

10
000

Cathy Sun
AP Stats - 4
9/5/13
Mr. Micek

Historical Notes on Statistics

1. In the beginning, statistics involved _____.
It involved summarizing data by means of charts and tables
2. What did the Chinese use stats for? How long ago?
They used it for keeping state records in 2000 B.C, which is 4000 years ago. They maintained records of revenue collection, government expenditure, and availability of warriors.
3. Who began the study of statistics?
John Graunt began the study the statistics with his book, *Natural and Political Observations Upon the Bills of Mortality*.
4. Why wasn't statistical theory commonly used prior to 1930? What changed?
The accumulation and analysis of statistical data was too time consuming and involved complicated computations. Things changed after the invention of the computer because it could do long and difficult calculations in a relatively short amount of time.
5. What is inference?
Inference is making generalizations on the basis of samples.
6. Where are the origins of the study of probability found?
The origins of probability are found in Cardan's book *The Book on Games of Chance*.
7. Who wrote the 1st book on the theory of randomness? What was the title?
Girolomo Cardan wrote *The Book on Games of Chance*.

1.1 # 1-3, 5, 7, 9

- 1 a) responses about how frequent surveyors eat at fast-food restaurant
b) qualitative
c) responses about fast-food restaurants from all US adults
- 2 a) miles per gallon
b) quantitative
c) mpg of all new 2003 cars
- 3 a) student fees at colleges/universities
b) quantitative
c) student fees of all colleges/universities in nation
5. a) time interval between receipt and check clearance
b) quantitative
c) time interval between receipt and check clearance in 5 state region
- 7 a) ratio
b) interval
c) nominal
d) ordinal
e) ratio
f) ratio
- 9 a) nominal
b) ratio
c) interval
d) ordinal
e) ratio
f) interval

Gertrude's situation

Gertrude wants to find out the most popular type of phone that AT&T users have. She goes into the database and uses a random ^{computer} generator to choose 200 customers. The population is the type of phone of the customers, and the sample consists of 200 people. After randomly selecting her sample, Gertrude goes through and finds out that the most popular phone choice is an iPhone 5. She therefore concludes that most AT&T customers prefer iPhones.

the next morning I went to the beach to look for shells. I found some very nice ones. I also found a dead crab. It was about 6 inches long. I took it home and cooked it for dinner. It was delicious.

After lunch, I went for a walk in the forest. I saw many birds and insects. I also found some mushrooms. I picked some wild onions and ate them raw. They were very spicy. In the afternoon, I went for a swim in the lake. The water was cool and refreshing. I swam for about an hour. When I got out, I was tired but happy. I had a great time at my vacation.

cathy sun
9/14/13
P4 - STATS

Planning a study 5.26-5.28

- 5.26 This is not an experiment because no treatments or groups are imposed; the subjects self-declare their votes. NO explanatory or response variables because it's not an experiment.
- 5.27 a) the 20 pieces of package liner
b) the temperature; 4 levels (250°F , 275°F , 300°F , 325°F)
c) the force needed to peel each seal
- 5.28 a) Yes, the educator imposes a treatment (different curriculums) on the students to observe a response.
b) explanatory variable is the computer reading software
response variable is the change or increase in reading ability

functionality of the system. In addition, the system's performance is evaluated by comparing it with other existing systems.

Overall, the proposed system is a promising solution for improving the efficiency and effectiveness of the supply chain.

The system's ability to handle large amounts of data and its ability to provide real-time information make it a valuable tool for decision-making.

By integrating data from various sources, the system provides a more complete picture of the supply chain.

Finally, the system's user-friendly interface makes it easy for users to access and analyze the data.

In conclusion, the proposed system offers significant benefits for the supply chain industry.

It is a valuable tool for improving efficiency and effectiveness, and it is likely to become an integral part of the supply chain industry in the future.

1.3 #2b-d, 3-6

- 2b) experiment
- c) simulation
- d) sampling
- 3a) sampling
- b) simulation
- c) census
- d) experiment
- 4a) NO, "over the last three years" is a more precise phrase
- b) Yes, if "Have you ever run a stop sign" was asked, many of those who respond yes would oppose the doubling of fines
- c) Many would say no because they don't want to seem lazy. ALSO, they self-dictate what is too much TV. Rarely, sometimes, and frequently are better responses because there are ~~distinct~~ differences among the three levels.
- 5a) Randomly pick 10 calves to get vaccinated but test all calves; no placebo
- b) Randomly pick 9 schools to send police officer to but survey all schools; no placebo
- c) Randomly pick 40 volunteers to get real patch, other 35 get no drugs on patch. Survey all subjects; yes placebo
- 6a) Randomly choose 25 cars to have new temperature-bona tires. Examine all tires after 35000 miles. Double-blind experiment used.
- b) Randomly choose 10 bags to carry weapons. Monitor and observe all bags with a security camera. Double-blind exp not used because agent knows if there is a weapon in the bag.
- c) Randomly pick 35 people to get new eyedrops. Measure eye pressure of all patients. Double-blind exp used.

162

163

164

165

166

167

168

169

170

171

172

173

174

175

176

177

178

179

180

181

182

183

184

185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

200

201

202

203

204

205

206

207

208

209

210

211

212

213

214

215

216

217

218

219

220

221

222

223

224

225

226

227

228

229

230

231

232

233

234

235

236

237

238

239

240

241

242

243

244

245

246

247

248

249

250

251

252

253

254

255

256

257

258

259

260

261

262

263

264

265

266

267

268

269

270

271

272

273

274

275

276

277

278

279

280

281

282

283

284

285

286

287

288

289

290

291

292

293

294

295

296

297

298

299

300

301

302

303

304

305

306

307

308

309

310

311

312

313

314

315

316

317

318

319

320

321

322

323

324

325

326

327

328

329

330

331

332

333

334

335

336

337

338

339

340

341

342

343

344

345

346

347

348

349

350

351

352

353

354

355

356

357

358

359

360

361

362

363

364

365

366

367

368

369

370

371

372

373

374

375

376

377

378

379

380

381

382

383

384

385

386

387

388

389

390

391

392

393

394

395

396

397

398

399

400

401

402

403

404

405

406

407

408

409

410

411

412

413

414

415

416

417

418

419

420

421

422

423

424

425

426

427

428

429

430

431

432

433

434

435

436

437

438

439

440

441

442

443

444

445

446

447

448

449

450

451

452

453

454

455

456

457

458

459

460

461

462

463

464

465

466

467

468

469

470

471

472

473

474

475

476

477

478

479

480

481

482

483

484

485

486

487

488

489

490

491

492

493

494

495

496

497

498

499

500

10-

1.2 # 5, 6, 10-13

5. a) I would assign each healthy subject a number 01 through 40. Then I would use the random number table; the first subject assigned to the 1st randomly selected # will be in group 1, and the second subject chosen will be in group 2. The subjects would be sorted so forth.

b) Do the same in part a except only number 01-22.

c) Number the two healthy groups 1 and 2, name the two sick groups 3 and 4. If the # from the RNT is odd, then group 1 will be in the first group. For the sick groups, if the # is odd, then group 3 will be in the first group.

