

Final Examination (150 minutes)

No calculators allowed. Just set up your answers, for example, $P = 49/52$. If you want extra time, you can buy time at a price of 1 point a minute; for example, if your test is handed in 10 minutes after the scheduled finish time, 10 points will be subtracted from the test score. Illegible answers will be ignored.

1. It was reported that,

Researchers at the University of Adelaide compared adverse events of St. John's wort and the antidepressant drug fluoxetine (Prozac). The team used information from doctors' reports to Australia's national agency on drug safety. Between 2000 and 2013, there were 84 adverse reaction reports for St. John's wort. There were 447 reports for Prozac.

As a statistician, what is the most important reason why you would advise caution in concluding that St. John's wort is safer than Prozac?

2. Identify the most appropriate null hypothesis and statistical test for each of these studies (for example,

H_0 : the average difference is zero

test: matched-pair t-test:

a. A black-box algorithm predicts whether the Dow Jones Industrial Average will go up or down each day.

H_0 :

test:

b. Housing prices depend on square footage, number of bedrooms, and number of bathrooms.

H_0 :

test:

c. Midterm and final exam scores in a statistics class are positively correlated.

H_0 :

test:

d. At-risk patients are less likely to have a fatal heart attack if they take aspirin daily instead of a placebo.

H_0 :

test:

e. Japan's soccer goalie is more likely to dive to the left than to the right on penalty kicks.

H_0 :

test:

3. Identify the most appropriate null hypothesis and statistical test for each of these studies (for example,
H₀: the average difference is zero
test: matched-pair t-test:
- a. An ESP enthusiast guesses the number that will be rolled with a fair six-sided die.
H₀:
test:
- b. Thirty people label the picture of Joe Biden shown on the front page of the *New York Times* and the picture of Joe Biden shown on the front page of the *New York Post* on the day after the 2020 presidential election as “favorable” or “unfavorable.”
H₀:
test:
- c. Essays written by 100 students in a college history class will be graded on a scale of 1 to 100 by the professor and by an AI algorithm.
H₀:
test:
- d. An experienced professor will grade 10 essays written by college students and 10 essays written by an AI algorithm.
H₀:
test:
- e. The prices Hollister and Abercrombie & Fitch charge for essentially the same clothes will be compared.
H₀:
test:
4. For golfers who played the final two rounds of the 2015 and 2016 Masters golf tournament, the correlation between their 2015 and 2016 scores was 0.38. For the top 15 golfers, the correlation between their 2015 and 2016 scores was 0.04. This is an example of
- the law of averages
 - the law of large numbers
 - the central limit theorem
 - self-selection bias
 - the paradox of luck and skill

5. A fair coin was flipped 10 times and the longest streak of consecutive heads or tails was recorded. It turned out to be 7 tails. What is wrong with the calculation? (Do not calculate the correct probability, just identify the error(s) here.)

The probability of 7 tails in a row is 0.5^7 .

6. John Gottman has written several books, given innumerable talks, and, with his wife, created The Gottman Institute for marriage consulting and therapist training. In a 2007 survey of psychotherapists, Gottman was voted one of the ten most influential members of their profession over the past 25 years. In his seminal study, 130 newlywed couples were videotaped while they had a 15-minute discussion of a contentious topic. Gottman went over the videotapes, frame by frame, recording detailed changes in facial expressions, tone of voice, and the like—for example, noting whether the corners of a person's mouth were upturned or downturned during a smile. He then kept in touch with each couple for six years and noted whether they had divorced during that time. After these six years, he estimated a statistical model for predicting divorce based on the codings he had made six years earlier. He reported that his model was 82.5 percent accurate in its predictions. Malcolm Gladwell gushed that, "He's gotten so good at thin-slicing marriages that he says he can be at a restaurant and eavesdrop on the couple one table over and get a pretty good sense of whether they need to start thinking about hiring lawyers and dividing up custody of the children."

Why are you skeptical of Gottman's procedure?

7. A replication study of Gottman's procedure by two psychology professors identified the best divorce predictors based on interviews with 204 couples and then applied these predictors to 204 different couples. They found that of the 167 couples that were still married, 123 had been predicted to be married and that of the 37 couples that were divorced 17 had been predicted to be divorced.

Overall, what was the probability that a couple predicted to become divorced actually became divorced?

