МС-23 Аудиторное задание
Сравнение генеральных средних и генеральных дисперсий
двух нормальных совокупностей.
Гипотеза о равенстве вероятностей успеха в двух сериях испытаний Бернулли.

1. По двум независимым выборкам, извлечённым из нормальных генеральных совокупностей X и Y (см. файл **МС-23_Двувыборочные критерии.xlsx,** лист **«Task 1»**) на уровне значимости 0,05 проверить:
1) нулевую гипотезу $H_0 : \mu_x = \mu_y$ при конкурирующей гипотезе $H_1 : \mu_x \neq \mu_y$ (без каких-либо предположений о равенстве дисперсий);
2) нулевую гипотезу $H_0 : \sigma_x^2 = \sigma_y^2$ против конкурирующей гипотезы $H_1 : \sigma_x^2 \neq \sigma_y^2$.

2. В двух независимых выборках представлены данные о размерах формы сотрудников двух компаний разных городов. При уровне значимости $\alpha = 0{,}05$ проверить нулевую гипотезу $H_0 : p_1 = p_2 = p$ о равенстве долей размера «S» в этих компаниях при конкурирующей гипотезе $H_1 : p_1 \neq p_2$.
(исходные данные в файле **МС-23_Двувыборочные критерии.xlsx**, лист **«Task 2»**)

3. Ряд совместных наблюдений независимых нормально распределенных случайных величин X и Y, описывающих некоторый финансовый показатель двух фирм, задан двумерной выборкой:

{(167.9, -225.541); (133, -227.0618); (172.4, NA); (114.4, -187.947); (182.1, NA); (146.6, -238.1706); (NA, -195.6855); (157.5, -226.3498); (163.1, -232.4315); (164.6, -219.3768); (139.1, -205.4677); (112.9, NA); (149.6, -221.0258); (166.8, -190.341); (153.8, -219.2795); (NA, -198.8605); (87.5, -207.1957); (175.2, NA); (198.5, -277.6407); (147.7, -215.5379); (186, -209.1277); (150.9, -252.0035); (178.7, -221.1615); (143.3, -264.381); (148, -200.406); (NA, -291.3722); (184.8, -209.8789); (151.5, NA); (151.3, NA); (159.8, -261.9098); (124.5, -248.9302); (140, NA); (164.7, NA); (186.4, -255.7522); (154.5, -259.0014); (182.9, -222.0292); (112.8, -209.1327); (132.1, -224.1615); (180.5, -178.7437); (141.6, -261.1121); (157.8, -247.9286); (211, -209.7416); (136.9, -241.0031); (124.6, -276.8816); (109.4, -233.4274); (162.9, -235.5742); (130.8, NA); (187.5, -231.0311); (183.5, -232.3752); (193.9, -188.5517); (165, -257.8477); (184.5, -236.9394); (164.4, -225.4218); (166.1, -216.091); (241.3, -197.7659); (141.8, -219.751); (NA, -207.4731); (NA, -240.3647); (NA, -258.889); (136.6, -217.16); (194.5, -261.1401); (157.6, NA); (149.6, -213.3036); (152.5, -288.5258); (170.4, -241.9711); (NA, -243.0995); (133.6, -232.4539); (139.1, -214.5584); (111.7, NA); (138.1, -271.9439); (166.3, -204.7177); (185.6, NA); (160.4, -229.6342); (152.4, -237.8129); (197.6, -207.0127); (149.8, NA); (180.7, -215.8441); (156.1, -221.4436); (130.5, -286.4889); (140, -235.5511); (NA, -229.0371); (143.1, -257.7442); (177.6, -220.4417); (124.7, -256.3137); (142, -218.7544); (143.6, -260.6194); (121.3, -186.2013); (78.2, -173.376); (155.9, -261.1379); (137.6, -237.259); (170.8, -204.3441); (156.8, -212.3563); (128.4, -200.0559); (NA, -238.497); (129.3, -238.3039); (147.1, -257.0837); (117.9, -205.2149); (174.3, -247.1452); (163.2, -194.3524); (151.5, -219.2332); (153.3, -192.9653); (148.4, -215.8789); (174.8, -205.3518); (84.2, -197.7495); (163.6, -227.4809); (205.5, -250.75); (169.8, -211.6129); (NA, -188.3579); (116.9, NA); (205.5, -180.5642); (181.1, -195.1596); (137.4, -222.561); (140.5, -255.2292); (125, -221.2531); (212.9, -196.9889); (152.7, -200.074); (137.4, NA); (142.8, -201.6862); (178.4, -232.8285); (165.1, -208.838); (NA, -240.8741); (134.3, -224.8478); (180.5, -229.5657); (122.8, -204.9998); (179.7, -272.7181); (163.8, -239.3508); (182.2, -232.8887); (172.8, -220.529); (NA, -221.5642); (NA, -195.5116); (151, -222.4601); (NA, -256.248); (204.2, -230.9828); (182.9, -234.9166); (219.3, -198.5935); (153.4, NA); (85.1, -201.3523); (214.6, -226.9573); (96.2, -245.2855); (153, -261.5914); (112.8, -212.7011); (NA, -244.1466); (NA, -213.4919); (153.3, -239.8558); (177.6, -272.8503); (158.6, -314.0774); (NA, -249.3596); (162.3, -216.9371); (123.8, -197.6739); (158.3, -235.9429)}.