6. T is odd 8 6 7 3 5 3 5 5 6 1 7 6 6 4 9 1 8 2 1 7 6 3 4 4 6

H is even H H T T T T T H T T H H H T T H T T H T H H H

10. a) Assign the headsets # 001 - 500. Use RNT to pick 6

three digit numbers. 496, 351, 383, 115, 456, 242 will be inspected.

b) There are 240 minutes between 8 and 12, so pick 10 times at which to inspect a shirt. From the RNT, get 10

3 digit numbers from 001-240. Min 113, 140, 154, 171, 126, 43, 119, 87, 100, 64

c) Use RNT to pick 9 5-digit numbers between 21942 and 98756. Radios 37818, 72142, 67140, 50785, 53362, 44940, 60430, 22834, and 96593 will be inspected.

d) Assign the cartons numbers 001-800 and pick 12 3 digit ~~going~~ [?] numbers from RNT. 594, 68, 84, 675, 530, 140, 634, 795, 139, 255, 887, 716 will be checked. ~~foess~~ 807 +

11. a) Label sheep 001 to 250 and choose 15 3-digit numbers in this range. Sheep 43, 111, 247, 70, 8, 72, 239, 204, 193, 7, 101, 88, 46, 189, 90 will be checked.

b) Pick 10 4 digit numbers from 1024 - 8942. 4045, 6892, 6848, 5624, 2103, 0278, 1739, 2056, 2977, 2678 will be looked out.

c) Pick 5 numbers from 01-25 minutes. Luggages at min 21, 03, 19, 16, and 23 will be checked.

d) Choose 12 two digit numbers from RNT that fall between 01-60. Adults at min 36, 13, 38, 27, 58, 42, 52, 28, 53, 45, 08, and 18 will be surveyed.

12. even=true, odd=false 9 3 7 1 2 5 1 2 8 7 0 5 7 5 4 7 9 3 9 6
F. F F F T F F T T F T F F F F F F F F
Q1 Q2 Q3 Q4 Q5 Q6 Q7 Q8 Q9 Q10 Q11 Q12 Q13 Q14 Q15 Q16 Q17 Q18 Q19 Q20

14 falses, 6 trues

13. choose 10 numbers in the interval 1-5, repeats are OK.

4 3 1 2 2 1 1 5 4 4
D C A B B A A E P P

- 5.39
- a) there is no control group and FIZZ will not know if responses are due to the medication or other factors
 - b) divide the 60 patients into three groups. one will get FIZZ's medication, one will get an aspirin, and one will get a placebo. Then make sure the patients and doctors do not know who is getting what. An hour later, ask all patients how much pain relief they experienced.
 - c) NO. If a patient knows they are getting a placebo, their mind will not think the body is getting better and the patient is aware that no reaction or difference should occur. If the patient does not know, then everybody thinks they have a chance of healing, and the mind responds positively.
 - d) Yes it should be double-blind also because then the doctors will treat each patient the same, showing no favorability or extra care towards those who get the pain relief medication. otherwise, the patients will figure out who has or has not been given the medication based on the doctor's behavior.

	①	②	③	④
5.40	a) Wiensky 42	Brown 34	Loren 32	Smith 29
	Nevesky 39	Cruz 34	Brunk 30	Mann 28
	Birnbaum 35	Jackson 33	Obrach 30	Kendall 28
	Tran 35	Stall 33	Rodriguez 30	Santiago 27
	⑤			
	Hernandez 25			
	Moses 25			
	Deng 24			
	Williams 22			

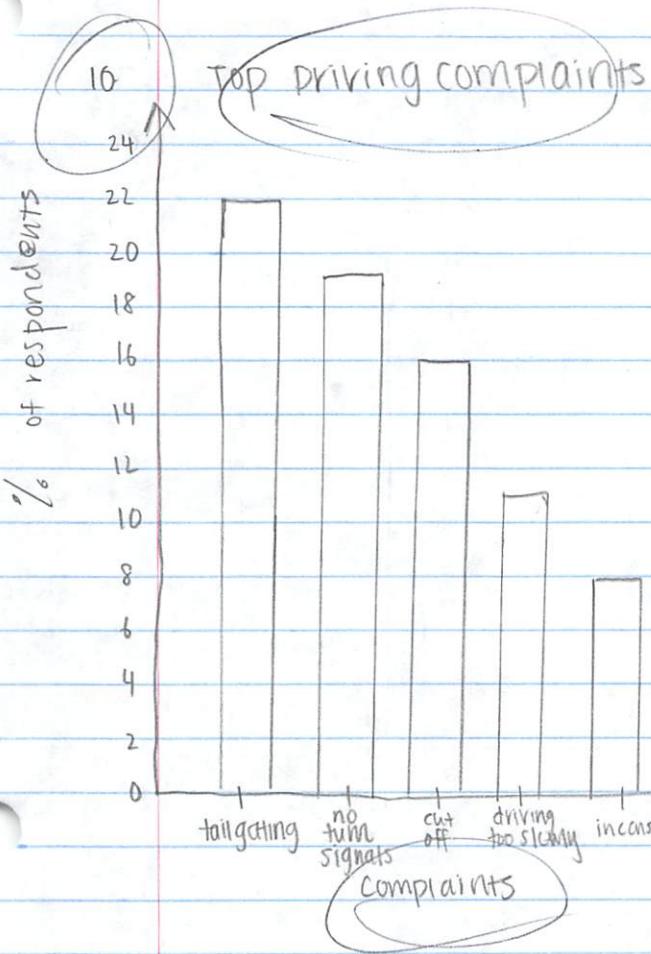
- b) I dropped an eraser and started counting to the right. First I am looking for a # 1-5 to pick a block. The first number in the interval that I come across is # 4. →

NOW IN BLOCK 4, I NEED TO ASSIGN WEIGHT LOSS TREATMENTS.

NOW I'M LOOKING FOR A NUMBER 1-8. FIRST PERSON IS 1+5, SECOND IS 2+6, THIRD IS 3+7, FOURTH IS 4+8. EACH TIME, THE FIRST NUMBER PICKED WILL REFER TO WHO GETS A, THE SECOND WILL REFER TO WHO GETS B, THIRD WITH C, AND FOURTH WITH D. MY NUMBER IS 1, SO SMITH GETS A. THEN 8, SO SANTIGO GETS B, THEN 2 SO MANN GETS C, AND LASTLY, KENDALL GETS D. REPEAT WITH OTHER BLOCKS.

I GOT THE #1, SO I WORK WITH BLOCK 1 NEXT. I DRAW 8, 6, 7, 1. TRAN GETS A, NEVESKY GETS B, BIMBAUM GETS C, WILANSKY GETS A. NEXT NUMBER 3 → BLOCK 3. I DRAW 5, 4, 3, 6. LOREN GETS A, RODRIGUEZ GETS B, ODRACH GETS C, BRUNK GETS D. NEXT NUMBER 2 → BLOCK 2. I DRAW 7, 1, 6, 8. JACKSON GETS A, BROWN GETS B, CRUZ GETS C, STOLL GETS D. ONE BLOCK LEFT, #5. I DRAW 2, 7, 1, 4. MOSES GETS A, DENG GETS B, HERNANDEZ GETS C, WILLIAMS GETS D.

5.42 EACH PERSON WILL EXERT FORCE WITH THE RIGHT AND LEFT HAND. TO DETERMINE WHICH HAND WILL USE THE SCALE FIRST, WE WILL FLIP A COIN. HEADS MEAN RIGHT HAND GOES FIRST; TAILS MEAN LEFT HAND GOES FIRST. DO THIS WITH ALL 10 SUBJECTS.



