Review Free Response Problems – Fall Final

KEY

1. Die A has two 5s, three 3s, and one 1 on its six faces; die B has two 2s and four 4s on its faces. Both dice are fair. Each player simultaneously rolls one of the dice, and the winner is the player with the higher number showing.
   a. If you want to win, would you rather roll die R or B? Explain.
   b. If the winner receives whatever shows on her winning die, what is the expected value for one roll to each player? Explain.

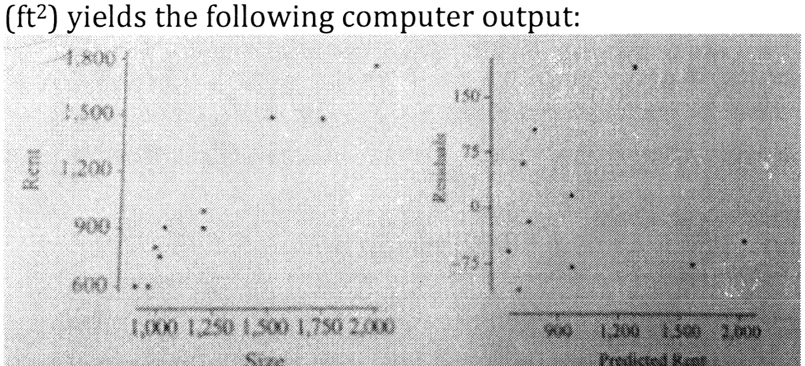
2. High Cholesterol levels can be reduced by low-fat diet or medication. Researchers would like to test the effectiveness of a new medication and to note whether the effectiveness if any, is enhanced by diet. One hundred volunteers with high cholesterol levels, on no special diets, and not on medication, are recruited for a study.
   a. Conclusions will apply to what population?
   b. Explain how you would design a completely randomized experiment.
   c. How might you incorporate blocking and for what purpose?
   d. How might blinding by incorporated in this study and for what purpose?

on back

3. Suppose that jumps by Olympic men high jumpers have a normal distribution with mean 2.12 meters and standard deviation .12 meters; women's jumps have a normal distribution with mean 1.80 meters and standard deviation .09 meters. A man and woman Olympic high jumper are ricked at random.
   a. What is the probability the sum of their jumps is over 4 meters?
   b. What is the probability that the man jumped higher than the woman?

4. Does cell phone radiation increase a person's risk of getting cancer?
   a. How can an observational study by performed to answer this question?
   b. How can an experiment be performed to answer this question?
   c. Which is more appropriate here? Explain.
   d. If performing the experiment, give an ethical consideration, which may halt the experiment early.

5. An SRS of apartment listings in a large northeastern city comparing monthly rent ($) versus size ($ft^2$) yields the following computer output:



Variable	Coef	s.e. Coef	t	p
Constant	-311.341	117.5	-2.65	0.0294
Size	1.07707	0.09047	11.9	0.0001

s = 102.4 R-squ = 94.7% R-sq(adj) = 94.4%

   a. Is a linear model appropriate for these data? Explain.
   b. Interpret the slope of the regression line.
   c. Interpret r^2 in context.

(a) Scatterplot is approximately linear, and residual plot has no pattern so yes!

(b) for each additional square foot in size, the average rise in rent is $1.08.

(c) 94.7% of the variation in rent is explained by variation in size (the least squares regression line.)

1. (a) $P(A\text{wins}) = \frac{1}{3} + \frac{1}{2}\left(\frac{1}{3}\right) = \frac{1}{2}$

$P(B\text{wins}) = \left(\frac{1}{3}\right)\left(\frac{1}{6}\right) + \left(\frac{2}{3}\right)\left(\frac{1}{2} + \frac{1}{6}\right) = \frac{1}{2}$

equal chance of winning

(b) $E(A) = 5\left(\frac{1}{3}\right) + 3\left(\frac{1}{6}\right) = 2\frac{1}{6}$

$E(B) = 2\left(\frac{1}{18}\right) + 4\left(\frac{4}{9}\right) = 1\,8/9$

2. (a) conclusions apply to only non-dieters w/ high cholesterol levels who are not taking medications.

