

## Stat 242 Quiz – Topics Drawn from Chapter 1

What's Your Name? \_\_\_\_\_

1. A study enrolled a group of individuals born at a Copenhagen hospital between 1959 and 1961 and followed them for 40 years. Among other variables, they recorded each individual's cholesterol level, the number of glasses of wine they drank, their education level, and their income. Their primary interest was in whether there was an association between the number of glasses of wine someone drank and their cholesterol level. Identify whether the researchers are thinking of each variable as an *explanatory variable*, *response variable*, or a potential *confounding variable*. Write a brief sentence for each explaining why.

**Cholesterol level** is a (circle one): response variable

Briefly, explain why:

The researchers want to see if cholesterol levels are different for people with different behavior; cholesterol levels may respond to wine consumption.

**Wine consumption** is a (circle one): explanatory variable

Briefly, explain why:

The researchers want to see if wine consumption may explain differences in cholesterol levels for different people.

**Education level** is a (circle one): potential confounder

Briefly, explain why:

A person's education level may affect both how much wine they are likely to drink and their overall health. So education level may be a confounder if our primary interest is to study the relationship between wine consumption and cholesterol levels.

**Income** is a (circle one): potential confounder

Briefly, explain why:

Same explanation as for education: A person's income may affect both how much wine they are likely to drink and their overall health.

(this page left blank intentionally)

## Stat 242 Quiz – Topics Drawn from Chapter 1

What's Your Name? \_\_\_\_\_

1. In each of the following studies, identify the *scope of conclusions* that can be drawn from the study design. This should include a description of whether the findings from the study can be applied to any particular population (and if so, describe the population), and whether a claim of a causal association can be made.

(a) Researchers at a university enrolled a sample of children in the metropolitan area where their university was located. They found statistically significant evidence that children who watch more than two hours of television each day tend to have higher cholesterol levels than children who watch less than two hours of television each day.

Inference to population (and if so, what is the population)?

No, this was not a random sample from any well-defined population so we cannot make inferences about a population.

Evidence of causal association?

No, study participants were not randomly assigned by researchers to watch certain amounts of TV, so the study does not provide evidence that watching more TV causes a child to have higher cholesterol levels.

(b) Researchers conducted a poll of randomly selected elderly American adults to investigate the relationship between education levels and emotional well-being; the sample was nationally representative in terms of gender, race and ethnicity, and socio-economic status. The study found that on average, elderly adults with a college degree or higher scored higher on the Emotional Health Index than elderly adults without a college degree.

Inference to population (and if so, what is the population)?

Yes, the sample was randomly selected and is representative. We can draw inferences to the population of elderly Americans.

Evidence of causal association?

No, the study researchers did not randomly assign study participants to different education levels, so the study does not provide direct evidence that graduating from college causes increased emotional health.

(c) Researchers used 7 red and 7 black playing cards to randomly assign 14 volunteer males with high blood pressure to one of two diets for four weeks: a fish oil diet and a standard oil diet. After four weeks, the study subjects assigned to the fish oil diet had a larger average reduction in blood pressure than the subjects assigned to the standard oil diet.

Inference to population (and if so, what is the population)?

No, the study participants were volunteers, so we cannot use the study data to make inferences about a population.

Evidence of causal association?

Yes, the subjects were randomly assigned to the treatment groups, so we can claim that *within this group of subjects*, there is evidence that the fish oil diet led to a larger reduction in blood pressure.

## Stat 242 Quiz – Topics Drawn from Chapter 1

What's Your Name? \_\_\_\_\_

### 1. Give the formal definition of a *sampling distribution*.

The **sampling distribution** of a statistic is the distribution of values for that statistic that we could obtain from all possible samples of size  $n$ .

2. Radon is a radioactive gas that is found in some homes, and can cause cancer. A county health department wants to estimate the average concentration of radon for homes in the county. They randomly select 25 homes from the local property tax list and measure the radon concentration in each home in the sample; they summarize the results by calculating the sample mean concentration. Explain, in the context of this example, what the sampling distribution of the sample mean is.

The sampling distribution of the sample mean is the distribution of values for the sample mean radon concentration that we could get from all possible samples of 25 houses in the county.

(this page left blank intentionally)

## Stat 242 Quiz – Topics Drawn from Chapter 1

What's Your Name? \_\_\_\_\_

### 1. Give the formal definition of a *p-value*.

The *p-value* is the probability of obtaining a test statistic as extreme or more extreme than the value of the statistic we obtained for our sample, if the null hypothesis is correct.

### 2. A study found that the use of bed nets was associated with a lower prevalence of malarial infections in the Gambia. Their test statistic measured the difference in prevalence of malarial infections when using bed nets and when not using bed nets. A report of the study stated that the p-value was 0.00043. Explain what this means in a way that could be understood by someone who has not studied statistics.

If in fact the bed nets had no effect on the prevalence of malarial infections in the Gambia, there would be a probability of 0.00043 that a study conducted in this way would find a reduction at least as big as the reduction they actually saw.

(this page left blank intentionally)