Analysis

top driving complaints
are the ones that cause
greatest anxiety and have
the greatest potential for
accidents

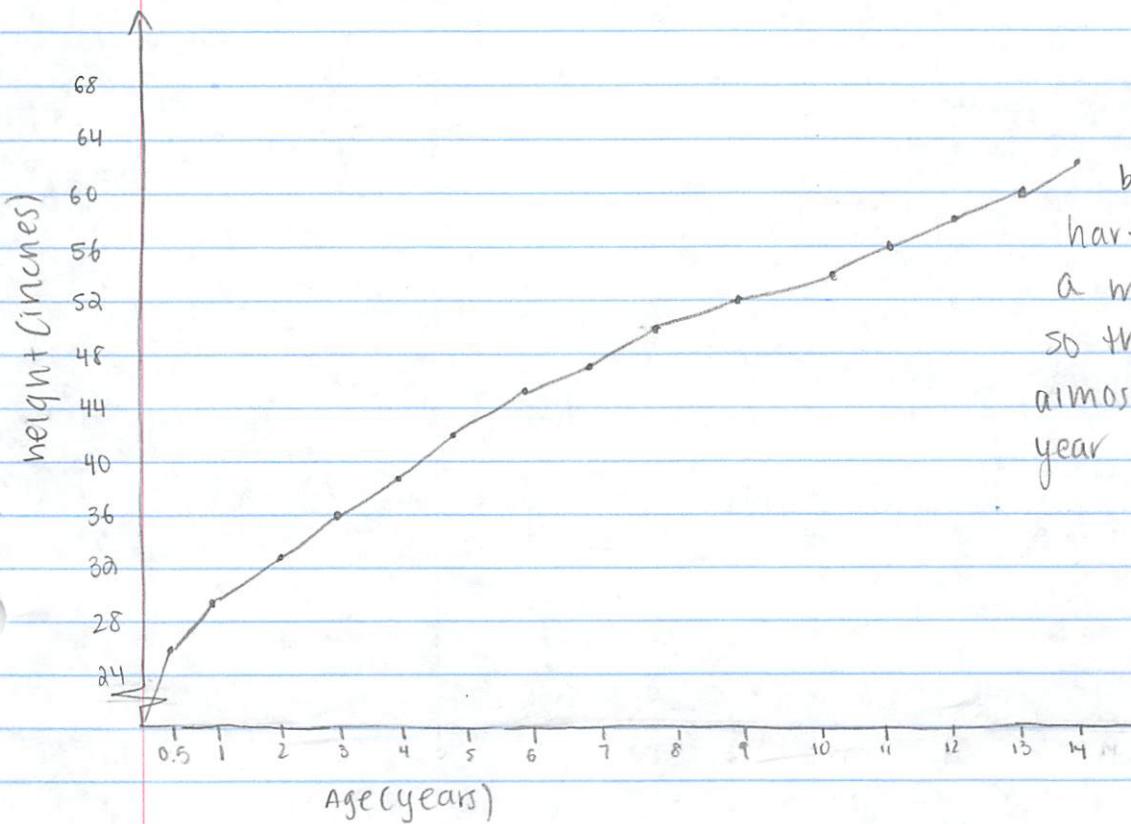
why?

why (lowest?)

* cannot be a pie chart b/c
%'s don't add up to 100

↳ "other" section??

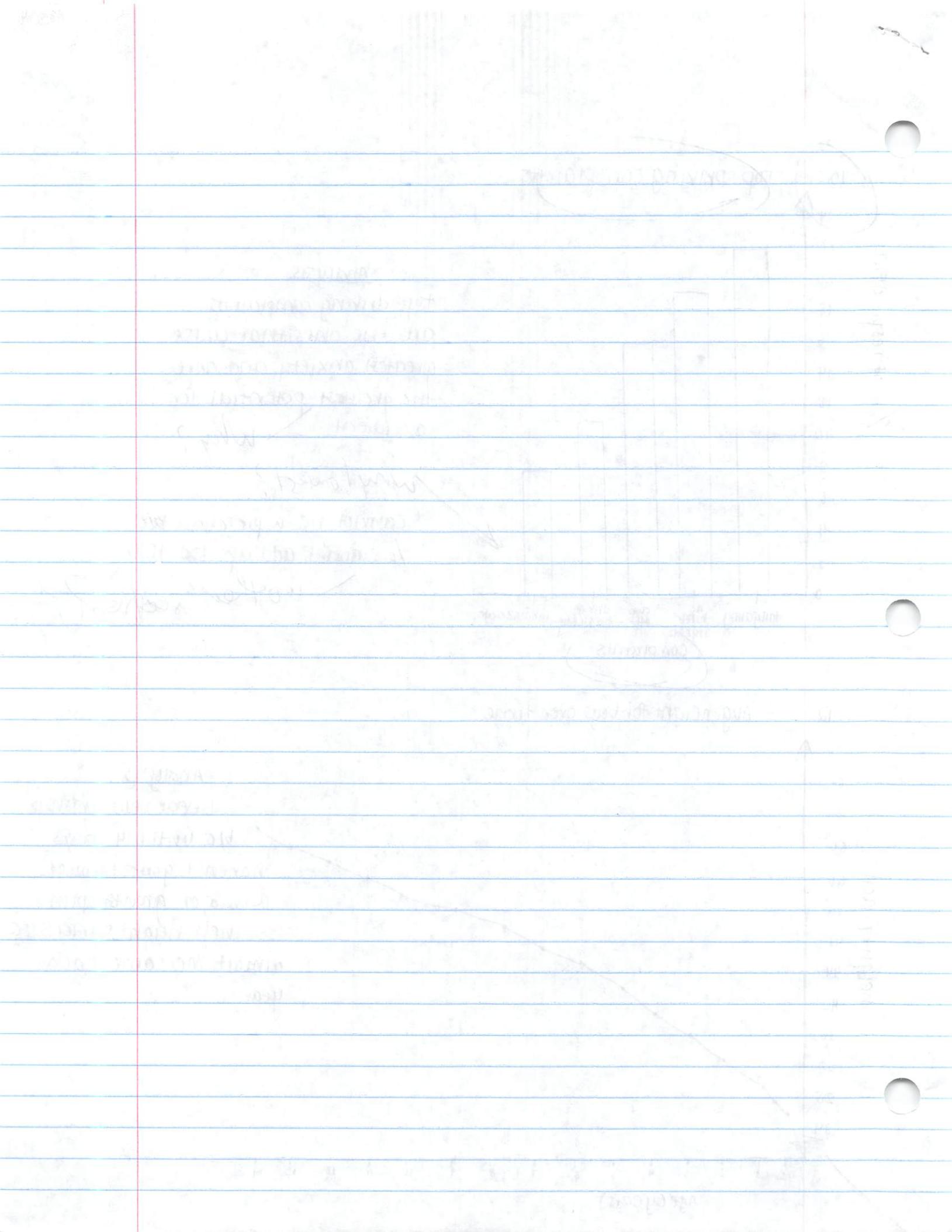
12. Avg height for boys over time



Analysis

- linear relationship

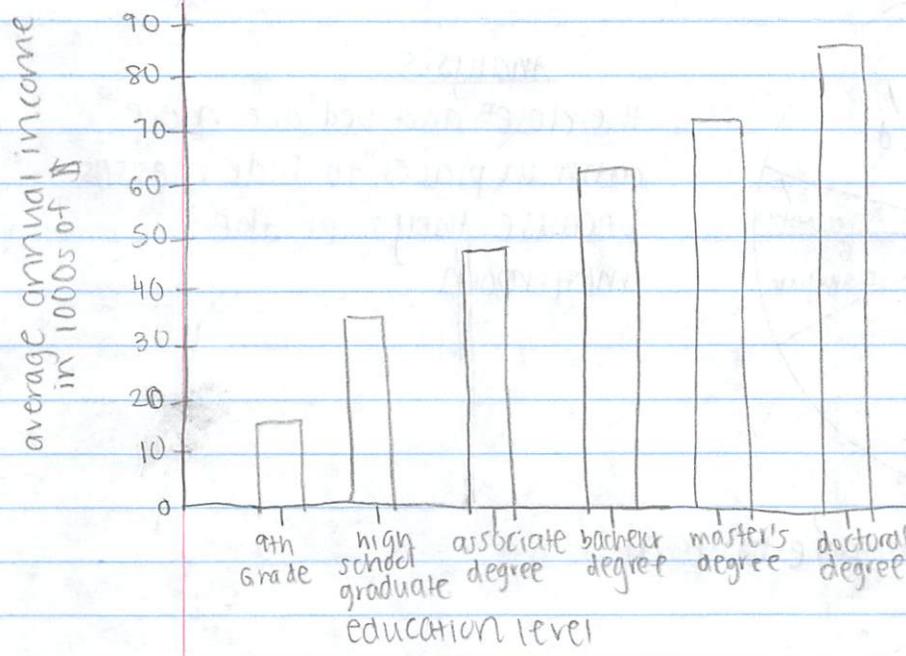
b/c until 14, boys
haven't gone through
a major growth spurt
so their heights increase
almost the same each
year



