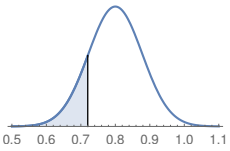
**Math 213 ‡ Practice Test #2 Solutions ‡ Fall 2021**

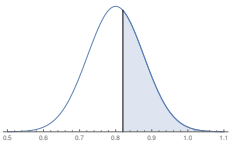
---

(1) Which of the following statements about probability distributions is/are <u>true</u>?
   (a) The area under the entire distribution curve is 1.
   (b) The distribution is never negative.
   (c) All distributions are symmetric.
   (d) 95% of the area under the distribution curve is within 2 standard deviations of the mean.

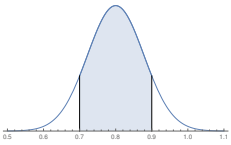
(2) The length of the thorax of a population of fruit flies is normally distributed with mean 0.8 mm and standard deviation 0.078 mm.
   (a) What proportion of the fruit flies have thorax length less than 0.72 mm?
   Start by drawing a picture, this helps us determine what calculation to make.



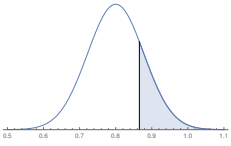

   The R command is `pnorm(0.72, mean = 0.8, sd = 0.078)`. The result is 0.1525304.
   (b) What proportion of the fruit flies have thorax length greater than 0.82 mm?
   Start by drawing a picture, this helps us determine what calculation to make.



   The R command is `1-pnorm(0.82, mean = 0.8, sd = 0.078)`. The result is 0.398817.
   (c) What proportion of the fruit flies have thorax length between 0.7 and 0.9 mm?
   Start by drawing a picture, this helps us determine what calculation to make.



   The R command is `pnorm(0.9,0.8,0.078)-pnorm(0.7,0.8,0.078)`. (As long as you keep the inputs in the order $x$, mean, standard deviation you don't have to include `mean =` and `sd =`.) The result is 0.8001753.
   (d) We wish to select the fruit flies with the highest 20% of thorax length. What is the shortest thorax length we should consider?
   Start by drawing a picture, this helps us determine what calculation to make.



   The R command is `qnorm(0.8, 0.8, 0.078)`. The result is 0.8656465.

(3) Which of the following statements about z-scores is/are <u>true</u>?

(a) larger z-scores are always better
(b) the z-score for an observation that is equal to the mean is 1
(c) if a z-score is 2 that means that the observation is two times the mean
(d) if a z-score is negative that means that the observation is less than the mean
(e) none of the above are true

(4) The distribution of rhesus monkey tail lengths is bell-shaped, unimodal, and approximately symmetric. The average tail length is 6.8 cm and the standard deviation is 0.44 cm. Roscoe has a tail that is 10.2 cm long. What conclusion can we make based on the information given?
(a) We can apply the empirical rule to conclude that Roscoe is a potential outlier because he falls more than three standard deviations away from the mean.
(b) We can apply the empirical rule to conclude that Roscoe is not a potential outlier because he falls within three standard deviations away from the mean.
(c) We cannot apply the empirical rule because the distribution does not fit the criteria for the empirical rule.
(d) There is not enough information given to make any conclusions about potential outliers.

(5) Based on a random sample of 120 rhesus monkeys, a 95% confidence interval for the proportion of rhesus monkeys that live in a captive breeding facility and were assigned to research studies is (0.67, 0.83). Which of the following is true?
(a) 95 of the sampled monkeys were assigned to research studies
This is false. The middle of the confidence interval is 0.75 and $0.75 \times 120 = 90$, so from the sample, 90 monkeys were assigned to studies.
(b) the margin of error for the confidence interval is 0.16
This is false. The width of the whole confidence interval is 0.16, the margin of error is half that, since it is the distance form the middle of the confidence interval to the ends.
(c) a larger sample size would yield a wider confidence interval
This is false. Larger samples yield tighter confidence intervals.
(d) if we used a different confidence level, the interval would not be symmetric about the sample proportion
This is false. Confidence intevals for proportions are symmetric. (This is somewhat by convention. When a normal model is not appropriate, a symmetric confidence interval might not be either.)
(e) none of the above are true
This is the answer.

