BIOSTATS 540 - Introductory Biostatistics
Fall 2015
Examination 1
Due: Monday October 5, 2015
Last Date for Submission with Credit: Monday October 12, 2015

Ground Rules:
This is an “open book” “take-home” exam. You are welcome to use any reference materials you wish. You are welcome to use the computer as you wish, too. However, you MUST work this exam by yourself and you may not consult with anyone.

Instructions and Checklist

___1. Start each problem on a new page.
___2. Write your name on every page
___3. Make a copy of your exam for safekeeping (sometimes a mailed exam is lost!)
___4. Submit a completed signature page.

How to submit your exam:

Worcester “in-class” Section:
___1. Bring your completed exam to class on Monday October 5, 2015, being sure that you have made a copy for safe keeping; OR
___2. Mail your completed exam to me with post-mark October 5, 2015 to my address below.

Online Section:
___1. Upload your completed exam to the ASSIGNMENT tab no later than 11:59 pm on Monday October 5, 2015. This must be a single pdf and be named using the convention lastname_exam1.pdf. OR
___2. Mail your completed exam to me with post-mark October 5, 2015 to my address below.

Address and Telephone Number for Mailing

Carol Bigelow
Biostatistics & Epidemiology/402 Arnold House
University of Massachusetts/Amherst
715 North Pleasant Street
Amherst, MA 01003-9304
Tel. 413-545-1319
Signature

This is to confirm that in completing this exam, I worked independently and did not consult with anyone.

Signature:  ___________________________________________________________

Printed Name:  _____________________________________________________

Date:   ___________________________
1. (15 points total)

1a. (5 points)
The U.S. Census Bureau reports the median family income in its summary of census data. Why do you suppose they use the median instead of the mean? What might be the disadvantages of reporting the mean?

1b. (5 points)
You have just bought a new car that claims to get a highway fuel efficiency of 31 miles per gallon. Of course, your mileage will “vary”. If you had to guess, would you expect the interquartile range (IQR) of gas mileage attained by all cars like yours to be 30 mpg, 3 mpg, or 0.3 mpg? Why?

1c. (5 points)
A company selling a new MP3 player advertises that the player has a mean lifetime of 5 years. If you were in charge of quality control at the factory, would you prefer that the standard deviation of life spans of the players you produce be 2 years or 2 months? Why?
2. (10 points total)

Illegal immigration was one (of many, actually!) “hot button” topics of the September 16, 2015 Republican candidate debate (oh my goodness, what a show that was!). The following table lists the numbers of deportable aliens (COUNT) caught by the U.S. Border Patrol between 1971 and 2009.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Count (1000's)</td>
<td>420</td>
<td>656</td>
<td>767</td>
<td>1042</td>
<td>1076</td>
<td>976</td>
<td>1251</td>
<td>1349</td>
<td>1190</td>
<td>954</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year</th>
<th>1991</th>
<th>1993</th>
<th>1995</th>
<th>1997</th>
<th>1999</th>
<th>2001</th>
<th>2003</th>
<th>2005</th>
<th>2007</th>
<th>2009</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count (1000's)</td>
<td>1198</td>
<td>1327</td>
<td>1395</td>
<td>1536</td>
<td>1714</td>
<td>1266</td>
<td>932</td>
<td>1189</td>
<td>877</td>
<td>556</td>
</tr>
</tbody>
</table>

2a. (5 points)
Summarize these data by constructing an appropriate graph of your choosing.

2b. (5 points)
In 2-5 sentences at most, what are the important facts that these data show?
3. **(15 points total)**

The Bureau of Transportation Statistics of the U.S. Department of Transportation collects and publishes statistics on airline travel. Below are three displays of the percent (%) of flights arriving on time from 1995 through 2010, overall and by month.

---

**3a. (5 points)**

In 1-3 sentences *at most*, describe what the histogram says about on-time arrivals.

---

**3b. (5 points)**

In 1-3 sentences *at most*, what does the box plot (middle plot) suggest that you *cannot* see in the histogram?

---

**3c. (5 points)**

In 1-3 sentences *at most*, describe what is shown in the month by month box plot (right hand plot). At what time of year are flights most likely to be on time? Can you suggest a reason for this pattern?
A survey is taken by the U.S. Bureau of the Census. It reports that the median incomes in the past 12 months were $52,168 for women and $61,975 for men. It is also reported that the mean earnings in the past 12 months were $59,890 for women and $76,724 for men.

4a. (5 points)
Does this data suggest that the distribution of income for each gender is symmetric, positive (right) skewed or negative (left) skewed? In 1-2 sentences at most, explain your reasoning.

4b. (5 points)
The results of this survey were obtained from 73.8 million women and 83.4 million men. Calculate the overall sample mean income. Show your work.
5. (10 points total)

The following are 9 observations of cost ($) of compact refrigerators.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$150</td>
<td>$150</td>
<td>$160</td>
<td>$180</td>
<td>$150</td>
<td>$140</td>
<td>$120</td>
<td>$130</td>
<td>$120</td>
</tr>
</tbody>
</table>

5a. (2 points)
By hand (no calculator!) calculate the value of the sample mean. Show your work.

5b. (2 points)
By hand (no calculator!) calculate the value of the sample standard deviation. Show your work.

5c. (2 points)
By hand (no calculator!) calculate the value of the sample median. Show your work.

