Variability, Randomness, Uncertainty.

Descriptive Statistics

Inferential Statistics

Population - entire set of interest

Sample - part of the population about which we have collected data

biased - a sample that is not representative.

parameter - what we're trying to learn about the population

statistic - value computed from a sample.
Variable

Categorical

Nominal

Ordinal

Numerical

Discrete

Continuous

Grade points for this class.

0, 4, 3, 2, 1, 0 - Discrete.

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>

What GPAs could you have after 2 classes?

0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4
controlled experiments and observational studies.

- causal effects.
- confounding factors

Randomized Controlled Double Blind Experiment.

<table>
<thead>
<tr>
<th>treatment group</th>
<th>control group</th>
</tr>
</thead>
<tbody>
<tr>
<td>remove placebo effect</td>
<td>psychological biases</td>
</tr>
</tbody>
</table>

overcome confounding factors
Observational Studies.

- shows association.

\[ \text{age} \]
\[ \text{shoe size} \quad \text{reducing score} \]

association is not causation. due to confounding factors.

- "control" for confound potential confounding factors.

Generating Representative Samples.

- simple random sample
- avoid selection bias response bias

Quota sampling
stratified sampling
cluster sample
convenience sample
Looking at data.

- range of values
- most likely values
- symmetry / tail behaviour
- uni/multimodality.

Pie chart
bar chart
dot chart

Histogram.

Area represents %.

Distribution of heights of people taking AMS 5

- define class intervals.
- decide which endpoint to include
- count frequency in each interval
- compute %.
- compute % - per-unit
Measures of center and spread:
- Mean
- Median

Interquartile range: \( \frac{1}{2} \) data is less than the mean.

\( \text{SD} = \text{RMS deviation from the mean.} \)

Root Mean Square.

Normal approximation:
- Distribution of a lot of data sets can be approximated by the normal curve.

Normal units:
- \( \text{Value - mean} / \text{SD} \)

Outliers.
area from the table in the book.

- once the values are in standard units.

\[ SD = \sqrt{\frac{(10.3 - 14.9)^2 + (11.7 - 14.9)^2 + \ldots}{20}} \]
Probability.

prob. - % of time an event is expected to occur when repeat independently under same conditions.

- # outcomes were interested in total # of outcomes.

\[ P = \text{opposite event went } 1 - P \]

\[ P(A \text{ and } B) = \text{prob of } A \times \text{prob of } B \]

given that \( A \) has happened.

mutually exclusive events.

\[ P(A \text{ or } B) = \text{prob of } A + \text{prob of } B \]

not mutually exclusive

\[ P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B) \]
\[ p(\text{red}) = \frac{1}{3}, \quad p(\text{green}) = \frac{2}{3} \]

Drew with replacement.

Order doesn't matter.

\[ \Pr(2 \text{ reds out of 5}) = \frac{1}{5} \times \frac{1}{3} \times \frac{2}{5} \times \frac{2}{5} \times \frac{2}{5} \]

\[ \Pr(2 \text{ reds out of 5}) \]

\[ \Pr(k \text{ successes in } n \text{ trials}) = \binom{n}{k} p^k (1-p)^{n-k} \]

-n! 
\[ \frac{k! (n-k)!}{k! (n-k)!} \]

- Binomial coefficient

- # trials

- # "successes"
Box Model.

- Sum of # on the tickets that are drawn.
- What numbers to put on the tickets.
- How many tickets of each type.
- How many draws.

Sum of draws
≡ # of heads.

Sum of draws
≡ # of republican voters.

sum of draws → distribution. Follows normal curve.

Expected value
≡ #draws \times \text{ave of box}

Standard error
≡ \sqrt{\#\text{draws} \times \text{SD of box}}

\text{Dist. of sum of draws}
Box with only 2 types of tickets.

$$SD_{box} = \sqrt{\left( \frac{\text{large number on the tickets}}{\text{small number on tickets}} \right) \times \left( \frac{\text{fraction of tickets with large #}}{\text{fraction of tickets with small #}} \right)}$$
A Normal Table is on the last page of this exam.

1. A professor has five different colored sweaters in his closet. During the day, his small children play in the closet, and enjoy rearranging the sweaters in a random order. Each morning, the professor picks the sweater off the top of the pile, and wears it to class.

   The professor’s students have noticed, however, that in 10 class sessions, the professor has always worn the same sweater.
   
