Математическая статистика
Вариант 10

Задача. *Служба маркетинга оценивает дилеров фирмы по объему продаж. Сведения об объеме ежедневных продаж товара (в тыс. ден. ед.) некоторым дилером за последние 100 дней приведены для каждого варианта в приложении.*
Требуется:

1. *Построить интервальный вариационный ряд, полигон и гистограмму (на одном рисунке), кумуляту (на другом рисунке).*
2. *Вычислить выборочные характеристики: среднее, дисперсию, среднее квадратичное отклонение, коэффициент вариации, асимметрию, эксцесс, моду медиану.*
3. *Заменив параметры нормального закона распределения их выборочными характеристиками, скорректированными на поправку Шеппарда, рассчитать и построить графики функции плотности и функции распределения нормального закона, «наложив» эти графики соответственно на полигон и кумуляту.*
4. *На 5% уровне значимости проверить гипотезу о нормальном законе распределения объема ежедневных продаж.*
5. *Предположив нормальность распределения объема продаж, построить 95%-ные интервальные оценки математического ожидания, дисперсии и среднего квадратичного отклонения.*
6. *Предположив нормальность распределения объема продаж, на 5%-ном уровне значимости проверить следующие гипотезы:*

а) $H_0: MX = [\overline{x}]$ *при альтернативной гипотезе* $H_1: MX \neq [\overline{x}]$ *(здесь* $[s]$ *- целая часть числа* s *), рассчитать вероятность ошибки второго рода, задавшись альтернативным числовым значением* MX *;*

б) $H_0: DX = [s_X^2]+1$ *при альтернативной гипотезе* $H_1: DX \neq [s_X^2]+1$*, рассчитать вероятность ошибки второго рода, задавшись альтернативным числовым значением* DX *.*

10,3	12,87	12,8	13,4	8,09	6,87	13,08	10,61	11,4	10,79	8,6	12,11	4,06	14,75	7,28	10,35	9,94	7,56	13,96	7,87
10,33	9,79	9,26	7,57	8,09	9,19	9,97	5,99	14,39	6,14	9,21	17,57	9,86	9,51	5,21	5,32	2,53	6,64	12,86	6,18
14,17	13,55	8,78	11,1	16,11	13,75	15,09	6,38	12,9	14,68	12,2	9,99	7,33	9,38	10,22	7,61	8,5	9,86	9,14	9,87
2	12,05	13,38	4,34	11,62	11,24	3,21	8,5	13,23	14,14	4,28	6,44	7,9	7,28	9,59	12,86	9,07	9,64	5,99	5,02
11,64	7,13	13,12	15,07	11,22	10,98	10,8	6,71	8,33	11,34	7,84	7,22	11,19	7,94	6,63	12,36	10,24	12,51	5,56	11,93

