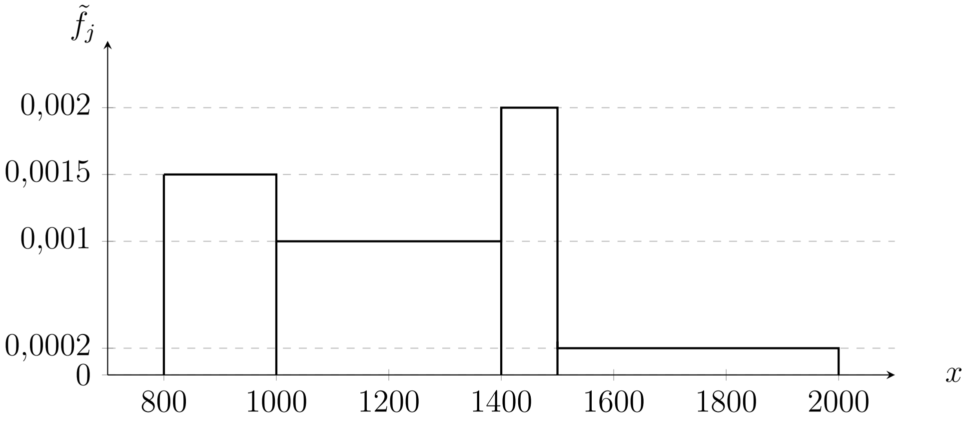
Aufgabe 1

In Berlin wurden 4000 Bewohner mit $50m^2$ Wohnfläche nach der Höhe ihrer Miete (in EUR) befragt. Um eine bessere Übersicht zu erhalten, wurden die Mietbeträge in Klassen eingeteilt. Es ergab sich folgendes Histogramm:



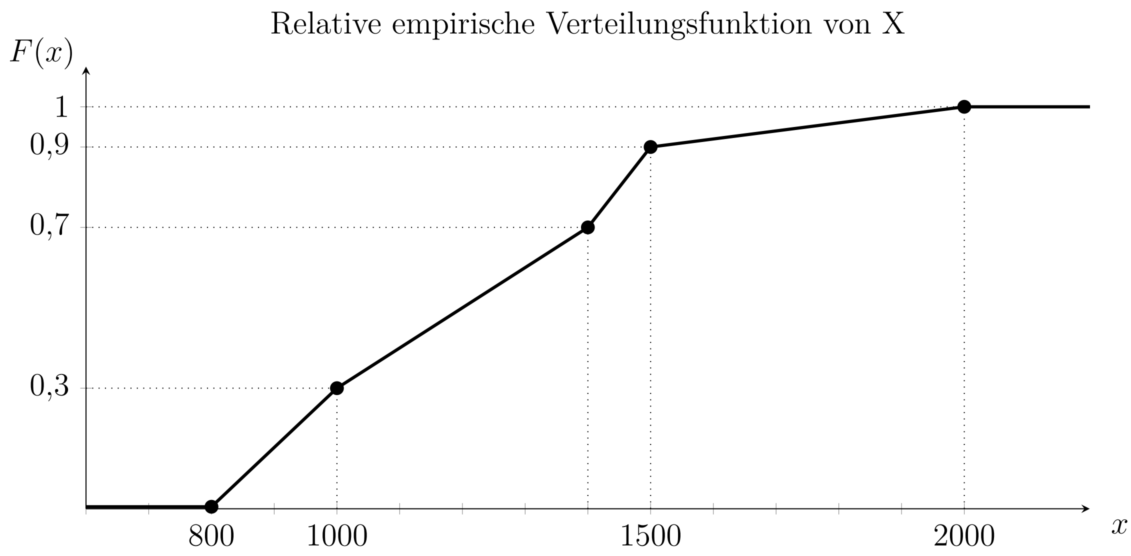
a) Wie lautet das *zugrundeliegende* statistische Merkmal und wie ist es skaliert?