10-

2.1 #1, 3, 5, 6, 8, 10, 12

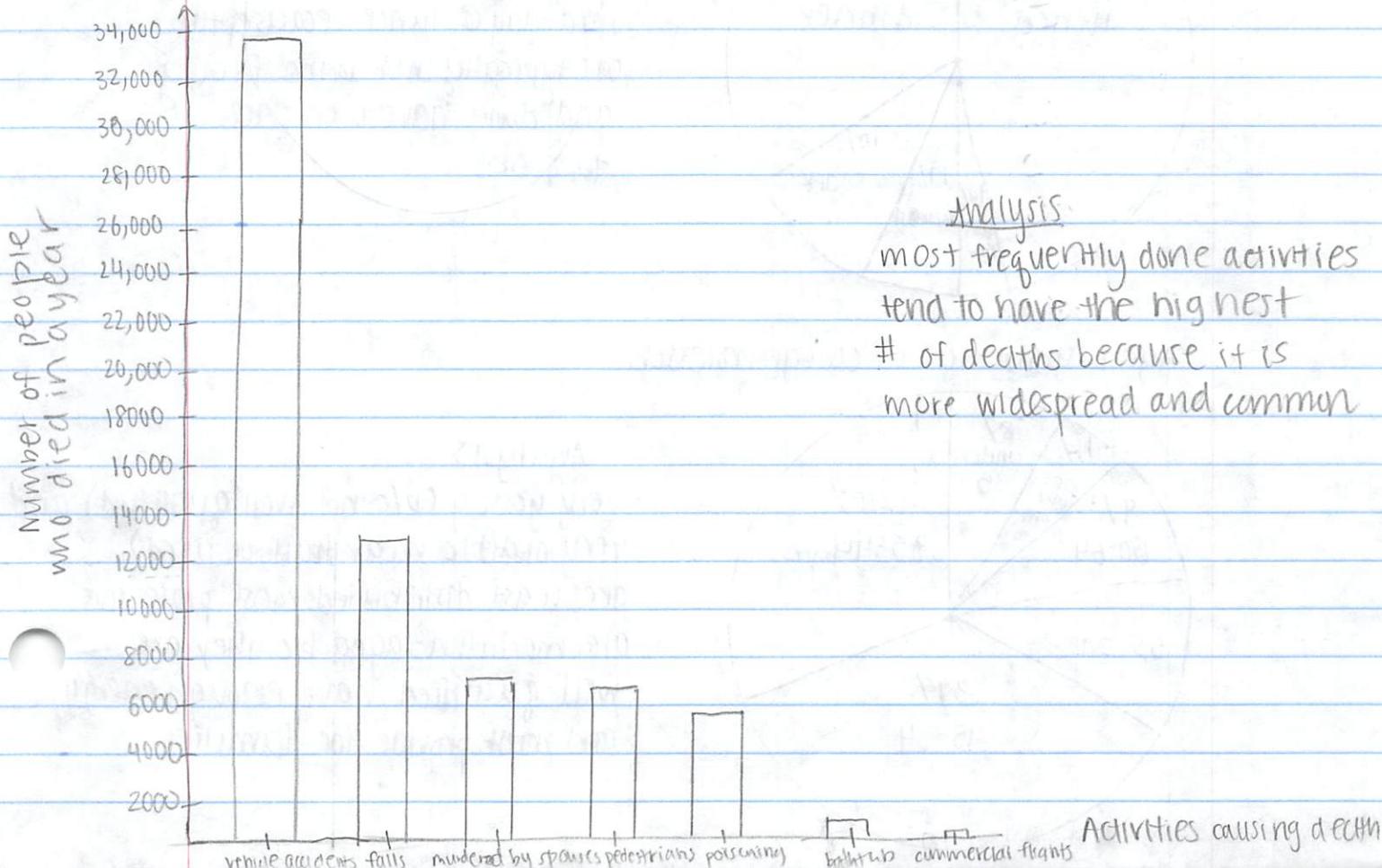
1) Household income for education levels



Analysis

as education level increases, income increases because the people can take more specific or higher quality jobs

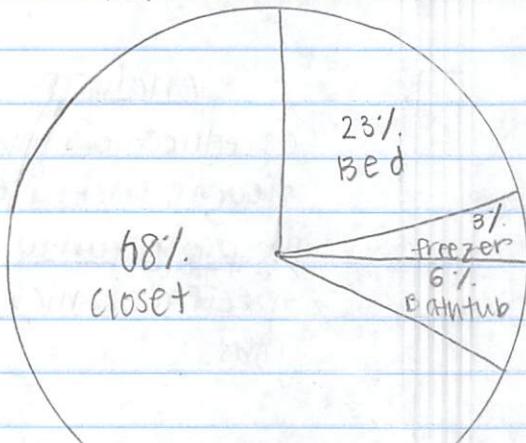
3) Activities causing death



Analysis

most frequently done activities tend to have the highest # of deaths because it is more widespread and common

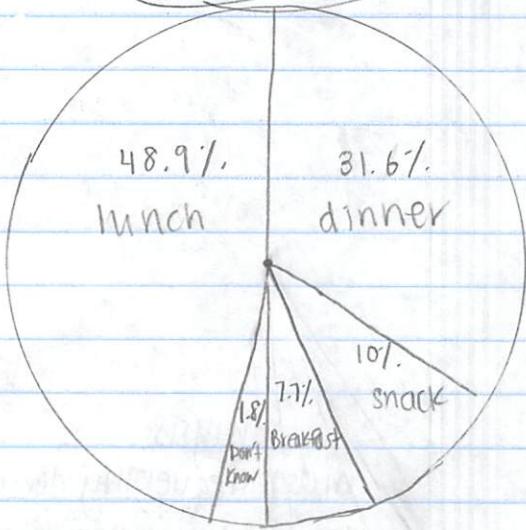
5. Places to hide a mess when a guest is present



Analysis

the closet and bed are more common places to hide messes because they offer the most room

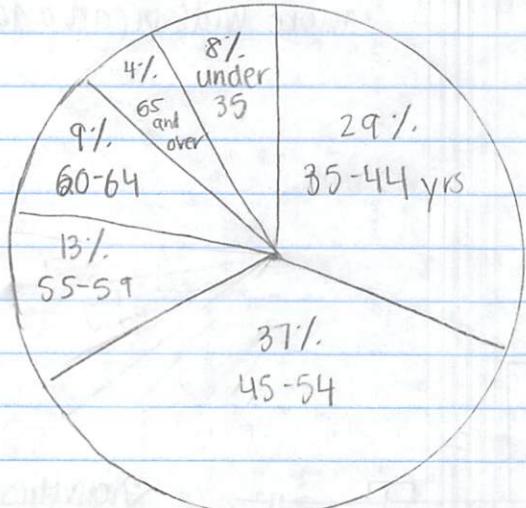
6. Meals to eat in a fast-food restaurant



Analysis

people tend to eat fast-food mid-day @ lunch because they are typically not with family and don't have a chance to cook

8. Age distribution of college professors



Analysis

- very young (b/c not well-qualified) and very old (b/c weak, frail, or tired) are least distributed. Most professors are medium-aged b/c they are well-qualified, have enough energy, and must provide for families

10

1. on a scale of 1-10 with 1 being absolutely not and 10 being absolutely yes, will there be more terrorist attacks on American soil?

