Sample Solutions

Name _____

Date _____

# MAS 170
# Elementary Statistics
# Spring 2020
# Exam 2

**Instructions:**

- Show work! Final answers given without supporting work receive no credit.
- All parts (a), (b), etc, are worth the same amount.
- A calculator is allowed, but no electronic devices with network capability are allowed.
- No books or notes are allowed.

1. Given the $X, Y$ data in the table below, calculate the 5 summary numbers: average and SD for $X$, average and SD for $Y$, and the correlation $r$. Show work! You can check you answers with a calculator, but show exactly what you calculate to find the 5 numbers.

Answers:

AVE($X$) = 2 SD($X$) = $\sqrt{2}$ ≈1.41 AVE($Y$) = 2 SD($Y$) = $\sqrt{2/3}$ ≈.82 $r$ = $\frac{\sqrt{3}}{2}$ ≈.87

| $X$ | $Y$ | devX | devY | (devX)$^2$ | (devY)$^2$ | StdX | StdY | product |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | -1 | -1 | 1 | 1 | $-1/\sqrt{2}$ | $-1/\sqrt{2/3}$ | $\frac{\sqrt{3}}{2}$ |
| 1 | 2 | -1 | 0 | 1 | 0 | $-1/\sqrt{2}$ | 0 | 0 |
| 4 | 3 | +2 | +1 | 4 | 1 | $2/\sqrt{2}$ | $+1/\sqrt{2/3}$ | $\sqrt{3}$ |
| total | 6 | 6 | | 6 | 2 | | | $\frac{3\sqrt{3}}{2}$ |
| $\mu$ | 2 | 2 | | 2 | 2/3 | | | $\sqrt{3}/2 = r$ |

$\mu_X = 2$ $\mu_Y = 2$ $\sigma_X = \sqrt{2}$ $\sigma_Y = \sqrt{2/3}$

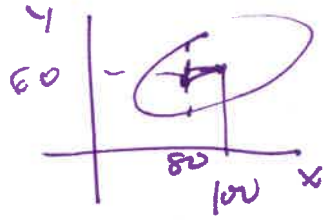
check

$\frac{1}{2} = r \cdot \frac{SD(Y)}{SD(X)}$

$= \frac{\sqrt{3}}{2} \cdot \frac{\sqrt{\frac{2}{3}}}{\sqrt{2}} \checkmark = \frac{1}{2}$

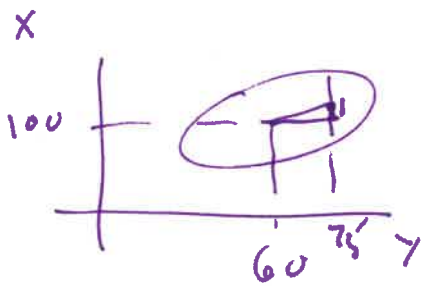
2. Two measurements called $X$ and $Y$ are made on a large population of objects. The list of $X$ measurements has an average of 100 and an SD of 15. The list of $Y$ measurements has an average of 60 and an SD of 10. The two lists have a correlation of 0.8 and the scatter diagram shows linear association.

(a) Use regression to estimate the average $Y$ value for objects whose $X$ value is near 80.



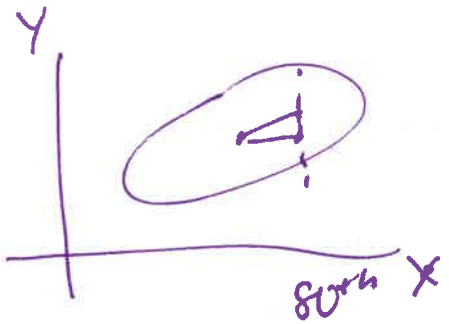

$Y_{est} = \left(\frac{80-100}{15}\right) \cdot (.8) \cdot 10 + 60 \approx$ **49.3** Y units

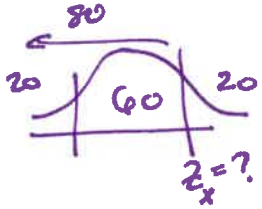
(b) Use regression to estimate the average $X$ value for objects whose $Y$ value is near 75.

