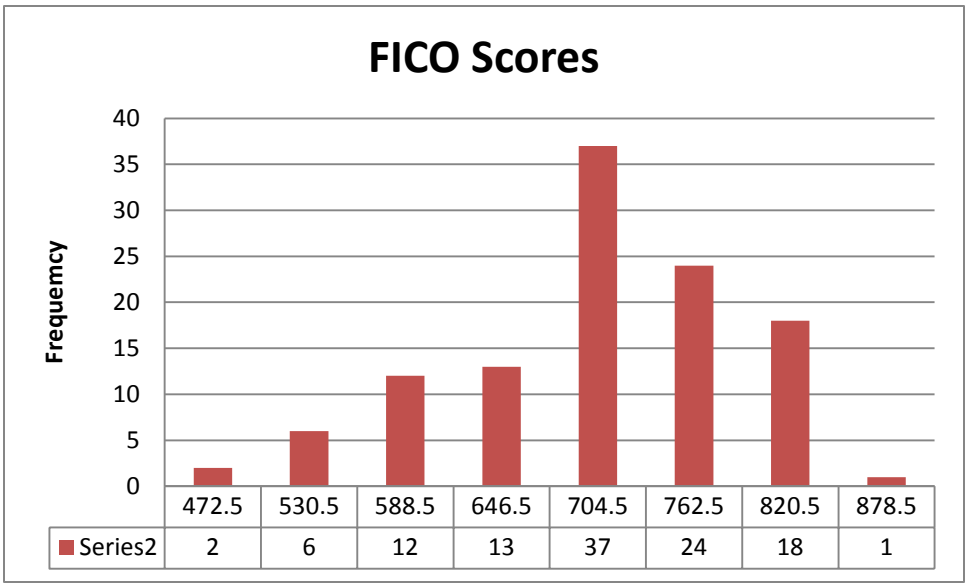
Final Exam Review
STAT 1450, Fall 2013

Instructions: Answer each question as thoroughly as possible. Be sure to use correct notation. To earn full credit, you must show work or the display from your TI-83/84 calculator screens. Round answers to three decimal places where appropriate (where the rounding rules discussed in class do not apply). Formulas are at the back. Note: The review is more comprehensive than the actual exam and will likely take longer to complete than the real test even working straight through.

1. In order to determine the average FICO score of Americans, a simple random sample of 100 scores were taken as reported by Experian. Use the data shown in the chart below to answer the following questions.



a. Population Americans (with credit ratings)

b. Sampling Method simple random sample

c. Data Variable FICO score

d. Is the data qualitative or quantitative?

e. Is the data nominal, ordinal, interval or ratio? (credit scores don't go to zero)

f. Class Width 58

g. Sample Size 100

h. Mean 704.5 1-VarStats L_1, L_2

i. Median 704.5

j. Modal Class 704.5

k. If a trade publication reported only the mean FICO score, would this be misleading? Why or why not?

No, it would not be misleading, because the mean/median/mode are all the same.

l. Standard Deviation 86.36

m. What does the standard deviation tell you about the data?

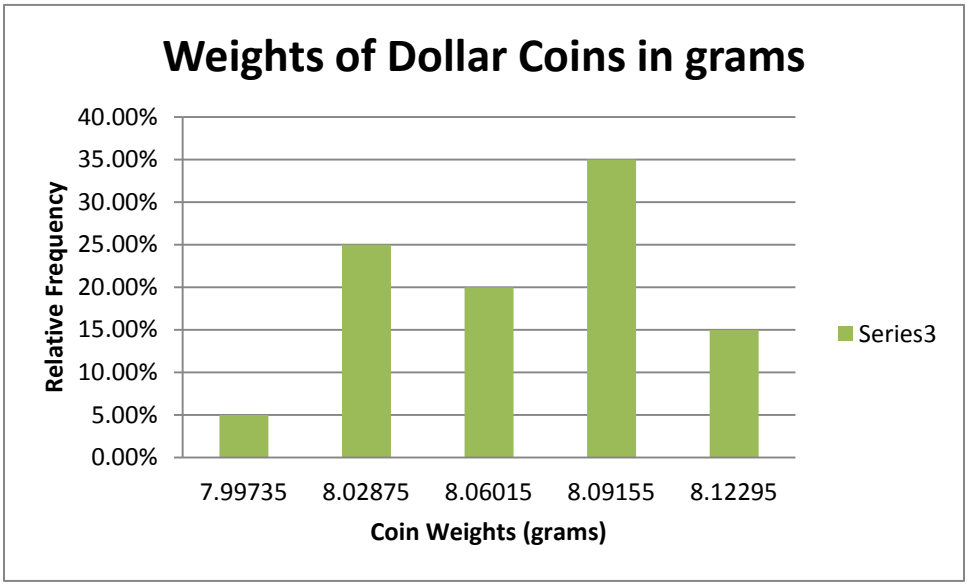
The spread of the data

n. Shape? Approximately normal

o. Explain how the descriptive statistics confirm the shape.

