Stats 7 (Baldi) – Practice Midterm 1

Problem 1

The culinary herb cilantro is very polarizing with some people loving it and others hating it, some saying that it has a soap-like flavor. A large survey of American adults of European ancestry asked whether they think that cilantro tastes soapy. The published results include the following table ("N" stands for number of people, and "SD" stands for standard deviation). There are three variables reported in this table. State what they are and, for each one, state whether it is quantitative or categorical.

	N	Female	Age (SD)
Tastes soapy	1994	0.566	49.0 (15.0)
Doesn't taste soapy	12610	0.489	48.3 (15.2)
Total	14604	0.500	48.4 (15.2)

1. Taste - categorical
2. Sex - categorical
3. Age - quantitative

Problem 2

A recent study reported that "dog owners tend to be healthier than non-dog owners." The conclusion was based on the fact that the percent of individuals in the study suffering from cardiovascular disease was 3.5 percentage points lower among participants who were dog owners than among those who did not own a dog.

(a) Name the response variable assessed in this study: cardiovascular disease

(b) The response variable is (circle one):

(Categorical) Quantitative Ranked (ordinal)

(c) Can we conclude that the cause for the lower rate of heart disease is dog ownership? Explain.

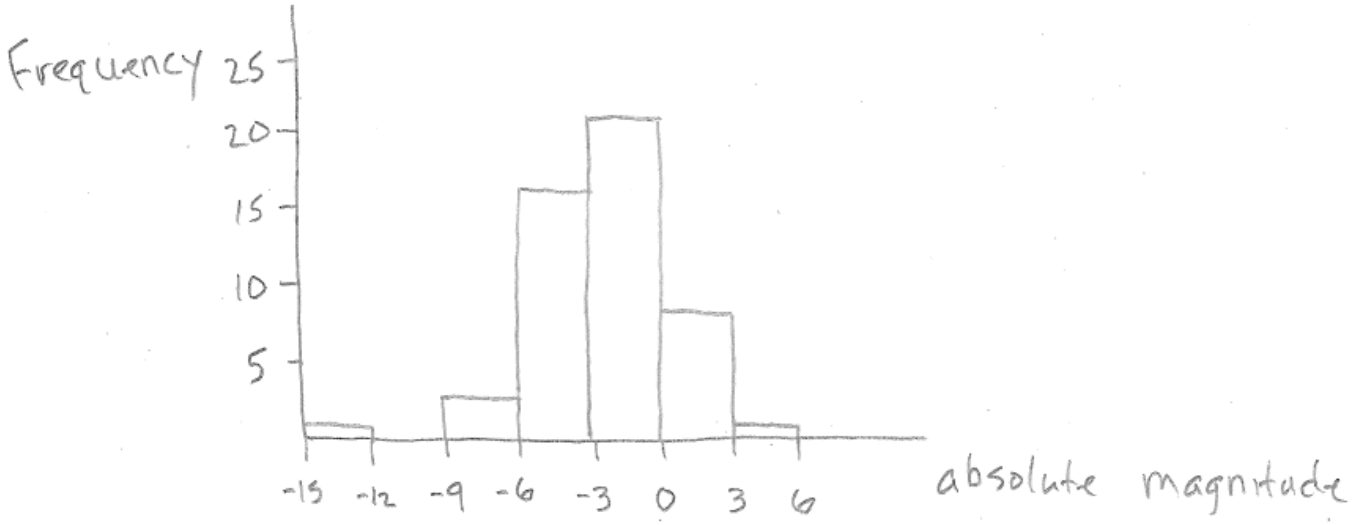
No. This is an observational study.

Problem 3

Absolute magnitudes in astronomy represent how bright stars would appear from 32.6 light years (10 parsecs) away. Negative values correspond to greater brightness. Here are the absolute magnitudes of 50 stars visible from Earth, sorted in ascending order:

Star Name	Abs. Mag.	Star Name	Abs. Mag.	Star Name	Abs. Mag.	Star Name	Abs. Mag.
R136a1	**-12.90**	Acrux	**-4.19**	Elnath	**-1.37**	Menkalinan	**-0.10**
Wezen	**-6.87**	Adhara	**-4.10**	Dubhe	**-1.08**	Koo She	**-0.01**
Rigel	**-6.69**	Mirzam	**-3.95**	Miaplacidus	**-0.99**	Hamal	**0.48**
Alnilam	**-6.38**	Mimosa	**-3.92**	Algieba	**-0.92**	Vega	**0.58**
Canopus	**-5.53**	Polaris	**-3.64**	Alnair	**-0.73**	Castor	**0.59**
Hadar	**-5.42**	Atria	**-3.62**	Aldebaran	**-0.63**	Pollux	**1.09**
Regor	**-5.31**	Spica	**-3.55**	Alkaid	**-0.60**	Sirius	**1.45**
Antares	**-5.28**	Achernar	**-2.77**	Alhena	**-0.60**	Fomalhaut	**1.74**
Alnitak	**-5.26**	Sargas	**-2.75**	Gacrux	**-0.56**	Altair	**2.20**
Betelgeuse	**-5.14**	Bellatrix	**-2.72**	Regulus	**-0.52**	Procyon	**2.68**
Shaula	**-5.05**	Peacock	**-1.81**	Capella	**-0.48**	Rigil Kent.	**4.34**
Avior	**-4.58**	Alphard	**-1.69**	Arcturus	**-0.31**		
Mirfak	**-4.50**	Kaus Aust.	**-1.44**	Alioth	**-0.21**		

(a) In the space below, draw a histogram of these data using classes of size 3 with the following boundaries: -15, -12, -9, -6, -3, 0, 3, 6. Be sure to label your graph correctly.