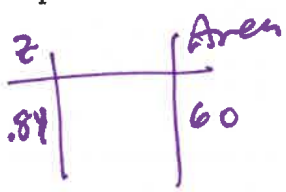



$X_{est} = \left(\frac{75-60}{10}\right)(.8) \cdot 15 + 100 \approx$ **118** X units

(c) Suppose that $X$ and $Y$ are both approximately normal. Use regression to estimate the average percentile rank of the $Y$ value for objects whose $X$ value is 80th percentile.

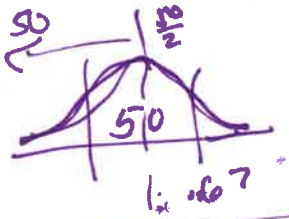


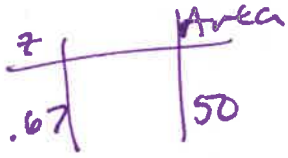




$z_y = r \cdot z_x$
$= (.8)(.84) = .672 \sim .67$

Check.
$(.84)(15) + 100 \sim 112.6$



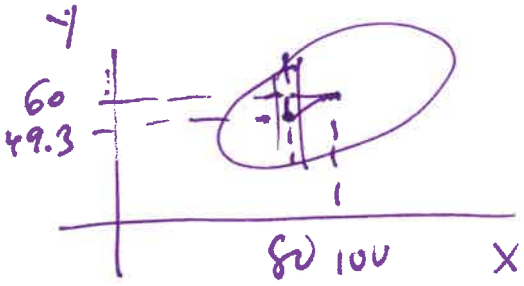




$Y_{est} \approx \left(\frac{112.6 - 100}{15}\right)(.8)(10) + 60$
$\approx 66.72$

$z_Y = \frac{6.72}{10} \approx .67$

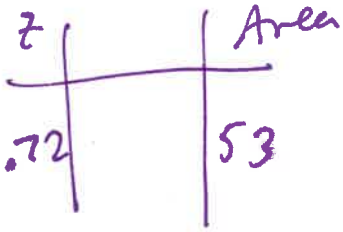
Ans: ~ 75-th percentile

(d) Suppose that $X$ and $Y$ are both approximately normal and that the scatter diagram is homoscedastic. Estimate the percent of $Y$ measurements that are 45 and higher for those objects whose $X$ measurement is near 80.




new ave = 49.3 (part(a))
new SD = $\sqrt{1-(.8)^2}(10) = 6$

+3


$z = \frac{45 - 49.3}{6} \sim -.72$

Ans: $50 + \frac{53}{2} \sim$ 76 / 76.5 / 77 percent

(e) A computer picks hundreds of the objects at random, and spits out the $X$ value for each one. You use the regression line to estimate the $Y$ value that goes with each of the reported $X$ values, while your friend guesses 60 for the $Y$ value every time. Estimate (i) the percent of actual $Y$ values that are within 5 units of your estimated $Y$ values, and (ii) the percent of actual $Y$ values that are within 5 units of your friend's consistent guess of 60.

Answer:

(i) 58% (ii) 38%

rms error regr = $\sqrt{1-(.8)^2} \cdot 10 = 6 \rightarrow z = 5/6 \approx .83$

SD(Y) = $10 \rightarrow z = \frac{5}{10} = .5$

| z | Area |
|---|---|
| .5 | 38 |
| .83 | 58 |

(f) Explain, using one or more complete sentences, how you can be confident, without doing any calculations, that percentage (i) will be higher than percentage (ii) in part (e) above.

The regression line has the lowest r.m.s. error among all possible lines. My friend uses the horizontal line $Y = 60$ for predictions so my guesses are better (lower error).

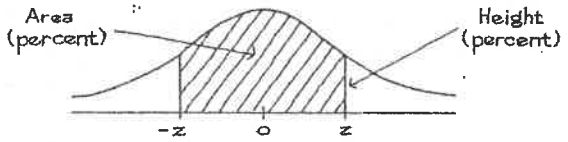
3. Give a definition (using one or more complete sentences) of the term *ecological correlation*, and give an example that illustrates the issue the text warns us of.

