

BIOS 600.001 Exam 1

September 20, 2012

Honor Statement. I pledge that I have not used any reference materials (including electronic materials) during this examination. I pledge that I have neither given nor received any aid from any other person during this examination, and that the work presented here is entirely my own. I furthermore pledge that I will not reveal any of the material on this examination, either in the form of the exact question or the topics covered, to any person for any reason until after exams have been returned. I pledge that I will report all Honor Code violations observed by me. I understand that if I have committed any of the above, I have violated the UNC Honor Code.

Name: _____

Solution example

Signature: _____

Date: _____

INSTRUCTIONS: No electronic devices are permitted. Full credit will be given for correct answers that are unsimplified, e.g. $\frac{1+2+3}{4+5+6}$ is an acceptable form. There are 100 points on this exam. Please ask Dr. Herring if you have any questions. Note that this exam will be videorecorded.

EXAMS ARE DUE AT 10:45. At 10:45am you should be signing the honor statement and turning in the exam. Exams turned in after 10:45 am will be penalized one point for the first minute delay (10:46) and 5 points per minute thereafter (starting at 10:47).

1. (15 points) Researchers in adolescent sexual development use many questions to attempt to characterize an individual's sexual identity. A subset of such questions from a recent survey is provided below. (NOTE: do not answer these questions yourself as part of the exam)

SAMPLE QUESTIONNAIRE (DO NOT FILL OUT FOR EXAM!)

Q1: Mark an 'X' in the location on the line below that best describes your current sexual orientation.

Homosexual

Heterosexual

Q2: How many romantic partners have you had of the same sex? _____

Q3: How many romantic partners have you had of a different sex? _____

Q4: In the past year, have you been romantically attracted to someone of the same sex?

YES

NO

Q5: In the past year, have you been romantically attracted to someone of a different sex?

YES

NO

(a) Is the measurement scale for Q1 continuous, count, nominal, or ordinal?

(b) Is the measurement scale for Q2 continuous, count, nominal, or ordinal?

- (c) Based on Q4 and Q5, researchers define a new variable for attraction that indicates whether a respondent reports no attraction at all (Q4=no, Q5=no), attraction only to individuals of the same sex (Q4=yes, Q5=no), attraction only to individuals of a different sex (Q4=no, Q5=yes), or attraction both to individuals of the same sex and to individuals of a different sex (Q4=yes, Q5=yes). Is the measurement scale for this variable continuous, count, nominal, or ordinal?

2. (15 points) A high school student conducted a small survey of mothers of toddlers in her little brother's day care class. She was interested in the number of doctor visits each toddler had from birth through age 1. She obtained the following visit counts for the 5 infants in the class: 7, 7, 11, 9, 6. Calculate the following from these data.

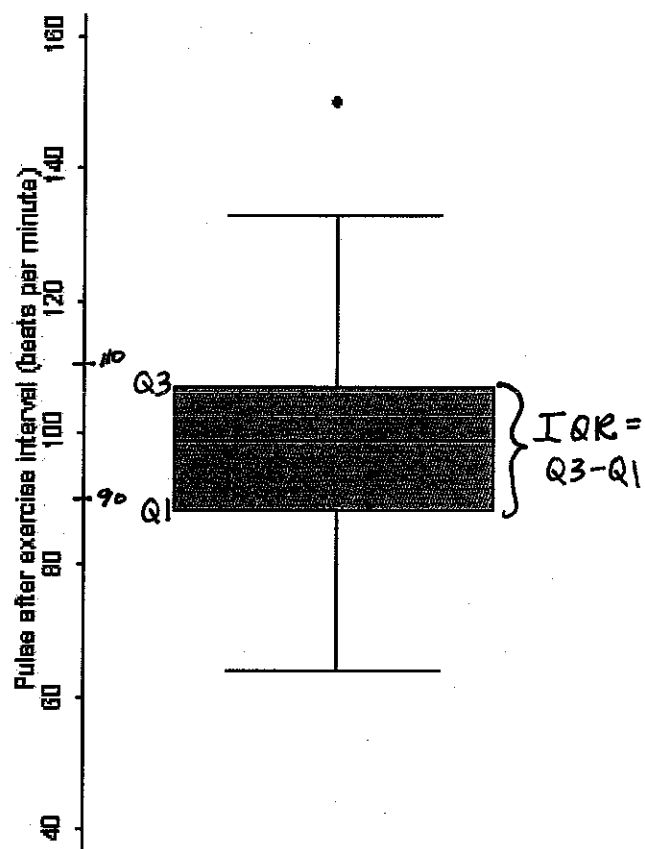
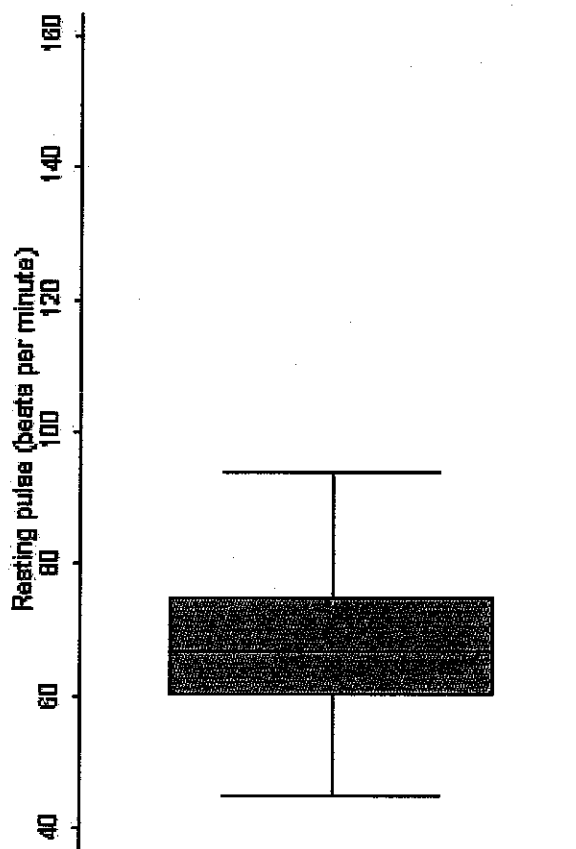
$$\frac{7+7+11+9+6}{5} = 8$$