(b) Random assignment to 4 treatments

- medication + diet
- medication and no diet
- placebo and diet
- placebo and no diet

then measure + compare

(c) Block based on age or gender b/c both could effect how volunteers respond to treatment.

(d) Placebo = Blinding

3. (a) $E(m+w) = E(m) + E(w) = 2.12 + 1.80 = 3.92$ meters

$\sigma_{m+w} = \sqrt{\sigma_m^2 + \sigma_w^2} = \sqrt{.12^2 + .09^2} = .15$

* Both → each jump is independent
→ both random variables are normal so sum is normal $N(3.92, .15)$

$P(\underset{m+w}{x} > 4) = \boxed{.2969}$

(b) $E(m-w) = E(m) - E(w) = 2.12 - 1.8 = .32$

$\sigma_{m-w} = \sigma_{m+w} = .15$

$N(.32, .15)$ $\quad P(\underset{m-w}{x} > 0) = \boxed{.9836}$

4. (a) random sample of people - compare prop. of cell phone users to prop. of those that have cancer.

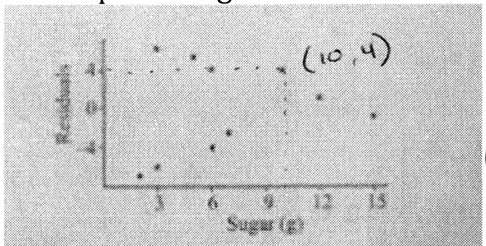
(b) group of non-cancerous people
→ random assignment to cell phone / no cell phone (over a period of time) (then compare proportions)

(c) Observational study = lurking variables / Experiment = cause + effect claim

(d) coming down w/ cancer at a rate statistically significant before it is completed.

←

6. The calories and sugar content per serving size of ten brands of breakfast cereal are fitted with a least squares regression line with computer output:




```
Dependent variable is: Calories
s = 6.544  R-sq = 77.9%  R-sq(adj) = 75.2%
Source      df    SS        MS       F
Regression   1    1210.08   1210.08  28.3
Residual     8    342.547   42.8184
Variable  Coef     s.e.Coef   t     P
Constant  78.3275  4.117      19    0.0001
Sugar     2.74239  0.5159     5.32  0.0007
```

ⓐ $r = .883$ = strong, positive, linear relationship
* residual plot might show a pattern.

ⓑ On average, the calorie content increases by 2.74 for each additional gram of sugar.

ⓒ the average calorie content for cereals w/ 0 grams of sugar is 78.3 calories.

ⓓ $\hat{y} = 2.74x + 78.3$
$= 2.74(10) + 78.3 = \boxed{\text{10 grams of sugar}}$

ⓔ residual = actual − predicted
$4 = x - 105.7$
$x = \boxed{\text{109.7 calories}}$

a. Is a line an appropriate model? Explain.
b. Interpret the slope of the regression line in context.
c. Interpret the y-intercept of the regression line in context.
d. What are the predicted calories for a brand with 10g of sugar per serving?
e. What were the actual calories for the brand with 10g of sugar per serving?

7. Simple random samples of young adults and older adults in a large city were surveyed as to credit card debt, and the data are summarized below:

	n	Mean	Median	Min.	Max.	Q1	Q3	SD
Younger Adults	100	1,500	1,500	1,300	1,700	1,400	1,600	120
Older Adults	100	1,520	1,500	1,000	2,100	1,400	1,655	200

a. Is there evidence that either of these samples were drawn from populations with normal distributions?
b. Assume that both samples are drawn from normally distributed populations. Would a greater percentage of younger or older adults more likely be able to pay off their credit debt with $1,550? Explain.

