Statistics 201 Exam 2 – Practice Exam
Spring 2024
Chapters 07 – Sampling Distributions
With Key

Disclaimer:

This practice exam is provided solely for the purpose of familiarizing you with the format and style of the Stat 201 exams. There is no explicit or implicit guarantee that the upcoming exam will ask similar questions. If you use the practice exam as your only tool to help you prepare for the upcoming exam, you most likely will not do well on the exam. You should still do the things you would have done if you did not have access to this practice exam, such as re-read the text, go over your class notes, re-work the online homework problems, and look at the list of exam topics provided and make sure that you understand all the concepts listed within it.

NOTE: These questions come from actual older exams. As such, the points on this practice exam may total more than 100 points.

HASLAM
COLLEGE OF BUSINESS
THE UNIVERSITY OF TENNESSEE, KNOXVILLE

Exam Grade: _____**100**__

1. Say we have a fair (two-sided) coin which we will flip.

   i) (2 points) If the first three coin flips come up tails, what is the probability that the fourth coin flip comes up tails?

   0.5

   ii) (2 points) What is the probability of getting five tails in a row?

   $$(0.5)^5 = 0.03125$$

2. Suppose we have two events A and B with P(A)=.3 and P(B)=.5.

   i) (3 points) Find P(A or B), assuming A and B are disjoint (i.e., mutually exclusive).

If disjoint,

$$\mathbf{P(A\ or\ B) = P(A) + P(B) = 0.3 + 0.5 = 0.8}$$

   ii) (3 points) Find P(A and B), assuming A and B are independent.

If independent,

$$\mathbf{P(A\ and\ B) = P(A) \times P(B) = 0.3 \times 0.5 = 0.15}$$

3. (3 points) Suppose we are considering a person's drive to work one day. Let Event C represent this person being rear-ended by another vehicle before they get to work. Let event D represent this person getting to work without having any sort collision with anything. Let's assume the following:

   P(C) = 0.0002
   P(D) = 0.9995

   Find P(C and D).

   C and D are mutually exclusive events!

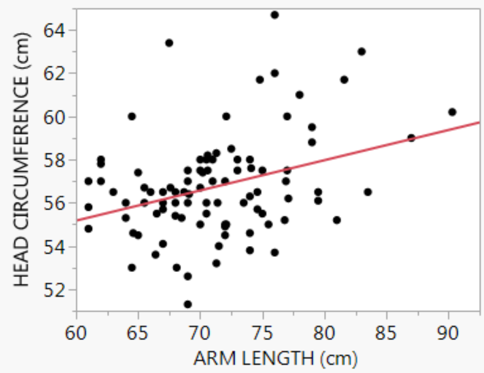
   $$\mathbf{P(C\ and\ D) = 0}$$

4. In Stat 201, a group of students measured their head circumference and their right arm length. Below is some output from JMP for a regression analysis, using:

X = right arm length (cm)

Y = head circumference (cm):

Bivariate Fit of HEAD CIRCUMFERENCE (cm) By ARM LENGTH (cm)



Linear Fit

Linear Fit

HEAD CIRCUMFERENCE (cm) = 46.784076 + 0.1398624*ARM LENGTH (cm)

Summary of Fit

RSquare	0.115974
RSquare Adj	0.106766
Root Mean Square Error	2.195403
Mean of Response	56.75612
Observations (or Sum Wgts)	98

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	46.784076	2.818699	16.60	<.0001*
ARM LENGTH (cm)	0.1398624	0.039411	3.55	0.0006*

i) (1 point) Based on the output above, is the linear relationship between these two variables"statistically significant"? Circle one: YES NO

ii) (2 points) Circle the ONE value in the output that lead to your conclusion in part (i).

iii) (3 points) Regardless of your answers above, briefly describe what it means for a linearrelationship to be "statistically significant". Limit your answer to two or less sentences.

The amount of linear associated in the data is unlikely due to pure chance.

5. The following regression equation was developed after a large number of nursing students agreed to share their latest medical records with us:

Height (in.) = 54.55 + 0.085*Weight (lbs.)

i) (2 points) Mary was one of the nursing students that were part of this study. Suppose Mary weighs 110 pounds. Use the regression equation above to predict Mary's height (in inches).

$$\widehat{\textbf{Height}} = \mathbf{54.55 + 0.085(110lbs) = 54.55 + 9.35 = 63.9\ inches}$$

ii) (2 points) If Mary is actually 65 inches tall, what would the residual for this observation be?[Note: if you have no answer for part (i), use 60.4 inches, which is NOT the correct answer for part (i).]

$$\boldsymbol{e = y - \hat{y} = 65 - 63.9 = 1.10}$$

iii) (3 points) Interpret the slope of this regression equation in the context of this problem.