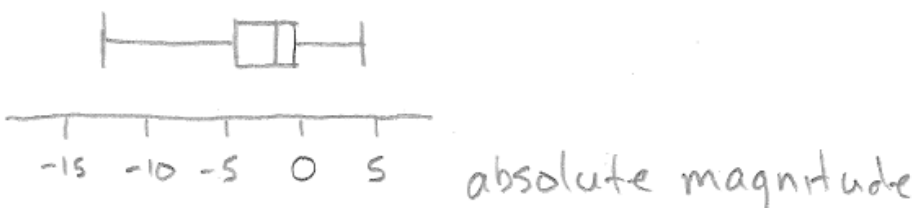


(b) The distribution of absolute magnitudes for these 50 stars is best described as

A) roughly symmetric with an outlier.
B) clearly left skewed.
C) clearly right skewed.
D) clearly bimodal (multiple-peaked).

(c) The median absolute magnitude for these 50 stars is: __**-1.565**___

(d) The first and third quartiles for these 50 stars are, respectively, -4.50 and -0.31. In the space below, draw a boxplot of these absolute magnitudes. Be sure to label your graph correctly.



Problem 4
Researchers studied the effect of regional haze from air pollution on the yields of rice grown in China. For a sample of 9 agricultural areas, they measured the accumulated surface solar irradiance naturally received during a period critical to rice growth and later recorded the relative rice yield for that area (relative to an optimal yield). Here are the data:

Solar irradiance(x)	Relative yield(y)
65	60
84	59
87	75
88	89
90	72
92	83
95	74
99	95
100	92

(a) A scatterplot of the data indicates a linear relationship between solar irradiance and relative yield. Give the equation of the least squares regression line describing relative yield (y) as a function of solar irradiance (x). Round the values of the slope and intercept to 4 decimal places.
y* = –6.9226 + 0.9516*x

(b) What percent of the variation in relative yield can be explained by this regression model?
57.4%

(c) Based on the least squares regression line you wrote above, what would be the predicted relative yield for an area receiving a solar irradiance value of 80? Show your work.
Predicted value = -6.9226 + 0.9516(80) = 69.2054

(d) Would it be appropriate to conclude that solar irradiance causes the variation in rice relative yields? Briefly explain.
No. This is an observational study.

(e) Consider these 9 relative yield data a random sample of agricultural areas in China. Use your calculator to obtain the standard deviation of this sample of 9 relative yields.
s = 13.1 *(this is not sigma 12.4)*

Problem 5
(a) Falling is very dangerous for elderly people. A sociologist studying eldercare and carpeting examines a random sample of 150 hospital records of elderly people who were admitted to the hospital after a fall and a random sample of 150 records of elderly people who were admitted for other reasons. He calls the families and asks whether the house in which the elderly person lives is carpeted. This is an example of

A) a case-control observational study.
B) a cohort observational study.
C) a survey with a probability sample.
D) an experiment with a completely randomized or with a block design.
E) an experiment with a matched pairs or repeated measures design.
F) an unscientific study using a voluntary response sample.

(b) Aging men participated in a study on the effect of a testosterone gel. Men could participate if they were 60 years or older. They were randomly assigned to the testosterone gel or a placebo. This is an example of

A) a case-control observational study.
B) a cohort observational study.
C) a survey with a probability sample.
D) an experiment with a completely randomized or with a block design.
E) an experiment with a matched pairs or repeated measures design.
F) an unscientific study using a voluntary response sample.

(c) Negative attitudes toward poor people are common. Are attitudes more negative when a person is homeless? A study interviewed by phone a random sample of 544 American adults. Based on random assignment, the participants were read one of the following two quotes: "Jim is a 30-year-old single man. He is currently living in a small single-room apartment." or "Jim is a 30-year-old single man. He is currently homeless and lives in a shelter for homeless people." Participants were then asked what they thought of Jim. This is an example of

A) a case-control observational study.
B) a cohort observational study.
C) a survey with a probability sample.
D) an experiment with a completely randomized or with a block design.
E) an experiment with a matched pairs or repeated measures design.
F) an unscientific study using a voluntary response sample.

(d) The Community Intervention Trial for Smoking Cessation asked whether a community-wide advertising campaign would reduce smoking. The researchers located 11 pairs of communities, each pair similar in location, size, economic status, and so on. Based on random assignment, one community in each pair participated in the advertising campaign and the other did not. Tobacco sales were monitored at all locations a month later. . This is an example of

A) a case-control observational study.
B) a cohort observational study.
C) a survey with a probability sample.
D) an experiment with a completely randomized or with a block design.
E) an experiment with a matched pairs or repeated measures design.
F) an unscientific study using a voluntary response sample.

Problem 6
A scatterplot indicates a linear relationship between two quantitative variables, x and y. The equation of the least-squares regression line is

$$y = 12.127 - 4.395x$$

and the regression model if found to explain 81% of the variations in y. The numerical value for the correlation coefficient is $r =$ __ **–0.9**___.