The mean/median/modal class are all the same.

2. In order to determine the average weight of dollar coins produced by the US Mint, random sample of 20 coins in circulation were taken from a single two banks in the Midwest. Use the data shown in the chart below to answer the following questions.



a. Population dollar coins made by US Mint

b. Sampling Method stratified sampling (though it may be characterized as convenience sampling)

c. Data Variable weight in grams

d. Is the data qualitative or quantitative?

e. Is the data nominal, ordinal, interval or ratio?

f. Class Width 0.0314

g. Sample Size 20

h. Mean 8.06957

i. Median 8.07585

j. Modal Class 8.09155

k. If a trade publication reported only the mean weight of dollar coins, would this be misleading? Why or why not?

Perhaps a little; the mean/median/modal class are all different, suggesting a skewed distribution, even though the absolute differences are quite small

l. Standard Deviation 0.03612

m. What does the standard deviation tell you about the data?

How spread out the data is.

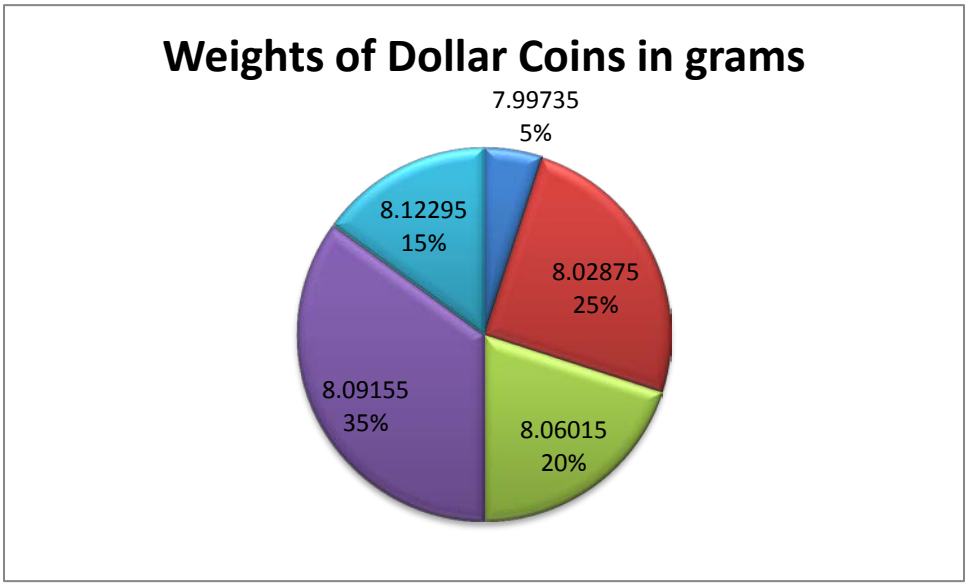
n. Shape? The graph appears to have two peaks (bimodal), but this may be a statistical anomaly. Ignoring that suggests left-skewed.

o. Explain how the descriptive statistics confirm the shape.

The relationship between the mean/median/modal class suggest left-skewed.

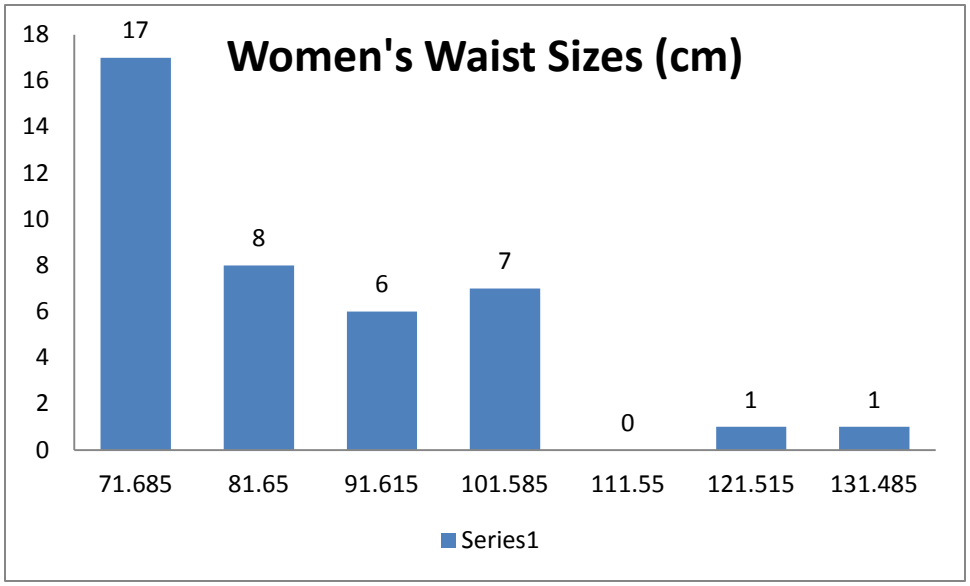
p. Use the information in the chart to build a pie chart for the data. Is this a better or worse method of visualizing the data? Explain your answer.