8. The 2022 World Cup quarterfinal match between Brazil and Croatia was decided by a penalty shootout in which each team takes five penalty kicks. Before the penalty shootout began, the betting odds were:
- (1) a bet on Brazil: win \$100 if Brazil wins, lose \$125 if Brazil loses
 - (2) a bet on Croatia: win \$110 if Croatia wins, lose \$100 if Croatia loses
- a. For what probability of a Brazil victory does wager (1) have a positive expected value?

b. For what probability of a Croatia victory does wager (2) have a positive expected value?

c. If you think the shootout is essentially just a coin flip with each team having a 50% chance of winning, which of these two wagers has the highest expected value?

9. Do you agree or disagree with the following statements?

- a. "If R-squared is not high enough, you should add more explanatory variables" yes no
- b. "If there is a multicollinearity problem, you should drop some explanatory variables" yes no
- c. "If the null hypothesis is true, the p-value is more likely to be less than 0.10 than 0.05." yes no
- d. "The least squares line goes through the sample means of X and Y." yes no
- e. "The probability of A or B is equal to the probability of A plus the probability of B." yes no

10. A multiple regression model was used to estimate the relationship between a Major League Baseball (MLB) player's salary and his batting average and home runs:

		Mean	Standard Deviation
S = annual salary, millions of dollars		2.5	3.5
B = batting average for season		0.263	0.032
H = number of home runs in season		12.5	10.4
	Coefficient	Standard Error	t-value
constant	-11.70	0.97	12.1
B	13.24	0.98	13.5
H	0.15	0.0031	48.4
R ² = 0.26			

Can we conclude from these results that salaries are affected more by batting average than by home runs, or vice versa? Explain your reasoning.

11. What is the most important problem with this specification of a multiple regression equation to investigate whether birth order affects college grades for English majors?

$$G = \alpha + \beta_1A + \beta_2H + \beta_3D + \varepsilon$$

- G = grade point average (0.0 - 4.0)
 A = age, years
 H = height, inches
 D = 0 if the person is an only child
 = 1 if the person is a first-born child
 = 2 if the person is a second-born child
 = 3 if the person is third-born or later

12. What is wrong with this argument? (Don't calculate the correct probability, just identify the error.)

If a presidential candidate has a 60% chance of winning Ohio, a 60% chance of winning Pennsylvania, and a 60% chance of winning Florida, then the chances of winning 2 of these 3 states is $3(0.6^2)(0.4^1)$

13. A total of 85 randomly selected college students were asked if they had a serious romantic relationship in the past two years and, if so, to identify the month in which the most recent relationship began.

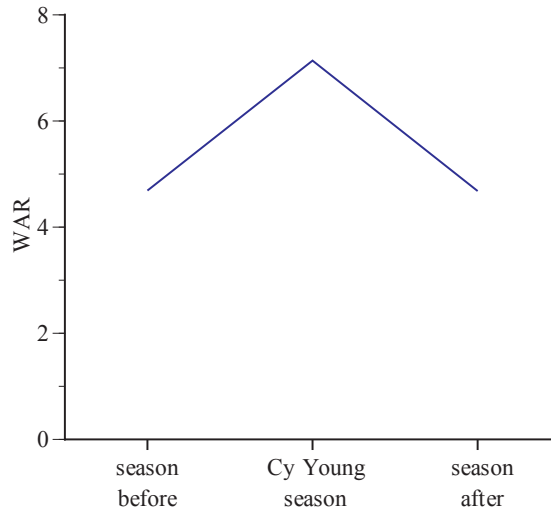
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
observed	10	13	7	3	4	4	6	4	6	11	8	9	85
expected	7.083	7.083	7.083	7.083	7.083	7.083	7.083	7.083	7.083	7.083	7.083	7.083	85

A chi-square test was then used to test the null hypothesis that each month is equally likely for the beginning of a romantic relationship. Identify two distinct errors in the conclusion (do not check any of the mathematical calculations):

The chi-square value is 15.66 and the two-sided p-value is 0.154. Therefore, a chi-square test yielded deviations between observed and expected values that are too small to be explained plausibly by random sampling error.

14. A Cy Young Award is given each year to the best pitcher in Major League Baseball’s American League and National League. Chicago White Sox pitcher LaMarr Hoyt won a Cy Young Award in 1983 and complained in 1984 that he had been jinxed: “I’ll tell you, there have been a lot of times this season I’ve felt jinxed. A lot happens to me that is unexplainable.” An article reporting his grievance noted that the four previous AL Cy Young winners had all had disappointing seasons the year after winning the award.