8. An MLB scout is checking out a minor league pitching prospect and, using a timer, finds the average speed of his fastball to be 92.3 mph with a standard deviation of 2.3 mph. Assume the pitcher never tires and there is a consistent normal distribution of fastball speeds.
a. Would you check the timer if it registered 95 mph on the next fastball? Explain.
b. Would you check the timer if the next 30 fastballs averaged 95 mph? Explain.

9. In a study of selling prices of new homes in an expensive development, regression analyses are run on price versus square footage of homes and on price versus square footage of the lots. The graphs of the residuals:

7. (a) Younger adults $m = 1{,}300$ $M = 1{,}700$

$\frac{1{,}300 - 1{,}500}{120} = -1.67$

$\frac{1{,}700 - 1500}{120} = 1.67$

not normal for min and max to be only 1.67 st. dev. from mean.

Older Adults

min z-score $= \frac{-520}{200} = -2.6$

$Q_1 = \frac{-120}{200} = -.6$

$Q_3 = \frac{135}{200} = .675$

max $= \frac{580}{200} = 2.9$

all between -3 and 3

about .67 z-scores!

roughly normal :)

(b) young adults $z = \frac{1{,}550 - 1{,}500}{120} = .4167 \rightarrow$ 66.2% w/ debt less than \$1,550.

older adults $z = \frac{1550 - 1520}{200} = .15 \rightarrow$ 56% w/ debt less than \$1,550.

* With \$1,550, a greater % of younger adults will more likely be able to pay off their debt.

8. (a) $\frac{95 - 92.3}{2.3} = 1.174$

$P(z > 1.174) = .1202$

$\boxed{12.02\% \text{ chance}}$

fairly reasonable

(b) $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2.3}{\sqrt{30}} = .4199$

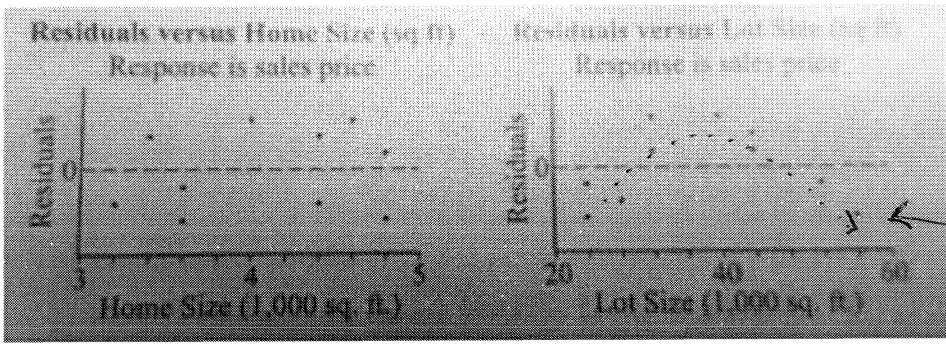
10% condition $10(30) = 300 <$ all fastballs

$30 \geq 30$ CLT ✓

$\frac{95 - 92.3}{.4199} = 6.430$

→ timer needs to be checked b/c probability of this happening is very low

9/9

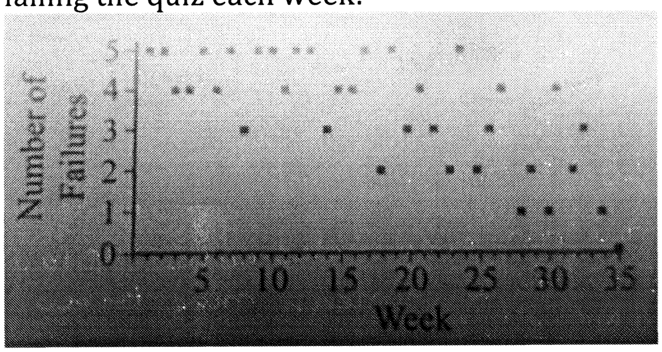


(a) Sales price v. home sq. ft is better linear fit than sales price vs. lot sq ft. because the second plot shows distinct pattern

