

STA102 Introduction to Biostatistics
Spring 2002
Quiz 1 (20 points)

Name: KEY

Question 1

Studies have suggested that low dietary intake or low plasma concentrations of retinol and beta-carotene might be associated with increased risk of developing certain types of cancer. One such study collected data from $n=20$ subjects for the purpose of investigating the effects of personal characteristics and dietary factors on plasma concentrations of retinol and beta-carotene. The variables measured are described below, followed by a listing of the first 3 values of each variable.

On the short line next to each variable name below, indicate whether the variable is (1 point each):

N (nominal)

O (ordinal)

D (discrete) or

C (continuous);

use only ONE letter per line.

Answer:

Some comments on the solutions: AGE (years) is a measure of time. Yes, it is recorded and discussed often in whole years, but we can think about age in terms of time measured to any desired level of precision. Just because a measurement is recorded as integers (without decimals), this doesn't mean it's Discrete. Smoking status (SMOK) is a tough one, I gave you a point if you said either Nominal or Ordinal. Note that you may argue that number of drinks per week (ALC) is Discrete. I might concede that number of drinks could be considered Discrete (although you might have 1.5 drinks) But, drinks *per week* is a rate, which, at least in the cases below, should be considered Continuous. I think the remaining variables are straightforward.

Variable Names and Descriptions:

 C AGE: Age (years)

 N SEX: Sex (1=Male, 2=Female).

N or O SMOK: Smoking status (1=Never, 2=Former, 3=Current Smoker)

 C QUET: Quetelet ($\text{weight}/(\text{height}^2)$) (kg/m^2)

 O VIT: Vitamin Use (1=Yes, fairly often, 2=Yes, not often, 3=No)

 C CAL: Number of calories consumed per day.

 C FAT: Grams of fat consumed per day.

 C FIB: Grams of fiber consumed per day.

 C ALC: Number of alcoholic drinks consumed per week.

 C CHOL: Cholesterol consumed (mg per day).

 C BETAD: Dietary beta-carotene consumed (mcg per day).

 C RETD: Dietary retinol consumed (mcg per day)

__C__ BETAP: Plasma beta-carotene (ng/ml)
__C__ RETP: Plasma Retinol (ng/ml)

First 3 records of the data set:

| AGE | SEX | SMOK | QUET | VIT | CAL | FAT | FIB | ALC | CHOL | BETAD | RETD | BETAP | RETP |
|-----|-----|------|---------|-----|--------|------|------|------|-------|-------|------|-------|------|
| 64 | 2 | 2 | 21.4838 | 1 | 1299.8 | 57.0 | 6.3 | 0.0 | 170.3 | 1945 | 890 | 200 | 915 |
| 76 | 2 | 1 | 23.8763 | 1 | 1032.5 | 50.1 | 15.8 | 0.0 | 75.8 | 2653 | 451 | 124 | 727 |
| 68 | 2 | 2 | 20.0108 | 2 | 2372.3 | 83.6 | 19.1 | 14.1 | 257.9 | 6321 | 660 | 328 | 721 |

Question 2

I was using S-Plus 2000 to analyze the data in Question 1. Unfortunately, I sneezed and must have accidentally deleted a data value for the FIB variable. I'm wondering how I can figure out the value I deleted. I remembered I have a summary statistics printout for FIB, which was produced **before** I accidentally deleted the value. Also, I produced a summary statistics printout **after** I deleted the value. The printouts are shown below. Read the printouts and answer the question at the bottom of the page.

The **printout before** I accidentally deleted one value follows. Note that the printout indicates there are no missing values (indicated by the NA's), so I know the calculations are based on n=20 values (Total N: 20.000000000).

```
*** Summary Statistics for data in: plasma2 ***  
  
                FIB  
Min:  6.300000000  
1st Qu.:  9.600000000  
Mean: 14.665000000  
Median: 14.950000000  
3rd Qu.: 18.425000000  
Max: 28.700000000  
Total N: 20.000000000  
NA's :  0.000000000  
Std Dev.:  6.247127761
```

The **printout after** I accidentally deleted one value follows. Note that S-Plus tells me that I have one missing value: NA's : 1.0000. So, I know the summary calculations are based on 19 values, NOT 20.

```
*** Summary Statistics for data in: plasma2 ***  
  
                FIB  
Min:  6.300000000  
1st Qu.:  9.950000000  
Mean: 14.931578947  
Median: 15.000000000  
3rd Qu.: 18.650000000  
Max: 28.700000000  
Total N: 20.000000000  
NA's :  1.000000000  
Std Dev.:  6.300357365
```

Using the mean and other information provided, what is the value I accidentally deleted? You MUST show how you arrived at your answer to receive credit. The space below should be enough room for your work, but the next sheet is blank just in case you need it. (4 points)

Answer:

Before I deleted the value, I have a mean of

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{\sum_{i=1}^{20} x_i}{20} = 14.665, \text{ which means } \sum_{i=1}^{20} x_i = 20(14.665) = 293.3.$$

After I deleted the value, I have a mean of

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{\sum_{i=1}^{19} x_i}{19} = 14.931578947, \text{ which means } \sum_{i=1}^{20} x_i = 19(14.931578947) = 283.699999993.$$

Thus, missing data value must be $293.3 - 283.7 = 9.6$ (g/day).

Question 3

Acute dilated cardiomyopathy is a heart condition where the heart muscle has become weakened and enlarged. When the left ventricle contracts, it ejects a smaller fraction of the blood than in healthy people. For 27 such patients, here are measures of their left ventricular ejection fraction (LVEF). The data are shown below, along with a box plot.

Give the median of the data. No work required; just your answer will suffice. (1 point)

Answer:

There are $n=24$ data points, not $n=27$. My mistake. **This is a freebee; everyone gets this point.** The data are already conveniently ordered and there's no need to do $n*k/100$ (although that's okay if you did). Just average the two middle values: $(0.20 + 0.22)/2 = 0.21$. By the way, this question and the next one are a part of a question/answer pair submitted by a student. (Thanks!) I like the "biostat" nature of the data.

Describe the overall shape of the distribution of data (circle ONE): (1 point)

A) skewed left

B) skewed right

C) roughly symmetric, perhaps slightly skewed

D) kurtosis

| | | |
|------|------|------|
| 0.07 | 0.19 | 0.24 |
| 0.12 | 0.19 | 0.24 |
| 0.13 | 0.20 | 0.28 |
| 0.14 | 0.20 | 0.30 |
| 0.14 | 0.22 | 0.30 |
| 0.17 | 0.23 | 0.32 |
| 0.17 | 0.23 | 0.32 |
| 0.18 | 0.24 | 0.40 |

