Stat 204

Exam 1 2/22/18 Name: SOLUTIONS

1. Here are some definitions from the first three chapters. For each definition, fill in the blank(s) with the word(s) being defined.

(a) Some variables are quantitative, taking numerical values; other variables are categorical/qualitative, taking category designations.

(b) A sample is a subset of the population on which we record data.

(c) The individual entries on which data are recorded are observational units

(d) The SD is a common measure of variability.

(e) The population is the entire collection of all possible sources of data and can be summarized by numbers known as parameters.

(f) The median is the middle data value when data are sorted from smallest to largest.

(g) A variable's distribution is its pattern of outcomes; this pattern is skewed if most values fall on one side, with a long tail on the other.

(h) An outlier is an observation that does not fit a variable's overall pattern of outcomes.

(i) The p-value is the probability of obtaining a result at least as extreme as that observed if the null hypothesis is true.

(j) A result is significant if it is unlikely to occur by random chance.

(k) A statistic is a number computed from a sample.

(l) The significance level is a value (often 5%) used to decide which hypothesis is better supported by the data.

2. IQ scores are normally distributed, with an average score of 100 and a standard deviation of 15. While revising a research manuscript, a professor finds summary statistics for the IQs of a sample of students. For this sample, the average IQ is 108, but the sample size is obscured by a coffee stain; the professor is certain, however, that the sample size is either 10 or 40. Which sample size is more likely? Compute and interpret standardized statistics to justify your answer.

$n=10$: $z = \dfrac{108-100}{\frac{15}{\sqrt{10}}} = 1.69$ ① ← far more likely! ①

3

$n=10$

$n=40$: $z = \dfrac{108-100}{\frac{15}{\sqrt{40}}} = 3.37$ ①

3. A study of college students finds that the men have an average weight of 165 pounds and a standard deviation of 10 pounds; the women have an average weight of 135 pounds and a standard deviation of 10 pounds. The weights for each gender are roughly normally distributed.

(a) What are the observational units?

① students

5

(b) What is the variable, and what kind of variable is it?

① weight – quantitative ①

(c) Is the standard deviation of *all* of the weights (men and women together) smaller than 10 pounds, just about 10 pounds, or bigger than 10 pounds? Why?

① bigger – combining data increases variability ①

8

4. Birth weights of babies born in the U.S. have a mean of 3250 grams and standard deviation of 550 grams. Based on this information, which of the following is less likely? Choose one and explain your answer.

   (a) A randomly selected baby has a birth weight greater than 4000 grams.

   (b) A random sample of 16 babies has an average birth weight greater than 4000 grams.

(b) is far less likely, as variability of sample means decreases with sample size (550 vs. $\frac{550}{4}$ for SDs). 2

3

choice: 1 explanation: 2

5. A multiple-choice test has 20 questions; each question has 3 possible answers, exactly one of which is correct. A student must answer 10 or more questions correctly to pass the test. If a student answers each question by guessing randomly, what is the probability of passing? Use a simulation to determine the answer, then briefly explain what you did.

simulation: probability of success = $\frac{1}{3}$, sample size = 20, count # samples with # successes ≥ 10. 2

$\Longrightarrow$ probability $\approx .094 = 9.4\%$ (1st simulation)

(or: 9.9%, 8.3%, 9.53%, ...

$\Longrightarrow$ 8 - 10% all possible)

1

3

6

6. Mixed-handed people favor one hand for some tasks and the other hand for other tasks; some research suggests that 30% of people are mixed-handed. To test this finding, a simple random sample of 400 people is selected, and 100 of the people in the sample are mixed-handed.

(a) State the relevant hypotheses and compute the observed statistic.

3

$H_0: \pi = .3 = 30\%$

$H_a: \pi \neq 30\%$ OR $H_a: \pi < 30\%$

$\hat{p} = \frac{100}{400} = 25\% = .25$

(b) Without using a simulation, compute the standardized statistic used to test these hypotheses.

2

$z = \frac{\hat{p} - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} = \frac{.25 - .3}{\cancel{\quad} \leftarrow \sqrt{\frac{.3(.7)}{400}}} = -2.18$

$z = -2.18$

(c) Use a simulation to estimate the p-value used to test these hypotheses.

2

1-sided p: 1%. - 2%.

2-sided p: 2-4%.

(d) Your answers to (b) and (c) should be consistent. Based on them, what is your conclusion?

1

reject the null!

8

7. An exam has 40 true-false questions. Bob takes the exam and answers 24 questions correctly. Do you think he prepared for the exam, or do you think he answered the questions by guessing randomly?

(a) State the relevant hypotheses and compute the observed statistic.

3 ① $H_0: \pi = \frac{1}{2} = 50\%$ vs. $H_a: \pi > 50\%$ ①

② $\hat{p} = \frac{24}{40} = 60\% = .6$

(b) Without using a simulation, compute the standardized statistic used to test these hypotheses.

2 $$z = \frac{\hat{p} - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} = \frac{.6 - .5}{\sqrt{\frac{\frac{1}{2}\cdot\frac{1}{2}}{40}}} = 1.26$$

(c) Use a simulation to estimate the p-value used to test these hypotheses.

2 probability of heads = .5
number of tosses = 40
repetitions = 1000

Then count # with successes ≥ .6

$\longrightarrow$ $p \approx 12-13\%$

1 (d) Your answers to (b) and (c) should be consistent. Based on them, what is your conclusion?

slight evidence against null; don't reject it, but no great evidence that he studied

8

8. Sheep brain weights are normally distributed, with an average weight of 150 grams and a standard deviation of 10 grams. As part of an experiment, 8 sheep are given a synthetic hormone, and their brains are weighed at the end of the study. Here are the resulting measurements:

146	148	143	144	129	131	152	154

Based on this sample, is there evidence that this hormone decreases brain weight?

(a) State the relevant hypotheses.

① $H_0: \mu = 150$ vs. ① $H_a: \mu < 150$ } 2

(b) Use an applet to compute the mean and standard deviation for this sample.

① $\bar{x} = 143.375$ ① $s = 9.07$ } 2

(c) Compute the relevant z-statistic and t-statistic.

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{143.375 - 150}{\frac{10}{\sqrt{8}}} = -1.87$$ ②

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{143.375 - 150}{\frac{9.07}{\sqrt{8}}} = -2.07$$ ②

4

(d) Use an applet to conduct a simulation to test these hypotheses, then provide the following p-values.

i. Simulated p-value based on null distribution: .027 = 2.7%

ii. Theoretical p-value based on null distribution: .0305 = 3.05%

iii. Simulated p-value based on simulated t-statistics: .037 = 3.7%

iv. Theoretical p-value based on t-distribution: .0386 = 3.9%

} 4

(e) All of these results should be consistent. Based on them, what is your conclusion?

moderate evidence against null — reject @ 5%. } 2

14