Задача на тестирование BP

Условие задачи

t	y_t	t	y_t	t	y_t	t	y_t
1	235	16	220	31	250	46	310
2	320	17	400	32	355	47	220
3	115	18	275	33	280	48	320
4	355	19	185	34	370	49	215
5	190	20	370	35	250	50	260
6	320	21	255	36	290	51	190
7	275	22	285	37	225	52	295
8	205	23	250	38	270	53	275
9	295	24	300	39	180	54	205
10	240	25	225	40	270	55	265
11	355	26	285	41	240	56	245
12	175	27	250	42	275	57	170
13	285	28	225	43	225	58	175
14	200	29	125	44	285	59	270
15	290	30	295	45	250	60	225

Имеются данные о размерах запасов компании А.

Требуется провести тестирование ряда на постоянство математического ожидания и дисперсии с помощью параметрических тестов на основе:

- 1. - критерия Стьюдента;
- 2. - критерия Фишера;
- 3. критерия Кокрена, основанного на распределении Фишер;
- 4. критерия Бартлетта.
- и непараметрических тестов:
- 5. Манна-Уитни;
- 6. Сиджела – Тьюки;

Критерий Стьюдента

- Для тестирования ряда на постоянство математического ожидания по критерию Стьюдента, разобьем ряд на 2 части, в первую из которых войдут наблюдения с 1 по 35, а во вторую – с 36 по 60.
- Определим оценки математических ожиданий:

$$\overline{y}_1 = \frac{\sum_{t=1}^{35} y_t}{35} = 265,8571 \qquad \overline{y}_2 = \frac{\sum_{t=36}^{60} y_t}{25} = 246$$

Критерий Стьюдента

- Рассчитаем дисперсии:

$$\sigma_1^2 = \frac{\sum_{t=1}^{T_1}\left(y_t - \overline{y}_1\right)^2}{T_1 - 1} = \frac{\sum_{t=1}^{35}\left(y_t - 265,8571\right)^2}{35 - 1} \frac{153674,2857}{34} = 4519,8319$$

$$\sigma_2^2 = \frac{\sum_{t=T_1+1}^{T}\left(y_t - \overline{y}_2\right)^2}{T_2 - 1} = \frac{\sum_{t=36}^{60}\left(y_t - 246\right)^2}{25 - 1} = \frac{42400}{24} = 1766,6667$$

Критерий Стьюдента

$$t=\frac{\left|\overline{y}_1-\overline{y}_2\right|}{\sqrt{\frac{\sigma_1^2}{T_1}+\frac{\sigma_2^2}{T_2}}}=\frac{265,8571-246}{\sqrt{\frac{4519,8319}{35}+\frac{1766,6667}{25}}}=1,4048$$

- Сравнивая с критическим значением

$$t(\alpha,T-2)=t(0,05;58)=2,0017$$

- приходим к выводу, что нельзя отклонить гипотезу, что математическое ожидание постоянно, т.к. $t<t(0,05;58)$

Критерий Фишера

- Проверка гипотезы о постоянстве дисперсии временного ряда в случае разбиения исходного интервала на две части осуществляется с использованием двухстороннего критерия Фишера. Расчетное значение критерия Фишера определяется:

$$F = \frac{\sigma_1^2}{\sigma_2^2}$$

- Для нашего ряда:

$$F = \frac{4519,8319}{1766,6667} = 2,5584$$

- Сравнивая его с табличным значением критерия Фишера с 34 и 24 степенями свободы:

$$F(0,025;34;24) = 2,1797$$

- можно сделать вывод, о том, что гипотеза о постоянстве дисперсии отвергается, так как

$$F > F(0,025;34;24)$$

Критерий Кокрена

- При разбиение ряда на несколько частей для проверки гипотезы о постоянстве дисперсий может быть использован критерий Кокрена, основанный на распределении Фишера. Он применяется в предположении, что объемы этих частей равны между собой. Расчетное значение этого критерия определяется:

$$K=\frac{\sigma^2_{\max}}{\sigma_1^2+\ldots+\sigma_n^2},$$

где

$$\sigma^2_{\max}=\max_i(\sigma_i^2)$$

- А критическое значение критерия рассчитывается по формуле:

$$K(\alpha,n,N-1)=\frac{F\left(\frac{\alpha}{n},\nu_1,\nu_2\right)}{(n-1)+F\left(\frac{\alpha}{n},\nu_1,\nu_2\right)}$$

Критерий Кокрена

- Где $\nu_1 = N - 1$ и $\nu_2 = (n-1)\cdot\nu_1$

Разобьем исходный ряд на 5 равных частей ($N = 12$).

