

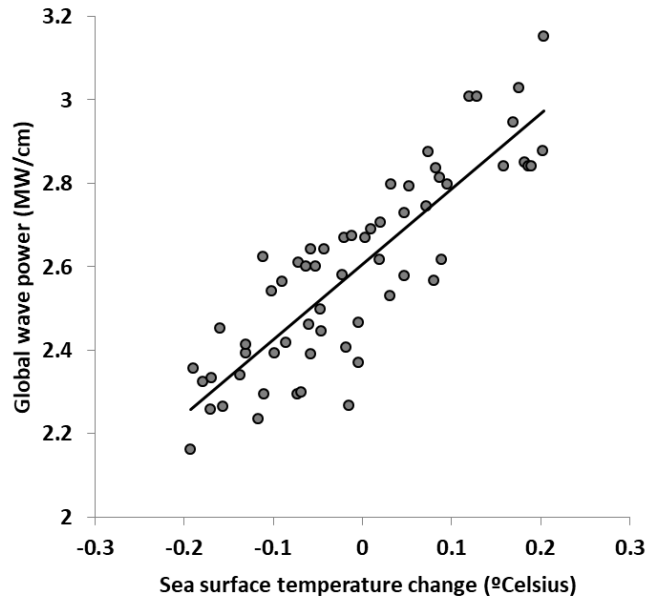
Statistics, winter 2019 (Baldi) – SOLUTIONS

Problem 1 (13 points: 4 + 2 + 3 + 4)

Climate change is noticeable in increasingly warmer oceans globally. A 2019 study examined the relationship between sea surface temperature and how powerful waves are globally, by analyzing worldwide historical records for 1948 to 2017.

“Global wave power” reflected the power of waves all over the globe in a given year, and was measured in megawatts per centimeter (MW/cm). Ocean temperature in a given year was assed as the overall “sea surface temperature change,” measured in degrees Celsius relative to the average sea surface temperature in the late 20th century.

The findings are displayed in this graph. The line represents the least-squares regression line modeling global wave power as a function of sea surface temperature change. The numerical value of the slope is 1.804 and the linear correlation coefficient r is 0.861.



a) Interpret the value of the slope in context.

We expect global wave power to increase by 1.804 MW/cm, on average, for every 1 degree Celsius increase in sea surface temperature change (in a given year).

b) Rounded to one decimal place, what is the value of the y intercept?

[Circle correct answer] [The y intercept is the value of y when $x = 0$]

A) -0.3	B) 0.0
C) 2.1	<u>D) 2.6</u>

c) What percent of the variation in global wave power is accounted for by sea surface temperature change in this regression model? [This is $r^2 = 0.861^2$]

74.1%

d) Circle all appropriate conclusions from this study (there could be more than one).

- A) Global wave power can be expected to exceed 4 MW/cm, on average, in years where the sea surface temperature would be a whole 1 degree Celsius higher than in late 20th century.
- B) Higher sea surface temperatures cause globally more powerful waves, on average.
- C) The association between global wave power and sea surface temperature change is reasonably strong (clearly noticeable).**
- D) On average, globally, more powerful waves tend to be recorded for years with higher sea surface temperatures.**

Problem 2 (8 points: 2 + 3 + 3)

The Nurses' Health Study has been collecting health and behavioral data every two years since 1976 using the same very large, representative sample of female registered nurses in the United States. In 2004, a measure of optimism was included in the questionnaire filled by 70,021 study participants. A 2017 scientific article evaluated the association between optimism and mortality from heart disease in the 8 years following the optimism questionnaire. It reported that the proportion of women dying from heart disease was smaller for the group of women with the highest optimism level than for the group of women with the lowest optimism level.

a) This study is

- A) a cross-sectional sample survey.
- B) a matched pairs/repeated measures experiment.
- C) a longitudinal cohort study.**
- D) a completely randomized experiment.
- E) a case-control observational study.
- F) an unscientific study using voluntary response.
- G) a census (population data).

b) Other than optimism level, what is the variable described in the 2017 study and is it quantitative or categorical? [The individuals are female nurses and we record whether or not they died of heart disease during the 8 years following the optimism questionnaire]

Variable: death from heart disease; quantitative / categorical

c) Can we conclude from the 2017 study that optimism helps prevent death from heart disease?

- A) Yes, because the study has a very large sample size.
- B) Yes, but only for the population of female registered nurses in the United States.
- C) Not necessarily, because preexisting health issues are likely confounded with optimism level.** [The study is observational]
- D) No, because there were also heart disease deaths among women with the highest optimism.