X: Miete einer $50m^2$ großen Wohnung in Berlin (in EUR)
X ist kardinal-stetig skaliert.

b) Vervollständigen Sie die folgende Tabelle.

j	von g_{j-1} bis g_j	f_j	$F(x)$	h_j	b_j	$\tilde{f}_j$	m_j	$m_j \cdot h_j$
1	[800 ; 1000)	0,3	0,3	1200	200	0,0015	900	1080000
2	[1000 ; 1400)	0,4	0,7	1600	400	0,0010	1200	1920000
3	[1400 ; 1500)	0,2	0,9	800	100	0,0020	1450	1160000
4	[1500 ; 2000]	0,1	1,0	400	500	0,0002	1750	700000
$\sum$	-	1,0	-	$n = 4000$	-	-	-	4860000

c) Zeichnen Sie die empirische relative Verteilungsfunktion.



d) Berechnen Sie die Werte der empirischen relativen Verteilungsfunktion für die folgenden Beobachtungen:

i) $x_1 = 1000$ ii) $x_2 = 1300$ iii) $x_3 = 1950$

Wie lässt sich der Wert für $F(1300)$ aus dem Histogramm ablesen? Kennzeichnen Sie den entsprechenden Bereich und tragen Sie den Wert in die kumulierte Verteilungsfunktion ein. Interpretieren Sie das Ergebnis von ii) im Kontext der Aufgabe.

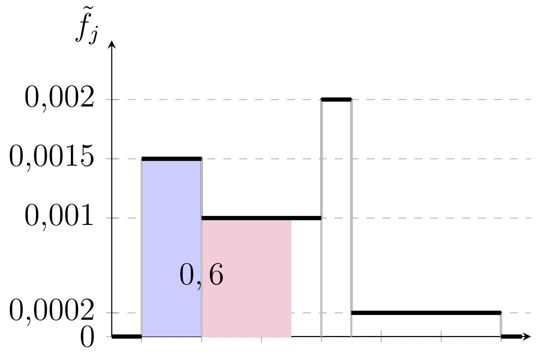
Für $x \in [g_{j-1}, g_j)$ gilt: $F(x) = F(g_{j-1}) + (x - g_{j-1}) \cdot \tilde{f}_j$ mit $f_j = \frac{f_j}{b_j}$, also

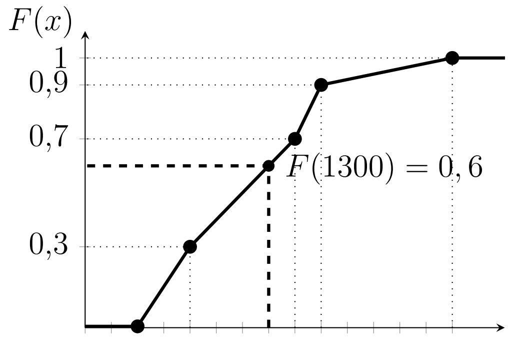
i) $F(1000) = F(1000) + (1000 - 1000) \cdot 0,0010 = 0,3$

ii) $F(1300) = F(1000) + (1300 - 1000) \cdot 0,0010 = 0,3 + 0,3 = 0,6$

iii) $F(1950) = F(1500) + (1950 - 1500) \cdot 0,0002 = 0,9 + 0,09 = 0,99$

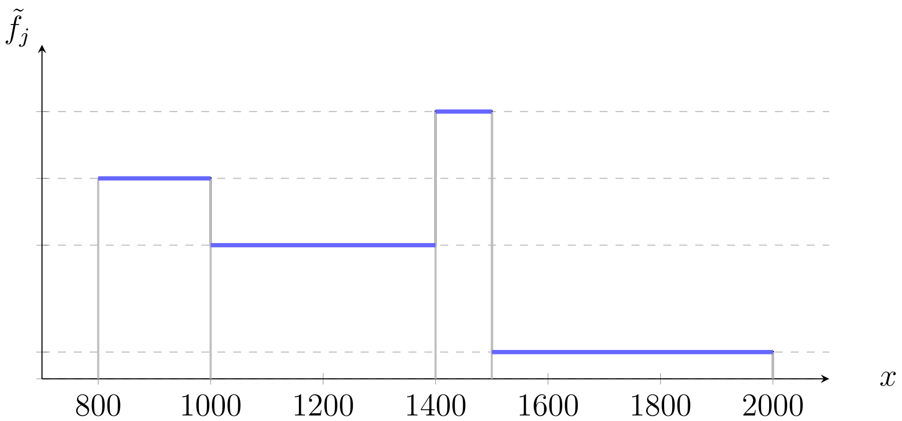
$F(1300)$ stellt die *kumulierte relative Häufigkeit* aller Beobachtungen bis zu $x = 1300$ dar und entspricht der Fläche unter dem Histogramm bis zu diesem Wert. Im Kontext der Aufgabe heißt das, dass 60% der befragten Bewohner eine Miete von höchstens 1300 EUR zahlen.





e) Welche Verteilungsannahme wird durch das Histogramm vorgegeben?