For additional pound a nursing student was, they were 0.085 inches taller, on average.

iv) (4 points) John, another nursing student that was part of this study, is 68 inches tall. We used the above regression equation to find John's predicted height, and then calculated theresidual for this observation, and found that the residual was -1.47. How much does John weigh (in pounds)?

$$\boldsymbol{y = 68\ \textbf{inches}, e = -1.47, x = ?\ \textbf{pounds}}$$
$$\boldsymbol{e = y - \hat{y}\ \textbf{and}\ \hat{y} = 54.55 + 0.085(\mathrm{x})}$$

Combine equations:
$$\boldsymbol{e = y - (54.55 + 0.085(\mathrm{x}))}$$

Plug in y and e:
$$\boldsymbol{-1.47 = 68 - (54.55 + 0.085(x))}$$

Solve for x:
$$\boldsymbol{-1.47 = 68 - 54.55 - 0.085(\mathrm{x})}$$
$$\boldsymbol{-1.47 = 13.45 - 0.085(\mathrm{x})}$$
$$\boldsymbol{-14.92 = -0.085(\mathrm{x})}$$
$$\boldsymbol{x = 175.53\ \textbf{pounds}}$$

6. A major airline wants to find out if additional charges for baggage affects the airline passengers choose to fly with and how often they fly. They decide to survey passengers who are flying on their airline.

i) (3 points) The airline decides to randomly select 30 of their flights over a one week period and survey all passengers on those flights. What sampling method is beingused?

a) Simple Random Sample
b) Convenience
c) Cluster
d) Systematic
e) Stratified

ii) (3 points) What is the sampling frame?

a) All passengers that fly on any airline.
b) All passengers that fly this particular airline.
c) All passengers that flew this particular airline during the week they did thesurvey.
d) All passengers on the 30 randomly selected flights.
e) All passengers that chose to participate in the survey.

iii) (4 points) Explain one potential source of bias using this sampling method.

If this airline does not charge an extra fee, many of their customers may have chosen this airline because of that. They have not sampled any airline passengers that fly as airline that does charge an extra fee for bags. Their data will probably be biased toward a large % of passengers choose an airline based on no baggage fee.

There are probably many other good answers here!

iv) (3 points) Suppose instead that the airline decides to survey passengers on all of its flights on a single day by selecting every 10th passenger entering each of its planes totake the survey. What sampling method is being used?

a) Simple Random Sample
b) Convenience
c) Cluster
d) Systematic
e) Stratified

7. (5 points) A high school student is interested in attending an "Ivy League" school. As long as a student has a high school GPA of 4.0 and an SAT score of at least 1300, there is a 38% chance of getting accepted to any one of these institutions. One can assume that being accepted or rejected is independent across the eight Ivy League colleges.

   What is the probability that this student is accepted to at least one of these schools, provided the student has met the acceptance requirements and the GPA and SAT criteria described above?

$$\mathbf{P(Accepted) = 0.38}$$

$$\mathbf{P(Not\ Accepted) = 1 - 0.38 = 0.62}$$

$$\mathbf{P(Accepted\ to\ AT\ LEAST\ ONE\ out\ of\ the\ 8)}$$
$$\mathbf{= 1 - P(Accepted\ to\ NONE\ of\ the\ 8)}$$

$$\mathbf{= 1 - [(0.62)^8] = 1 - 0.0218 = 0.9782}$$

8. The owner of a new apartment building must install 50 water heaters. Let's say that the proportion of all Always Hot brand water heaters that will last 10 or more years is 0.25. We can think of the 50 Always Hot brand water heaters that this owner actually buys as a random sample from all Always Hot brand water heaters. After 10 years go by, the owner will know the proportion of water heaters (out of the 50 she purchased) that actually lasted 10 years. This proportion can be thought of as a "sample proportion".

   i) (3 points) In this example, the "sampling distribution of the sample proportion" refersto (circle the best answer):

      a) The distribution of all possible sample proportions from samples of size n=50.
      b) The exact value of the sample proportion this owner will calculate.
      c) The distribution of all possible sample proportions for all possible sample sizes.
      d) The exact value of the population proportion.
      e) The distribution of the population proportion.

   ii) (6 points) To be able to know what the mean, standard deviation and shape of the sampling distribution of the sample proportion is, three conditions must be checked and met. List each condition, and briefly comment on whether or not each condition is metin this case.

