Frequentist Statistics DA 1 report

Made by Ihor Markevych.

Data Set 1 (n=30).

Q1. Do traits X,Y,Z predict the purchasing habits of customers regarding items A-G?

1) For each response identify the appropriate test

Wilcox(paired=FALSE) for combinations of trait-item.

2) Evaluate any statistical assumptions for the test.

a) *independent samples* - data is independent as there is no evidence that for each item ‘yes’ category affects ‘no’ category.

b) *2 categories* - for each item we have two categories 1 and 0, or yes and no.

c) *1 time interval* - we should take trait as an interval/ratio, as traits might (and should) depend on each other. And we use 1 trait for each test.

3) Present your data and results using descriptive statistics and tables as well as graphs if appropriate.

- *Data:*
  - *traitX, traitY, traitZ - numerical data*
  - *ItemA-ItemG - categorical data (0/1)*

	traitX	traitY	traitZ	itemA	itemB	itemC	itemD	itemE	itemF	itemG
median	50.0000000	49.5000000	50.0000000	1.00000000	0.00000000	1.00000000	1.00000000	1.00000000	0.00000000	0.00000000
mean	50.5666667	50.4666667	50.7000000	0.60000000	0.40000000	0.60000000	0.53333333	0.53333333	0.46666667	0.46666667
SE.mean	1.0019330	1.1633615	1.3772787	0.09097177	0.09097177	0.09097177	0.09264111	0.09264111	0.09264111	0.09264111
CI.mean.0.95	2.0491831	2.3793414	2.8168512	0.18605815	0.18605815	0.18605815	0.18947235	0.18947235	0.18947235	0.18947235
var	30.1160920	40.6022989	56.9068966	0.24827586	0.24827586	0.24827586	0.25747126	0.25747126	0.25747126	0.25747126
std.dev	5.4878130	6.3719933	7.5436660	0.49827288	0.49827288	0.49827288	0.50741626	0.50741626	0.50741626	0.50741626
coef.var	0.1085263	0.1262614	0.1487903	0.83045480	1.24568220	0.83045480	0.95140549	0.95140549	1.08732056	1.08732056

- *Results:*

Trait:	Item:	p-value:						
traitX	itemA	0.67029	traitY	itemA	0.4834	traitZ	itemA	0.58015
traitX	itemB	0.89837	traitY	itemB	0.03195	traitZ	itemB	0.27789
traitX	itemC	0.59459	traitY	itemC	0.56636	traitZ	itemC	0.70175
traitX	itemD	0.67589	traitY	itemD	0.10831	traitZ	itemD	0.07238
traitX	itemE	0.28637	traitY	itemE	0.63147	traitZ	itemE	0.98333
traitX	itemF	0.22534	traitY	itemF	0.55926	traitZ	itemF	0.16157
traitX	itemG	0.81813	traitY	itemG	0.0334	traitZ	itemG	0.95002

As we can see, next pairs are significantly dependant:

Trait:	Item:	p-value:
traitY	itemB	0.03195
traitY	itemG	0.03340

4) Analyze your results.

There is sufficient evidence to claim that *trait Y* can predict the purchasing habits of customers regarding *items B and G*.

Q2: Does the purchasing of one of the items (A-G) predict purchasing other items?

1) For each response identify the appropriate test

Fisher exact test for combinations of item - item.

2) Evaluate any statistical assumptions for the test.

a) *independent samples* - data is independent as there is no evidence that measuring traits would change measurements for customer buying habits.

b) *2 variables of categorical data* - we have item and item, 2 binary variables (yes-no category).

3) Present your data and results using descriptive statistics and tables as well as graphs if appropriate.

- *Data (first rows)*

itemA	itemB	itemC	itemD	itemE	itemF	itemG
1	1	1	1	1	1	1
0	1	1	0	0	0	0
1	0	1	0	0	0	0
0	1	1	0	1	1	1
1	0	0	1	1	1	1
0	1	0	1	0	0	1
1	1	1	1	0	1	1
0	0	1	0	0	1	1
0	1	0	1	1	0	1
0	0	0	0	1	1	0

- *Results:*

Item:	Item:	p-value:			
itemA	itemB	0.45817	itemC	itemD	0.28385
itemA	itemC	0.13618	itemC	itemE	1
itemA	itemD	0.72197	itemC	itemF	1
itemA	itemE	1	itemC	itemG	1
itemA	itemF	1	itemD	itemE	0.08126
itemA	itemG	1	itemD	itemF	0.73001
itemB	itemC	0.70859	itemD	itemG	0.08126
itemB	itemD	0.28385	itemE	itemF	1
itemB	itemE	1	itemE	itemG	1
itemB	itemF	0.28385	itemF	itemG	0.46425
itemB	itemG	0.45717			

As we can see, there are no significantly dependant pairs:

4) Analyze your results.

There is NO sufficient evidence to claim that the purchasing of one of the items (A-G) predict purchasing other items.

Data Set 2 (n=1000).

Q1. Do traits X,Y,Z predict the purchasing habits of customers regarding items A-G?

1) For each response identify the appropriate test

Wilcox(paired=FALSE) for combinations of trait-item.

2) Evaluate any statistical assumptions for the test.

a) *independent samples* - data is independent as there is no evidence that for each item ‘yes’ category affects ‘no’ category.

b) *2 categories* - for each item we have two categories 1 and 0, or yes and no.

c) *1 time interval* - we should take trait as an interval/ratio, as traits might (and should) depend on each other. And we use 1 trait for each test.

