

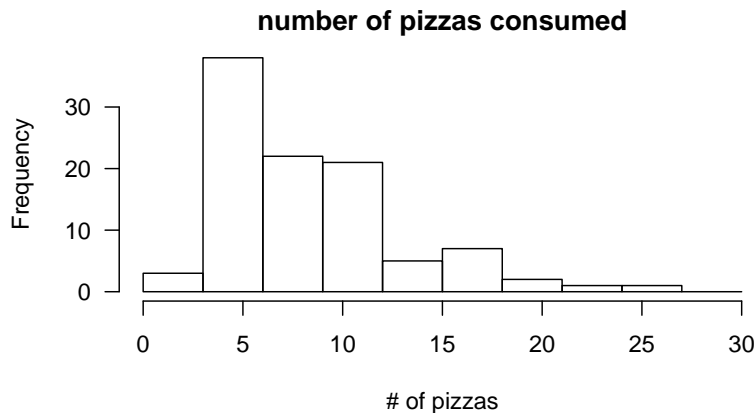
Review for Final (6/5, 3:30-6:20p, Young Hall 2200)

David Diez

- (1) The table shows data from a sample of students from UCLA.
- (a) What percent of the sampled students are male?
 - (b) What proportion of sampled students are social science majors or male?
 - (c) If we wanted to check if the proportion of males who are social science majors is different from the proportion of females that are social science majors, should we look at column or row proportions?
 - (d) Are gender and whether a student is a social science major independent variables?

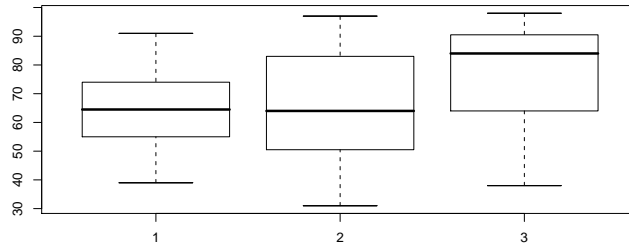
	social science major	other major	TOTAL
female	25	35	60
male	30	27	57
TOTAL	55	62	117

- (2) Describe the histogram of the number of pizzas consumed by each sampled UCLA student last year. Include shape, any unusual features, and if it is unimodal, bimodal, or multimodal. Would the mean or median be more appropriate as a measure of center here? Would the standard deviation or IQR be a more appropriate measure of variation?



- (3) Match the shape of the distribution with the proper case:
- (a) symmetric distribution.
 - (b) right skewed.
 - (c) left skewed.
 - (i) mean greater than the median
 - (ii) mean and median are the same
 - (iii) mean less than the median

- (4) Identify the boxplot with:
- the highest median,
 - the highest mean,
 - the least variability,
 - skew (and for the skewed plot, what direction is it skewed?).



(5) Chipotle burritos are said to have a mean of 22.1 ounces, a st. dev. of 1.3 ounces, and they follow a normal distribution. Additionally, the weight was found to be independent of the type of burrito ordered. Chipotle advertises their burritos as weighing 20 ounces.

- What proportion of burritos weigh at least 20 ounces?
- Your other friend gets a burrito and finds that it weighs 25.5 ounces. (Your friend secretly carries a scale everywhere.) What proportion of burritos weigh at least this much?
- Suppose your friend says that the last 6 times she has been at Chipotle, her burrito weighed between 22 and 23 ounces. What is the chance of it happening again the next time she goes?
- What is the chance exactly 3 of your next 5 burritos from Chipotle will weigh more than 24 ounces?
- If your friend said she got a burrito at the 97th percentile, how much did her burrito weigh?

(6) Below is a proposed probability distribution for the number of bags that get checked by individuals at American Airlines. No one checks more than 4 bags.

number of bags checked	0	1	2	3	4
probability	0.42	0.29	0.22	0.05	0.02

- Is this a probability distribution? (Explain.) If you confirm it is, assume it is accurate.
- Picking out a passenger at random, what is the chance s/he will check more than 2 bags?
- American Airlines charges \$15 for the first checked bag, \$25 for the second checked bag, and \$50 for each additional bag (beginning this summer). How much does American Airlines expect to take in revenue for baggage costs for a single passenger? How much would they expect to get for 20 passengers? (Note: the cost of 2 bags is $\$15 + \$25 = \$40$).

