Name:KEY ID:

Stat 218 - Exam 1 Practice Test

Formula sheet

Sample standard deviation: $\sqrt{\dfrac{\sum(x-\overline{x})^2}{n-1}}$

Sample proportion: $\hat{p} = \dfrac{\text{Number of successes in the sample}}{n}$

Standardized statistic: $\dfrac{\text{statistic} - \text{mean of null distribution}}{\text{SD of null distribution}}$

Long-run proportion standard deviation: $\sqrt{\dfrac{\pi(1-\pi)}{n}}$

Standardized statistic for a single mean: $t = \dfrac{\overline{x} - \mu_0}{s/\sqrt{n}}$

Concepts

Circle T (true) or F (false) for each of the following statements:

F A p-value is the proportion of failures in the sample

T A statistic is the quantity of interest computed for the sample

T π is the symbol for the population proportion

F A standardized statistic is the raw distance between the sample statistic and the population mean

F A parameter is the quantity of interest computed for the sample

T The validity conditions for a hypothesis test on a quantitative variable are to have 20 observations and a not-strongly-skewed distribution

F A categorical variable has at least three possible options

F Normal distributions have a right skewed distribution

T A sampling frame is the set of all possible observational units in the population

T The standardized statistic is commonly denoted by the variable t or z

F The central limit theorem predicts the behavior of the null distribution when validity conditions are met

T α is the probability of a Type I error

T A two-sided test estimates the p-value by considering results that are as extreme as the observed result in either direction

F $\hat{p}$ is the symbol for the sample mean

F A *plausible* model is one in which there is no reasonable explanation for the data we observed

T All binary variables are also categorical variables

Weather Predictions

Weather forecasts contain many different variables - high temperature, low temperature, wind speed, and precipitation. The National Weather Service would like to evaluate whether their precipitation predictions are significantly better than the expected success rate of 75%. The NWS defines a success as predicting that precipitation will occur on a day when precipitation actually falls, or predicting that precipitation will not occur on a day when there is no precipitation. That is, we are not considering here the *amount* of precipitation; only whether or not it occurred.

Evaluating predictions made over 2019 in Lincoln, the weather service correctly predicted the presence or absence of precipitation on 290 of the 365 days of the year.

1. What is the research question?

   Is the NWS's model for precipitation better than expected?

2. What is the observational unit?

   A single day of the year.

3. What is the parameter?

   The long-run proportion of days with correctly predicted precipitation.

4. Calculate the sample statistic
   $\hat{p} = \frac{290}{365} \approx 0.795$

5. What is H_0?
   $\pi = 0.75$

6. What is H_A?
   $\pi > 0.75$

7. What type of hypothesis test is this?
   A one-sided hypothesis test

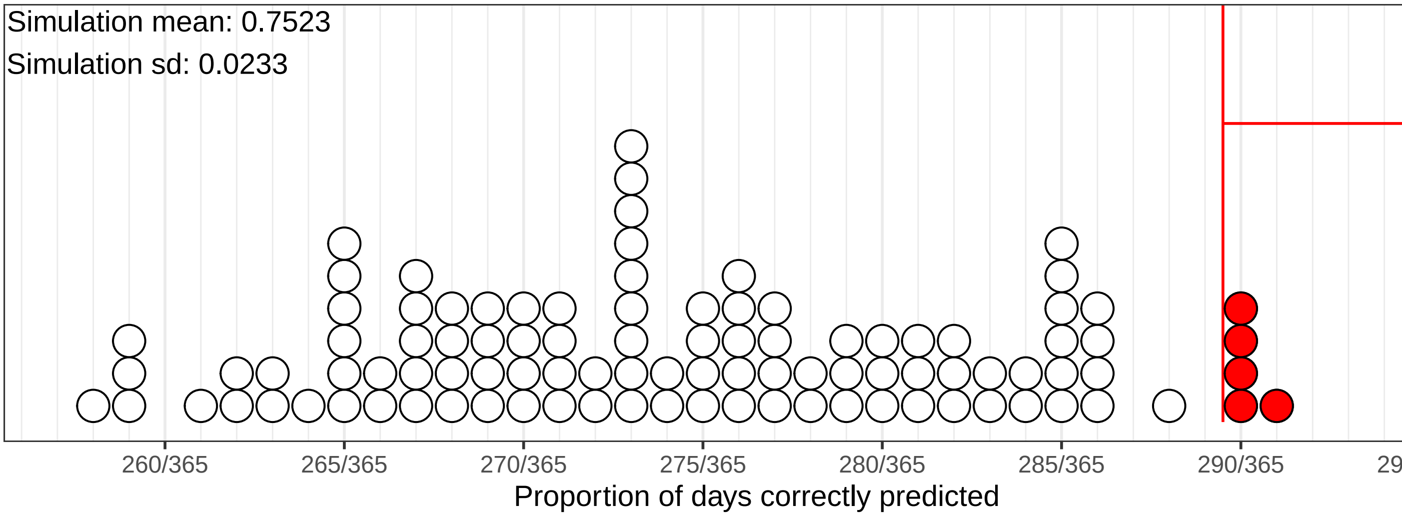
8. You want to conduct a simulation to answer the research question. To test the hypotheses above, you set the one-proportion applet to have the following values:
   a. Probability of heads: *0.75*