1 2 3 4 5 6 7 8 9 10

2. How often do you and your parents regularly discuss the news?

0 1 2 3 4 5 6 7 8 9 10 > 10
times a week times a week

3. DO you favor or oppose high schools offering daily current events classes?

1 2 3 4 5 6 7 8 9 10
oppose favor

01

the best in terms of mean error is LSTM-CNN to predict the

the second highest in terms of mean error is GRU-LSTM-DNN.

Since 0.0079

is the best result in terms of mean error, it is adopted as the final model.

The results of the proposed model are compared with the other models.

The results show that the proposed model has the best performance.

The results show that the proposed model has the best performance.

The results show that the proposed model has the best performance.

The results show that the proposed model has the best performance.

The results show that the proposed model has the best performance.

The results show that the proposed model has the best performance.

The results show that the proposed model has the best performance.

The results show that the proposed model has the best performance.

The results show that the proposed model has the best performance.

The results show that the proposed model has the best performance.

The results show that the proposed model has the best performance.

The results show that the proposed model has the best performance.

The results show that the proposed model has the best performance.

The results show that the proposed model has the best performance.

The results show that the proposed model has the best performance.

The results show that the proposed model has the best performance.

The results show that the proposed model has the best performance.

The results show that the proposed model has the best performance.

The results show that the proposed model has the best performance.

The results show that the proposed model has the best performance.

The results show that the proposed model has the best performance.

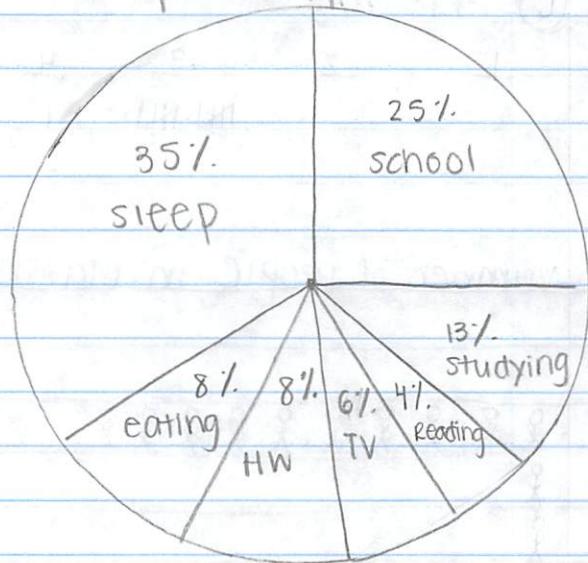
The results show that the proposed model has the best performance.

(A)

- 25%. 6 hours of school
 8%. 2 hours of eating
 6%. 1.5 hours of watching TV
 35%. 8.5 hours of sleeping
 8%. 2 hours of homework
 13%. 3 hours of studying
 4%. 1 hour of reading

Math Lab

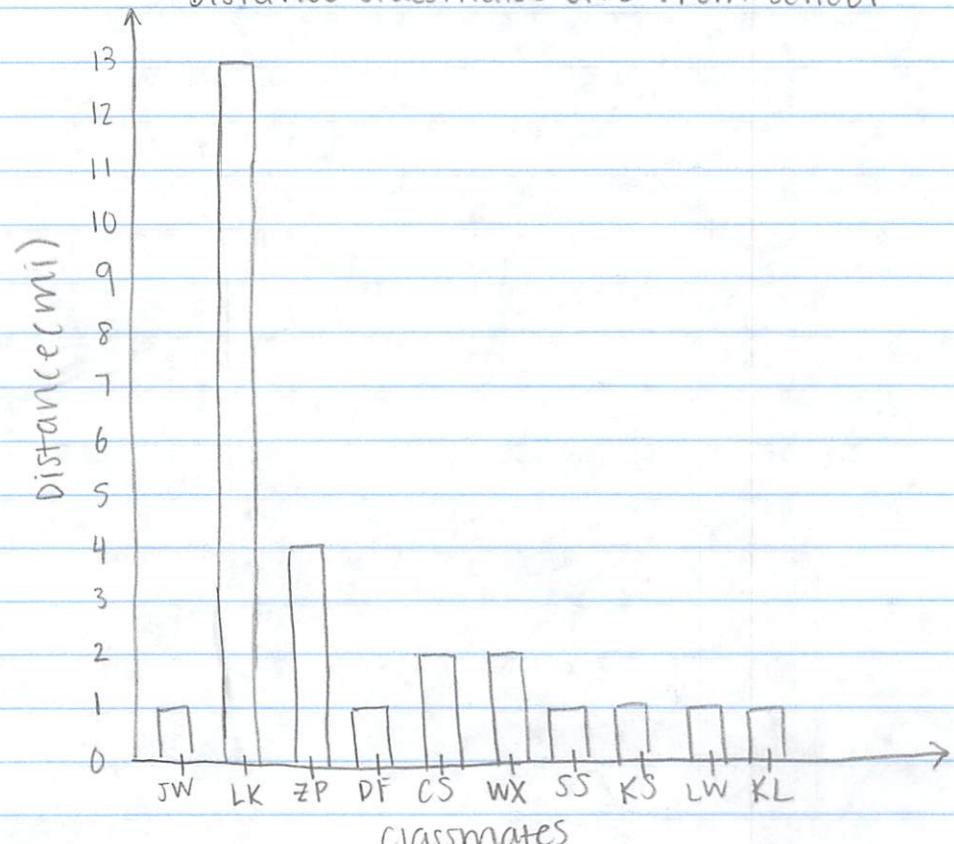
How I spend a typical school day



(B)

JW	JESSICA	1 mi
LK	Leah K.	13 mi
ZP	Zainab	4 mi
DF	Dorinda	1 mi
CS	Catherine	2 mi
WX	wendy	2 mi
SS	shirani	1 mi
KS	Kyra	1 mi
LW	Lean W.	1 mi
KL	Kuenhee	1 mi

Distance classmates live from school



Analysis: A majority lives close to the school b/c going to DPHS is convenient. For LK, the reputation of the school is incentive for attendance.

(c)

1	2	3	4	5	6
NN	NN		1	1	

Number of people in classmates' immediate family

1	
2	
3	○ ○ ○ ○ ○ ○ ○ ○ ○
4	○
5	○
6	

key - ○ = 1 student

(1, 365, 41)

- 310, 203, 308, 59, 161, 87, 339, 146, 265, 361, 15, 80, 225, 185, 86, 326, 36, 321, 61, 239, 251, 25, 175, 161, 104, 120, 147, 144, 168, 83, 33, 318, 31, 234, 167, 290, 329, 37, 291, 165, 319
 - 293, 178, 233, 91, 249, 254, 243, 239, 25, 153, 281, 178, 182, 75, 5, 328, 154, 164, 198, 245, 246, 35, 134, 160, 300, 223, 213, 238, 104, 247, 230, 92, 268, 337, 232, 252, 99, 217, 65, 275, 20
 - 291, 274, 129, 74, 277, 165, 59, 110, 186, 153, 293, 337, 49, 148, 347, 259, 51, 62, 67, 119, 212, 308, 50, 103, 49, 317, 276, 348, 188, 265, 161, 29, 226, 160, 323, 181, 326, 265, 180, 74, 14
 - 151, 42, 353, 195, 291, 161, 43, 247, 32, 290, 184, 62, 81, 167, 294, 202, 181, 274, 251, 96, 67, 342, 152, 314, 62, 351, 34, 130, 347, 96, 115, 216, 94, 226, 351, 124, 253, 94, 111, 160, 348
- 2/4 = 50%.

