Math 151, Elementary Statistics Sample Exam 2, BPS4e Normal, ch. 3,4 & (part of) 5 Name Solutions

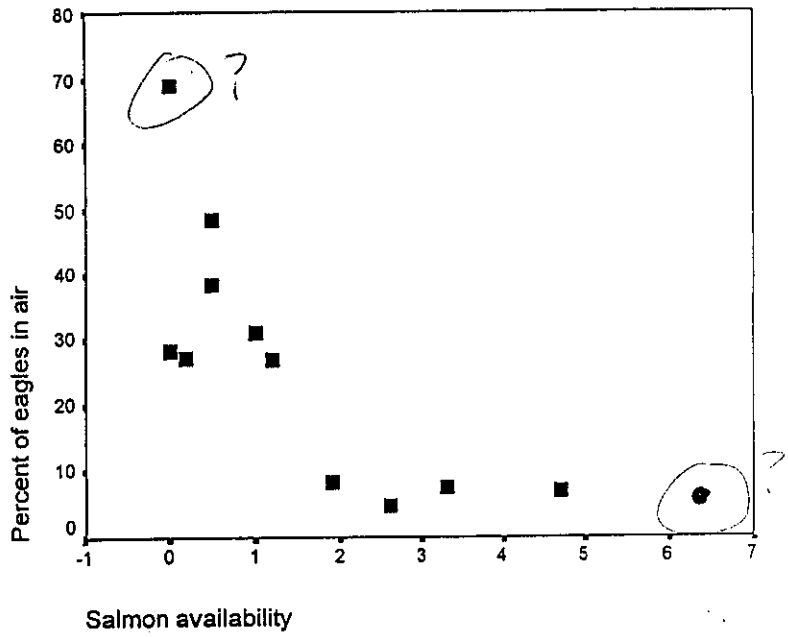
One sheet of notes. Nothing else but simple calculator. Get Normal tables from me.

Space collapsed for #3 on sample exam. "Real" exam will have space for answers. Watch linked version from Day 15, 16, 17 to see if anything changes (unlikely!) in regard to what's covered in Ch. 5 (problem 7).

---

1) In winter, when food is scarce, eagles may have to spend much time and energy in the air searching for prey. This data set explores the relationship between the availability of salmon, a preferred food, and the percent of a population of eagles soaring in the air at that time.

Salmon	Pct in air
.0	28.2
.0	69.0
.2	27.0
.5	38.5
.5	48.4
1.0	31.1
1.2	26.9
1.9	8.2
2.6	4.6
3.3	7.4
4.7	7.0
6.5	6.8



a) The last pair in the table (6.5, 6.8) has been left off the graph. **Add it to the graph** (as well as you can)

b) The "explanatory" variable here is salmon (Absent salmon → birds fly -- in search of other food?)

The "response" variable is % birds flying

c) Look at the Scatterplot:

Is the association **Weak**, **Moderately strong**, **Very strong**?

Is the association **linear**? NO **Other**?(Describe briefly) Drops steeply then flattens out at a low level

Circle outliers, if any. Maybe, maybe not

Is there a clear **direction**? Yes, at least for lower salmon levels - then goes pretty flat

If so, **Positive**, or **Negative**?

d) Describe, *for someone who can't see the scatterplot*, what you see as the relationship between salmon availability and percent of eagles in the air.

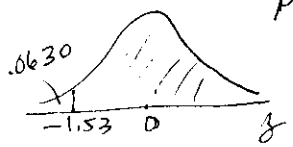
When salmon availability is very low, the birds are in the air a lot - A high of 70% - though there is variability here - 28% at the same level of salmon. Then the % in air drops as salmon increase. When salmon availability reaches about "2", the % of birds in air essentially flattens out at less than 10%. It appears that above about "2", more salmon can't decrease level of birds in the air any more.

2 Show your work for full (or partial) credit.

A) Use Table A to find the proportion of observations in a Standard Normal distribution where:

a) $z > -1.53$

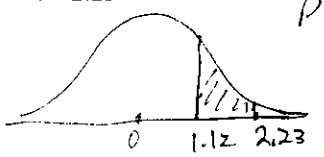
$P(z > -1.53) = 1 - P(z < -1.53) = 1 - .0630 = .9730$

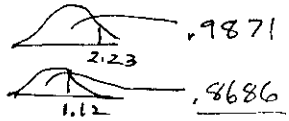


b) $1.12 < z < 2.23$

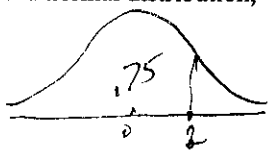
$P(1.12 < z < 2.23) = P(z < 2.23) - P(z < 1.12)$

$= .9871 - .8686 = .1185$





2B) a) In a standard normal distribution, what is the z value which is at the third quartile? = 75th percentile



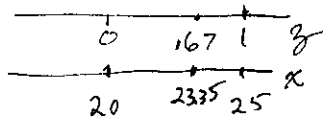
$z = .67$ or $.68$ ($.7486$ to the left of it) ($.7517$ to the left)

b) In a N(20,5) distribution, where the units are inches, the third quartile is 23.35 or 23.4 inches.