(a) Mean: _____

(b) Median: 7

(c) Mode: 7

3. (15 points) Consider the boxplots below, which illustrate data on pulse rate collected in class at rest and after a one-minute exercise interval.



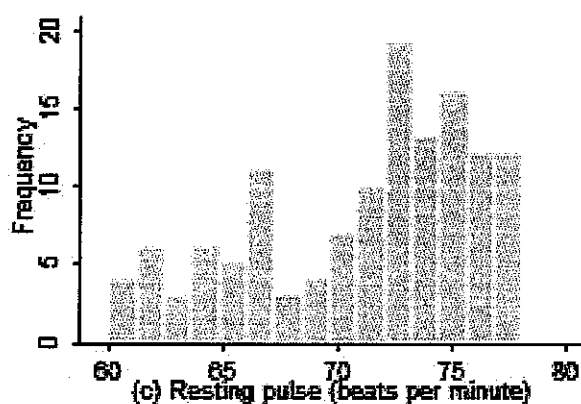
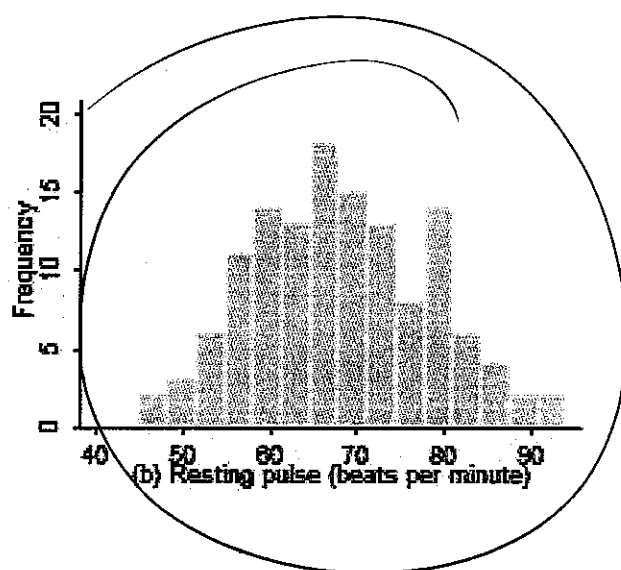
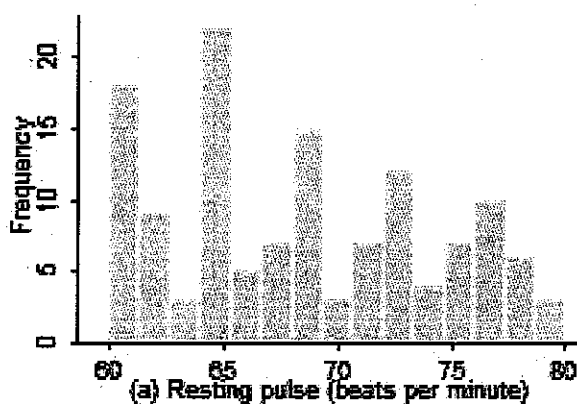
- (a) What is the interquartile range of pulse rate after exercise?

$$Q3 - Q1 \approx 108 - 88 = 20$$

(b) In general, when were pulse rates higher: at rest or after exercise?