This is worse since the data is continuous, this is rather inappropriate. It also makes it harder to see the shape of the data distribution.



3. The data below was collected from 40 women measuring the sizes of their waists in centimeters. Use the information to answer the questions that follow.
67.2, 82.5, 66.7, 93.0, 82.6, 75.4, 73.6, 81.4, 99.4, 67.7, 100.7, 72.9, 85.0, 85.7, 126.0, 74.5, 74.5, 94.0, 92.8, 105.5, 75.5, 126.5, 70.0, 98.0, 104.7, 67.8, 99.3, 91.1, 74.5, 95.5, 79.5, 69.1, 105.5, 78.8, 85.7, 92.8, 72.7, 75.9, 68.6, 68.7

a. Construct a histogram of the data. Be sure to clearly label your class midpoints and the frequency of each class.



b. Population women

c. Data Variable waist sizes in centimeters

d. Is the data qualitative or quantitative?

e. Is the data nominal, ordinal, interval or ratio?

f. Class Width 9.965

g. Sample Size 40

h. Mean 85.03

i. Median 81.95

j. Modal Class 71.685

k. If a beauty magazine reported only the mean waist size, would this be misleading? Why or why not?

Very misleading since the data is heavily right-skewed.

l. Standard Deviation 15.401

m. What does the standard deviation tell you about the data?

spread

n. Shape? Very right-skewed

o. Explain how the descriptive statistics confirm the shape.

The modal class is well below the mean or the median, with the mean the largest of the three.

4. Use the following data on the length of R movies to construct a stem-and-leaf plot. Use the graph to answer the questions below.
R: 110, 96, 120, 113, 108, 94, 90, 154, 127, 150, 136, 119.

9	046
10	8
11	039
12	07
13	6
14	
15	04

a. Shape of the graph. Difficult to say from this little data. Multimodal?

b. Percentile rank of 130 minutes.

75 percentile (there are 9 datapoints out of 12 less than 130).

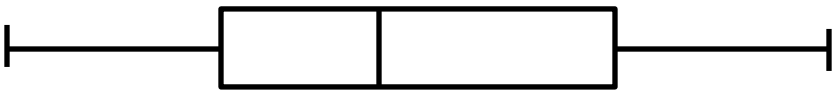
c. Write a statistically true statement about the length of R movies.

The median length of R movies is under 120 minutes. (answers may vary)

d. Give the five number summary of the data above.

Min: 94, Q1: 108, Median: 119, Q3: 136, Max: 154

e. Use that information to construct a box plot of the data.



f. Are there any outliers?
no

g. Suppose we also included data from PG-13 movies, PG movies and G movies. What type of graph would be best used to compare the results?

pareto

5. Use the data below on Brown and Yellow M&Ms plain candy weights (grams) to construct a stem-and-leaf plot comparing the two data sets. Use the graph to answer the following questions.
Yellow: 0.883, 0.769, 0.859, 0.784, 0.824, 0.858, 0.848, 0.851
Brown: 0.696, 0.876, 0.855, 0.806, 0.840, 0.868, 0.859, 0.982

Yellow		Brown
	.6	9
	.7	
87	.7	
2	.8	14
86655	.8	6678
	.9	
	.9	9

a. What is the shape of each graph?

Basically normal, given the small amount of data

b. What is the percentile rank of 0.830 grams for each color of M&M candy?

38th for both sets (3 of 8 values are less)

c. Write a statistically true statement about the claim that both candies weigh the same.

The mean weight of both candies is within 0.01 grams of each other (answers will vary)

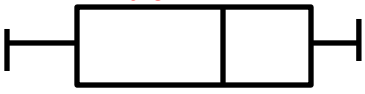
d. Which color candy has a more consistent weight?

yellow

e. Use your TI-83/84 to construct a boxplot for each data set. Note any outliers on each.