   b. Number of tosses: *365*

   c. Number of repetitions: 100

9. In this simulation, what does a heads represent?
   Heads represents a day with correctly predicted precipitation.

10. Your simulation produces the data shown in the plot below:



a. What does a dot represent in the plot?
A set of 365 simulated observations.

b. On the plot, draw one or two vertical line(s) to indicate the cutoff(s), and an arrow(s) indicating the direction(s) of H_A. Shade in the dots corresponding to the simulation p-value calculation.

c. Calculate the simulation p-value:
$$P(\hat{p} >= 290/365) \approx P(p^* >= .795) = \frac{6}{100} = 0.06$$

d. What is your conclusion?
With a p-value of 0.06, I have moderate evidence against the null hypothesis that $\pi = 0.75$. I reject H_0 and conclude that the national weather service correctly predicts whether or not there will be precipitation over the long run with a probability of $\pi > .75$

11. Is it appropriate to use theory-based inference on this data? Why or why not?

Yes, we could use theory-based inference on this data. We would expect that we would have $365 \times 0.75 = 273.75$ successes and $365 \times 0.25 = 91.25$ failures. As both the expected successes and failures under H_0 are greater than 10, we can use theory-based inference.

12. Calculate the theory-based standardized statistic for this test.
$$z = \frac{\hat{p} - \pi}{\sqrt{\pi(1-\pi)/n}} = \frac{.7945 - .75}{\sqrt{.75(1-.75)/365}} = \frac{0.0445205}{0.0226649} = 1.96$$
Note that the theory-based standardized statistic does not use the simulation mean or standard deviation. You must use the formula for σ from the formula sheet.

13. What should the national weather service conclude about their precipitation model?

The national weather service should conclude that there is moderate evidence against the null hypothesis. The standardized statistic is almost 2, which indicates that the sample value is unlikely to be observed by chance alone. The NWS should reject H_0 and conclude that $\pi > 0.75$

3D Printing Mixup

Dr. Vanderplas has been experimenting with her new 3d printer, and has used it to print both fair and unfair dice to use in Stat 218. The unfair dice she printed are weighted, causing some faces to be more likely to appear.

Dr. Vanderplas's son, Alex, mixed the two die together so that it is now impossible to tell which die is fair or unfair by sight. To separate the two die out, Dr. Vanderplas decides to test one of the die by rolling it 60 times and recording how many times each face appears. Using her data, can you determine whether the die she rolled is unfair?