   (a) What is the chance of this occurring?

   \[
   \left( \frac{1}{5} \right)^{10}
   \]

   (b) What is the chance that during the next 6 class sessions, the students will see this sweater exactly 4 times?

   \[
   \binom{6}{4} \left( \frac{1}{5} \right)^4 \left( \frac{4}{5} \right)^2
   \]

2. People who get lots of vitamins by eating five or more servings of fresh fruits and vegetables each day have much lower death rates from colon cancer and lung cancer, according to many observational studies. These studies were so encouraging that two randomized controlled experiments were done. The treatment groups were given large doses of vitamin supplements, while the control group just ate their usual diet. One experiment looked at colon cancer; the other at lung cancer.

   The first experiment found no difference in the death rate from colon cancer between the treatment group and the control group. The second experiment found that beta carotene (as a diet supplement) increased the death rate from lung cancer. True or false, and explain:

   (a) The experiments confirmed the results of the observational studies.

   (b) The observational studies could easily have reached the wrong conclusions, due to confounding – people who eat lots of fruit and vegetables have lifestyles that are different in many other ways too.
(c) The experiments could easily have reached the wrong conclusions due to confounding — people who eat lots of fruit and vegetables have lifestyles that are different in many other ways too.

3. A statistics professor buys a loaf of bread each week from the same bakery. The baker advertises the weight of the loaf as one pound (16 oz). Over the course of 20 weeks the professor records the weight of the loaves. They are (in oz, sorted into increasing order)

\[10.3, 11.7, 12.5, 13.4, 13.6, 14.0, 14.4, 14.5, 14.7, 15.0, 15.2, 15.4, 15.5, 15.6, 16.0, 16.4, 17.0, 17.4, 17.6, 18.7\]

(a) Complete the table below, where the class interval includes the lower limit, but not the upper limit.

<table>
<thead>
<tr>
<th>Class Interval (oz)</th>
<th>Frequency</th>
<th>percentage</th>
<th>percent-per-oz</th>
</tr>
</thead>
<tbody>
<tr>
<td>10-12</td>
<td>11</td>
<td>10%</td>
<td>5</td>
</tr>
<tr>
<td>12-14</td>
<td>3</td>
<td>15%</td>
<td>7.5</td>
</tr>
<tr>
<td>14-16</td>
<td>9</td>
<td>45%</td>
<td>22.5</td>
</tr>
<tr>
<td>16-18</td>
<td>5</td>
<td>25%</td>
<td>12.5</td>
</tr>
<tr>
<td>18-20</td>
<td>1</td>
<td>5%</td>
<td>2.5</td>
</tr>
</tbody>
</table>

(b) Sketch the histogram on the graph at the top of the page 3. Label the axes.
(c) Does the histogram appear to follow the normal curve? Explain briefly.

(d) The mean weight of the 20 loaves is 14.9 oz, and the SD is 2.0 oz. What is the median?

\[
\text{average of the } 10^\text{th} + 11^\text{th} \text{ weights} = \frac{15.0 + 15.2}{2} = 15.1 \text{ oz}
\]
(e) The professor tells the baker that he is selling underweight loaves. Assuming the weight of the loaves follows the normal curve, what percentage of loaves weigh at least the advertised 16oz?

\[ \text{16 in standard units } \frac{16 - 14.9}{2} = 0.55 \]

\[ \frac{1}{2} (100 - 42) = 29\% \]

(f) The baker tells the professor that he has increased the weight of the loaves, but instead weighs loaves each morning until he finds one that weighs at least 16oz, which he then puts aside for the professor.

Sketch the histogram of the weights of the loaves the professor now buys.

(g) Does the professor believe the baker? Explain why/why not.
4. True or false, and explain briefly –

(a) If you add 7 to each entry on a list, that adds 7 to the average.

(b) If you add 7 to each entry on a list, that adds 7 to the SD.

(c) If you double each entry on a list, that doubles the average.

(d) If you double each entry on a list, that doubles the SD.

(e) If you change the sign of each entry on a list, that changes the sign of the average.

(f) If you change the sign of each entry on a list, that changes the sign of the SD.

5. Read the abstract of the paper “Alcohol Consumption and Male Erectile Dysfunction: An Unfounded Reputation for Risk” printed at the end of this exam paper.

(a) Was this a controlled experiment or an observational study?

(b) The investigators used reply paid questionnaires sent to the subjects. Give two possible problems with this methodology.
(c) What confounding factors did the investigators consider?

\[ \text{smoking} \quad \text{age} \quad \text{CVD} \]

(d) Can you conclude that alcohol consumption causes erectile dysfunction? Explain briefly.

\[ \text{No. - observational study cannot overcome confounding factors.} \]

6. A box contains red and blue marbles; there are more red marbles than blue ones. Marbles are drawn one at a time from the box, at random, with replacement. You win a dollar if a red marble is drawn more often than a blue one. There are two choices.