Problem 3 (2 points)

Deep brain stimulation (DBS) is emerging as a potential treatment for mood disorders unresponsive to conventional therapy. A 2018 study examined the effectiveness of orbitofrontal cortex DBS in a sample of 25 subjects suffering from depression who were implanted with intracranial electrodes to treat epileptic seizures. During the procedure, the subject's mood was quantified in real time with a composite mood score (a higher value indicating a more positive mood). The researchers compared the composite mood score for each patient when receiving, in random order, a real DBS (a very low and painless 6 mA current) and a sham DBS (no electric current).

This study is

- A) a cross-sectional sample survey.
- B) a matched pairs/repeated measures experiment.**
- C) a longitudinal cohort study.
- D) a completely randomized experiment.
- E) a case-control observational study.
- F) an unscientific study using voluntary response.
- G) a census (population data).

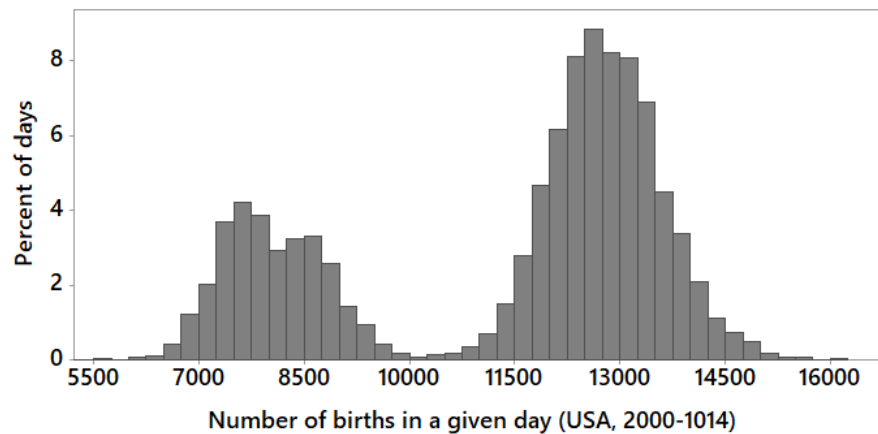
Problem 4 (8 points: 2 + 3 + 3)

The U.S. Social Security Administration keeps a record of all births in the United States. The following is an analysis based on births records for the years 2000 to 2014.

a) The data set is displayed graphically below. Use this graph to fill in the blanks.

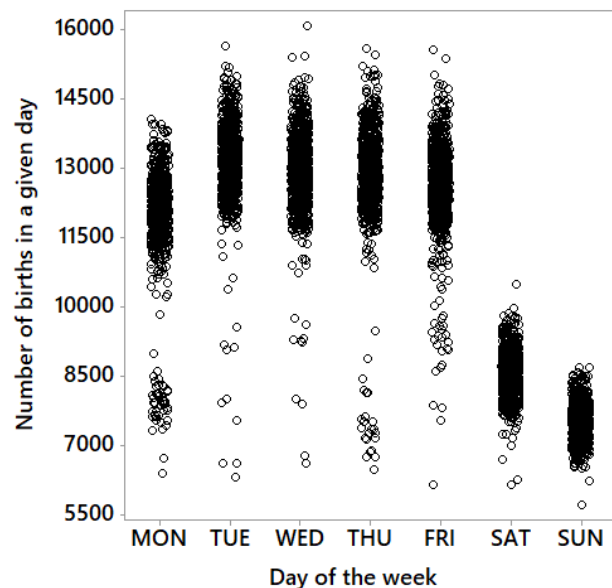
For the 2000 to 2014 time period in the United States, the number of births in a day ranged approximately from **5500** to **16000** births, and the distribution's shape is best described as **bimodal**.

[The range asked for rough values; *minor* differences from those reported here are fine]



b) The data set can be broken down by the day of the week that each birth occurred (Monday through Sunday), as shown here.

Day	Min	Q1	Median	Q3	Max
MON	6409	11616	12087	12506	14053
TUE	6325	12690	13198	13590	15645
WED	6620	12452	12932	13404	16081
THU	6471	12460	12958	13439	15590
FRI	6160	12179	12633	13170	15555
SAT	6159	8255	8574	8862	10499
SUN	5728	7259	7512	7779	8701



Circle all correct interpretations (there could be more than one).

In the United States, over the 2000–2014 period,

- A) The minimum number of births was fairly similar for all seven days of the week, but the maximum varied noticeably across days of the week.
- B) The maximum number of births on weekend days was smaller than the first quartile for any of the non-weekend days. [weekend: sat, sun]
- C) Three-quarters of all Tuesdays had more than 13590 births.
- D) The data set probably has typos or recording errors because the number of births cannot possibly be so much smaller on weekend days than non-weekend days.
- E) There is a clear association between number of births in a day and day of the week.

c) Briefly describe how the second graph helps explain the pattern seen in the first graph.