Face	1	2	3	4	5	6
# Rolls	4	5	7	12	14	17

1. Identify the research question.
   Is the die fair?

2. What is the probability of rolling a 6 under random chance?
   $1/6 \approx 0.1667$

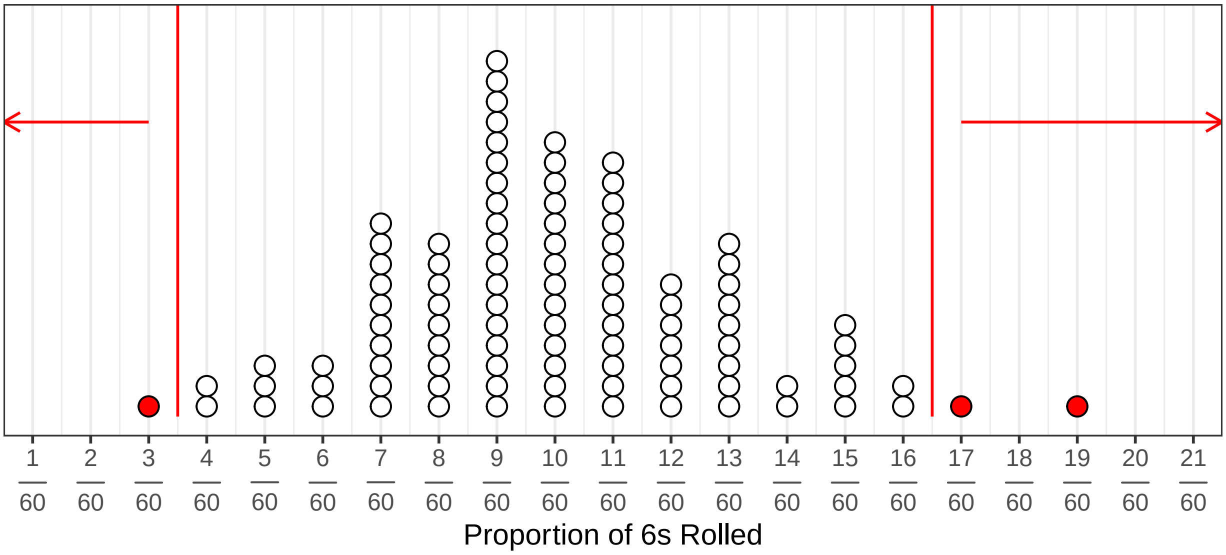
3. You decide to define a success as rolling a 6. Using this definition, state the null and alternative hypotheses.
   $H_0 : \pi = 1/6$
   $H_A : \pi \neq 1/6$

4. What does π mean for this problem?
   π is the long-run probability of rolling a 6 with the selected die.

5. What is $\hat{p}$ for this problem?
   $\hat{p} = \frac{17}{60} = 0.3$

6. What type of hypothesis test is this?
   A two-sided hypothesis test

7. Using your selected probability under random chance, you conduct 100 simulation trials, with the results shown in the plot below.



a. Describe the distribution's shape, center, and variability.

The distribution is centered at approximately 1/6, and has a slight right skew. Most of the values are concentrated between 5/60 and 15/60 (or 1/12 and 1/4).
Note: It is also ok to say this distribution is relatively symmetric. The skew is very slight.

b. On the plot, draw one or two vertical line(s) to indicate the cutoff(s), and an arrow(s) indicating the direction(s) of H_A. Shade in the dots corresponding to the simulation p-value calculation.

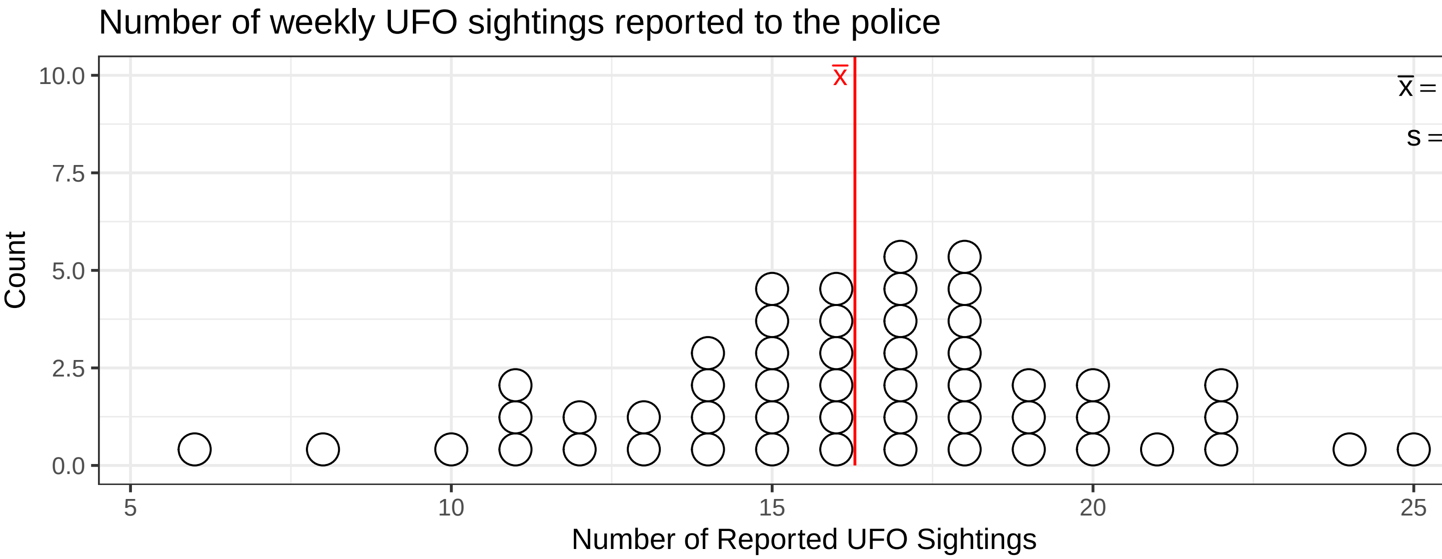
c. Calculate the simulation p-value.
$P(|\hat{p} - 1/6| \geq 7/60) \approx P(|p^* - 1/6| \geq 7/60) = \dfrac{3}{100} = 0.03$