- (A) 100 draws are made from the box.
- (B) 200 draws are made from the box.

Choose one of the four options below; explain your answer.

(a) A gives a better chance of winning.
(b) B gives a better chance of winning.
(c) A and B give the same chance of winning.
(d) Can't tell without more information

7. In roulette, once in a while, someone will bet $1 on red; and, at the same time, someone else will bet $1 on black. (Recall that there are 18 red slots, 18 black slots, and two green slots. You win if the ball ends up in one of the slots of the color you have chosen.) Suppose this pair of bets is made 100 times in the course of an evening.

(a) The house will make money on \( \frac{5}{3} \cdot \frac{3}{2} \) of the 100 pairs of bets, give or take \( 2.2 \) or so.

(b) The net gain for the house from the 100 pairs of bets will be around \( 10.6 \) give or take \( 4.4 \) or so.

\[ \text{EV} = 100 \times \frac{36 \times 0 + 2 \times 1}{38} = 5.4 \]

\[ \text{SE} = \sqrt{100} \times \text{SD}_{\text{box}} = 2.2 \]

\[ \text{Amount of money the house makes.} \]

\[ \text{SD}_{\text{box}} = (1-0) \sqrt{\frac{2}{36} + \frac{2}{36}}. \]
Alcohol Consumption and Male Erectile Dysfunction: An Unfounded Reputation for Risk?

Kew-Kim Chew, MBBS, FRCPEdin, FRCPGlasc,1† Alexandra Bremner, BSc(Hons), GradDipAppStats, PhD,1 Bronwyn Stuckey, MBBS, FRACP,1,5 Carolyn Earle, BSc, PGradDipHlthSci,1 and Konrad Jamrozik, MBBS, D Phil1

1Keogh Institute for Medical Research, Queen Elizabeth II Medical Centre, Nedlands, Western Australia, Australia; 1School of Population Health, School of Medicine and Pharmacology, University of Western Australia, Nedlands, Western Australia, Australia; 1School of Medicine and Pharmacology, University of Western Australia, Nedlands, Western Australia, Australia; 1Department of Endocrinology and Diabetes, Sir Charles Gairdner Hospital, Nedlands, Western Australia, Australia; 1School of Population Health and Clinical Practice, University of Adelaide, Adelaide, South Australia, Australia

DCI: 10.1111/j.1743-6109.2008.01115.x

ABSTRACT

Introduction. Alcohol consumption is a contentious social topic and is often assumed to have deleterious effects on sexual performance. There is a lack of consensus on whether alcohol consumption may in fact be beneficial to erectile function.

Aim. We examined the data from a population-based cross-sectional study of men's health to assess the association between usual alcohol consumption and erectile dysfunction (ED).

Method. Reply-paid questionnaires were posted to a randomly selected age-stratified male population sample obtained from the Western Australian (WA) Electoral Roll.

Main Outcome Measures. The survey questionnaire included sociodemographic details, self-reported clinical information, and drinking habits. The 5-item International Index of Erectile Function (IIEF-5) was used to assess erectile function.

Results. Most (87%) participants were current alcohol drinkers, with binge drinking, as defined by the Australian National Health and Medical Research Council (NHMRC), reported by 20% of drinkers. Compared with never-drinkers, the age-adjusted odds of ED were lower among current, weekend, and binge drinkers and higher among ex-drinkers. Among current drinkers, the odds were lowest for consumption within the NHMRC guidelines of between 1 and 20 standard drinks a week. On further adjustment for cardiovascular disease (CVD) or for cigarette smoking, age-adjusted odds of ED were reduced by 25–30% among alcohol drinkers.

Conclusions. Our findings suggest a modest negative association between alcohol consumption and ED and confounding of the association by CVD and cigarette smoking. The Western Australia Men’s Health Study certainly provides no justification for advising men with ED whose drinking habits are consistent with NHMRC guidelines that they should cease or reduce their consumption of alcohol. Chew K-K, Bremner A, Stuckey B, Earle C, and Jamrozik K. Alcohol consumption and male erectile dysfunction: An unfounded reputation for risk? J Sex Med **;**:**_**.

Key Words. Alcohol Consumption; Erectile Dysfunction; Cardiovascular Disease; Cigarette Smoking

Introduction

Consumption of alcohol has been associated with felicitous and celebratory occasions in many cultures since time immemorial. Alcohol is the most commonly used recreational drug worldwide.

Although excessive chronic intake of alcohol and binge drinking may result in serious adverse effects on health, there are cardio-protective benefits associated with daily light to moderate alcohol consumption [1,2]. As erectile dysfunction (ED) is mostly vasculogenic and intimately linked to cardiovascular disease (CVD) [3,4], it is possible