$\mu = 20, \sigma = 5$

$x = \text{mean} + z(\text{sd}) = 20 + .67 \times 5 = 20 + 3.35 = 23.35$

(using $z = .68$ gives 23.4)



3) Ann lives in the Plumtown suburb of Metropolis, where the distribution of incomes is approximately normal with mean 65,000 and standard deviation $5000. Ann's income is $60,000. :

Betty lives in the Appleville suburb, where the distribution is also approximately normal, but with mean $40,000 and standard deviation $10,000. Betty's income is $45,000.

Taken from Normal Probability Practice, p. 2. Answers there. Handout or Link from Weblinks or Days 11, 12

a) For each woman, find the z-score of her income (in relation to her suburb).
Who is richer compared to her own neighbors?

b) What proportion of the incomes in Appleville are *higher* than Betty's?

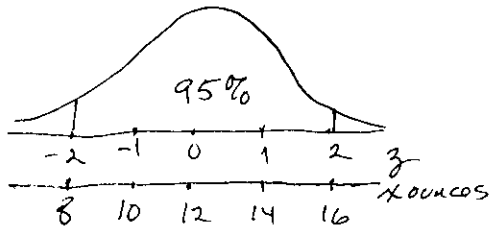
c) What proportion of the incomes in Appleville are between $25,000 and $45,000?

d) Appleville decides to give 20 pounds of apples free to the poorest 20% of its occupants. What is the cutoff income level for getting free apples?

Note these suburbs are very homogeneous, to have normal distributions of income. Most distributions of people's incomes are highly skewed right!

(hint for sample exam: 4, 5 can/should be done without the Normal table)---

4) The Russet potatoes that Farmer Brown harvested last fall had weights that were approximately normally distributed with **mean 12 ounces** and **standard deviation 2 ounces**.



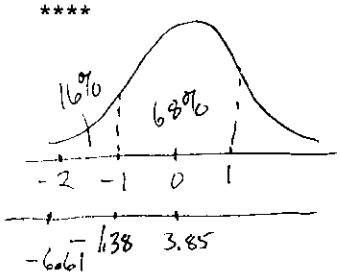
A. **Approximately half** of all his potatoes **weigh less than**
(a) 10 ounces
(b) 12 ounces
(c) 14 ounces
(d) Can't tell, because the median weight is not given

Normal is symmetric so mean = median

B. **About 5%** of his potatoes have weights **outside the range**
(a) 10 ounces to 14 ounces
(b) 8 ounces to 16 ounces – ±2 s.d's. 95% within this range.
(c) 6 ounces to 18 ounces
(d) 6 ounces to 12 ounces

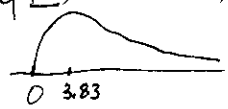
---

5) Suppose data is compiled for the size of purchases from the convenience store (soda, beer, newspapers, popcorn, some groceries) on Aurora St. in downtown Ithaca. Suppose the **mean sale is for $3.85** (on average a person spends $3.85 on a visit) and the **standard deviation is $5.23.**
Why do I know, if these are really the mean and standard deviation, that the data **CAN'T** have a **Normal** distribution? **Explain.**
****



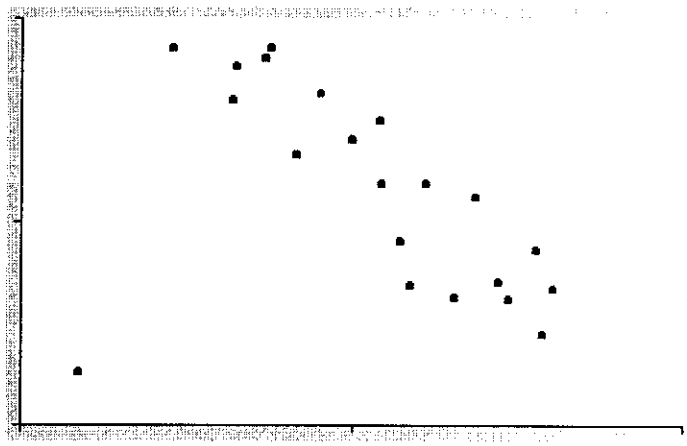
3.85
−5.23 mean
−1.38 −1 s.d.
−6.61 −2 s.d.