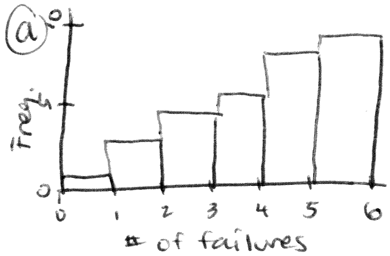
(b) residual = Actual - predicted
home size 40,000 and lot size 40,000 both have (+) residuals, actual is greater than predicted, so regression line gave underestimate

a. Which of the regression lines indicates a better linear fit? Explain.
b. The two regression lines are used to find estimates for the sales price of a 4,000 sq. ft. home sitting on a 40,000 sq. ft. lot. One of the homes in the sample has 4,000 sq. ft. and is sitting on a 40,000 sq. ft. lot. Which of the estimates were underestimates and which were overestimates of the sales price of this home? Explain.

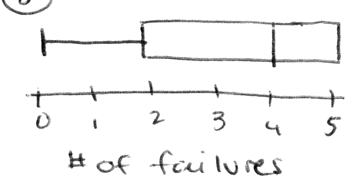
10. An AP teacher gives a short weekly quiz. Following is a scatterplot of the number of students failing the quiz each week:



(a)



(b)



(c) roughly linear trend, neg. slope, associated w/ # of failures each week!

(d) # of failures is skewed left.

a. Draw a histogram of the frequencies of the number of failures.
b. Draw a boxplot of the number of failures.
c. Name a feature apparent in the scatterplot but not in the histogram of boxplot.
d. Name a feature that is shown by the histogram and boxplot but is not as obvious in the scatterplot.

11. Suppose we calculate the melting points of eight metal rods and find the mean to be 1,400^{0}F with a standard deviation of 90^{0}F. What would be the mean and standard deviation of the melting points of the eight rods if the eight measurements were converted to Celsius? [^{0}F = (1.8)(^{0}C) +32]

12. Suppose a police breathalyzer unit will correctly test positive for 98% of legally drunk drivers. But will also give a false positive reading for 3% of drivers whose actual blood alcohol concentration is under the limit. If 5% of late Saturday night drivers are legally drunk, and an officer arrests two drivers with positive readings, what is the probability that exactly one of the two is guilty of driving drunk?

on back

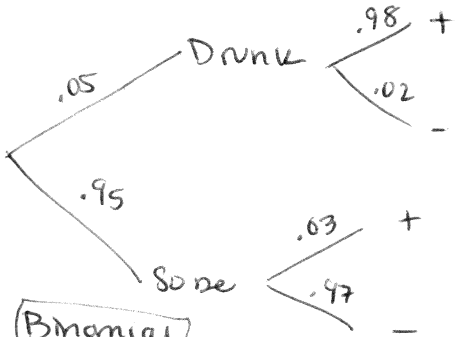
13. Suppose the number of minutes per day high school students spend text messaging is normally distributed with mean of 21 minutes.
a. Which is more likely; an SRS of 30 students text messaging an average of less than 20 minutes per day, or an SRS of 100 students text messaging an average of less than 20 minutes per day. Explain.

11. $C° = \frac{1}{1.8}(°F) - \frac{32}{1.8}$

$\mu = \frac{1}{1.8}(1400) - \frac{32}{1.8} = \boxed{760°C}$

$\sigma = \frac{1}{1.8}(90) = \boxed{50°C}$ ← you only multiply/divide to measures of spread (no adding/subtracting)

12.



$P(D \cap +) = (.05)(.98) = .049$

$P(S \cap +) = (.95)(.03) = .0285$

$P(+) = .049 + .0285 = .0775$

$P(D \mid +) = \frac{P(D \cap +)}{P(+)} = \frac{.049}{.0775} = .6323$

$P(S \mid +) = 1 - .6323 = .3677$

Binomial

P(exactly 1 success in 2 trials) = $\underbrace{2}_{{}_2C_1}(.6323)(.3677) = \boxed{.4650}$

13. ⓐ $\sigma_{\bar{x}} = \frac{1}{\sqrt{n}}$ (smaller for larger n) → the prob. of a sample mean more than 1 minute less than the pop. mean decreases for larger n. So sample of 30 is more likely to have a mean more than 1 minute less than the pop. mean.

