

The following frequency distribution will be used in the next three questions. The variable is assumed to be continuous. Show your work.

value	%	% per unit	%-ile of right endpoint
0–6	10	$10/6 \approx 1.67$	10
6–10	10	$10/4 = 2.5$	20
10–13	9	3	29
13–15	8	4	37
15–17	9	4.5	46
17–20	11	$11/3 \approx 3.67$	57
20–25	15	3	72
25–30	10	2	82
30–40	12	1.2	94
40–50	6	.6	100

1. Which bracket contains the median?

SOLUTION: The right endpoint crosses the 50-th percentile in the 17–20 bracket, so that is where the median is.

2. Which bracket contains the mode?

SOLUTION: The highest value in the % per unit column occurs in the 15–17 bracket. The % per unit column gives the height of the histogram, so the mode will be in the 15–17 bracket.

3. Which bracket contains the 65-th percentile?

SOLUTION: The right endpoint crosses the 65-th percentile in the 20–25 bracket, so that bracket will contain the 65-th percentile.

Exam 1 Solutions

The following is data for a scatter diagram. It is used in the next two questions.

	x	y	$x - \bar{x}$	$y - \bar{y}$	SU(x)	SU(y)	SU(x)SU(y)
	1	5	-3	0	$-\frac{3}{3}$	$\frac{0}{2}$	0
	1	8	-3	3	$-\frac{3}{3}$	$\frac{3}{2}$	$-\frac{9}{6}$
	3	3	-1	-2	$-\frac{1}{3}$	$-\frac{2}{2}$	$\frac{2}{6}$
	3	6	-1	1	$-\frac{1}{3}$	$\frac{1}{2}$	$-\frac{1}{6}$
	7	6	3	1	$\frac{3}{3}$	$\frac{1}{2}$	$\frac{3}{6}$
	9	2	5	-3	$\frac{5}{3}$	$-\frac{3}{2}$	$-\frac{15}{6}$
avg	4	5					$\frac{-20}{36} = \frac{-5}{9}$

Here,

$$SD(x) = \sqrt{\frac{(-3)^2 + (-3)^2 + (-1)^2 + (-1)^2 + 3^2 + 5^2}{6}} = 3$$

$$SD(y) = \sqrt{\frac{0^2 + 3^2 + (-2)^2 + 1^2 + 1^2 + (-3)^2}{6}} = 2.$$

4. What is the correlation coefficient for this data?

SOLUTION: From the last column, we see that $r = -5/9$.

5. What is the formula for the SD line in slope-intercept form (i.e., in the form $y = mx + b$).

SOLUTION: The point-slope formula gives

$$\frac{y - 5}{x - 4} = (-1)\frac{2}{3}$$

$$y - 5 = -\frac{2}{3}(x - 4) = -\frac{2}{3}x + \frac{8}{3}$$

$$y = -\frac{2}{3}x + \frac{23}{3}.$$

Suppose given a binormal distribution with

$$\begin{aligned}\bar{x} &= 120 & \bar{y} &= 90 \\ \text{SD}(x) &= 50 & \text{SD}(y) &= 40 & r &= -.85\end{aligned}$$

We study it in the next five questions.

6. What is the regression estimate for $x = 80$?

SOLUTION:

$$\begin{aligned}\text{SU}_y(\text{reg. est.}) &= (-.85)\frac{80 - 120}{50} \\ &= (-.85)(-.8) = .68,\end{aligned}$$

so

$$\text{reg. est.} = 90 + .68 \cdot 40 = 117.2 .$$

7. What is the slope-intercept formula for the regression line ($y = mx + b$)?

SOLUTION:

$$\begin{aligned}\frac{y - 90}{x - 120} &= (-.85)\frac{40}{50} = -.68 \\ y - 90 &= -.68(x - 120) = -.68x + 81.6 \\ y &= -.68x + 171.6\end{aligned}$$

8. Suppose the x -value is in the 80-th percentile. What is the percentile value of its regression estimate?

SOLUTION: We first consult the normal table, and see that the SU value of the 80-th percentile is approximately .84. We now have

$$\begin{aligned}\text{SU}_y(\text{reg. est.}) &= (-.85)\text{SU}_x(80\text{th \%ile}) \\ &\approx (-.85)(.84) \approx -.71 .\end{aligned}$$

We now go back to the normal table, and see that the corresponding percentile value is approximately 23.89%.

9. Suppose $y = 120$. What is the regression estimate for predicting the value of x from that of y ?

SOLUTION:

$$\text{SU}_x(\text{reg. est.}) = (-.85)\text{SU}_y(120) = (-.85)\frac{120 - 90}{40} = -.6375,$$

so

$$\text{reg. est.} = 120 + (-.6375)(50) = 88.125 .$$

Exam 1 Solutions

10. What percent of the x -values are between 100 and 150?

SOLUTION: $SU_x(150) = .6$, which has a percentile value of 72.57% (from the normal table). $SU_x(100) = -.4$, which has a percentile value of 34.46%. Subtracting 34.46% from 72.57%, we see 38.11% of the x -values are between 100 and 150.

Suppose given a binormal distribution with

$$\begin{aligned} \bar{x} &= 200 & \bar{y} &= 150 \\ SD(x) &= 80 & SD(y) &= 50 & r &= .8 \end{aligned}$$

Control for $x = 300$.

11. What is the average after controlling?

SOLUTION: The average after controlling is the regression estimate for $x = 300$:

$$SU_y(\text{reg. est.}) = .8 \frac{300 - 200}{80} = .8(1.25) = 1,$$

so

$$\text{reg. est.} = 150 + 1 \cdot 50 = 200.$$

12. What is the standard deviation after controlling?

$$\text{SOLUTION: } \sqrt{1 - r^2} SD(y) = \sqrt{1 - .64} \cdot 50 = .6 \cdot 50 = 30.$$

13. What is the 40-th percentile of the controlled y -distribution?

SOLUTION: From the normal table, we see that the 40-th percentile of a distribution has an SU value of about $-.25$. In the controlled y -distribution this gives a numerical value of about

$$\text{reg. est.} + (-.25) \cdot SD_{\text{cont}} = 200 + (-.25) \cdot 30 = 192.5 .$$

14. What percent of the controlled y -values are between 140 and 180?

SOLUTION:

$$\begin{aligned} SU_{\text{cont}}(140) &= \frac{140 - 200}{30} = -2 \\ SU_{\text{cont}}(180) &= \frac{180 - 200}{30} \approx -.67 . \end{aligned}$$

Thus, from the normal table, the percentile values of 180 and 140 are approximately 25.24% and 2.28%, respectively. Subtracting, we see that about 22.96% of the distribution is between 140 and 180.

15. What percent of the controlled y -values are above 180?

SOLUTION: Since 25.24% is below 180, 74.76% is above it.