- (7)** You roll a die several times in a row. What is the probability that
- (a) your sixth roll is the first 3 you get.
 - (b) five rolls give only even numbers.
 - (c) the fourth roll is the first 3 you get.
 - (d) the fourth roll is a 2 or a 3.
 - (e) What is your expected number of rolls before getting the first '1'?
- (8)** It is known that 80% of people like peanut butter, 89% like jelly, and 78% like both.
- (a) If we pick one person at random, what is the chance s/he likes peanut butter or jelly?
 - (b) How many people like either peanut butter or jelly, but not both?
 - (c) Suppose you pick out 8 people at random, what is the chance that exactly 1 of the 8 likes peanut butter but not jelly?
- (9)** The Dow Jones Industrial Average and NASDAQ composite are used to measure the state of the stock market. During 2007, the Dow rose on 58% of trading days. On the days the Dow did rise, the NASDAQ rose 74% of the time. On the days the Dow fell, the NASDAQ also fell 76% of the time. For this example, suppose that 2007 was a typical year in the sense that these numbers are accurate over the long run.
- (a) Are the Dow and NASDAQ independent?
 - (b) Compute the percent of days we expect the NASDAQ to rise.
 - (c) If we hear that the NASDAQ rose, what is the probability that the Dow also rose?
- (10)** It is known that 27% of Americans smoke.
- (a) If we pick one person (American) at random, what is the probability s/he does not smoke?
 - (b) If we pick twelve people at random, what is the chance precisely one of them smokes?
 - (c) If we randomly pick five people, what is the probability the last one is the first one that we find to be a smoker?
 - (d) If we select seven people, what is the chance exactly two of them smoke?
 - (e) Suppose we select seven people. What is the chance we find no more than one smoker?
 - (f) If we select seven people, what is the chance at least two of them smoke?

(11) When reserving a room for a review session at the university, you need to specify how many seats are needed. Suppose that there are 60 people who may attend, and from past experience you know the chance a single individual shows up is 75%.

- (a) What is the expected number of people who will show up? What is the standard deviation of the number of people who will show up? Verify that we can use the normal model.
- (b) If a room seating fifty people is reserved, what is the chance that not everyone will get a seat who attends the review session?
- (c) To be 95% sure the room will be big enough, at least how many seats must there be?

(12) Remember our Chipotle burritos in problem 5? The burritos had mean weights of 22.1 ounces, standard deviation 1.3 ounces, and were normally distributed. Suppose you are setting up a meeting of 6 people and you are ordering Chipotle...

- (a) What is the probability that your burrito weighs less than 21 ounces?
- (b) What is the chance that the average weight of the five burritos is less than 21 ounces?

(13) We would like to poll voters on their preference between John McCain and Barack Obama (making the assumption Obama is the Democratic candidate for president). As a pollster, you know that for a poll to be published by news networks, it must have a margin of error no larger than 5% using 95% confidence.

- (a) Determine the sample size that will ensure this accuracy.
- (b) Using the sample size you found in part (a), you conduct a sample. Suppose you find that 54.1% of likely voters intend to vote for Barack Obama. Find the 95% confidence interval for voters who intend to vote for Barack Obama.

(14) Suppose you conduct a larger sample of Montana with a sample size of 1600 people, which gives a 95% confidence interval for Obama's support as 0.506 to 0.554. What can you say about your confidence interval? (possibly more than one correct choice)

- (a) 95% of voters support Barack Obama between 50.6% and 55.4% of the time.
- (b) 50.6% to 55.4% of voters support Barack Obama 95% of the time.
- (c) One pollster said John McCain would win Montana if the election were held today. The confidence interval supports this claim.
- (d) Another pollster said 51% of people in Montana support Barack Obama. The confidence interval supports this claim.
- (e) We are 95% confident that our sample proportion is between 0.506 and 0.554.
- (f) We are 95% confident Obama's support in Montana is between 0.506 and 0.554.

(15) Al wants to assess job satisfaction within his company. He takes a sample of 6 people and finds that on his survey, the employees averaged 4.67 out of 10 in job satisfaction. Wanting to improve this, Al gives his employees an additional 15 minute break each day. One month later he repeats the survey for 6 people. The data is below:

Person #	Sample 1	Sample 2	Difference
1	4.2	5.2	-1.0
2	5.7	6.5	-0.8
3	4.4	5.7	-1.3
4	5.3	6.6	-1.3
5	3.7	4.6	-0.9
6	4.7	5.9	-1.2
mean	4.67	5.75	-1.08
st. dev.	0.73	0.77	0.21

- If the same employees are sampled and asked about their job satisfaction one month later, is the data paired?
- If a new set of employees is randomly selected for the second job satisfaction survey, is the data paired?
- Run hypothesis tests for (a) and (b), including all steps.

