Sample Exam

September 29, 2018

- You have ∞ minutes to complete this exam.
- The exam packet is **7** pages: **1** cover sheet, **14** questions, **1** codebook, and **0** pages of R code. Make sure you have everything.
- The exam is out of **0** points.
- You are allowed to have your course packet as notes only. No additional notes may be used.
- All phones must be put away during the entire exam. If your phone is out you will be asked to turn in your exam and leave immediately.
- Use the back pages as scratch paper, more can be made available. All scratch paper is to be submitted with your exam. If you want me to review notes on your scratch paper specifically, tell me so and make sure it's findable.
- The exam will not start until all materials are put away.
- Bathroom breaks are allowed, but you must leave all materials at your desk, and your exam with me.
- Show all your work. Partial credit can be awarded.
- Be sure to use summary statistics and full english sentences in your answer where appropriate.

Name: _____

Do not turn this page until instructed to do so.

Data in this sample exam come from the NCbirths data set. The codebook is provided

1. Identify the data type for the variables listed below. If the variable is categorical, state what the levels are.

    - lowbirthweight
    - visits
    - marital
    - gained
    - weight
    - gender

2. What variable(s) would it be appropriate to create a scatterplot of?

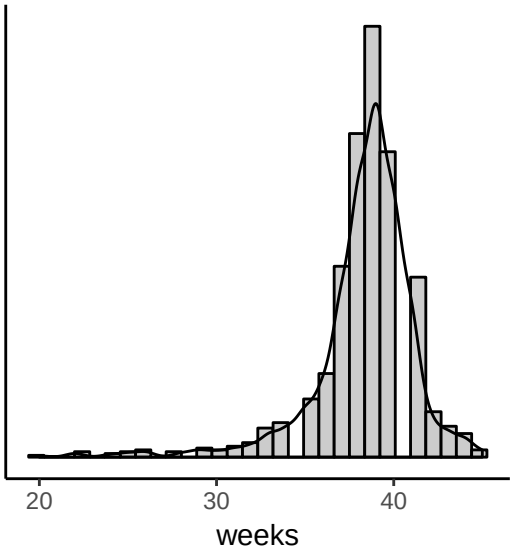
3. What variable(s) would it be appropriate to create a bar chart of?

4. What variable(s) would it be appropriate to create a violin plot of?

5. What variable(s) would it be appropriate to create side by side bar charts of?

6. Describe the distribution of `weeks` in context of the problem. Be sure to use summary statistics and full english sentences in your answer.

mean	median	sd	IQR	var
38.33	39	2.93	3	8.59

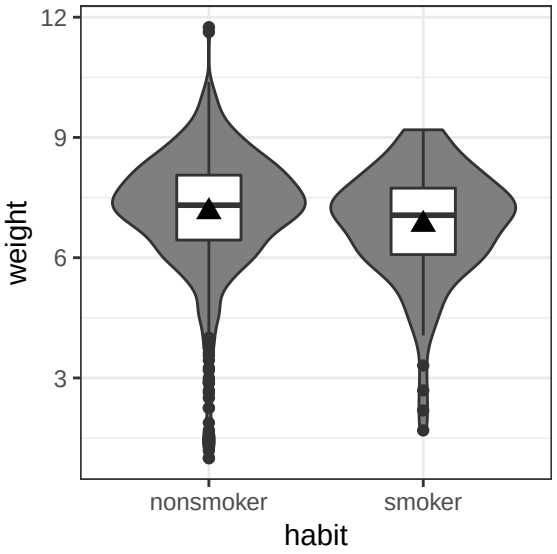


7. Describe the distribution of `gender` in context of the problem. Be sure to use summary statistics and full english sentences in your answer.