One of the tools of modern sabermetrics for assessing baseball players is wins above replacement (WAR)—the higher, the better. The figure below shows the average WAR for all Cy Young winners since 2000 during their award-winning season and during the seasons before and after they won the award. Are these data consistent with the jinx theory? How would you explain these data?



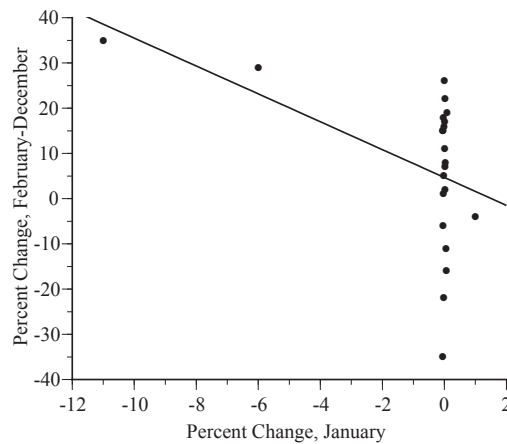
15. A multiple regression model of home prices was specified in two ways:

$$(1): P = \alpha + \beta_1 S + \beta_2 B + \varepsilon$$

$$(2): P = \alpha + \beta_1 S + \beta_2 D_1 + \beta_3 D_2 + \beta_4 D_3 + \varepsilon$$

Where P is the market price, S is the livable square footage, B is the number of bathrooms, $D_1 = 1$ if 2 bathrooms, 0 otherwise; $D_2 = 1$ if 3 bathrooms, 0 otherwise; and $D_3 = 1$ if 4 bathrooms, 0 otherwise. In these data, the square footage ranged from 1,800 to 3,900 and number of bathrooms was either 1, 2, 3, or 4. What is the difference in the assumptions made by these two models?

16. Two researchers used annual data for 2000-2021 to estimate a simple regression model where Y = percentage change in the S&P 500 index of stock prices from February 1 through December 31 and X = percentage change in the S&P 500 from January 1 through January 31:



The authors reported that

$$Y = 4.6695 - 3.0814X, R^2 = 0.22$$

where $b = -3.08$, $t\text{-value} = -2.3898$, $p\text{-value} = 0.1917$

The beta is not substantial because the $t\text{-value}$ is less than 2

We can reject the null hypothesis. Our results show that February-December % change in the S&P 500 does not depend on the % change in January.

Identify 5 distinct problems with this analysis.

-
-
-
-
-

17. What is wrong with the answer to this Car Talk Puzzler?

Monty Hall. Let's Make a Deal. Doors 1, 2, and 3... You select doors without knowing what is behind them. One of the doors has a good prize behind it, the others have crummy prizes. Monty says, "You picked door #1, but you don't know what's behind it. I've shown you what's behind door #2, and it is a bunch of stale Eskimo pies... Would you like to switch to door #3?" What do you do, and why?

Answer: The answer is, you should switch. Monty Hall is always going to show you a crummy door. Right? So you have a 50/50 chance of getting a good door, if you switch. So statistically, if you do this enough times, by switching doors, you up the chance of getting a good outcome.

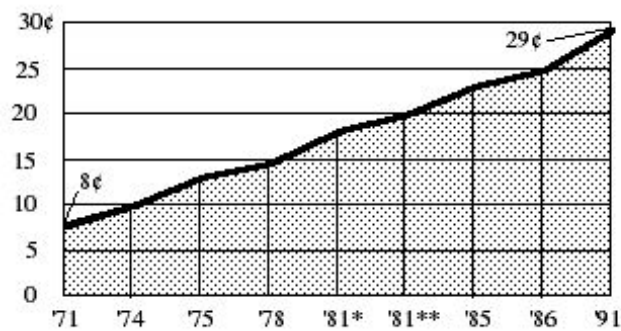
18. Consider a country that has a “one-child-plus” policy. A woman is only allowed to have one baby; however, if the baby is a girl, the woman is allowed to have another baby. No woman is allowed to have more than 2 babies. Assuming that girl and boy babies are equally likely and independent, would the “one-child-plus” policy increase or reduce the ratio of boy babies to girl babies?

19. Explain why you either agree or disagree with this argument:

Don't be discouraged if your job application is rejected. Ninety out of 100 job applications are rejected, so every rejection makes a future job offer more likely.

20. Identify two distinct reasons why this figure from *USA Today* (March 7, 1994) is misleading:

Rising Postal Rates



* March
** November