(6) Approximately 19% of physics majors in the US are women. A random sample of 50 physics majors from all Colorado universities with majors in physics includes 23 females.
(a) What is your point estimate for the proportion of Colorado physics majors who are female?
$23/50 = 0.46$

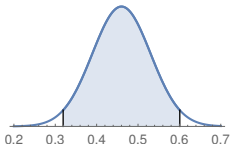
(b) Using a normal model for the proportion, what is the standard error in your estimate?
Using the formula:

$$\sqrt{\frac{0.46(1-0.46)}{50}} \approx 0.07$$

(c) Give a 95% confidence interval for the proportion of potential physics majors at Western who are female.
For 95% confidence our $z^*$ is about 1.96. See page 103 in your text.

$$0.46 \pm 1.96 \times 0.07 = 0.46 \pm 0.138 = (0.322, 0.598)$$

We can draw a picture here too.




(d) If you would like your margin of error to be at most ±5% how many physics majors would you have to include in your sample?
We use the formula that gives margin of error, set it to 0.05, and include the sample size as a variable.

$$0.05 = 1.96\sqrt{\frac{0.46(1-0.46)}{n}}$$

Solve for $n$.

$$n = 1.96^2 \frac{0.46(1-0.46)}{0.05^2} = 381.7$$

So we will need to sample 382 physics majors.

(7) The World Bank reports that 1.7% of the US population lives on less than $2 per day. A policy maker claims that this number is misleading because of variation from state to state and rural to urban. To investigate this, she takes a random sample of 100 households in Atlanta to compare with the national average and finds that 2.1% of the Atlanta population live on less than $2/day. Select the null and alternative hypothesis to test whether Atlanta differs significantly from the national percentage.
(a) $H_0$: $p = 2.1$, $H_a$: $p \neq 2.1$
(b) $H_0$: $\mu = 0.021$, $H_a$: $\mu \neq 0.021$
(c) $H_0$: $p = 1.7$, $H_a$: $p \neq 1.7$
(d) $H_0$: $p = 0.017$, $H_a$: $p \neq 0.017$
This is the correct answer. The null hypothesis is that Atlanta has the same percentage as the rest of the United States. This is a better answer than (c) because we prefer to express proportions as decimals.
(e) $H_0$: $\mu = 2$, $H_a$: $\mu \neq 2$

(8) Complete the following sentence: When conducting a hypothesis test, we _____ and then evaluate the test results to determine if there is enough evidence to _____.
(a) Assume that the null hypothesis is false; accept the null hypothesis
(b) Assume that the null hypothesis is true; reject the null hypothesis

(c) Assume that the alternative hypothesis is true; reject the null hypothesis
(d) Assume the alternative hypothesis is false; reject the alternative hypothesis

(9) Approximately 8% of Colorado residents have been infected with COVID-19. Which of the following are true?
(a) If we take samples of size 20, the sampling distribution for the proportion of Coloradans who have been infected with COVID-19 will be right skewed.
(b) If we take samples of size 200, the sampling distribution for the proportion of Coloradans who have been infected with COVID-19 will be right skewed.
(c) A sample of 200 Coloradans of whom 50 have been infected with COVID-19 would be considered unusual.
(d) A sample of 200 Coloradans of whom 20 have been infected with COVID-19 would be considered unusual.