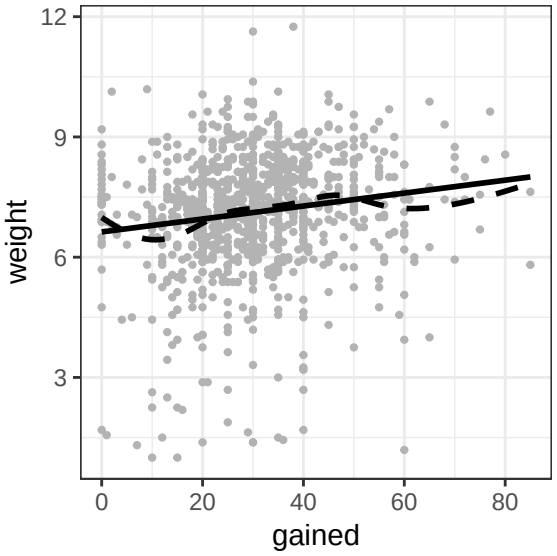
	Frequency	Percent
female	503	50.3
male	497	49.7
Total	1000	100.0

8. Describe the association of `weight` and `habit` in full context of the problem. Be sure to use summary statistics and full english sentences in your answer.

habit	n	mean	sd	median	IQR
nonsmoker	873	7.144273	1.518681	7.31	1.6200
smoker	126	6.828730	1.386180	7.06	1.6575



9. Describe the association of `weight` and `gained` in full context of the problem. Be sure to use summary statistics and full english sentences in your answer. Here is a piece of information: r =0.1541716



10. Use the Univariate and bivariate tables for the combination of smoking status and marital status shown below to answer the following questions. Show your work as needed, Single number answers will suffice.

Frequency Tables

married	not married
386	613

married	not married
0.39	0.61

nonsmoker	smoker
873	126

nonsmoker	smoker
0.87	0.13

Proportion Tables

	married	not married
nonsmoker	0.32	0.55
smoker	0.07	0.06

	married	not married
nonsmoker	0.37	0.63
smoker	0.52	0.48

	married	not married
nonsmoker	0.83	0.9
smoker	0.17	0.1

(a) What percent of the sample are married?

(b) What percent of the sample are non smokers?

(c) What percent of smokers are not married?

(d) What percent of married women are not smokers?

(e) What is the percent of non smoking married women?

The following code chunks created the plots and tables you saw in this exam, but with 4 pieces missing. Fill in the appropriate code for the 4 blank spots in each question.

11. Problem 8: Calculating grouped summary statistics

```
ncbirths %>% select(_____, _____) %>% na.omit() %>%
               group_by(_____) %>% summarise(n=n(),
                                             mean=mean(weight, na.rm=TRUE),
                                             sd = sd(weight, na.rm=TRUE),
                                             median = median(weight),
                                             IQR = IQR(_____, na.rm=TRUE)) %>%
  kable()
```

12. Problem 8: The name of this type of plot is _____ (fill this in too)

```
ggplot(ncbirths, aes(x=_____,  y=_____)) +
        geom__[____](fill="grey50") +
        ____________(width=.3) +
        theme_bw() +
        stat_summary(fun.y=mean, geom="point", shape=17, size=3)
```

13. Problem 9: The name of this type of plot is _____ (fill this in too)

```
ggplot(________, aes(x=gained,  y=________)) +
        geom_point(col="grey70",size=.8) +
        geom_smooth(se=FALSE, lty="dashed", col="black")+
        geom_smooth(method="__", se=FALSE, col="black") + theme_bw()
```

14. Problem 10, the last two-way table.

```
____.table(table(________$habit,ncbirths$_______), ______=2) %>%
  kable(digits=2)
```

Codebook for NCbirths

Usage

This data is available in the `openintro` package as the data frame `ncbirths`.

Description

In 2004, the state of North Carolina released to the public a large data set containing information on births recorded in this state. This data set has been of interest to medical researchers who are studying the relation between habits and practices of expectant mothers and the birth of their children. This is a random sample of 1,000 cases from this data set.

Format

A data frame with 1000 observations on the following 13 variables.

variable	description
`fage`	father's age in years.
`mage`	mother's age in years.
`mature`	maturity status of mother.
`weeks`	length of pregnancy in weeks.
`premie`	whether the birth was classified as premature (premie) or full-term.
`visits`	number of hospital visits during pregnancy.
`marital`	whether mother is `married` or `not married` at birth.
`gained`	weight gained by mother during pregnancy in pounds.
`weight`	weight of the baby at birth in pounds.
`lowbirthweight`	whether baby was classified as low birthweight (`low`) or not (`not low`).
`gender`	gender of the baby, `female` or `male`.
`habit`	status of the mother as a `nonsmoker` or a `smoker`.
`whitemom`	whether mom is `white` or `not white`.

This information was copied from `?ncbirths` on 2015-09-27