Disclaimer: This review does not necessarily cover *every* topic that could be on the test. Similarly, questions asked on the test may not follow the exact format of the questions on this review. I STRONGLY ENCOURAGE you to also study the Test 1 and Test 2 Reviews, the homeworks, the videos, the tests, etc.

1. A soda-bottling machine is calibrated to dispense 20 ounces of soda to each bottle. The actual amount of soda dispensed varies according to a normal distribution. In a random sample of 30 bottles, we observe a mean soda volume of 19.91 ounces and standard deviation 0.12 ounces. Does this sample provide evidence that the mean volume of soda dispensed to bottles differs from 20 ounces? Use the output below to help you, and specify which output (A, B, or C) you used.

A.

A screenshot of a cell phone

Description automatically generated

B.

A screenshot of a social media post

Description automatically generated

C.

A screenshot of a social media post

Description automatically generated

A screenshot of a cell phone

Description automatically generated

2. At right is a scatterplot (from StatKey) of body temperature and pulse (bpm) for a sample of 25 healthy US women. We are investigating the correlation between body temperature and pulse.

(a) Is there an overall straight-line relationship between body temperature and pulse?

(b) What is the correlation between body temperature and pulse? Is this correlation strong, weak, or moderate?

(c) Report the equation of the regression line for predicting pulse from body temp. What does the slope of the regression line mean in context of this problem?

(d) The 95% confidence interval for the correlation is (0.221, 0.723). Interpret this confidence interval.

3. In 1980, the active ingredient in marijuana was shown to outperform a placebo in reducing nausea in chemotherapy patients. Many studies are now underway to see if the drug has additional medicinal uses. In one controlled, randomized trial, 67 patients with HIV were randomly assigned to two groups, with one group getting cannabis (marijuana) and the other getting a placebo. All of the patients had severe neuropathic pain, and the response variable is whether or not pain was reduced by 30% or more (a standard benchmark in pain measurement). The results are shown in the table below. The question of interest is whether marijuana is more effective than a placebo in relieving pain.

|  |  |  |
| --- | --- | --- |
|  | Pain Reduced (by at least 30%) | Pain Not Reduced |
| Cannabis | 14 | 13 |
| Placebo | 17 | 23 |

1. Let be the proportion of patients with reduced pain while on the cannabis treatment and be the proportion of patients with reduced pain while on the placebo. State the hypotheses of the experiment.
2. Calculate the sample proportions of patients with reduced pain in each group.

1. The p-value for the test stated in part (a) is approximately 0.4. State the conclusion of the experiment in context.
2. Based on your conclusion, what type of error *might* you have made? (choose one)

1) Type I

2) Type II

3) Both Type I and Type II

4) No errors are possible.

4. Some researchers have investigated whether the air in carpeted rooms contains more bacteria than air in uncarpeted rooms. For an experiment, we have 25 rooms that are carpeted and 25 that are uncarpeted. The rooms are similar in size. The concentration of bacteria in the air is measured in all of the rooms (units are bacteria per cubic foot). For carpeted rooms, the mean concentration was 204 and SD was 27. For uncarpeted rooms, the mean concentration was 172 and SD was 17.9.   
Construct and interpret a 90% confidence interval for the difference in mean concentration of bacteria in the air between carpeted rooms and uncarpeted rooms. The margin of error for this interval is 10.899. Do you conclude that the air in carpeted rooms contains a higher average concentration of bacteria than that of uncarpeted rooms?

5. Give the definition of a p-value.

6. Suppose you weigh an object 100 times and observe a sample mean weight of 31.345 grams and a standard deviation of 0.2 grams.

(a) Construct and interpret a 90% confidence interval for the weight of this object. The margin of error for this interval is 0.033.

(b) In context of this problem, explain what is meant by “90% confidence.”

7. Local officials want to know how a community feels about building a new community center. A sample of households in a community is selected at random from the telephone directory. In this community, 4% of households have no telephone and another 35% have unlisted telephone numbers.