5.66

- a) Responses from residents in the Province of Ontario: Responses from 61,239 Ontario residents are the sample.
- b) No, perhaps those who didn't respond are those with lower income and their opinions / data are not represented.

5.67

- a) explanatory: breast cancer treatment
response: survival times after treatment
- b) the researcher does not impose a treatment; he merely observes the effects and information
- c) aspects like different health, exercises, diet, stress, and severity of the cancer can all affect different survival times

5.68

Yes, because a treatment (the psychological test) is imposed on the two groups. However, there is no control group to compare the results to.

5.69

- a) Number students 1-3478 and use RNT to find 100 #s from 0001-3478. send a questionnaire to the students with the designated labeled numbers
- b) Students # 2940, 769, 1481, 2975, 1315.

5.61

ASSUMPTION:

- no twins, non-leap year, independent - equally likely
randint(1, 365, 23) which represents dates. look for doubles.
 $\hookrightarrow \text{stat 1} \rightarrow L, \text{stat 2} = \text{sort } A(L, 1)$

Rep 1: N

Rep 2: Y

Rep 3: N

Rep 4: Y

Rep 5: Y

Rep 6: Y

Rep 7: N

$$15/27 \approx 56\%$$

8: Y

9: Y

10: N

11: N

12: Y

13: N

14: N

15: N

16: N

17: Y

18: Y

19: Y

20: Y

21: Y

22: Y

23: N

24: Y

25: Y

26: N

27: N

Simulation Packet

- 5.56 a) First, randomly choose 10 students to question and ask them "DO YOU favor or oppose abolishing evening exams?" I would then repeat it many times to see the likelihood that 100% of 10 favor to abolish.

b) 0-7 = yes 8-9 = no

c) 3675958984	6828822913	1863854303	0079508727
YYYYNYYYYY	YNUNNYYNYY	YNUYNYYYYY	YYYNYYNYYY
6905164817	8717409517	8453406489	8720197245
YNYYYYYYNY	NYYUNNYNYX	NYYYYYYYNN	NYYYYNYYYY
0500716632	8119414873	0419785576	4519596565
*YYYYYYYYYY	NYUNYYUNYY	YUUUNYUXXX	YYUNYNYYYY
6873255259	8429208796	4316593739	3168597160
YNYYYYYYYN	NYYNYYNYNY	VYYYYNYYN	YXUNYNYYYY
4574041807	6556133302	6705193623	1813209547
VYYYYYYNY	*YYYYYYYYYY	YYYYYNYYYY	VNYYYYNYYY
2781678416	1832921337	3521337741	0431268508
YYNYYYYYYY	YNYUNYYYYY	*YYYYVYYYYY	YYYYYYNYYN
6692555658			
YUNYYYYYYN		3/25 = 12%	

- 5.57 a) Assign numbers 1-7 to mean she makes it, and assign 8,9,0 to be a miss. Then use RNT to get 5 numbers and associate it with a make or miss. 7/10 digits = 70% make of her free throw shots.

b) 96746 12149 37823 71868 18442 35119 62103 39244 96927 19931
YYYYY YYYYN YYNYY YYNNY VNYYY YYYYYN VYYYY YNYYY NYNYY VNMYY
*36809 74192 77567 88741 *48409 41903 *43909 99477 25330 64359
YNNNN YYNNY YYYYY NNYYY YNANN YNNNY YNNNN NNYYY VYYYY YYYYY
*40085 16925 85117 36071 15689 14227 06565 14374 13352 49367
YNNNY YYNYY NYYYY YYNYY VYYNN YYYYY NYYYY YYYYY YYYYY YNYYY
11/50 = 22% *81982 *87209 36759 58984 68288 22913 18638 54303 00795 08727
YNNNY NYNNN VYYYN YNNNN YNUNN YNYYX YNYYN XYYNY NNYYN NNYYY
69051 64817 87174 09517 84534 06489 87201 97245 05007 16632
YNNYY VYNYX NYYYY NNYYY NYYYY NYNNN NYNY YYYYY NYNNY YYYYY

5.60 a) use TI-84 to choose 20 numbers between 1 and 100. rand(1, 100, 20).

All numbers from 1-32 chosen means a hit, from 33-100 means a miss. Repeat 20 times.

- b)
- $\frac{H \ H \ H \ H \ H \ M \ H \ H \ M \ H \ M \ M \ H \ M \ H \ M \ M \ H \ M \ H \ M \ H \ M \ H \ M}{29 \ 12 \ 12 \ 12 \ 13 \ 54 \ 24 \ 9 \ 46 \ 25 \ 94 \ 53 \ 28 \ 83 \ 26 \ 86 \ 72 \ 2 \ 85 \ 36 \ 23} = 12/20$
 - $\frac{M \ H \ H \ M \ H \ H \ M \ H \ M \ M \ M \ H \ M \ M \ M \ M \ M \ M \ M \ M \ M}{81 \ 15 \ 1 \ 100 \ 11 \ 8 \ 52 \ 5 \ 83 \ 95 \ 72 \ 93 \ 61 \ 15 \ 100 \ 57 \ 66 \ 78 \ 75 \ 98} = 6/20$
 - $\frac{M \ M \ H \ M \ M \ M \ M \ H \ M \ M \ M \ H \ M \ H \ M \ M \ H \ M \ M}{83 \ 99 \ 9 \ 42 \ 90 \ 51 \ 36 \ 29 \ 93 \ 52 \ 82 \ 10 \ 90 \ 3 \ 90 \ 97 \ 18 \ 4 \ 64 \ 43} = 6/20$
 - $\frac{H \ H \ M \ H \ M \ M \ H \ H \ M \ M \ M \ H \ M \ M \ H \ M \ H \ H \ M}{26 \ 24 \ 68 \ 20 \ 43 \ 100 \ 17 \ 23 \ 69 \ 5 \ 36 \ 8 \ 87 \ 72 \ 28 \ 42 \ 78 \ 22 \ 4 \ 57} = 9/20$
 - $\frac{H \ M \ M \ M \ M \ H \ M \ M \ H \ M \ M \ M \ M \ M \ H \ M \ M \ H \ M}{1 \ 33 \ 67 \ 90 \ 66 \ 6 \ 39 \ 46 \ 26 \ 16 \ 90 \ 89 \ 60 \ 53 \ 74 \ 35 \ 66 \ 54 \ 10 \ 99} = 6/20$
 - $\frac{M \ M \ M \ H \ M \ M \ M \ M \ H \ M \ M \ M \ H \ M \ M \ M \ M \ M}{56 \ 38 \ 90 \ 28 \ 43 \ 91 \ 69 \ 47 \ 15 \ 95 \ 84 \ 53 \ 10 \ 80 \ 80 \ 39 \ 85 \ 37 \ 86 \ 86} = 3/20$
 - $\frac{M \ M \ M \ H \ M \ M \ H \ M \ M \ H \ M \ M \ M \ M \ H \ M \ M \ H \ M}{70 \ 92 \ 65 \ 16 \ 91 \ 47 \ 11 \ 44 \ 11 \ 42 \ 69 \ 63 \ 39 \ 33 \ 38 \ 25 \ 59 \ 55 \ 28 \ 70} = 5/20$
 - $\frac{M \ M \ M \ H \ M \ M \ M \ M \ H \ M \ M \ H \ M \ M \ H \ M \ M \ H \ M}{42 \ 49 \ 57 \ 28 \ 40 \ 48 \ 87 \ 90 \ 29 \ 47 \ 69 \ 28 \ 67 \ 69 \ 29 \ 38 \ 37 \ 86 \ 13 \ 62} = 5/20$
 - $91 \ 25 \ 20 \ 44 \ 78 \ 10 \ 24 \ 8 \ 37 \ 85 \ 47 \ 82 \ 80 \ 38 \ 2 \ 40 \ 28 \ 31 \ 67 \ 64 = 7/20$
 - $58 \ 70 \ 20 \ 81 \ 32 \ 50 \ 41 \ 74 \ 63 \ 63 \ 68 \ 38 \ 40 \ 53 \ 53 \ 55 \ 82 \ 54 \ 67 \ 86 = 2/20$
 - $9 \ 35 \ 82 \ 70 \ 80 \ 67 \ 74 \ 2 \ 88 \ 25 \ 82 \ 38 \ 1 \ 38 \ 19 \ 12 \ 37 \ 81 \ 89 \ 31 = 7/20$
 - $68 \ 34 \ 94 \ 30 \ 39 \ 74 \ 96 \ 1 \ 41 \ 89 \ 51 \ 62 \ 3 \ 13 \ 8 \ 89 \ 94 \ 41 \ 36 \ 33 = 5/20$
 - $86 \ 33 \ 22 \ 47 \ 38 \ 89 \ 22 \ 25 \ 18 \ 69 \ 59 \ 48 \ 99 \ 26 \ 38 \ 25 \ 51 \ 66 \ 70 \ 57 = 6/20$
 - $21 \ 10 \ 53 \ 80 \ 24 \ 56 \ 92 \ 74 \ 4 \ 58 \ 73 \ 20 \ 61 \ 17 \ 67 \ 84 \ 1 \ 36 \ 93 \ 8 = 8/20$
 - $68 \ 80 \ 49 \ 77 \ 86 \ 99 \ 30 \ 40 \ 91 \ 65 \ 58 \ 62 \ 89 \ 38 \ 30 \ 18 \ 76 \ 72 \ 31 \ 34 = 4/20$
 - $82 \ 60 \ 51 \ 76 \ 61 \ 64 \ 34 \ 52 \ 99 \ 25 \ 23 \ 93 \ 1 \ 83 \ 84 \ 82 \ 68 \ 50 \ 71 \ 27 = 4/20$
 - $78 \ 99 \ 42 \ 11 \ 50 \ 31 \ 32 \ 32 \ 72 \ 91 \ 2 \ 15 \ 91 \ 4 \ 58 \ 95 \ 43 \ 40 \ 71 \ 79 = 6/20$
 - $52 \ 16 \ 7 \ 80 \ 44 \ 2 \ 53 \ 12 \ 82 \ 71 \ 28 \ 38 \ 90 \ 96 \ 79 \ 87 \ 93 \ 58 \ 90 \ 63 = 5/20$
 - $46 \ 81 \ 16 \ 98 \ 50 \ 3 \ 34 \ 62 \ 12 \ 27 \ 78 \ 48 \ 95 \ 25 \ 47 \ 10 \ 1 \ 53 \ 81 \ 2 = 7/20$
 - $80 \ 11 \ 81 \ 19 \ 38 \ 19 \ 43 \ 45 \ 70 \ 70 \ 93 \ 36 \ 46 \ 52 \ 56 \ 78 \ 31 \ 5 \ 72 \ 76 = 4/20$
- c) $117/400 = 0.293$, which is slightly less than 0.320