(16) We would like to determine the difference in average femur lengths for birds that died and those that survived. If we suppose Bumpus' study is a simple random sample of all birds that died and survived, use the *Fathom output* from below to

- find the confidence interval for the difference.
- describe whether there is a significant difference for the mean femur length of birds that died and those that survived.

Difference of Means ▾

Estimate of bumpus

First attribute (numeric): Femur_in
 Second attribute (numeric or categorical): Survival

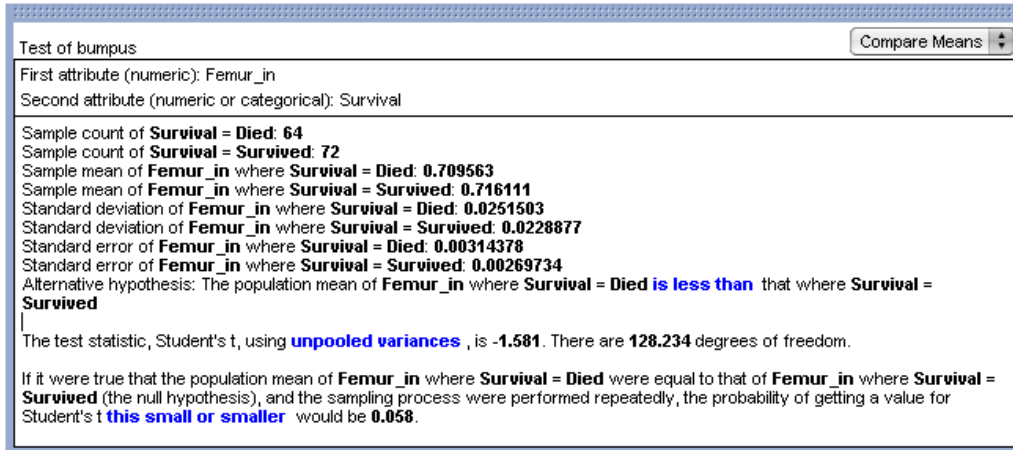
Interval estimate for the difference of means of **Femur_in** for **Survival = Died** and **Survived**.

Sample count of **Survival = Died**: **64**
 Sample count of **Survival = Survived**: **72**
 Sample mean of **Femur_in** where **Survival = Died**: **0.709563**
 Sample mean of **Femur_in** where **Survival = Survived**: **0.716111**
 Standard deviation of **Femur_in** where **Survival = Died**: **0.0251503**
 Standard deviation of **Femur_in** where **Survival = Survived**: **0.0228877**
 Standard error of **Femur_in** where **Survival = Died**: **0.00314378**
 Standard error of **Femur_in** where **Survival = Survived**: **0.00269734**

Based on the samples and using **unpooled variances**, the **95.0** % confidence interval for the mean(**Femur_in** where **Survival = Died**) - mean(**Femur_in** where **Survival = Survived**) is **-0.00654861** plus or minus **0.00819619** ranging from **-0.0147448** to **0.00164758**.

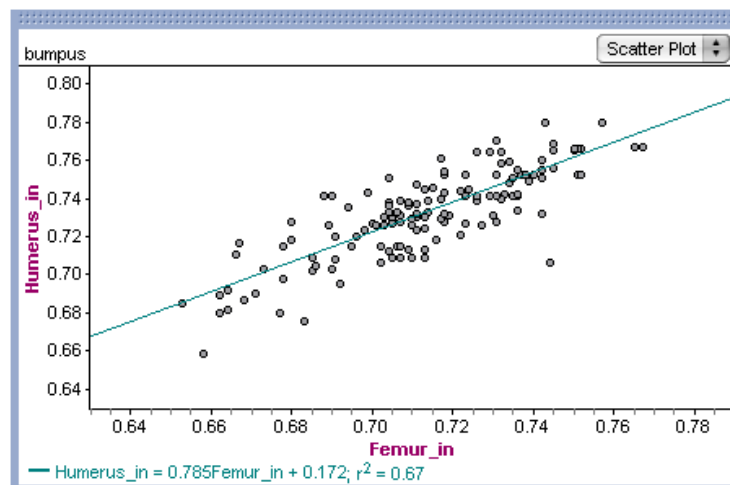
If the sampling process were performed repeatedly, the confidence intervals generated would capture the population difference of means **95.0** % of the time.