Use terms and say facts:

* based on rates or averages for groups of data
* tends to overestimate correlations

Give an example (picture okay).

# Tables

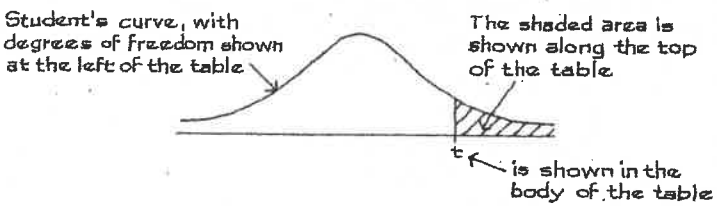



## A NORMAL TABLE

| z | Height | Area | z | Height | Area | z | Height | Area |
|---|---|---|---|---|---|---|---|---|
| 0.00 | 39.89 | 0 | 1.50 | 12.95 | 86.64 | 3.00 | 0.443 | 99.730 |
| 0.05 | 39.84 | 3.99 | 1.55 | 12.00 | 87.89 | 3.05 | 0.381 | 99.771 |
| 0.10 | 39.69 | 7.97 | 1.60 | 11.09 | 89.04 | 3.10 | 0.327 | 99.806 |
| 0.15 | 39.45 | 11.92 | 1.65 | 10.23 | 90.11 | 3.15 | 0.279 | 99.837 |
| 0.20 | 39.10 | 15.85 | 1.70 | 9.40 | 91.09 | 3.20 | 0.238 | 99.863 |
| 0.25 | 38.67 | 19.74 | 1.75 | 8.63 | 91.99 | 3.25 | 0.203 | 99.885 |
| 0.30 | 38.14 | 23.58 | 1.80 | 7.90 | 92.81 | 3.30 | 0.172 | 99.903 |
| 0.35 | 37.52 | 27.37 | 1.85 | 7.21 | 93.57 | 3.35 | 0.146 | 99.919 |
| 0.40 | 36.83 | 31.08 | 1.90 | 6.56 | 94.26 | 3.40 | 0.123 | 99.933 |
| 0.45 | 36.05 | 34.73 | 1.95 | 5.96 | 94.88 | 3.45 | 0.104 | 99.944 |
| 0.50 | 35.21 | 38.29 | 2.00 | 5.40 | 95.45 | 3.50 | 0.087 | 99.953 |
| 0.55 | 34.29 | 41.77 | 2.05 | 4.88 | 95.96 | 3.55 | 0.073 | 99.961 |
| 0.60 | 33.32 | 45.15 | 2.10 | 4.40 | 96.43 | 3.60 | 0.061 | 99.968 |
| 0.65 | 32.30 | 48.43 | 2.15 | 3.96 | 96.84 | 3.65 | 0.051 | 99.974 |
| 0.70 | 31.23 | 51.61 | 2.20 | 3.55 | 97.22 | 3.70 | 0.042 | 99.978 |
| 0.75 | 30.11 | 54.67 | 2.25 | 3.17 | 97.56 | 3.75 | 0.035 | 99.982 |
| 0.80 | 28.97 | 57.63 | 2.30 | 2.83 | 97.86 | 3.80 | 0.029 | 99.986 |
| 0.85 | 27.80 | 60.47 | 2.35 | 2.52 | 98.12 | 3.85 | 0.024 | 99.988 |
| 0.90 | 26.61 | 63.19 | 2.40 | 2.24 | 98.36 | 3.90 | 0.020 | 99.990 |
| 0.95 | 25.41 | 65.79 | 2.45 | 1.98 | 98.57 | 3.95 | 0.016 | 99.992 |
| 1.00 | 24.20 | 68.27 | 2.50 | 1.75 | 98.76 | 4.00 | 0.013 | 99.9937 |
| 1.05 | 22.99 | 70.63 | 2.55 | 1.54 | 98.92 | 4.05 | 0.011 | 99.9949 |
| 1.10 | 21.79 | 72.87 | 2.60 | 1.36 | 99.07 | 4.10 | 0.009 | 99.9959 |
| 1.15 | 20.59 | 74.99 | 2.65 | 1.19 | 99.20 | 4.15 | 0.007 | 99.9967 |
| 1.20 | 19.42 | 76.99 | 2.70 | 1.04 | 99.31 | 4.20 | 0.006 | 99.9973 |
| 1.25 | 18.26 | 78.87 | 2.75 | 0.91 | 99.40 | 4.25 | 0.005 | 99.9979 |
| 1.30 | 17.14 | 80.64 | 2.80 | 0.79 | 99.49 | 4.30 | 0.004 | 99.9983 |
| 1.35 | 16.04 | 82.30 | 2.85 | 0.69 | 99.56 | 4.35 | 0.003 | 99.9986 |
| 1.40 | 14.97 | 83.85 | 2.90 | 0.60 | 99.63 | 4.40 | 0.002 | 99.9989 |
| 1.45 | 13.94 | 85.29 | 2.95 | 0.51 | 99.68 | 4.45 | 0.002 | 99.9991 |