5-61

use calc to generate 23 numbers from 1-365 and run 8 simulations

(on last page)
 $3/8 = 38\%$

57 152 29 85 255 23 217 314 305 309 216 250 252 167 204 212 287 256 24 92 260 295
 261, 243, 39, 196, 277, 39, 338, 209, 157, 142, 206, 358, 62, 226, 302, 148, 209, 119, 110, 51, 284, 236, 230
 19, 47, 107, 128, 124, 282, 17, 96, 98, 241, 128, 274, 55, 125, 207, 131, 114, 131, 121, 98, 254, 211, 76
 168, 223, 195, 61, 64, 123, 5, 321, 188, 271, 121, 350, 253, 198, 191, 348, 86, 279, 125, 18, 141, 144,
 194, 352, 34, 10, 118, 322, 239, 321, 244, 18, 27, 348, 125, 221, 339, 343, 289, 364, 56, 210, 89, 147, 76
 51, 190, 255, 275, 167, 340, 141, 272, 173, 83, 335, 318, 20, 26, 284, 119, 251, 321, 349, 1195, 49, 268, 191
 103, 289, 11, 55, 90, 157, 175, 26, 31, 262, 113, 113, 335, 87, 198, 49, 36, 76, 45, 8, 140, 310, 4
 340, 15, 71, 155, 75, 183, 1190, 286, 302, 248, 242, 269, 340, 31, 183, 212, 290, 86, 310, 177, 120

CHI RW 6-8

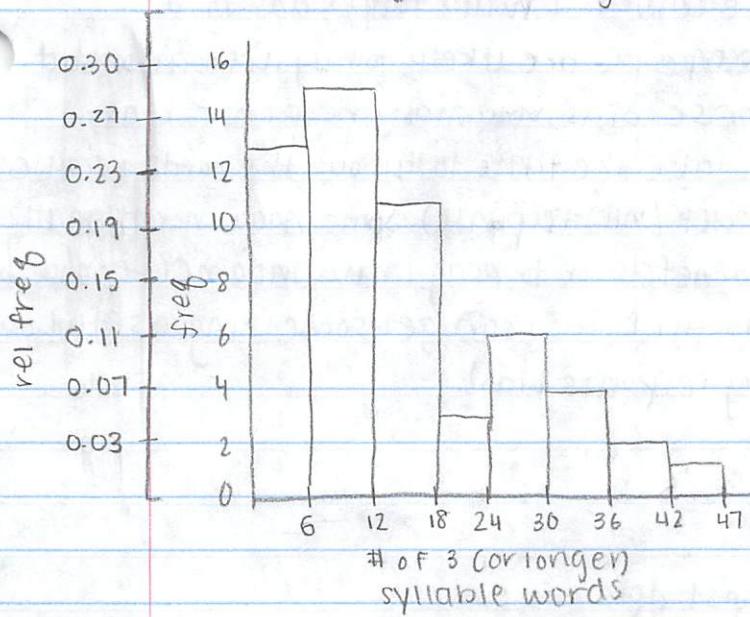
- 6a) cluster
- b) convenience
- c) systematic
- d) simple random
- e) stratified

- 7a) observational, no treatment from the researcher was imposed and the population self-assigns age
- b) experiment, the treatment is a multiple choice or essay exam and the results of the 2 groups are compared
- 8a) use RNT or calculator to assign half to receive literature and be asked for donations and the other half to be contacted by phone and be asked for donations. There is no placebo because both groups know exactly what treatment it is receiving.
- b) Randomly assign using RNT or calculator for 43 to be given actual tooth-whitening chemicals. The dentists/facilitators and the subjects should not know which toothpaste they are getting. Then, compare the results of the whiteness of the teeth.

6a) $43 - 2 = 6$ (width)

b) class limits	midpoints	freq	rel freq
0 - 5	2.5	13	0.24
6 - 11	8.5	15	0.27
12 - 17	14.5	11	0.20
18 - 23	20.5	3	0.05
24 - 29	26.5	6	0.11
30 - 35	32.5	4	0.07
36 - 41	38.5	2	0.04
42 - 47	44.5	1	0.02

c+d) # of 3 (or longer) syllable words
randomly selected magazine ads



Analysis

- center: 0-18
- shape: skewed right
- spread: mean - 14.4

med - 11

Magazines are targeted mostly to the common folk, so they don't want to use fancy or large vocabulary. Most use relatively few 3 or longer syllable words so the audience can understand the content.