(10) A psychologist wants to determine if socioeconomic status is related to game playing preferences. Sixty children, in total, were identified from families of low, middle, and high socioeconomic status (20 each), and then the children were asked to select one of Monopoly, Battleship, or Connect Four. The psychologist computed the test statistic $\chi^2 = 5.2$. The proportion of a theoretical $\chi^2$ distribution with 4 degrees of freedom that is greater than 5.2 is approximately 0.2674. What can we say about the $p$-value, $H_0$, and the conclusion at the $\alpha = 0.05$ level of significance?
(a) $0.05 < p$-value $< 0.1$; reject $H_0$; there is evidence of an association between socioeconomic status and game preference
(b) $p$-value $> 0.3$; fail to reject $H_0$; no evidence of an association between socioeconomic status and game preference
(c) $0.2 < p$-value $< 0.3$; fail to reject $H_0$; no evidence of an association between socioeconomic status and game preference
(d) $0.2 < p$-value $< 0.3$; fail to reject $H_0$; there is evidence of an association between socioeconomic status and game preference
(e) $0.05 < p$-value $< 0.1$; fail to reject $H_0$; no evidence of an association between socioeconomic status and game preference

(11) A coin is flipped 1000 times. It comes up heads 532 times. Is this a fair coin?
(a) Give appropriate null and alternative hypotheses.
$H_0 : p_{heads} = 0.5$
$H_a : p_{heads} \neq 0.5$
(b) Give the test statistic and $p$-value for the test.
If we do a proportions test using a normal model, the test statistic is the observed proportion of heads: 0.532. The $p$-value is 0.04635 using `prop.test(532, 1000)`
(c) Give a 95% confidence interval for the probability that the coin comes up heads.
Also from the `prop.test(532, 1000)`: $(0.5005103, 0.5632409)$.
(d) Clearly interpret your results in a sentence.
The coin seems slightly biased towards heads. At the $\alpha = 0.05$ significance level, we reject the null hypothesis that the coin is fair.

(12) The table below describes residents of an Atlanta neighborhood based on their car ownership and public transportation usage.

| | Owns car | Does not own car | Total |
|---|---|---|---|
| Uses public transport | 34 | 94 | 128 |
| Does not use public transport | 126 | 17 | 143 |
| Total | 160 | 111 | 271 |

(a) If there is no association between car ownership and public transportation usage, how many individuals would we expect to *not* own a car and *not* use public transport?

If these two variables are independent, then

$P(\text{no car AND doesn't use public transit}) = P(\text{no car}) \times P(\text{doesn't use public transit})$

From the Total row and column:

$P(\text{no car}) = 111/271, \qquad P(\text{doesn't use public transit}) = 143/271$

So we expect $111/271 \times 143/271 \times 271 = 58.57196$ if "owns car" and "uses transit" are independent.

(b) Perform a hypothesis test to analyze if car ownership and public transportation usage are independent.

The following test of equality of proportions shows that with a $p$-value less than 0.05 (actually much less) the proportion of public transit users is different between car owners and non car owners.

```
 >prop.test(c(34,94),c(160,111))
2-sample test for equality of proportions with continuity correction
data: c(34, 94) out of c(160, 111)
X-squared = 103.28, df = 1, p-value < 2.2e-16
alternative hypothesis: two.sided
95 percent confidence interval:
-0.7342060 -0.5344877
sample estimates:
prop 1 prop 2
0.2125000 0.8468468
```

(13) An ecologist hypothesizes that a lake's fish population is stable when the ratios of two types of fish are 3:2. The ecologist samples the fish in the lake collects the following data.

| **fish type** | **count** |
|---|---|
| fish A | 58 |
| fish B | 22 |
| **total** | 80 |

Do a hypothesis test to evaluate this model.

(a) State your null hypotheses in words.

The ratio of fish A to fish B in this lake is 3:2.

(b) What test statistic could you calculate from the sample to assess the validity of your null hypothesis?

We could calculate what proportion of the fish are type A. We would expect this to be 3/5 if the null hypothesis is correct. There are other correct answers. For example, we could look at the proportion of fish B.

(c) How many of fish A do we expect to find out of 80 total fish if the 3:2 model is correct?
$3/5 \times 80 = 45$

(d) State your null hypothesis as a mathematical expression. ($H_0 : \ldots$)
$H_0 : p_{fishA} = 0.6$

(e) What is the expected sampling distribution of your test statistic if your null hypothesis is true?
We can assume a normal model for the proportion, in which case expect to see $p_{fishA}$ distributed as $N(0.6, 0.055)$. (Using $\sqrt{\frac{0.6(1-0.6)}{80}} \approx 0.055$.)