ⓑ $z = \frac{23 - 21}{.8} = 2.5 \quad P(z > 2.5) = .0062$

ⓒ Calculation would not change, b/c central limit theorem, the sampling distribution of $\bar{x}$ is approximately normal.

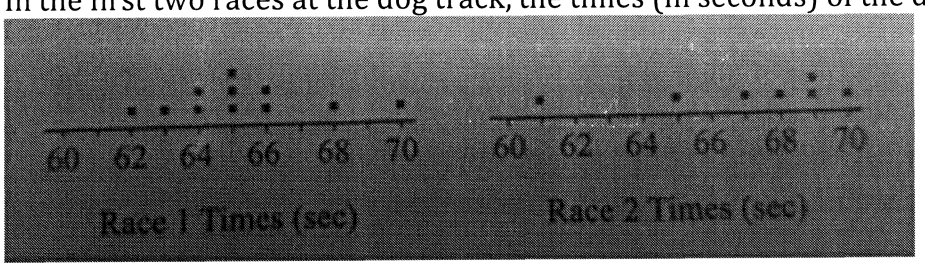
b. Suppose the sampling distribution of $\bar{x}$ for samples of size 100 has a standard deviation of .8 minutes. What is the probability an SRS of 100 students text messaging an average of more than 23 minutes?
c. Suppose the original population is not normal, but is skewed right (to the higher values). How would your calculation in (b) change?

14. A high school math department conducts a study to determine whether a new AP statistics textbook will lead to higher AP exam scores than the textbook currently in use. Two AP statistics classes were scheduled, each teacher has 18 students, and it is randomly decided which class will use which book. At the end of the academic year, the 18 students in each class the AP statistics exam, and the department notes the scores.
    a. Identify the response variable, the treatments, and the experimental units.
    b. Was randomization properly used? Explain.
    c. Was replication properly used? Explain.
    d. Teacher is a confounding variable. Explain.

15. An Apartment building elevator has a carrying capacity of 2,000 lb. Suppose the men living in the building have a mean weight of 185 lb. with a standard deviation of 15 lb., the women have a mean weight of 135 lb. with a standard deviation of 10 lb., and both weight distributions are normal.
    a. What are the mean μ_{sum} and standard deviation of σ_{sum} of the combined weight of seven men and five women assuming all weights are independent?
    b. What is the probability that the seven men and five women will overload the elevator?

on back

16. In the first two races at the dog track, the times (in seconds) of the dogs are given by:



    a. In either race are any of the times considered to be outliers? Explain,
    b. Each dog ran in one race but not both. If all the times are combined, does the new set of times have any outliers? Explain.
    c. Explain the reason for the answer to (b) in relation to the answer to (a).

17. The probability that Bonds hits a homer on any given at-bat is .12, and each at-bat is independent.
    a. What is the probability the next homer will be on his fifth at-bat?
    b. What is the probability he has exactly one homer in five at-bats?
    c. What is the expected number of homers in every ten at-bats?
    d. What is the expected number of at-bats until the next homer?

18. You are asked to choose between two envelopes, one of which has twice as much money as the other. You arbitrarily pick one, open it, and find $100. You are then given the chance to switch envelopes. You reason that the other envelope has either $50 or $200, each with a probability of .5. Applying your understanding of expected value, you calculate .5($50) + .5($200) = $125 and conclude that you should switch envelopes. Comment on this reasoning.

1st ~~one~~ envelope = $x   original choice Expect value = .5($x) + .5($2x) = $1.5x
2nd ~~two~~ envellpe = $2x   Switch Expected value = .5($2x) + .5($x) = $1.5x

Doesn't matter whether you switch or not.

14. ⓐ response = Exam scores
treatments = 2 books
Experimental units = two class (Not the students)

ⓑ Yes! two books (treatments) were randomly assigned

ⓒ No! Each treatment was applied to only one experimental unit.