- 9a)
- skewed left
 - uniform
 - symmetrical
 - bimodal
 - skewed right
- b) skewed left - too many get high grades, uniform - everyone gets equal distribution so a 1% difference could determine an A or a B. symmetrical is good because the most people will receive the average score. Test 4 is either you

get the material or you don't. Test S (skewed right) causes too large of a number of students to fail. I would want test A so that the number of A's is large and relatively easy to do well on the test. I do not like Test S, because then a disproportionate number of students fail because the test is too hard.

100) a - uniform

b - skewed right

c - skewed left

d - bimodal

e - symmetrical

b) I would target microwaves to everyone, electric fans for lower income families, computers for upper class, TVs for low + upper classes, and snowblowers to medium income families. I would target ads to a specific income-distribution so more people are likely to buy a certain product.

c) Not valid, there is non-response bias, voluntary response bias, and response bias. Some do not take the time to fill out the cards because they cannot waste \$ to send it back (non-response). Some may intentionally lie, thinking they will get more benefits or to brag about income (response bias). The responders are the ones who think they can get something useful out of filling out the cards (voluntary response bias).

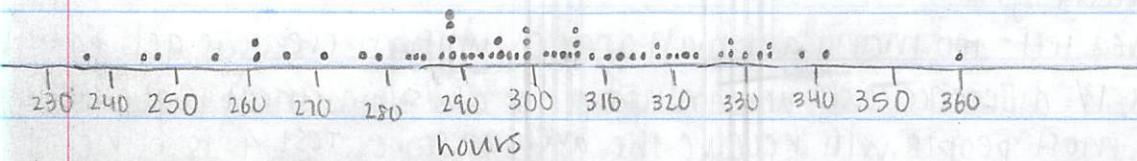
13) a) 1

b) $S/S_1 \approx 10\%$.

c) 650-750; there are the most dots.

14) can see overall spread, center, outliers, and specific data points on the dot plot, but a histogram doesn't reveal individual points or outliers well

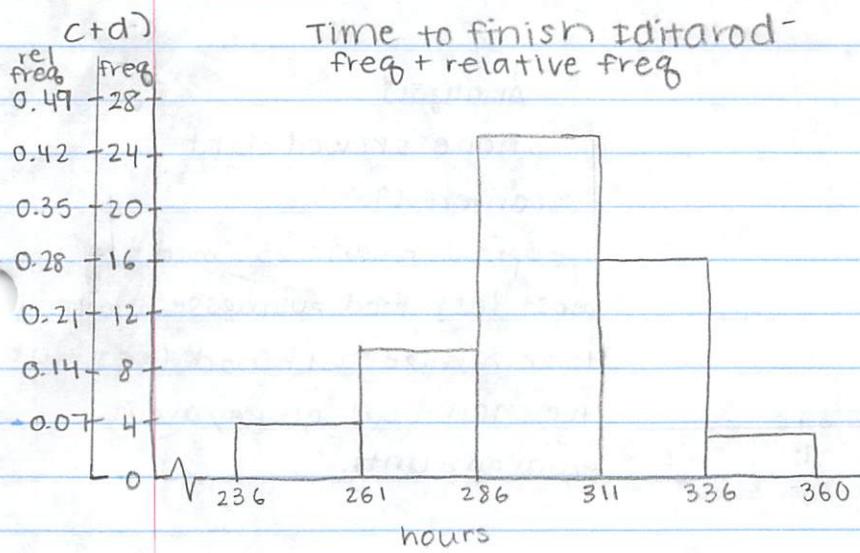
Finish time (hrs) for Iditarod Dog Sled Race



2.2 # 1, 2, 4-6, 9, 16, 13, 14

a) $\frac{360 - 236}{5} = 25$

b) class limits	Midpoint	Freq	Relative Freq
236 - 260	248	4	0.07
261 - 285	273	9	0.16
286 - 310	298	25	0.44
311 - 335	323	16	0.28
336 - 360	348	3	0.05

Analysis

center: 286 - 311

shape: symmetrical

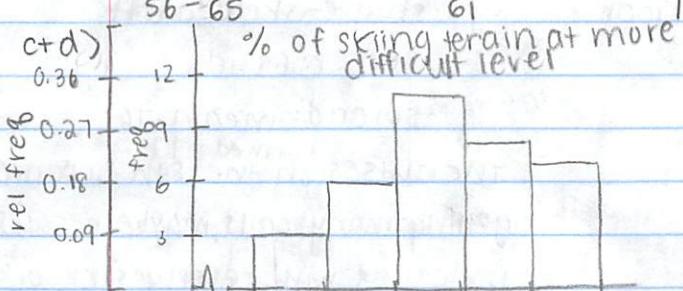
spread: mean - 300

median - 299

The dogs run about the same speed and can endure about the same time and weather. Thus, a majority of the times will be very close.

a) $\frac{65 - 20}{5} = 9$

b) class limits	Midpoint	Freq	Relative Freq
20 - 28	24	3	0.09
29 - 37	33	6	0.17
38 - 46	42	11	0.31
47 - 55	51	8	0.23
56 - 65	61	7	0.20

Analysis

spread: mean - 45, med - 45

center: 38 - 47

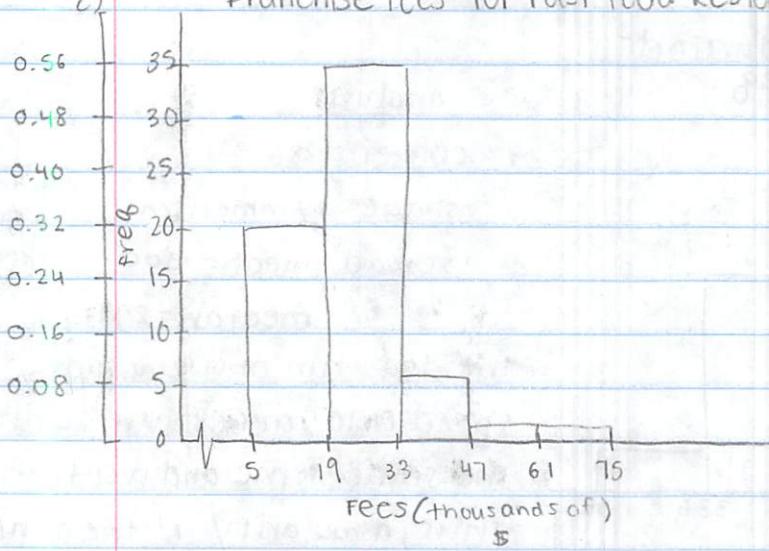
shape: skewed left

Ski areas want about equal distribution among easy, med, and hard to most have about 38-50% designated as more difficult

4a) $\frac{75-5}{5} = 14$ (width)

b) Class Limits	Midpoint	Freq	Rel Freq
5 - 18	11.5	20	0.32
19 - 32	25.5	35	0.56
33 - 46	39.5	6	0.10
47 - 60	53.5	1	0.016
61 - 75	67.5	1	0.016

c) Franchise fees for fast-food Restaurants



Analysis

- shape: skewed right
- center: 19-33
- spread: mean - 23 med - 24
- most fast-food businesses want to minimize fast-food fees, but the chains will all pay about equal amounts.

5a) $\frac{102-18}{5} = 17$ = (width)

b) Class Limits	Midpoint	Freq	Rel Freq
18 - 34	26	1	0.03
35 - 51	43	2	0.06
52 - 68	60	5	0.14
69 - 85	77	15	0.43
86 - 102	94	12	0.34

Analysis

- shape: skewed left
- center: 69-102
- spread: mean - 76 med - 77
- The nurses on the 8th floor are getting many calls, maybe because it is dealing with severely sick or old people. Either way, there needs to be more staff.

c+d) Room calls to Nurses on 8th Floor on 35 Random Nights