## A *t*-TABLE




| Degrees of freedom | 25% | 10% | 5% | 2.5% | 1% | 0.5% |
|---|---|---|---|---|---|---|
| 1 | 1.00 | 3.08 | 6.31 | 12.71 | 31.82 | 63.66 |
| 2 | 0.82 | 1.89 | 2.92 | 4.30 | 6.96 | 9.92 |
| 3 | 0.76 | 1.64 | 2.35 | 3.18 | 4.54 | 5.84 |
| 4 | 0.74 | 1.53 | 2.13 | 2.78 | 3.75 | 4.60 |
| 5 | 0.73 | 1.48 | 2.02 | 2.57 | 3.36 | 4.03 |
| 6 | 0.72 | 1.44 | 1.94 | 2.45 | 3.14 | 3.71 |
| 7 | 0.71 | 1.41 | 1.89 | 2.36 | 3.00 | 3.50 |
| 8 | 0.71 | 1.40 | 1.86 | 2.31 | 2.90 | 3.36 |
| 9 | 0.70 | 1.38 | 1.83 | 2.26 | 2.82 | 3.25 |
| 10 | 0.70 | 1.37 | 1.81 | 2.23 | 2.76 | 3.17 |
| 11 | 0.70 | 1.36 | 1.80 | 2.20 | 2.72 | 3.11 |
| 12 | 0.70 | 1.36 | 1.78 | 2.18 | 2.68 | 3.05 |
| 13 | 0.69 | 1.35 | 1.77 | 2.16 | 2.65 | 3.01 |
| 14 | 0.69 | 1.35 | 1.76 | 2.14 | 2.62 | 2.98 |
| 15 | 0.69 | 1.34 | 1.75 | 2.13 | 2.60 | 2.95 |
| 16 | 0.69 | 1.34 | 1.75 | 2.12 | 2.58 | 2.92 |
| 17 | 0.69 | 1.33 | 1.74 | 2.11 | 2.57 | 2.90 |
| 18 | 0.69 | 1.33 | 1.73 | 2.10 | 2.55 | 2.88 |
| 19 | 0.69 | 1.33 | 1.73 | 2.09 | 2.54 | 2.86 |
| 20 | 0.69 | 1.33 | 1.72 | 2.09 | 2.53 | 2.85 |
| 21 | 0.69 | 1.32 | 1.72 | 2.08 | 2.52 | 2.83 |
| 22 | 0.69 | 1.32 | 1.72 | 2.07 | 2.51 | 2.82 |
| 23 | 0.69 | 1.32 | 1.71 | 2.07 | 2.50 | 2.81 |
| 24 | 0.68 | 1.32 | 1.71 | 2.06 | 2.49 | 2.80 |
| 25 | 0.68 | 1.32 | 1.71 | 2.06 | 2.49 | 2.79 |