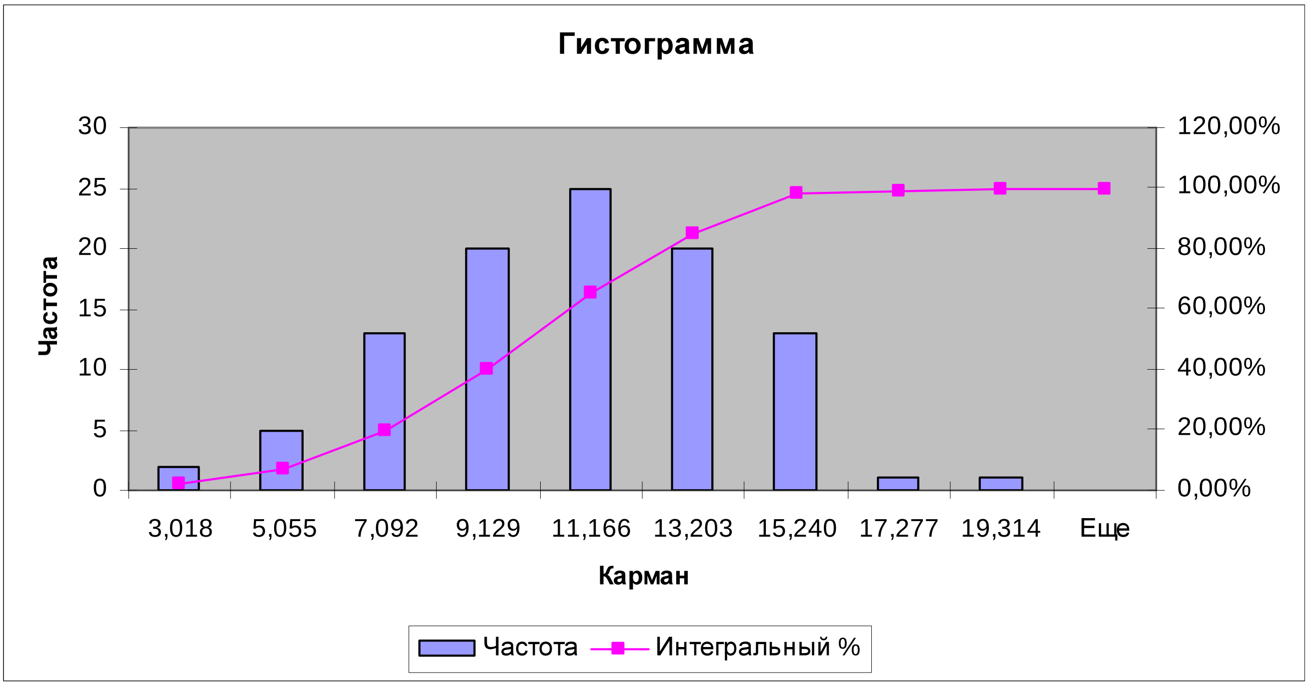
Решение

1. Построим интервальный вариационный ряд. Ширина интервала равна $\Delta = \frac{x_{(\max)} - x_{(\min)}}{1 + 3,322 \lg n}$. Расчеты будем делать в Excel:

Объем продаж за 100 дней	Названия параметров	Значения	
10,3	Объем выборки	100	
10,33	xmin	2	
14,17	xmax	17,57	
2	ширина интервала	2,037	
11,64	левые и правые границы интервалов	0,982	3,018
12,87		3,018	5,055
9,79		5,055	7,092
13,55		7,092	9,129
12,05		9,129	11,166
7,13		11,166	13,203
12,8		13,203	15,240
9,26		15,240	17,277
8,78		17,277	19,314

Для расчета интервальных частот и построения полигона и гистограммы воспользуемся программой «Гистограмма» из надстройки «Анализ данных». Получим следующие данные:

Карман	*Частота*	*Интегральный %*
3,018	2	2,00%
5,055	5	7,00%
7,092	13	20,00%
9,129	20	40,00%
11,166	25	65,00%
13,203	20	85,00%
15,240	13	98,00%
17,277	1	99,00%
19,314	1	100,00%
Еще	0	100,00%