1. Describe a source of bias that sample will *certainly* suffer from.

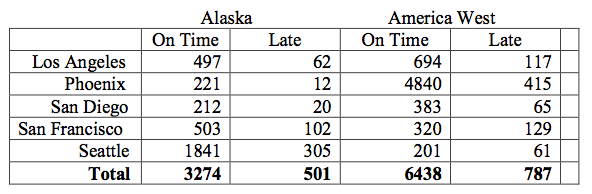
(b) Building the community center will require raising taxes. Two local officials conduct the survey. One local official (Bob) asks the question this way: “I’m sure that, like me, you think that our taxes are already too high. Would you support raising our taxes yet again to build a community center that may or may not get used?” The other local official (Julia) asks the question this way: “Would you support a local levy, the proceeds of which will go to build a new community center?”

1) T or F: Bob’s version will lead to inaccurate responses.

2) T or F: Julia’s version will lead to inaccurate responses.

3) T or F: The officials should have agreed ahead of time how they would ask the question.

8. The following table classifies all flights on Alaska Airlines and America West Airlines departing from 5 western United States Cities during June 1991. Flights are classified by airline (Alaska or America West), by City of Departure (Los Angeles, Phoenix, San Diego, San Francisco, Seattle), and by whether the departure was on time or late.



(a) For each *airline*, compute the proportion of flights that were late to depart.

(b) For each *city*, compare the proportion of late departs.

(c) What would be a good way to visualize the percentages in part (b) to compare them to each other?

9. A recent study examined the impact of the color blue on how attractive women perceive men to be. In the study, women were divided into two groups using a fair coin and were asked to rate the attractiveness of men on a scale of 1 (not at all attractive) to 9 (extremely attractive). One group of women were shown pictures of men on a white background and the other group were shown the same pictures of men on a blue background. The women who saw men on the blue background rated them as significantly more attractive. All participants and those showing the pictures and collecting the data were not aware of the purpose of the study.

(a) What is the explanatory variable? Is the explanatory variable categorical, ordinal, or quantitative?

(b) What is the response variable? Is the response variable categorical, ordinal, or quantitative?

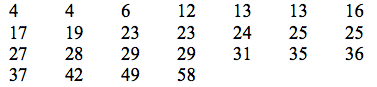
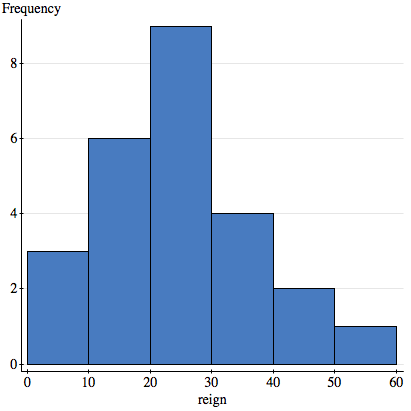
(c) Was randomization used in this investigation?

(d) Is this an experiment or an observational study?

(e) Was this study single blind, double blind, or not blinded?

(f) Can we conclude that using a blue background color instead of white increases women’s attractiveness rating of men’s pictures?

10. During the years 1058 to 1714, there were 25 different Kings and Queens of an independent Scotland. The durations of their reigns (rounded to the nearest whole year) are given below, along with a histogram.



(a) Describe the distribution of reign durations.

(b) For a brief summary of this distribution, would it be better to use the five-number summary or the mean and standard deviation? Explain your choice.

(c) Suppose I made a typo when entering the data into R, and I recorded the value “58” as “580”. Which of the following statistics would be affected by this error? (choose all that apply)  
1) mean 2) median 3) standard deviation 4) min 5) max

11. In a recent survey, 91 of 149 randomly selected college-aged men indicated that they would be willing to marry a woman of “lower social class.” Of 236 randomly selected college-aged women, 117 indicated a willingness to marry a man of lower social class. Can we conclude that men and women differ in a willingness to marry somebody of lower social class? Carry out a test of hypotheses to answer this question. The p-value is 0.0275.

12. A poll asked people if college was worth the financial investment. They also asked the respondent’s gender. The table shows a summary of the responses.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Female | Male | Total |
| No | 65 | 59 | 124 |
| Unsure | 96 | 96 | 192 |
| Yes | 517 | 432 | 949 |
| Total | 678 | 587 | 1265 |

1. If a person is randomly chosen, what is the probability that the person said “No”?
2. If a person is randomly chosen, what is the probability that the person is male and said “No”?
3. If a person is randomly chosen, what is the probability that the person is male or said “No”?
4. If a male is chosen at random, what is the probability that he said “No”?
5. Is being male independent of answering “No”? Explain using probabilities.