(f) What are the observed value of the test statistic and the $p$-value from your hypothesis test?
We observe $p = 58/80 = 0.725$ and `2*(1-pnorm(0.725, 0.6, 0.055))` gives a $p$-value of 0.02304262.

(g) What is your conclusion based on your test?
It seems that the ratio of fish A to fish B is not as expected. We can reject our null hypothesis at the $\alpha = 0.05$ level.

(14) An ecologist wants to know if the distributions of two types of fish are the same in two lakes. She collects the following data.

| **fish type** | **Blue Lake** | **Green Lake** | **totals** |
|---|---|---|---|
| fish A | 65 | 40 | 105 |
| fish B | 41 | 34 | 75 |
| **totals** | 106 | 74 | 180 |

Do a hypothesis test to answer the ecologist's question.

(a) State your null hypotheses in words.
Fish type and lake are independent. In other words, the ratio of fish A to fish B or the proportion of either fish are the same in Green Lake and Blue Lake.

(b) What test statistic could you calculate from the sample to assess the validity of your null hypothesis?
We could calculate the difference in the proportion of fish that are type A in Green Lake and Blue Lake.

(c) How many of fish A do we expect to see in Green Lake if the distributions are the same in both lakes?
Looking at the totals row and column:

$$P(\text{fish A}) = 105/180 \approx 0.583$$

$$P(\text{Green Lake}) = 74/180 \approx 0.411$$

So we expect $105/180 \times 74/180 \times 180 = 43.17$ fish A in Green Lake.

(d) State your null hypothesis as a mathematical expression. ($H_0 : \ldots$)

$$H_0 : p_{\text{Green}} - p_{\text{Blue}} = 0$$

Where $p_{\text{Green}}$ and $p_{\text{Blue}}$ are the proportions of fish A in Green and Blue Lakes.

(e) What is the expected sampling distribution of your test statistic if your null hypothesis is true?
We can assume a normal model for this difference in proportions. The mean of our sampling distribuition is 0 and the standard error can be calculated using the formula on page 129.

$$SE_{\text{diff}} = \sqrt{\frac{0.583(1-0.583)}{74} + \frac{0.583(1-0.583)}{106}} = 0.07469126$$

(f) What are the observed value of the test statistic and the $p$-value from your hypothesis test?
We observe $p_{\text{diff}} = 40/74 - 65/106 = -0.07266701$. Calculating `2*(pnorm(-0.07266701, 0, 0.07569126))` gives a $p$-value of 0.3370326.

(g) What is your conclusion based on your test?
This large $p$-value means that we can keep our null hypothesis - there is no evidence that the ratios of the fish differ between the two lakes.

(15) In many sports, teams compete in seven game series, where games are played until one team has won four total games. A seven game series takes at most seven games. A theoretical model used to predict the length of a seven game series says that evenly matched teams will conclude the series in 4 games 12.5% of the time, in 5 games 25% of the time, in 6 games 31.25% of the time and in 7 games 31.25% of the time.

During the years 1990-2019, 87 semifinal and final (World Series) baseball series were played in Major League Baseball. Of those series, 15 ended in 4 games, 21 in 5, 29 in 6 games, and 22 took all 7 games to conclude the series.

(a) If the theoretical model is correct, how many of the 87 series would we expect to have ended in 5 games?
$0.25 \times 87 = 21.75$

(b) A chi-squared test comparing these counts to the counts expected from the theoretical model gives a p value of 0.44. Which of the following is/are <u>true</u>?

(i) The theoretical model does not apply to Major League Baseball.
(ii) The difference in the number of series that took five games to complete and the number predicted by the theoretical model is statistically significant.
(iii) The distribution of series lengths from Major League Baseball is not inconsistent with the theoretical model.
(iv) The theoretical model is useful for predicting series lengths.
(v) None of the above are true.