Im Histogramm wird die auf die Gruppenbreite normierte relative Häufigkeit pro Gruppe als konstant dargestellt (waagrechte Linie innerhalb einer Gruppe). Das bedeutet, dass *innerhalb dieser Gruppe eine Gleichverteilung der Beobachtungen* angenommen wird. Diese Annahme wird *für alle Gruppen* getroffen.



f) Berechnen Sie das arithmetische Mittel und interpretieren Sie das Ergebnis im Kontext der Aufgabe.

$$\overline{x} = \frac{1}{n}\sum_{j=1}^{4} m_j \cdot h_j = \frac{1}{4.000} \cdot 4.860.000 = 1.215,$$

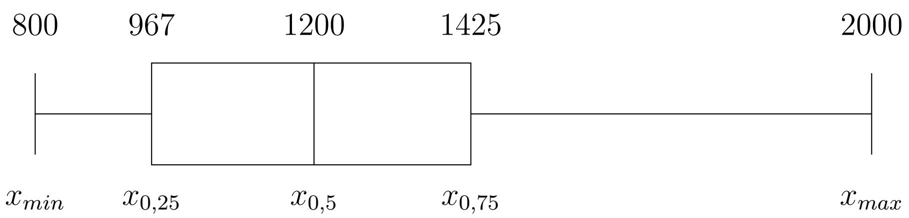
Die durchschnittliche Monatsmiete für eine $50m^2$ Wohnung der 4.000 Befragten in Berlin beträgt 1.250 EUR.

g) Berechnen Sie die empirischen Quartile, welche notwendig für die Erstellung eines Boxplots sind. Zeichen Sie den Boxplot. Runden Sie auf ganze Zahlen.

Für das p-te (empirische) Quartil x_p gilt: $x_p := g_{j-1} + [p - F(g_{j-1})]\frac{1}{\tilde{f}_j}$ mit $\frac{1}{\tilde{f}_j} = \frac{b_j}{f_j}$, also

$$x_{0,25} = 800 + [0,25 - 0] \cdot \frac{1}{0,0015} \approx 967,$$
$$x_{0,5} = 1000 + [0,5 - 0,3] \cdot \frac{1}{0,0010} = 1200,$$
$$x_{0,75} = 1400 + [0,75 - 0,7] \cdot \frac{1}{0,0020} = 1425.$$

Boxplot



h) Berechnen Sie den empirischen Quartilskoeffizienten der Schiefe.

$g_{0,25} := \frac{(x_{0,75}-x_{0,5})-(x_{0,5}-x_{0,25})}{x_{0,75}-x_{0,25}}$, also hier

$g_{0,25} = \frac{(1425-1200)-(1200-967)}{1425-967} = -\frac{8}{458} \approx -0,0175.$

i) Welche Aussage bezüglich der Schiefe der vorliegenden empirischen Verteilung lässt sich aus

   i) dem empirischen Quartilskoeffizienten der Schiefe,

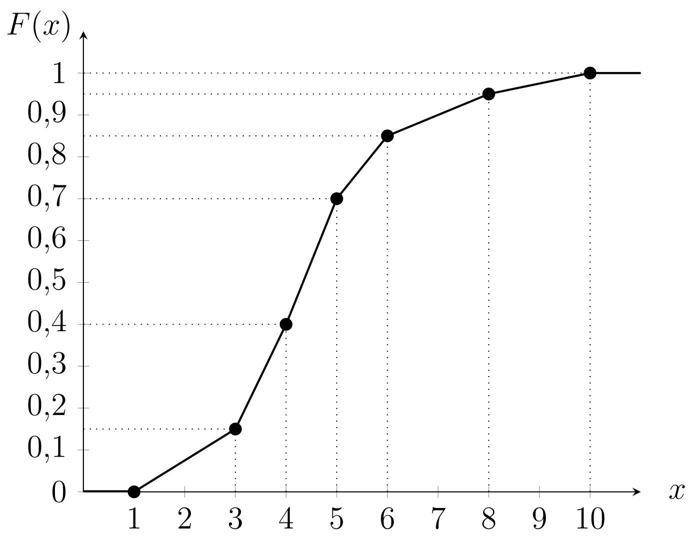
   ii) der Lageregel

   ableiten? Geben Sie kurz eine Erklärung für die unterschiedlichen Ergebnisse.