Condition 1 is: **Randomization**

Is condition 1 met? (Circle one) Yes No

Briefly explain your answer:

The problem states that we can treat these 50 water heaters as a random sample

Condition 2 is: **10% Condition**

Is condition 1 met? (Circle one): Yes No

Briefly explain your answer:

As long as the company has made more than 500 water heaters, we are sampling <u>less</u> than 10% of their water heaters.

Condition 3 is: **Success/Failure Condition**

Is condition 1 met? (Circle one): Yes No

Briefly explain your answer:

$$\mathbf{n = 50, p = 0.25, q = 0.75}$$
$$\mathbf{np = 50(0.25) = 12.5}$$
$$\mathbf{nq = 50(0.75) = 37.5}$$

Both are above 10, so the condition is met

Question 8 (continued):

iii) (6 points) Regardless of your answers above, assume the appropriate conditions aremet. What is the (a) mean, (b) standard deviation and (c) shape of the sampling distribution of the sample proportion in this case?

a) = $\mu(\hat{p}) = p = 0.25$

b) = $SD(\hat{p}) = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.25(0.75)}{50}} = 0.06124$

c) = **Normal**

iv) (2 points) Suppose that after 10 years pass, she finds that 16 of the 50 water heatersstill work. Calculate the sample proportion.

$$\hat{p} = \frac{X}{n} = \frac{16}{50} = 0.32$$

v) (3 Points) Using your answers from parts (iii) and (iv), calculate the z-score for thisobservation.

[Note: if you have no answers for part (iii) and/or part (iv), make up answers for thoseparts here, and use those values to answer this question.]

$$Z = \frac{\hat{p} - p}{SD(\hat{p})} = \frac{0.32 - 0.25}{0.06124} = 1.143$$

vi) (4 points) Given your answer to part (v), would you consider having 16 or more outof 50 Always Hot brand water heaters still working after 10 years an “unusual” event?

Circle one: YES NO

Briefly explain your “yes” or “no” choice below. Limit your answer to two or less sentences.

The sample proportion is just barely more than 1 standard deviation above average. Anything close to 1 standard deviation from the mean is considered “typical”

9. Suppose one takes many, many repeated random samples of size n=50 from a very large population (of quantitative observations) that is slightly skewed right, has a mean of 127.5 and a standard deviation of 28.5.

i) (2 points) What would be the average of the sampling distribution of the sample mean?

$$\boldsymbol{\mu(\bar{y}) = \mu = 127.5}$$

ii) (2 points) What would be the standard deviation of the sampling distribution of thesample mean?

$$\boldsymbol{SD(\bar{y}) = \frac{\sigma}{\sqrt{n}} = \frac{28.5}{\sqrt{50}} = 4.0305}$$

iii) (2 points) Assume all the of necessary conditions are met for the Central Limit Theorem to apply to the sampling distribution of the sample mean (which, they ARE met in this case). Given that the Central Limit Theorem applies, what would be the shape of the sampling distribution of the sample mean?

Meeting all 3 conditions (Randomness, 10% Condition, Large Enough Condition) means the sampling distribution of the sample mean will be approximately Normal.

iv) (6 points) In one particular sample of n=50, a sample mean of 150.7 was obtained. Would you consider this to be a "rare" or "unusual" value of the sample mean? Use your answers above to help you answer this question. Provide whatever arithmetic you need to help you answer this question, and be sure to show your work.

Circle one: "rare event" "not a rare event"

Briefly justify the choice you circled above. Limit your answer to two or less sentences.

(**NOTE 1**: If you have NO answers to parts [i] and/or [ii], make up answers for thoseparts here, and use that to answer this question. **NOTE 2**: If you have a calculator that can calculate the probability of this event, do NOT use that capability!)

$$\boldsymbol{t = \frac{\bar{y} - \mu_0}{SD(\bar{y})} = \frac{150.7 - 127.5}{4.0305} = 5.756}$$

The z-score is well over 3 which indicates this is a very rare event

True/False

Circle the best answer

(2 points each)

T F Extrapolation occurs when someone uses a linear regression model to make a prediction outside the original scope of the independent variable, x.

T F Event A is “a randomly selected student from Brian’s class makes an A in Stat 201”. Event B is “a randomly selected student from Brian’s class is a sophomore”. Events A & B are mutually exclusive events.

T F If Event A and Event B are independent, knowing the outcome of Event A changes the probability of Event B occurring.

T F Stratified Random Sampling occurs when the researcher picks a random starting point, and then samples every kth element or observation from the sampling frame.

T F Two events that are independent must have equal probabilites of occuring.

T F Comment cards at a restaurant are an example of convenience sampling.