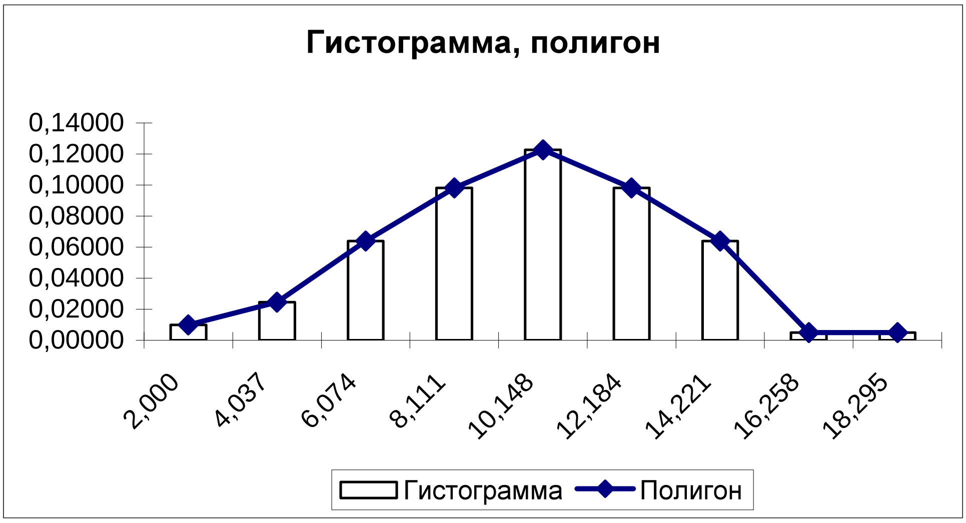


Дополним таблицу, добавив левые границы интервалов $[a_j; a_{j+1}]$, середины интервалов $x_j' = (a_j + a_{j+1})/2$, интервальные частости $\hat{p}_j = m_j / n$ и оценки плотности внутри интервалов $\hat{f}_X(x_j') = \hat{p}_j / \Delta$.

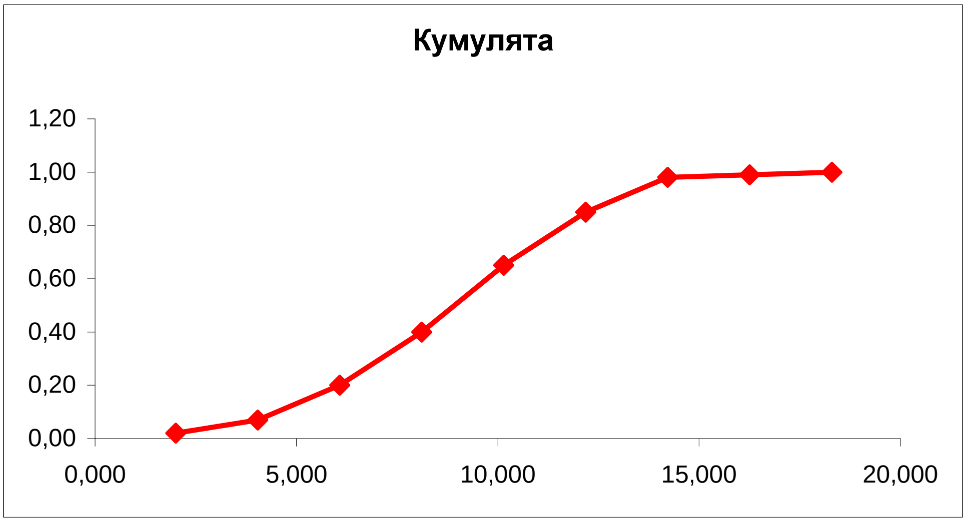
Получим новую таблицу:

Границы интервалов		Середина интервала	Частота	Накопл. относит. частота	Частость	Плотность
Левая	Правая					
0,982	3,018	2,000	2	0,02	0,02	0,00982
3,018	5,055	4,037	5	0,07	0,05	0,02455
5,055	7,092	6,074	13	0,20	0,13	0,06382
7,092	9,129	8,111	20	0,40	0,2	0,09819
9,129	11,166	10,148	25	0,65	0,25	0,12274
11,166	13,203	12,184	20	0,85	0,2	0,09819
13,203	15,240	14,221	13	0,98	0,13	0,06382
15,240	17,277	16,258	1	0,99	0,01	0,00491
17,277	19,314	18,295	1	1,00	0,01	0,00491

Строим гистограмму и полигон частот:



Строим кумуляту:



2. Для вычисления выборочных характеристик воспользуемся программой «Описательная статистика». Получаем:

Объем продаж за 100 дней	
Среднее	9,7445
Стандартная ошибка	0,318212253
Медиана	9,86
Мода	8,09
Стандартное отклонение	3,182122529
Дисперсия выборки	10,12590379
Эксцесс	-0,378933646
Асимметричность	-0,086836489
Интервал	15,57
Минимум	2
Максимум	17,57
Сумма	974,45
Счет	100

Получили следующие выборочные характеристики ежедневного объема продаж:

Выборочное среднее $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 9,7$.