ⓓ books (treatment) is confounded w/ teachers
↳ you don't know which caused the response

5. ⓐ
$\mu_{sum} = 7(185) + 5(135) = 1,970$

$\sigma_{sum} = \sqrt{7(15^2) + 5(10^2)} = 45.55$

ⓑ individual distributions are normal so sum is normal

$P\left(z > \frac{2000-1970}{45.55}\right) = \boxed{.255}$

16. ⓐ 1st race

$Q_1 = 64$ med $= 65$ $Q_3 = 66$
$IQR = 66 - 64 = 2$
$Q_1 - 1.5(2) = 61$
$Q_3 + 1.5(2) = 69$

70 is only outlier

2nd race

$Q_1 = 65$ med $= 68$ $Q_3 = 69$
$IQR = 69 - 65 = 4$
$Q_1 - 1.5(4) = 59$
$Q_3 + 1.5(4) = 75$

No outliers

ⓑ all values

$Q_1 = 64$ $Q_3 = 68$
$Q_1 - 1.5(IQR) = 58$
$Q_3 + 1.5(IQR) = 74$

No outliers

ⓒ the outlier, 70, is 1st race was on the high side, but the second race had values around 70, so 70 was no longer an outlier.

7. ⓐ geometric $p = .12$ $k = 5$ → $(1-.12)^4(.12) = \boxed{.0720}$

ⓑ Binomial $n = 5$ $p = .12$ $k = 1$ $\binom{5}{1}(.12)(1-.12)^4 = \boxed{.3598}$

ⓒ $\mu_{binomial} = np = (10)(.12) = \boxed{1.2}$ * the mean number of homers for every 10 at-bats is 1.2.

ⓓ $\mu_{geometric} = \frac{1}{p} = \frac{1}{.12} = \boxed{8.33}$ * the mean number of at-bats before hitting a homer is 8.33.

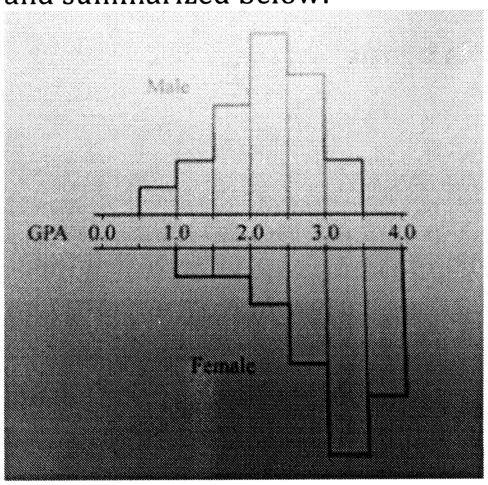
19. The standard "Olympic Distance" triathlon is a 1.5 km swim, 40 km bike ride, and 10 km run. In a random sampling of recent competitions, the mean and standard deviations of the participants' times for each event were:

	Swim	Bike Ride	Run
Mean (min.)	29	90	68
Standard Deviation (min.)	5	10	12

Assume the times for the three legs of the race are each normally distributed and independent. What is the probability that a participant will complete the triathlon in less than 175 minutes?

on back

20. The GPAs of random samples of 50 male and 50 female students at a large university are noted and summarized below:



Write a few sentences comparing the distributions of GPAs of male and female students at this university.

19. $\mu_{sum} = 29 + 90 + 68 = \boxed{187 \text{ min}}$

$\sigma_{sum} = \sqrt{5^2 + 10^2 + 12^2} = \boxed{16.4 \text{ min}}$

individual legs are normal so sum is normal

$P\left(z < \frac{175 - 187}{16.4}\right) = \boxed{.232}$

20. Shape: female GPA is skewed left, while male GPA is more Bell-Shaped, Both unimodal

Center: male GPA is less than female GPA distribution.

Spread: the ranges of the two distributions are the same

$3.5 - .5 = 3.0$

$4.0 - 1.0 = 3.0$

Outliers: no outliers in either distribution.