5d. (2 points)
By hand (no calculator!) calculate the values of the 1st (25th percentile) and 3rd (75th percentile) quartiles (Q1 and Q3). Show your work.

5e. (2 points)
By hand (no calculator!) calculate the value of the interquartile range (IQR). Show your work.
6. (10 points total)

Ozone levels (in parts per billion, ppb) were recorded monthly, every year, for 46 years. Shown below are boxplots of the data for each month over the 46 years, lined up in order (January = 1).

6a. (2 points)
In what month was the highest ozone level ever recorded?

6b. (2 points)
Which month had the largest interquartile range (IQR)?

6c. (2 points)
Which month had the smallest range?

6d. (2 points)
In 1-3 sentences at most write a comparison of the ozone levels in January (month=1) and June (month =6).

6e. (2 points)
In 2-5 sentences at most write a report on the monthly patterns you see in the ozone levels.
7. (10 points total)

A school system employs teachers at salaries between $40,000 and $70,000. The mean and standard deviation are $44,000 and $3,000 respectively. The teachers’ union and the school board are negotiating the form of next year’s increase in the salary schedule.

For Questions #7a – #7d ONLY: Suppose that every teacher is given a flat $1000 raise.

7a. (2 points)
By how much will the mean salary increase?

7b. (2 points)
By how much will the median salary increase?

7c. (2 points)
By how much will the interquartile range (IQR) increase?

7d. (2 points)
By how much will the sample variance increase? Tip – I said sample variance, not sample standard deviation.

For Question #7e ONLY: Now suppose that, instead of a flat raise, every teacher is given a 5% raise.

7e. (2 points)
By how much will the sample variance increase? Again – I said sample variance, not sample standard deviation.

9-28-2015 Update:
Express your answer as a percent (eg – “The sample variance increases by 25%). Note – it is NOT possible to give a number as your answer.
8. (20 points total)

A study examining the health risks of smoking measured the cholesterol (mg/dL) levels of people in two independent groups: 1) SMOKERS: those who had smoked for at least 25 years; and 2) NON-SMOKERS: persons of similar ages who had never smoked. The following are the data.

<table>
<thead>
<tr>
<th>Smokers</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>225</td>
<td>211</td>
<td>209</td>
<td>284</td>
</tr>
<tr>
<td>258</td>
<td>216</td>
<td>196</td>
<td>288</td>
</tr>
<tr>
<td>250</td>
<td>200</td>
<td>209</td>
<td>280</td>
</tr>
<tr>
<td>225</td>
<td>256</td>
<td>243</td>
<td>200</td>
</tr>
<tr>
<td>213</td>
<td>246</td>
<td>225</td>
<td>237</td>
</tr>
<tr>
<td>232</td>
<td>267</td>
<td>232</td>
<td>216</td>
</tr>
<tr>
<td>216</td>
<td>243</td>
<td>200</td>
<td>155</td>
</tr>
<tr>
<td>216</td>
<td>271</td>
<td>230</td>
<td>309</td>
</tr>
<tr>
<td>183</td>
<td>280</td>
<td>217</td>
<td>305</td>
</tr>
<tr>
<td>287</td>
<td>217</td>
<td>246</td>
<td>351</td>
</tr>
<tr>
<td>200</td>
<td>280</td>
<td>209</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NON-Smokers</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>250</td>
<td>213</td>
<td>300</td>
<td></td>
</tr>
<tr>
<td>249</td>
<td>213</td>
<td>310</td>
<td></td>
</tr>
<tr>
<td>175</td>
<td>174</td>
<td>328</td>
<td></td>
</tr>
<tr>
<td>160</td>
<td>188</td>
<td>321</td>
<td></td>
</tr>
<tr>
<td>213</td>
<td>257</td>
<td>292</td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>271</td>
<td>227</td>
<td></td>
</tr>
<tr>
<td>238</td>
<td>163</td>
<td>263</td>
<td></td>
</tr>
<tr>
<td>192</td>
<td>242</td>
<td>249</td>
<td></td>
</tr>
<tr>
<td>242</td>
<td>267</td>
<td>243</td>
<td></td>
</tr>
<tr>
<td>217</td>
<td>267</td>
<td>218</td>
<td></td>
</tr>
<tr>
<td>217</td>
<td>183</td>
<td>228</td>
<td></td>
</tr>
</tbody>
</table>
8a. (5 points)
By any means you like, construct two stem and leaf diagrams of the cholesterol values, so as to compare smokers and non-smokers.

8b. (5 points)
Complete the following table.

<table>
<thead>
<tr>
<th></th>
<th>Smokers</th>
<th>NON-Smokers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number in group, n =</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P_{25} = Q1 = $ Lower Quartile =</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P_{50} = Q2 = $ Median Quartile =</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P_{75} = Q3 = $ Upper Quartile =</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interquartile Range (IQR) =</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$1.5* IQR=$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Value of Lower Fence =</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Value of Upper Fence =</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outliers (if any) below lower fence (LIST) =</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outliers (if any) above upper fence (LIST) =</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

8c. (5 points)
By any means you like, construct a side-by-side box plot of the cholesterol values, so as to compare smokers and non-smokers.

8d. (5 points)
In 3-8 sentences at most, write a brief report comparing the cholesterol value distributions of smokers and non-smokers.