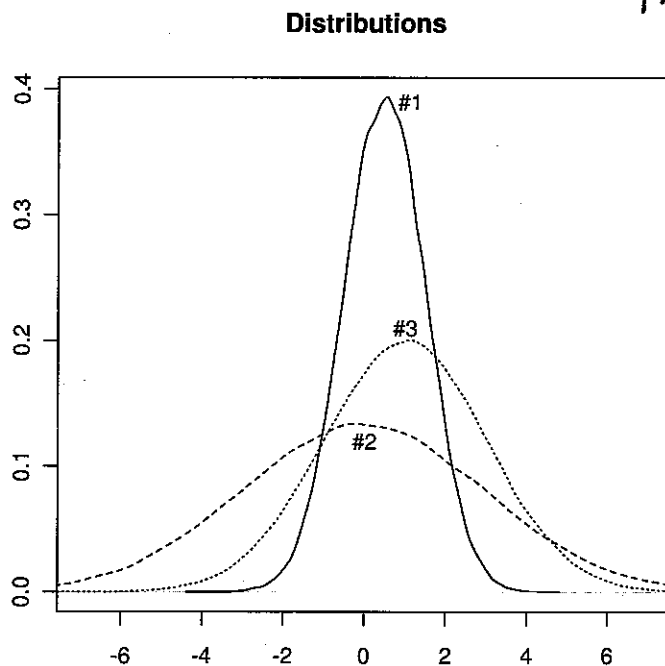
after exercise

(c) Which histogram describes the same data as the boxplot for resting pulse?



4. (5 points) Consider the distributions below. Which one has the largest standard deviation, #1, #2, or #3?

#2



5. (10 points) In the 2000 presidential election, George Bush won 48% of the popular vote. In the 2004 election, he won 51% of the popular vote. Based on these data, what is the probability that a randomly chosen person who voted in both elections chose President Bush both times? Describe how you would calculate this probability. If there is not enough information provided to answer this question, explain what additional information you would need.

Let B_1 = event voted for Bush in 2000

B_2 = event voted for Bush in 2004

Assume $\Pr(B_1) + \Pr(B_2)$ same in this group (voting in both elections) as overall.

$$\Pr(\text{voted for Bush in 2000 + 2004}) = \Pr(B_1 \cap B_2)$$

$$= \Pr(B_1) \Pr(B_2 | B_1) \quad \text{using multiplicative rule}$$

$$= 0.48 \Pr(B_2 | B_1)$$

We know $\Pr(B_2) = 0.51$ but need $\Pr(B_2 | B_1)$. Because those who chose Bush over Gore (and Nader) in 2000 were likely to choose Bush again (over Kerry) in 2004, B_1 & B_2 are not independent (and $\Pr(B_2 | B_1)$ is probably very high).

6. (20 points) The $\epsilon 4$ allele of the gene encoding apolipoprotein E (APOE) is strongly associated with Alzheimer's disease, but its value in making the diagnosis remains uncertain. A study was conducted among 2188 patients who were evaluated at autopsy for Alzheimer's disease by previously established pathological criteria. Patients were also evaluated clinically (while alive) for the presence of Alzheimer's disease. Suppose we consider the combination of both a clinical diagnosis for Alzheimer's disease *and* the presence of ≥ 1 $\epsilon 4$ allele as a screening test for Alzheimer's disease, with individuals screening positive if they have a clinical diagnosis of Alzheimer's and ≥ 1 $\epsilon 4$ alleles.