Для каждой из подвыборок рассчитаем дисперсию по формуле:

$$\sigma_i^2 = \frac{\sum_{t=1+T_{i-1}}^{T_i} \left(y_t - \overline{y}_i\right)^2}{N-1}$$

$$\sigma_1^2 = \frac{\sum_{t=1}^{12} \left(y_t - \overline{y}_1\right)^2}{11} = \frac{63766{,}6667}{11} = 5796{,}9697 \qquad \sigma_2^2 = \frac{\sum_{t=13}^{24} \left(y_t - \overline{y}_2\right)^2}{11} = \frac{43456{,}25}{11} = 3950{,}5682$$

$$\sigma_3^2 = \frac{\sum_{t=25}^{36} \left(y_t - \overline{y}_3\right)^2}{11} = \frac{44716{,}6667}{11} = 4065{,}1515 \qquad \sigma_4^2 = \frac{\sum_{t=37}^{48} \left(y_t - \overline{y}_4\right)^2}{11} = \frac{17891{,}6667}{11} = 1626{,}5152$$

Критерий Кокрена

$$\sigma_5^2 = \frac{\sum_{t=49}^{60}\left(y_t - \overline{y}_5\right)^2}{11} = \frac{19225}{11} = 1747{,}7273 \qquad \sigma_{\max}^2 = \max_i(\sigma_i^2) = \sigma_1^2 = 5796{,}9697$$

$$K = \frac{\sigma_{\max}^2}{\sum_{i=1}^{n}\sigma_i^2} = \frac{5796{,}9697}{5796{,}9697 + 3950{,}5682 + 4065{,}1515 + 1626{,}5152 + 1747{,}7273} = 0{,}3373$$

$$K\left(0{,}05;5;11\right) = \frac{F\left(\dfrac{1-0{,}95}{5};11;44\right)}{4 + F\left(\dfrac{1-0{,}95}{5};11;44\right)} = \frac{2{,}6804}{4 + 2{,}6804} = 0{,}4012$$

Поскольку расчетное значение меньше критического значения, то нельзя отвергнуть гипотезу о постоянстве дисперсии.

Критерий Бартлетта

- В нашем примере разобьем ряд на 3 части: первая – с 1 по 20, вторая – с 21 по 40, третья – 41 по 60. Рассчитаем дисперсии для подвыборок:

$$\sigma_1^2 = \frac{\sum_{t=1}^{20}\left(y_t - \overline{y}_1\right)^2}{19} = \frac{107623,75}{19} = 5664,4079 \qquad \sigma_2^2 = \frac{\sum_{t=21}^{40}\left(y_t - \overline{y}_2\right)^2}{19} = \frac{55363,75}{19} = 2913,8816$$

$$\sigma_3^2 = \frac{\sum_{t=41}^{60}\left(y_t - \overline{y}_3\right)^2}{19} = \frac{34513,75}{19} = 1816,5132$$

- Общая дисперсия для всей выборки:

$$\overline{\sigma}^2 = \frac{\sum_{i=1}^{3}\nu_i\sigma_i^2}{\sum_{i=1}^{3}\nu_i} = \frac{107623,75 + 55363,75 + 34513,75}{57} = 3464,9342$$

Критерий Бартлетта

- Т.к. $\nu_1 = \nu_2 = \nu_3 = \nu = 19$, то значение критерия находится по формуле:

$$\lambda = -\frac{1}{c}\sum_{i=1}^{n}\nu_i \ln\frac{\sigma_i^2}{\bar{\sigma}^2} = \frac{1}{c}n\nu\left(\ln\bar{\sigma}^2 - \frac{1}{n}\sum_{i=1}^{n}\ln\sigma_i^2\right)$$

- где
- или $c = 1 + \frac{1}{3(n-1)}\left(\sum_{i=1}^{n}\frac{1}{\nu_i} - \frac{1}{\sum_{i=1}^{n}\nu_i}\right)$ $= 1 + \frac{n+1}{3n\nu}$