> i) $g_{0,25} < 0 \Rightarrow$ Die Verteilung ist linksschief/rechtssteil.
>
> ii) $\tilde{x} = 1200 < 1215 = \overline{x} \Rightarrow$ Die Verteilung ist rechtsschief/linkssteil.
>
> Der Quartilskoeffizient der Schiefe berücksichtigt lediglich den *mittleren Datenkörper*, d.h. alle $x \in [x_{0,25}; x_{0,75}]$, während die Lageregel auch Beobachtungen außerhalb berücksichtigt. Dadurch kann es zu unterschiedlichen Ergebnissen kommen.

Aufgabe 2

Für das monatliche Einkommen (in Tsd. EUR) der 2000 Angestellten eines Unternehmens erhielt man folgende empirische relative Verteilungsfunktion $F(x)$.



a) Wie heißt die *zugrundeliegende* statistische Größe und wie ist sie skaliert?

> X: Monatliches Einkommen eines Angestellten eines Unternehmens (in Tsd. EUR)
> X ist kardinal-stetig skaliert.

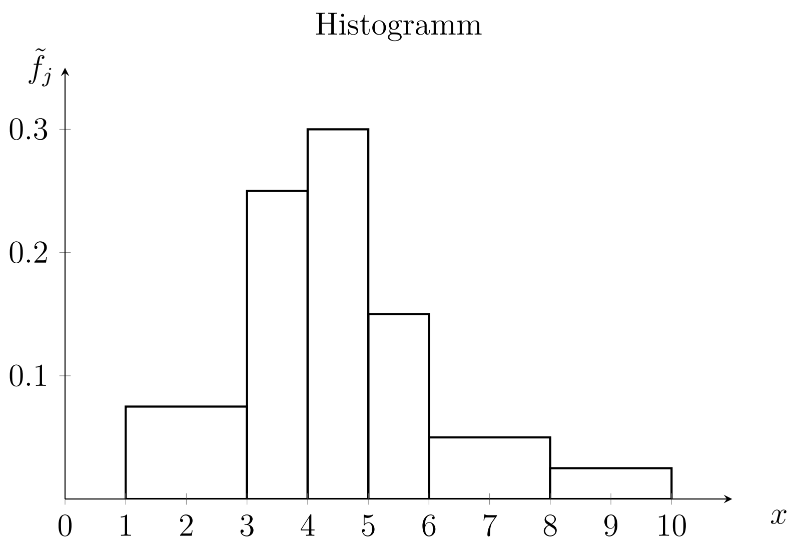
b) Welche Annahme treffen wir, um die empirische relative Verteilungsfunktion zwischen den Gruppengrenzen stückweise linear interpolieren zu dürfen?

Gleichverteilung innerhalb der Gruppen

c) Vervollständigen Sie die folgende Tabelle.

j	von g_{j-1} bis g_j	f_j	$F(x)$	h_j	b_j	$\tilde{f}_j$	m_j	$m_j \cdot h_j$
1	[1 ; 3)	0,15	0,15	300	2	0,075	2	600
2	[3 ; 4)	0,25	0,40	500	1	0,250	3,5	1750
3	[4 ; 5)	0,30	0,70	600	1	0,300	4,5	2700
4	[5 ; 6)	0,15	0,85	300	1	0,150	5,5	1650
5	[6 ; 8)	0,10	0,95	200	2	0,050	7	1400
6	[8; 10]	0,05	1,00	100	2	0,025	9	900
$\sum$	-	1,00	-	$n = 2000$	-	-	-	9000

d) Zeichnen Sie das zur empirischen relativen Verteilungsfunktion gehörende Histogramm.



e) Wie groß ist der relative Anteil der Angestellten, die monatlich

i) genau 6400 EUR ii) weniger als 6400 EUR iii) mehr als 6400 EUR

verdienen?

i) 0, weil X stetig (keine Fläche unter Einpunkt).

ii) $F(6,4) = F(6) + (6,4 - 6) \cdot 0,050 = 0,87$.

iii) $1 - F(6,4) = 1 - 0,87 = 0,13$

f) Lesen Sie das durchschnittliche monatliche Einkommen in diesem Unternehmen nach dem Median approximativ ab.

$x_{0,5} \approx\ 4,3$ (in Tsd. EUR)