Table 1: Influence of the APOE genotype in diagnosing Alzheimer's disease (AD)

APOE genotype	Both clinical and pathological criteria for AD	Only clinical criteria for AD	Only pathological criteria for AD	Neither clinical nor pathological criteria for AD
≥ 1 $\epsilon 4$ allele	1076	66	66	67
No $\epsilon 4$ allele	567	124	61	161
Total	1643	190	127	228

- (a) Complete Table 2 (remember it is ok to leave expressions unsimplified, e.g. $1+2+3$ instead of 6)

Table 2: Relationship between clinical and pathological diagnoses of Alzheimer's disease

Screening test	Pathological Diagnosis		
	Alzheimer's disease	Other dementia	Total
Positive	1076	66	1076+66
Negative	66+567+61	67+124+161	567+124+61+161+66+67
Total	1076+567+66+61	66+124+67+161	2188

Provide the following characteristics of the screening test.

$$(b) \text{ Sensitivity: } \frac{1076}{1076 + 567 + 66 + 61} = \frac{1076}{1770}$$

$$(c) \text{ False positive rate: } \frac{66}{66 + 124 + 67 + 161} = \frac{66}{418}$$

$$(d) \text{ Positive predictive value: } \frac{1076}{1076 + 66} = \frac{1076}{1142}$$

7. (20 points) The rate of major congenital malformations is roughly 2.5 per 100 deliveries. A study is set up to investigate whether offspring of Iraq and Afghanistan war veteran fathers are at special risk.

Suppose 100 infants of such fathers are identified in a birth registry. The probability of 0-6 infants having a malformation out of a group of 100 is provided in the table below, which was calculated from a binomial distribution with $n = 100$ and $p = 0.025$.

Number of malformed infants	Probability	Pr(given # or more malformed)
0	0.08	1.00
1	0.20	0.92
2	0.26	0.72
3	0.22	0.46
4	0.13	0.24
5	0.07	0.11
6	0.03	0.04
\vdots	\vdots	0.01

- (a) What is the probability of seeing 4 or more infants with major malformations in a randomly-selected group of 100 infants?

$$\begin{aligned}
 \Pr(4+ \text{ malformed}) &= 1 - \Pr(0-3 \text{ malformed}) \\
 &= 1 - [0.08 + 0.20 + 0.26 + 0.22] \\
 &= 1 - 0.76 = 0.24
 \end{aligned}$$

- (b) Suppose that if the number of infants x with congenital malformations is unusually large, we will conclude that children whose dads are veterans are at increased risk. If the probability of observing x or more malformed infants is ≤ 0.10 , we will consider x an unusually large number. Following this procedure, for what values of x in a sample of size 100 would we conclude that babies of men who are veterans are at increased risk?

For 6 or more (6-100) infants with malformations, we would conclude male veterans are a high risk group.

- (c) Another graduate student is using the same registry to examine the association between maternal marijuana use and major congenital malformations. She finds that of 75 infants born to mothers who reported using marijuana, 8 have major malformations. What can you say about the probability of this occurrence (without doing extensive new calculations)?

In our group of 100 infants the probability of more than 6 malformations is only 0.01. In our group of 100 infants, the probability of 8 or more malformations would be

$$\Pr(6 \text{ or more}) - \Pr(\text{exactly } 7),$$

which is ≤ 0.01

In a smaller group (say 75), the corresponding probability is even lower as you expect fewer malformations in a smaller group with fixed π .

So marijuana using¹⁰ moms are probably high risk. (we would conclude they are high risk using our rule).

Some comments

#1 (a) This is a visual analogue scale. As provided, it is a continuous measure, as you can place a mark anywhere on the line.

(c) Why not ordinal? You could order $Q4 = Q5 = \text{no}$ as no attraction and $Q4 = Q5 = \text{yes}$ as attraction to both groups, but how does one order $Q4 = \text{yes}, Q5 = \text{no}$ and $Q5 = \text{no}, Q4 = \text{yes}$?

#3 (a) The IQR is a value (one number), e.g.
 $\approx 108 - 88 = 20$.

#5 These are not independent events! You need to know $\Pr(\text{Bush 2004} | \text{Bush 2000})$. One bonus point to those who noted that the $\Pr(\text{Bush 2004})$ in the group who voted twice may be different from $\Pr(\text{Bush 2004})$ marginally and who got answer right.

#6(c) The false positive rate is $\Pr(T+|D-)$ or 1 minus the sensitivity. Many gave instead the false discovery rate (we didn't discuss this) which is $\Pr(D-|T+)$.

6(d) This was much faster to pull from table than to use a formula.

7(a) Note the table entries sum to 0.99 so that $\Pr(Y \geq 6) = 0.01$.

7(b) We will start thinking this way more as we begin hypothesis testing. In the general population, $\Pr(Y \geq 6) = 0.04$. So if in any group of 100 infants, seeing 6 or more malformed infants would cause us to worry.

7(c) We'll start thinking like this more too. Seeing 8 malformed infants in a group of 75 is even more rare than 8 in 100 (which has prob < 0.01) so we would worry that ~~pot~~ pot-smoking moms are higher risk.