(17) The hypothesis test uses the same data as above. Describe the hypothesis test (H_0 and H_A) along with other relevant test information (t , p-value) and whether you would reject or would not reject H_0 . Describe your result in context.

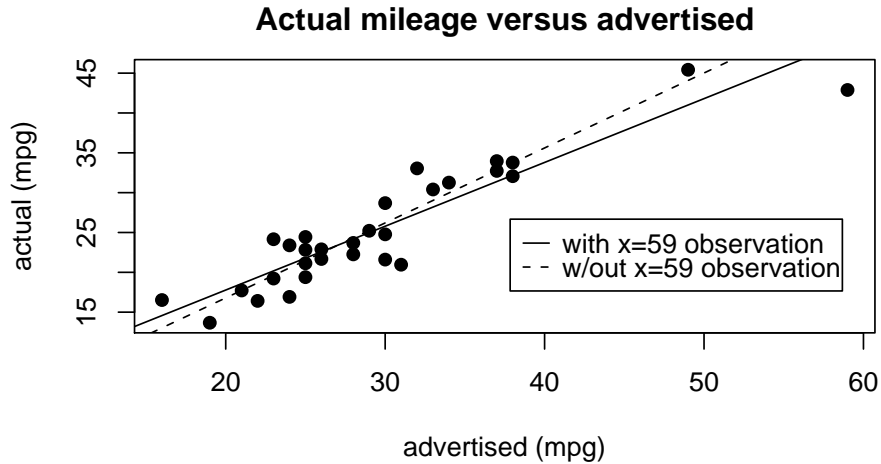


(18) Below is a scatterplot and linear model for Bumpus' data, looking at the humerus and femur length.

- What is the equation for the linear model?
- If a bird has a femur length of 0.704 inches, what is its predicted humerus length? (remember units!)
- If the bird from (b) has an actual humerus length is 0.733 inches, what is its residual? Was the predicted value an overestimate or underestimate?
- What is the correlation for this fit?
- What does r^2 represent?



(19) Data was collected on 30 drivers regarding their vehicle mileage. The advertised fuel economy of the drivers' vehicles was noted, and each driver was asked for the actual fuel economy. The average advertised mileage was 29.6 with a standard deviation of 8.76. The actual mileage achieved was 25.4 with standard deviation 7.61. The correlation between actual and advertised mileage was 0.92.



- (a) Find the equation of the LSR line using the summary information provided.
- (b) What mileage would you predict a driver with a car advertising 19 mpg actual gets?
- (c) The actual mileage for our driver from (b) is 13.7 mpg. What is her residual? Did the model over- or under-predict her mileage?
- (d) Does the point located at (59, 42.9) have high leverage?
- (e) Is the point (59, 42.9) an influential point? The LSR line that includes this observation may be compared with the LSR line that does not include it on the plot above.

(20) While reading comments in an online forum regarding Solomon Asch's study on conformity, Ted came across an interesting comment from a teacher. The teacher noticed that if one student said 'thank you' when she handed a paper back, the following student would almost always say 'thank you' as well.

Although interesting, this result is not exactly surprising. To take it one step further, Ted wondered if the conformity effect was the same between men and women. For his informal study, Ted collected information regarding students saying 'thank you' and did a full statistical analysis using MCMC methods. Although he found a very strong conformity effect as the teacher had mentioned in the forum, he did not find a significant difference between males and females in his classes, testing at $\alpha = 0.05$.

- (a) In Ted's study, what is the explanatory variable?
- (b) What is the response? (pick one)
 - (i) Whether the student attended class to get his/her assignment or not.
 - (ii) If the student gave the same response as the previous student.
 - (iii) The student's gender.
 - (iv) That Ted found no significant difference between men and women in his class.
- (c) Which of the following statements are true?
 - (i) Ted would have made the same conclusion if $\alpha = 0.10$.
 - (ii) Ted would have made the same conclusion at $\alpha = 0.01$.
 - (iii) This was an experiment.
- (d) If there is an actual difference in conformity of this type between men and women, what type of error did Ted make?