Критерий Бартлетта

- получаем, при

$$c = 1 + \frac{3+1}{3 \cdot 3 \cdot 19} = 1{,}0234$$

$$\lambda = \frac{3 \cdot 19}{c}\left(\ln \bar{\sigma}^2 - \frac{1}{3}\sum_{i=1}^{3} \ln \sigma_i^2\right) =$$

$$= \frac{57}{1{,}0234} \cdot \left(\ln 3464{,}9342 - \frac{1}{3}\left(\ln 5664{,}4079 + \ln 2913{,}8816 + \ln 1816{,}5132\right)\right) = 6{,}0798$$

- так как $\lambda = 6{,}0798 < \chi^2\left(\frac{0{,}05}{2}; 2\right) = 7{,}3778$, нельзя отклонить гипотезу о постоянстве дисперсии.

Отсортированные значения ряда	Принадлежность к выборке	Ранги критерия Манна - Уитни	Ранги критерия Сиджела - Тьюки	Отсортированные значения ряда	Принадлежность к выборке	Ранги критерия Манна - Уитни	Ранги критерия Сиджела - Тьюки	Отсортированные значения ряда	Принадлежность к выборке	Ранги критерия Манна - Уитни	Ранги критерия Сиджела - Тьюки
115	1	1	1	235	1	21	41	285	1	41	40
125	1	2	3	240	1	22	43	285	1	42	38
170	2	3	5	240	2	23	45	285	1	43	36
175	1	4	7	245	2	24	47	285	2	44	34
175	2	5	9	250	1	25	49	290	1	45	32
180	2	6	11	250	1	26	51	290	2	46	30
185	1	7	13	250	1	27	53	295	1	47	28
190	1	8	15	250	1	28	55	295	1	48	26
190	2	9	17	250	2	29	57	295	2	49	24
200	1	10	19	255	1	30	59	300	1	50	22
205	1	11	21	260	2	31	60	310	2	51	20
205	2	12	23	265	2	32	58	320	1	52	18
215	2	13	25	270	2	33	56	320	1	53	16
220	1	14	27	270	2	34	54	320	2	54	14
220	2	15	29	270	2	35	52	355	1	55	12
225	1	16	31	275	1	36	50	355	1	56	10
225	1	17	33	275	1	37	48	355	1	57	8
225	2	18	35	275	2	38	46	370	1	58	6
225	2	19	37	275	2	39	44	370	1	59	4
225	2	20	39	280	1	40	42	400	1	60	2

Критерий Манна - Уитни

- Сумма рангов для первой подвыборке равна:

$$R_1 = 1148$$

- Тогда стандартизованная переменная, рассчитанная по формуле:

$$z = \frac{R_1 - \dfrac{T_1 \cdot (T_1 + T_2 + 1)}{2}}{\sqrt{\dfrac{T_1 \cdot T_2 (T_1 + T_2 + 1)}{12}}}$$

- Будет равна:

$$z = \frac{1148 - \dfrac{35 \cdot (35 + 25 + 1)}{2}}{\sqrt{\dfrac{35 \cdot 25 \cdot (35 + 25 + 1)}{12}}} = 1{,}207$$

Критерий Манна - Уитни

- Статистика Манна – Уитни имеет стандартное нормальное распределение.

- Так как ,

$$-1,96 < z = 1,207 < 1,96$$

- то гипотеза о постоянстве математического ожидания принимается.

Критерий Сиджела - Тьюки

- Сумма рангов критерия Сиджела –Тьюки для первой подвыборки равна:

$$R_1 = 959$$

- Тогда стандартизованная переменная, рассчитанная по формуле:

$$z = \frac{R_1 - \frac{T_1 \cdot \left(T_1 + T_2 + 1\right)}{2}}{\sqrt{\frac{T_1 \cdot T_2 \left(T_1 + T_2 + 1\right)}{12}}}$$

- Будет равна:

$$z = \frac{959 - \frac{35 \cdot \left(35 + 25 + 1\right)}{2}}{\sqrt{\frac{35 \cdot 25 \cdot \left(35 + 25 + 1\right)}{12}}} = -1,627$$

Критерий Сиджела - Тьюки

- Статистика Сиджела – Тьюки, так же как и Манна - Уитни, имеет стандартное нормальное распределение.

- И так как ,

$$-1,96 < z = -1,627 < 1,96$$

- то гипотеза о постоянстве дисперсии не отклоняется.