Исправленная выборочная дисперсия $s_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = 10,1$.

Выборочная дисперсия $\hat{\sigma}_X^2 = \frac{n-1}{n} s_X^2 = \frac{99}{100}\cdot 10,1 = 10,02$. С учетом поправки Шеппарда $\tilde{\sigma}_X^2 = \hat{\sigma}_X^2 - \frac{1}{12}\Delta^2 = 9,7$.

Стандартное отклонение $s_X = \sqrt{s_X^2} = 3,18$. Выборочное среднее квадратичное отклонение $\hat{\sigma}_X = \sqrt{\hat{\sigma}_X^2} = 3,17$, с учетом поправки Шеппарда $\tilde{\sigma}_X = \sqrt{\tilde{\sigma}_X^2} = 3,11$.

Исправленная выборочная асимметрия $\tilde{A}_X = \frac{\sqrt{n(n-1)}}{n-2}\hat{A}_X = -0,087$. Выборочная асимметрия: $\hat{A}_X = \frac{n-2}{\sqrt{n(n-1)}}\tilde{A}_X = -0,086$.

Исправленный выборочный эксцесс $\tilde{E}_X = \frac{(n-1)\left[(n+1)\hat{E}_X + 6\right]}{(n-2)(n-3)} = -0,38$. Выборочный эксцесс $\hat{E}_X = \frac{(n-2)(n-3)}{(n-1)(n+1)}\tilde{E}_X - \frac{6}{n+1} = -0,42$.

Выборочный коэффициент вариации $V_X = \frac{s_X}{\bar{x}} = 0,3266 = 32,66\%$.

Выборочную медиану и моду вычислим самостоятельно по интервальному ряду.

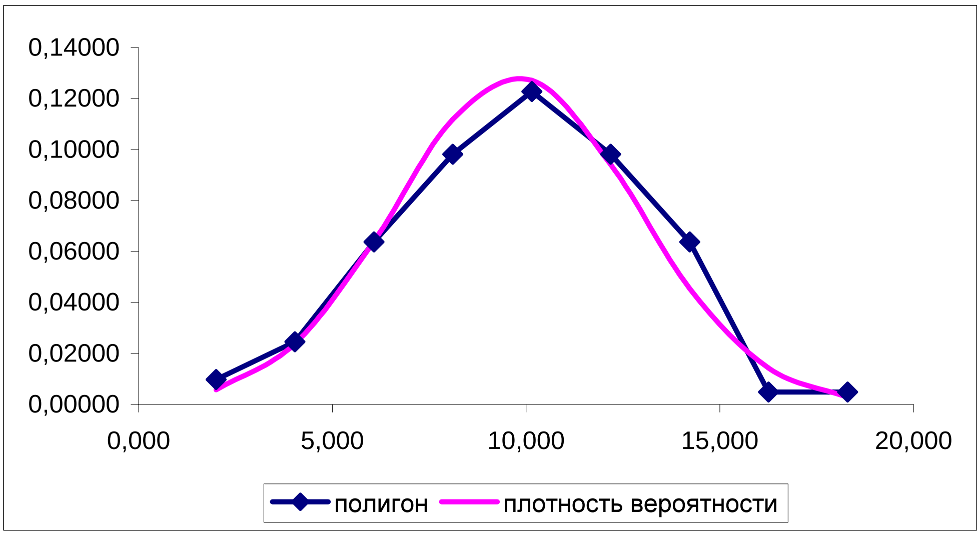
$\hat{x}_{med} = a_l + \Delta\frac{0,5 - \hat{F}(a_l)}{\hat{p}_l}$, где a_l - начало медианного интервала, т.е. такого интервала, что $\hat{F}(a_l) < 0,5$, а $\hat{F}(a_{l+1}) \ge 0,5$. В нашем случае $a_l = 9,129$, поэтому $\hat{x}_{med} = 9,129 + 2,037\frac{0,5-0,4}{0,25} = 9,94$.