If it were normal, about 16% of purchases would be more negative than $−1.38, that is, the sales would involve money being given out from the store, not taken in! Stores sell stuff and get money in, as a rule, so (almost?) all "sales" have to be >0! So the distribution must be More like this:



---

6) **Note the outlier** in the lower left corner of this scatterplot.



Below is a list of potential values for the correlation coefficient r of this data set.

a) Mark with an **A** the value you think is closest to the true r if the outlier is **removed**. A ≈ .9, actually

b) Mark with a **B** the value you think is closest to the true r if the outlier is **included**. B ≈ .5, actually

c) Put an **X** through any values which can **never** be a correlation coefficent for **any** data set. −2, 1.2 it can't be outside the range

d) Circle O any values which cannot possibly be the correlation coefficient for this data set (with or without the outlier.) −1 to +1

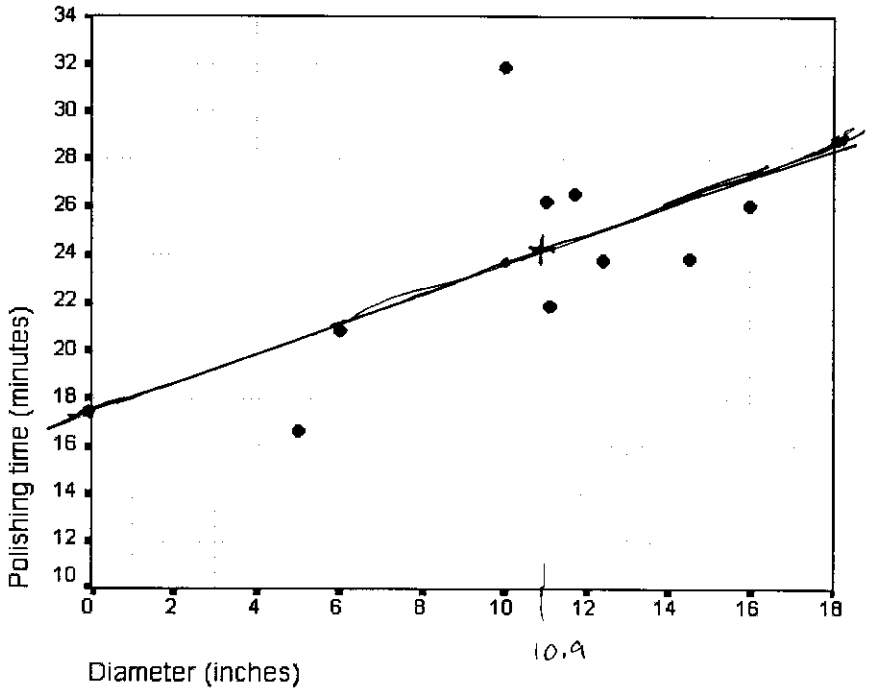
Values for r: ~~X~~ (-1) −.9 (A) −.5 (B) −.2 (0) (.2) (.5) (.9) (1) ~~1.2~~

why? r = 1 or −1 only for data in a perfect straight line!

Even with the outlier, r can't be very positive. You might plausibly think that the outlier would have enough effect to counteract the generally negative relationship that r might be around 0 or even +.2.

7) Nambe Mills, of Santa Fe, NM, makes decorative and functional bowls, casseroles, plates by sandcasting a special metal alloy. After casting, the surfaces must be polished smooth. How long it takes should of course depend to some extent on how big the item is.

Average polishing times for 9 styles of plate are plotted here, against the diameter of the plate.



(a) The correlation coefficient is (one is correct: circle the correct one)

–.522 or (+.522)

positive association

The equation of the least squares regression line of y on x (polishing time on diameter) is

$\hat{y} = 17.411 + .625x$

(b) Graph the line on the scatterplot

x	$\hat{y}$
0	$17.411 + .625 \times 0 = 17.411$
10	$17.411 + .625 \times 10 = 17.411 + 6.25 = 23.661$
18	$17.411 + .625 \times 18 = 17.411 + 11.25 = 28.661$

(c) The mean diameter of the 9 styles is $\bar{x} = 10.9$ inches. This is enough information for you to mark the $(\bar{x}, \bar{y})$ point on the plot. Mark it with a $+$. It's on the line!

(d) How much does the polishing time increase per inch increase in diameter, according to the equation?

.625 minutes (the slope!)

(e) What is the polishing time predicted for a plate 5 inches in diameter? Show your calculation.

$x = 5$

$\hat{y} = 17.411 + .625 \times 5 = 17.411$

$+3.125$

20.536 minutes

(f) What fraction of the variation in polishing times is explained by knowing the diameter and using the regression line?

$R^2 = r^2 = (.522)^2 = .2725 \approx (27\%)$