

H_1 : среднее значение показателя респираторного обмена отличается в двух группах.

Кофеин и респираторный обмен

Ранг	Наблюдение	Номер наблюдения	Наблюдение	Ранг
16.5	105	1	96	9
18	119	2	99	13
14	100	3	94	5.5
11	97	4	89	3
9	96	5	96	9
15	101	6	93	4
5.5	94	7	88	1.5
7	95	8	105	16.5
12	98	9	88	1.5

Статистика R — сумма рангов в одной из групп.

R: $p = 0.0521$.

Matlab: $p = 0.0468$.

Matlab при исключении наблюдения RER=119: $p = 0.0865$.

Зачем переставлять ранги, если можно переставлять сами наблюдения?

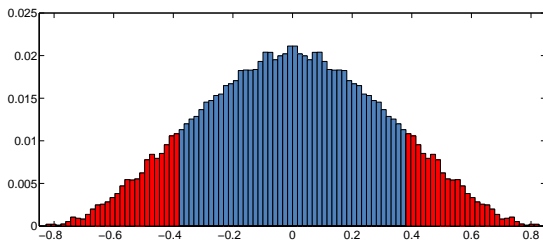
Зеркала в клетках мышей

Уточнённая постановка-3:

H_0 : в клетке с зеркалом мыши проводят в среднем половину времени.

H_1 : в клетке с зеркалом мыши проводят в среднем не половину времени.

Статистика: $T = \sum_i^n (x_i - 0.5)$; $t = -0.3784$.



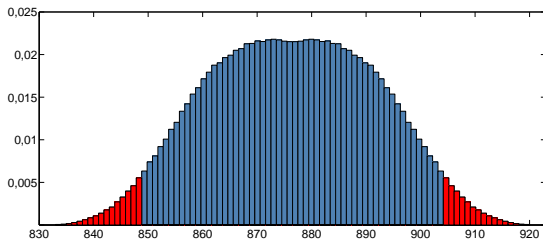
$$p = \frac{\#\{|T| \geq |t|\}}{2^n}.$$
$$p = 0.2292.$$

Кофеин и респираторный обмен

H_0 : среднее значение показателя респираторного обмена не отличается в двух группах.

H_1 : среднее значение показателя респираторного обмена отличается в двух группах.

Статистика: $T = \sum_i^{n_1} x_i$; $t = 905$.



$$p = \frac{\#\{|T - \bar{T}| \geq |t - \bar{T}|\}}{C_{n_1+n_2}^{n_1}}.$$

$$p = 0.0578.$$

Прикладная статистика
Семинар 3. Непараметрическая проверка гипотез.

Рябенко Евгений
riabenko.e@gmail.com