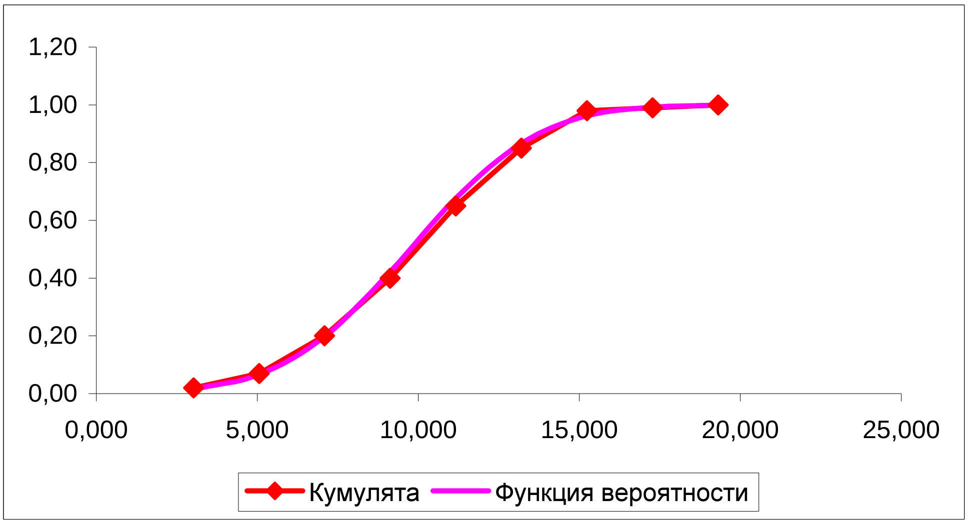
$\hat{x}_{\text{mod}} = a_m + \Delta\frac{\hat{p}_m - \hat{p}_{m-1}}{2\hat{p}_m - \hat{p}_{m-1} - \hat{p}_{m+1}}$, где a_m - начало модального интервала (с наибольшей частотой), $a_m = 9,129$, поэтому $\hat{x}_{\text{mod}} = 9,129 + 2,037\frac{0,25-0,2}{0,5-0,2-0,2} = 10,15$.

3. Заменим параметры нормального закона a и σ их выборочными оценками: $a=\overline{x}=9,74$, $\sigma=\widetilde{\sigma}_X=3,11$ и рассчитаем значения функции плотности нормального закона $f_N(x)=\dfrac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}$ в серединах интервалов и функции распределения $F_N(x)=\dfrac{1}{\sigma\sqrt{2\pi}}\int\limits_{-\infty}^{x}e^{-\frac{(t-a)^2}{2\sigma^2}}dt$ на правых концах интервалов через встроенную функцию НОРМРАСП. Результаты расчетов внесем в основную таблицу.

Границы интервалов		Середина интервала	Частота	Оценка функции распред.	Частость	Оценка плотности вероятности	Плотность распред.	Функция распред.
Левая	Правая							
0,982	3,018	2,000	2	0,02	0,02	0,00982	0,0058	0,0153
3,018	5,055	4,037	5	0,07	0,05	0,02455	0,0239	0,0660
5,055	7,092	6,074	13	0,20	0,13	0,06382	0,0640	0,1973
7,092	9,129	8,111	20	0,40	0,2	0,09819	0,1118	0,4221
9,129	11,166	10,148	25	0,65	0,25	0,12274	0,1272	0,6767
11,166	13,203	12,184	20	0,85	0,2	0,09819	0,0942	0,8672
13,203	15,240	14,221	13	0,98	0,13	0,06382	0,0454	0,9615
15,240	17,277	16,258	1	0,99	0,01	0,00491	0,0143	0,9923
17,277	19,314	18,295	1	1,00	0,01	0,00491	0,0029	0,9990

Достроим графики функций.