d. What do you conclude?
With a p-value of 0.03 I reject H_0 and conclude that it is unlikely that the long-run probability of rolling a 6 with the selected die is 1/6. It is more plausible that $\pi \neq 1/6$, which would indicate that Dr. Vanderplas rolled the unfair die.

e. Given your conclusion in d, what type of error could you be making? Explain.
(Assuming d is correct) It is possible that we have falsely rejected H_0 when in fact $\pi = 1/6$. This would be a Type I error.
(If d is not correct) It is possible that we have failed to reject H_0 when it is in fact false. This would be a Type II error.

UFO Sightings

Roswell, NM is home to a number of stories about sightings of unidentified flying objects (UFOs). The local police department has historically recorded an average of 10 reported sightings per month. Over the past year, a TV show about aliens has been popular on Netflix. During the period when the show was available, there have been, on average, 16.29 reported sightings per week (sample SD = 3.79). The police officers would like to know if there has been an increase in the number of reported UFO sightings since the TV show became available on Netflix.



1. Describe the shape of the distribution of the number of UFO sightings. Will the median be less than, greater than, or approximately equal to the sample mean?

   The distribution is relatively symmetric, with a mean around 16 and a standard deviation of almost 4. The sample median (which is 16.5) is approximately equal to the sample mean.

2. Formulate an appropriate research question

   Are there more reported UFO sightings since the show became popular on Netflix?

3. What is the parameter, in words?
   The long-run average number of reported UFO sightings since the show became popular on Netflix?

4. Using appropriate mathematical notation, write out H_0 and H_A.

   $H_0 : \mu = 10$
   $H_A : \mu > 10$

5. What are the values of the following variables? If the value is not known, write "unknown".

You can use values under H_0 where applicable.

$\mu =$ *10* $\sigma =$ *unknown*

$\overline{x} =$ *16.29* $s =$ *3.79*

6. Calculate the standardized statistic.

$$t = \frac{\overline{x} - \mu}{s/\sqrt{n}} = \frac{16.29 - 10}{3.79/\sqrt{52}} = \frac{6.29}{0.53} = 11.97$$

7. What are the validity conditions for a one-sample t-test?

The variable should have a symmetric distribution, or there should be at least 20 observations and the sample distribution should not be strongly skewed.

8. Do you think the validity conditions for theory-based inference are met? Why or why not?

There are more than 20 observations, and the sampling distribution is not strongly skewed, so the validity conditions are met. We can perform theory-based inference.

9. (Assume the validity conditions are met) What is your conclusion about the number of reported UFO sightings?

As t = 11.97, I reject H_0 and conclude that there is very strong evidence that the long-run average number of reported UFO sightings is greater than 10.

10. Which type of error is described?
    - Concluding that the number of UFO sightings reported to the police is higher than normal when in fact it is not.

      $\boxed{\text{Type 1 error}}$ Type 2 error

    - Concluding that we do not have enough evidence to reject H_0, when in fact the number of accidents is higher than normal.

      Type 1 error $\boxed{\text{Type 2 error}}$