Скопируйте данную выборку и преобразуйте ее в столбцы "A" и "B" соответственно для первой и второй фирмы. При этом связанные значения показателей должны располагаться в одной строке.

Очистите исходную выборку от пропущенных данных, обозначенных как "NA", и вычислите требуемые ниже величины:

1) выборочный коэффициент корреляции Пирсона между X и Y P-value его значимости;

2) **а)** значение P-value в проверке гипотезы о равенстве средних значений показателей фирм при альтернативной гипотезе о том, что среднее значение показателя больше у первой фирмы (без каких-либо предположений о равенстве дисперсий); **б)** на уровне значимости 0.05 можно ли утверждать, что среднее значение показателя больше у первой фирмы? Введите 1 - если да, и 0 - если нет

3) **а)** значение P-value в проверке гипотезы о равенстве дисперсий показателей двух фирм при альтернативной гипотезе об их неравенстве; **б)** на уровне значимости 0.05 можно ли утверждать, что дисперсии показателей фирм различны? Введите 1 - если да, и 0 - если нет

4. По двум независимым выборкам (данные представлены в файле **МС-23_Двувыборочные критерии.xlsx**, лист «**Task 4**») извлечённым из нормальных генеральных совокупностей X и Y, на уровне значимости $\alpha = 0{,}05$ проверяется гипотеза $H_0: E(X) = E(Y)$ против альтернативной гипотезы $H_1: E(X) < E(Y)$. Генеральные дисперсии известны $Var(X) = 2{,}7^2$ и $Var(Y) = 17{,}2^2$. 1) Найдите значение статистики критерия $z = Z_{\text{набл.}}$, критическое множество K_α и на 5% уровне значимости проверить основную гипотезу H_0 против H_1; 2) Найдите P-значение критерия и также сделать выводы; 3) Найдите вероятность ошибки второго рода β для $\Delta = \mu_X - \mu_Y = -0{,}01$ и сделать выводы.
Ответ: 1) $z = -0.745257$; 2) $pv(\vec{z}) = 0.228058$; 3) $\beta = 0.949153$.

5. По выборке объема $m = 32$ найден средний вес изделий, равный 140 г, изготовленных на первом станке; по выборке объема $n = 45$ найден средний вес изделий, равный 135 г, изготовленных на втором станке. Генеральные дисперсии известны: σ_x^2=70 г^2, σ_y^2 =90 г^2. Требуется на уровне значимости 0,05 проверить нулевую гипотезу $H_0: \mu_x = \mu_y$ при конкурирующей гипотезе $H_1: \mu_x \neq \mu_y$. Предполагается, что случайные величины распределены нормально и выборки независимы.

6. По двум независимым выборкам (данные представлены в файле **МС-23_Двувыборочные критерии.xlsx**, лист «**Task 6**»), извлечённым из нормальных генеральных совокупностей X и Y с неизвестными, но равными дисперсиями, проверяется гипотеза $H_0: E(X) = E(Y)$ против альтернативной гипотезы $H_1: E(X) \neq E(Y)$. 1) Найдите значение статистики критерия $t = T_{\text{набл.}} = T(\vec{z})$, критическое множество K_α и на 4% уровне значимости проверить основную гипотезу H_0 против H_1; 2) Найдите P-значение критерия и также сделать выводы.
Ответ: 1) $t = -2.27576$; **2)** $pv(\vec{z}) = 0.0391047$;

7. Пусть $\hat{f}$ – оценка числа степеней свободы f, т.е.

$$\hat{f} = \frac{\left(\frac{s_X^2}{n} + \frac{s_Y^2}{m}\right)^2}{\frac{s_X^4}{n^2(n-1)} + \frac{s_Y^4}{m^2(m-1)}}.$$

Показать, что $\min(n-1; m-1) \leq \hat{f} \leq n+m-2$.