4. Для проверки гипотезы о нормальном законе распределения воспользуемся критерием χ^2. Вычислим значения интервальных частот для нормального закона

$np_j = n\left[F_N(a_{j+1}) - F_N(a_j)\right]$,

предварительно приняв $F_N(a_1) = 0$ и $F_N(a_{v+1}) = 1$. Объединим те интервалы, в которых $np_j \le 5$ (в нашем случае это первые два и последние три интервала), при этом соответствующие выборочные интервальные частоты m_j (и теоретические частоты np_j) складываются. Затем в каждом из интервалов (с учетом объединения) вычислим значение величины: $\dfrac{(np_j - m_j)^2}{np_j}$ и просуммируем эти значения по интервалам. Получим выборочное (наблюдаемое) значение статистики $\chi^2_{v^*-l-1} = \sum_{j=1}^{v^*} \dfrac{(np_j - m_j)^2}{np_j} = 0,58$.

Часть расчетов приведена ниже (полная версия в Excel):

Функция распределения	p_j	np_j	После объединения интервалов		$\dfrac{(np_j - m_j)^2}{np_j}$
			m_j	np_j	
0	0,0153	1,53			
0,0153	0,0507	5,07	7	6,60	0,02434
0,0660	0,1313	13,13	13	13,13	0,001265
0,1973	0,2249	22,49	20	22,49	0,274806
0,4221	0,2546	25,46	25	25,46	0,008215
0,6767	0,1905	19,05	20	19,05	0,047016
0,8672	0,0943	9,43			
0,9615	0,0308	3,08			
0,9923	0,0077	0,77	15	13,28	0,224085

0,579727

Здесь v^* - число интервалов после их объединения, $v^* = 6$, $l = 2$ - число параметров нормального распределения, точные значения которых неизвестны. Вычисляем критическую точку по уровню значимости $\alpha = 0,05 = 5\%$ и числу степеней свободы $k = 6-2-1=3$, $\chi^2_{0,05;3} = 7,8$. Так как наблюдаемое значение статистики меньше критической точки, нет оснований отвергнуть гипотезу о нормальном законе распределения объема ежедневных продаж.

5. В предположении нормальности распределения объема продаж интервальная оценка математического ожидания объема продаж задается формулой:

$$P\left\{\overline{x} - t_{1-\gamma;n-1}\frac{s_X}{\sqrt{n}} < MX < \overline{x} + t_{1-\gamma;n-1}\frac{s_X}{\sqrt{n}}\right\} = \gamma.$$

При $\gamma = 0,95$, $n = 100$ критическая точка $t_{1-\gamma;n-1} = 1,98$, $\overline{x} = 9,74$, $s_X = 3,18$. Получаем оценку:

$$9,74 - 1,98\frac{3,18}{10} < MX < 9,74 + 1,98\frac{3,18}{10},$$

$$9,11 < MX < 10,37.$$

Интервальная оценка дисперсии задается формулой:

$$P\left\{\frac{(n-1)s_X^2}{\chi^2_{(1-\gamma)/2;n-1}} < DX < \frac{(n-1)s_X^2}{\chi^2_{(1+\gamma)/2;n-1}}\right\} = \gamma$$