3) Present your data and results using descriptive statistics and tables as well as graphs if appropriate.

a) Data:

	traitX	traitY	traitZ	itemA	itemB	itemC	itemD	itemE	itemF	itemG
median	50.0000000	50.0000000	50.0000000	1.00000000	0.00000000	0.00000000	0.0000000	1.00000000	0.00000000	0.00000000
mean	49.9630000	50.0770000	49.6660000	0.50300000	0.48400000	0.49000000	0.4860000	0.51300000	0.47700000	0.49800000
SE.mean	0.2322963	0.2278618	0.2185731	0.01581902	0.01581120	0.01581614	0.0158131	0.01581395	0.01580255	0.01581917
CI.mean.0.95	0.4558448	0.4471426	0.4289151	0.03104231	0.03102697	0.03103666	0.0310307	0.03103237	0.03101001	0.03104262
var	53.9615926	51.9209920	47.7742182	0.25024124	0.24999399	0.25015015	0.2500541	0.25008108	0.24972072	0.25024625
std.dev	7.3458555	7.2056222	6.9118896	0.50024118	0.49999399	0.50015013	0.5000541	0.50008107	0.49972064	0.50024619
coef.var	0.1470259	0.1438909	0.1391674	0.99451527	1.03304544	1.02071455	1.0289178	0.97481691	1.04763237	1.00451041

- *Results:*

Trait:	Item:	p-value:
traitX	itemA	0.3543725
traitX	itemB	0.632612
traitX	itemC	2.093716e-09
traitX	itemD	0.8428057
traitX	itemE	7.700772e-14
traitX	itemF	3.204237e-10
traitX	itemG	1.658687e-08
traitY	itemA	0.4454349
traitY	itemB	1.488059e-07
traitY	itemC	0.2990253
traitY	itemD	4.656348e-10
traitY	itemE	2.964637e-07
traitY	itemF	0.003676207
traitY	itemG	8.69643e-12
traitZ	itemA	2.777927e-05
traitZ	itemB	0.9458497
traitZ	itemC	0.4526224
traitZ	itemD	2.349502e-12
traitZ	itemE	0.08283914
traitZ	itemF	1.037367e-08
traitZ	itemG	2.537681e-07

As we can see, next pairs are significantly dependant:

Trait:	Item:	p-value:
traitX	itemC	2.093716e-09
traitX	itemE	7.700772e-14
traitX	itemF	3.204237e-10
traitX	itemG	1.658687e-08
traitY	itemB	1.488059e-07
traitY	itemD	4.656348e-10
traitY	itemE	2.964637e-07
traitY	itemF	0.003676207
traitY	itemG	8.69643e-12
traitZ	itemA	2.777927e-05
traitZ	itemD	2.349502e-12
traitZ	itemF	1.037367e-08
traitZ	itemG	2.537681e-07

4) Analyze your results.

There is sufficient evidence to claim that

- *trait X* can predict the purchasing habits of customers regarding *items C, E, F, G*,
- *trait Y* can predict the purchasing habits of customers regarding *items B, D, E, F, G*,
- *trait Z* can predict the purchasing habits of customers regarding *items A, D, F, G*,

Q2: Does the purchasing of one of the items (A-G) predict purchasing other items?

1) For each response identify the appropriate test

Chi square test for combinations of item - item.

2) Evaluate any statistical assumptions for the test.

a) *independant samples* - data is independant as there is no evidence that measuring traits would change measurments for customer buying habits.

b) *2 variables of categorical data* - we have item and item, 2 binary variables (yes-no category).

c) *no expected frequencies less than 5* - minimum expected frequency for all of the combinations is >200 (>>5).

3) Present your data and results using descriptive statistics and tables as well as graphs if appropriate.

- *Data (first rows)*

itemA	itemB	itemC	itemD	itemE	itemF	itemG
1	0	0	0	1	1	1
1	1	0	0	1	1	0
1	1	1	1	1	0	0
0	1	0	1	0	0	0
1	0	0	0	1	0	0
0	0	0	1	1	0	0
1	0	1	1	1	1	1
0	0	1	0	1	1	1
1	0	1	1	1	1	1
0	0	0	0	0	0	0
0	0	0	0	0	0	1

- *Results:*

Item:	Item:	p-value:			
itemA	itemB	0.0189	itemC	itemD	0.19936
itemA	itemC	0.66065	itemC	itemE	0.01838
itemA	itemD	0.03632	itemC	itemF	0.12018
itemA	itemE	0.16545	itemC	itemG	0.31268
itemA	itemF	0.52095	itemD	itemE	0.83139
itemA	itemG	0.03681	itemD	itemF	0.00091
itemB	itemC	0.24626	itemD	itemG	0.00365
itemB	itemD	0.82209	itemE	itemF	0.00698
itemB	itemE	0.82881	itemE	itemG	0.06606
itemB	itemF	0.3029	itemF	itemG	0.01945
itemB	itemG	0.61554			

As we can see, there next pairs are significantly dependant:

Item:	Item:	p-value:
itemA	itemB	0.01890487
itemA	itemD	0.03632468
itemA	itemG	0.03680537

Contingency tables:

```
      itemB
itemA   0   1
    0 275 222
    1 241 262
```

```
      itemD
itemA   0   1
    0 272 225
    1 242 261
```

```
      itemG
itemA   0   1
    0 266 231
    1 236 267
```

4) Analyze your results.

There is sufficient evidence to claim that

- the purchasing of *item A* predict purchasing *items B, D, G (and other way dependency also)*.