The histogram has two lumps on either side of roughly 10,000 births. The number of births on weekend days fell exclusively in the left part of the histogram whereas the number of births on weekdays fell mostly in the right part. [The most likely explanation is scheduled rather than natural pregnancies, with scheduling for convenience rather than biological necessity.]

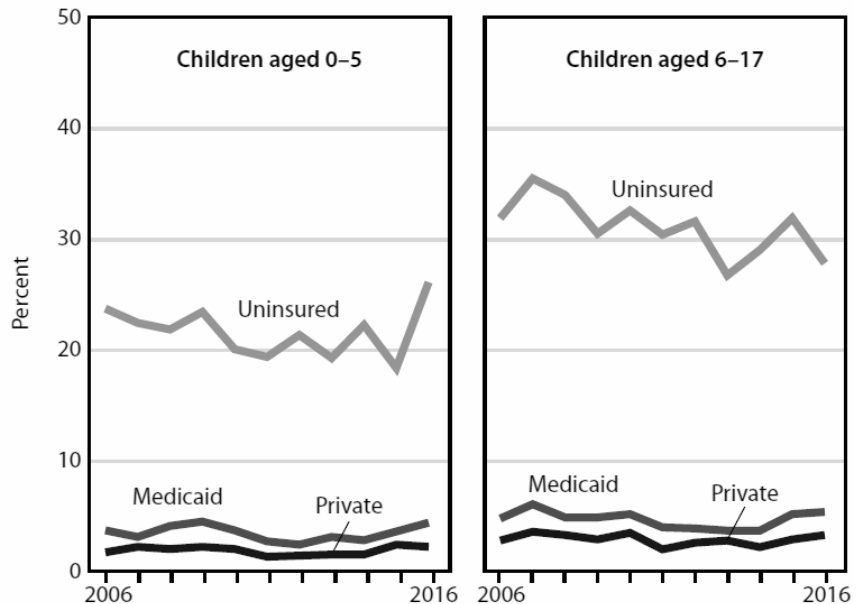
Problem 5 (8 points: 2 + 6)

Children benefit from having a “usual source of care” (such as a designated doctor) for preventive health services and the treatment of medical conditions. So it is important for a nation to identify the children who have no usual source of care.

Figure 15. No usual source of care among children under age 18 years, by type of health insurance coverage: United States, 2006–2016

The National Health Interview Survey, conducted yearly in the United States, collected data on that topic. The findings were published in this figure.

[Note: “Medicaid” is a public health insurance coverage.]



a) Not counting time (in years), how many variables are displayed in these time plots to describe the population of U.S. children? [Circle correct answer]

1	2	<u>3</u>	4	5	6	7	8	9
---	---	----------	---	---	---	---	---	---

[age group; type of health insurance; having no usual source of care]

b) Circle all correct interpretations of these graphs (there could be more than one).

[Note that the “percent” on the vertical axis is the percent of children who have no usual source of care, for each group each year]

Overall, during that period,

- A) there were more uninsured children aged 6–17 than uninsured children aged 0–5.
- B) children aged 6–17 were more likely to be uninsured than children aged 0–5.
- C) uninsured children aged 6–17 were more likely to have no usual source of care than uninsured children aged 0–5.**
- D) in both age groups, uninsured children were more likely to have no usual source of care than children with private or Medicaid health insurance.**
- E) in both age groups, children were more likely to be uninsured than to have private or Medicaid health insurance.
- F) in both age groups, rates of uninsured children showed cyclical variations whereas rates of children with private or Medicaid health insurance did not.

Problem 6 (< 3 min)

A 2018 Beliefs and Values survey by Gallup used a random sample of 1024 American adults to examine American's attitudes about the morality of alcohol and marijuana use. Respondents were asked, "Regardless of whether or not you think that it should be legal, for each one, please tell me whether you personally believe that in general it is morally acceptable or morally wrong."

Here are some findings displayed in the published report:

	Drinking Alcohol		Smoking Marijuana	
	% Acceptable	% Wrong	% Acceptable	% Wrong
U.S. adults	78	19	65	31
Gender				
Men	84	15	70	26
Women	73	23	59	36

List all the variables reported here, and specify whether each variable is quantitative or categorical.

Gender, categorical [Man/Woman]

Opinion on morality of drinking alcohol, categorical [Acceptable/Wrong]

Opinion on morality of smoking marijuana, categorical [Acceptable/Wrong]

END