При $\gamma = 0,95$, $n = 100$ критические точки $\chi^2_{(1-\gamma)/2;n-1} = \chi^2_{0,025;99} = 128,42$ и $\chi^2_{(1+\gamma)/2;n-1} = \chi^2_{0,975;99} = 73,36$, и поскольку $s_X^2 = 10,13$, получаем следующую интервальную оценку:

$$\frac{99 \cdot 10,13}{128,42} < DX < \frac{99 \cdot 10,13}{73,36},$$

$$7,81 < DX < 13,67.$$

Тогда оценка среднего квадратичного отклонения:

$$\sqrt{7,81} < \sigma_X < \sqrt{13,67},$$

$$2,79 < \sigma_X < 3,7.$$

6. а) В предположении нормальности распределения объема продаж проверим на 5%-ном уровне значимости справедливость гипотезы $H_0 : MX = 9$ при альтернативной гипотезе $H_1 : MX \neq 9$. Наблюдаемое значение статистики:

$$T_{n-1} = \frac{(\overline{x} - a_0)\sqrt{n}}{s_X} = \frac{(9,74-9)\sqrt{100}}{3,18} = 2,327.$$

При $\alpha = 0,05$ значение критической точки $t_{\alpha;n-1} = t_{0,05;\,99} = 1,98$. Поскольку 2,327>1,98, следует отвергнуть нулевую гипотезу $H_0 : MX = 9$.

Пусть альтернативное значение математического ожидания равно $a_1 = 10$, тогда вероятность ошибки второго рода равна

$$\beta = P\left\{T_{n-1} > \frac{(a_0 - a_1)\sqrt{n}}{s_X} - t_{\alpha;n-1}\right\} - P\left\{T_{n-1} > \frac{(a_0 - a_1)\sqrt{n}}{s_X} + t_{\alpha;n-1}\right\} =$$

$$= P\left\{T_{99} > \frac{(9-10)\sqrt{100}}{3{,}18} - t_{0{,}05;99}\right\} - P\left\{T_{99} > \frac{(9-10)\sqrt{100}}{3{,}18} + t_{0{,}05;99}\right\} =$$

$$= P\{T_{99} > -3{,}14 - 1{,}98\} - P\{T_{99} > -3{,}14 + 1{,}98\} = P\{T_{99} > -5{,}12\} - P\{T_{99} > -1{,}16\} =$$

$$= P\{T_{99} > -5{,}12\} - P\{T_{99} > -1{,}16\} = 0{,}999 - 0{,}876 = 0{,}123.$$

б) Проверим справедливость гипотезы $H_0 : DX = b_0 = 11$ при альтернативной гипотезе $H_1 : DX \neq 11$. Вычислим наблюдаемое значение статистики:

$$\chi^2_{n-1} = \frac{(n-1)s_X^2}{b_0} = \frac{99 \cdot 10{,}13}{11} = 91{,}17$$

При $\alpha = 0{,}05$ значения критических точек $\chi^2_{\alpha/2;n-1} = \chi^2_{0{,}025;99} = 128{,}42$ и $\chi^2_{1-\alpha/2;n-1} = \chi^2_{0{,}975;99} = 73{,}36$. Поскольку значение статистики не попадает в критическую область $91{,}17 \notin (0;73{,}36) \cup (128{,}42;+\infty)$, нет оснований отвергнуть нулевую гипотезу $H_0 : DX = b_0 = 11$.

Пусть альтернативное значение дисперсии равно $b_1 = 12$, тогда вероятность ошибки второго рода равна

$$\beta = P\left\{\chi^2_{n-1} > \frac{b_0 \chi^2_{1-\alpha/2;n-1}}{b_1}\right\} - P\left\{\chi^2_{n-1} > \frac{b_0 \chi^2_{\alpha/2;n-1}}{b_1}\right\} =$$

$$= P\left\{\chi^2_{99} > \frac{11 \cdot 73{,}36}{12}\right\} - P\left\{\chi^2_{99} > \frac{11 \cdot 128{,}42}{12}\right\} =$$

$$= P\{\chi^2_{99} > 67{,}25\} - P\{\chi^2_{99} > 117{,}72\} = 0{,}994 - 0{,}097 = 0{,}897.$$

Чтобы уменьшить вероятности ошибок второго нужно сужать критическую область, то есть уменьшать уровень значимости.