8. По двум независимым выборкам, объемы которых $m = 10$ и $n = 18$, извлеченным из нормальных генеральных совокупностей, найдены исправленные выборочные дисперсии s_x^2=24,02 и $s_y^2 = 17{,}14$. На уровне значимости 0,02 проверить нулевую гипотезу $H_0 : \sigma_x^2 = \sigma_y^2$ против конкурирующей гипотезы $H_1 : \sigma_x^2 > \sigma_y^2$.

Домашнее задание

1. Пусть $\vec{x} = (x_1, \ldots, x_{25})$ – реализация случайной выборки $\vec{X} = (X_1, \ldots, X_{25})$ из нормального распределения $N(\mu_x; 0{,}7^2)$, а $\vec{y} = (y_1, \ldots, y_{30})$ – реализация случайной выборки $\vec{Y} = (Y_1, \ldots, Y_{30})$ из нормального распределения $N(\mu_y; 1{,}4^2)$. Известно, что $\vec{X}$ и $\vec{Y}$ независимы. Проверяется гипотеза $H_0 : \mu_x = \mu_y$ против альтернативной гипотезы $H_1 : \mu_x > \mu_y$. При уровне значимости α применяется критерий с критической областью $\{Z > A\}$, где статистика критерия $Z = Z(\vec{X}, \vec{Y})$ – это нормированная разность $\vec{X} - \vec{Y}$, $A = A_\alpha$ – зависящее от α критическое значение. Соответствующее критическое множество имеет вид $K_\alpha = (A_\alpha; \infty)$. 1) Найдите значение статистики критерия $Z_{\text{набл.}} = Z(\vec{x}, \vec{y})$. 2) Найдите P-значение критерия. 3) Найдите критическое значение A, критическое множество K_α и проверьте гипотезу H_0 при $\alpha = 0{,}02$. 4) Найдите мощность критерия W в случае $\mu_x - \mu_y = 0{,}1$ и $\alpha = 0{,}02$. Исходные данные:

$\vec{x}$ = (3,842; 3,374; 4,18; 4,5; 4,247; 4,412; 3,756; 3,946; 3,729; 3,948; 3,631; 2,992; 4,324; 3,919; 3,059; 4,524; 3,565; 4,236; 4,71; 4,29; 4,998; 3,336; 4,482; 3,721; 3,59);

$\vec{y}$ = (3,19; 3,564; 4,079; 2,369; 5,261; 4,652; 1,849; 6,084; 6,654; 5,65; 3,748; 2,501; 5,476; 3,436; 5,711; 4,292; 5,367; 4,499; 4,989; 4,015; 6,5; 4,178; 4,563; 6,636; 2,113; 2,221; 5,357; 2,358; 6,721; 3,421).

2. В предположении, что $\frac{f \cdot s_w^2}{\sigma_w^2} \approx \chi^2(f)$ (Welch-Satterthwaite,(1938),(1941)), вывести формулу для числа степеней свободы

$$\frac{1}{f} = \frac{1}{\sigma_w^4}\left[\frac{\sigma_X^4}{n^2(n-1)} + \frac{\sigma_Y^4}{m^2(m-1)}\right].$$

Здесь σ_w^4 – несмещённая оценка статистики $s_w^2 = \frac{\sigma_X^4}{n} + \frac{\sigma_Y^4}{m}$.

3. По двум независимым выборкам (данные представлены в файле **MS-23_hw_task_3_XY.csv**), извлечённым из нормальных генеральных совокупностей X и Y**,** проверяется гипотеза $H_0: Var(X) = Var(Y)$ против альтернативной гипотезы $H_1: Var(X) \neq Var(Y)$. 1) Найдите значение статистики критерия $f_{\text{набл.}} = \mathbb{F}_{\text{набл.}} = \mathbb{F}(\vec{z})$, критическое множество K_α и на 2% уровне значимости проверить основную гипотезу H_0 против H_1; 2) Найдите P-значение критерия и также сделать выводы; 3)* Исследовать на несмещённость построенный критерий.

Ответ:
1) $\mathbb{F}_{\text{набл.}} = 0.266623;$
2) $pv(\vec{z}) = 0.0185793;$
3) Критерий с критической областью из п.1):
$K_{0.02} = \{(0; 0.270934) \cup (4.5204482; +\infty)\}$ **- оказался смещённым,**
а критерий с критической областью:
$K'_{0.02} = \{(0; 0.2638721586) \cup (4.389869488; +\infty)\}$ уже **несмещённый**
(проверить, т.е. надо показать, что вероятность ошибки первого рода осталась прежней, т.е. $\alpha = P_{H_0}\big((X_1, \ldots, Y_m) \in K'_{0.02}\big) = 0.02$, а мощность критерия $W(\lambda) \geq \alpha$ для $\forall \lambda = \frac{Var(X)}{Var(Y)} \neq 1)$!!!