Yellow

brown



6. A sample of 1504 American adults were recently surveyed by Pew Research (http://www.people-press.org/2013/11/01/gop-deeply-divided-over-climate-change/) to determine how party affiliation affected views of climate change. The data from the survey is briefly summarized below.

	Humans primary cause	Other causes primary/not occurring
Democrat/Lean Democratic	482	271
Republican/Lean Republican	172	579

a. What is the probability that a randomly selected American agrees that humans are causing climate change?

654/1504=0.4348… or 43.48%

b. What is the probability that a randomly selected American is both Republican (or Republican leaning) and believes that humans are not causing climate change?

579/1504=0.38497… or 38.50%

c. What is the probability that a randomly selected American agrees that climate change is caused by humans, given that they are Republican (or Republican leaning)?

172/751 = 0.229… or 22.9%

d. What is the probability that 4 randomly selected Americans will agree that humans are causing climate change given that they are Democrats (or Democratic leaning)? (without replacement)

(check for independence: 4/1504<0.003)

$$\left(\frac{482}{753}\right)^4 = 0.16788\ldots or\ 16.8\%$$

e. Construct a hypothesis test to determine if party affiliation affects one's views on climate change. Use a 1% level of significance.
(contingency table)

i. H_0: views on climate change are independent of party affiliation

H_1: views on climate change are dependent on party affiliation

ii. Test Statistic: $\chi^2 = 259.404243\ldots$

iii. P-Value: 2.31×10^{-58}

iv. Reject or Fail to Reject the null hypothesis.

v. State your conclusion in the context of the problem.
Party affiliation does affect one's views on climate change.

7. A sample of 1660 American adults were recently surveyed by Business Insider-Survey Monkey (http://www.businessinsider.com/poll-the-major-differences-between-how-single-men-and-women-approach-sex-2013-9) to determine how men and women behaved on the first date. The data from the survey is briefly summarized below.

	Slept with someone on the first date	Never slept with someone on the first date
Men	333	406
Women	258	663

a. What is the probability that a randomly selected American says they have slept with someone on their first date?

591/1660 or 35.6%

b. What is the probability that a randomly selected American is both a woman and has slept with someone on a first date?

258/1660 or 15.5%

c. What is the probability that a randomly selected American has slept with someone on a first date given that they are a woman?

258/921 or 28.01%

d. What is the probability that 3 randomly selected Americans will admit to sleeping with someone on a first date given that each is a man? (without replacement)

(check for independence 3/1660<5%)

$$\left(\frac{333}{739}\right)^3 \approx 9.15\%$$

e. Construction a hypothesis test to determine if gender plays a role in determine how likely someone is to be willing to sleep with a partner on a first date. Use a 5% level of significance.

i. H_0: men and women behave the same when it comes to sleeping with someone on the first date

H_1: men and women behave differently when it comes to sleeping with someone on the first date

ii. Test Statistic: $\chi^2 = 51.974$

iii. P-Value: $5.623\ldots \times 10^{-13}$

iv. Reject or Fail to Reject the null hypothesis.

v. State your conclusion in the context of the problem.

The likelihood of sleeping with someone on a first date depends on the gender of the person being asked.

8. About 10% of people are left-handed. Suppose that we choose 40 people randomly.
   a. Find the mean.
   np=4

   b. What is the standard deviation?

$$\sqrt{npq} = 1.897$$

   c. What is the probability that exactly 5 people in the sample will be left-handed?

Binomialpdf(40,.1,5)=16.47%

9. On average, suppose that 42 people arrive at a drive-thru window during lunch hour, between 11:30 and 12:30 p.m. The arrival times follow a Poisson distribution.
   a. What is the mean number of cars between 11:45 and 12:00 p.m.?

   10.5

   b. Calculate the probability that at least 15 new cars will arrive between 11:45 and 12:00 p.m.

   1-poissoncdf(10.5,14)=11.21%

10. According to a Harris Interactive Poll (http://www.harrisinteractive.com/NewsRoom/HarrisPolls/tabid/447/mid/1508/articleId/970/ctl/ReadCustom%20Default/Default.aspx), about 1-in-5 people have tattoos. Suppose that we choose a sample of 88 people.
    a. Calculate the mean number of people in the sample with tattoos.

    88/5=17.6

    b. What is the standard deviation?

    3.75

    c. What is the probability that 30 people or more in the sample have tattoos?

    1-bionomialcdf(88,.2,29)=0.00139

11. The average lifetime of an LED light bulb is about 50,000 hours, with a standard deviation of around 20,000 hours.
    a. Determine the lifetime of a light bulb in the 10^{th} percentile.

    invNorm(.1,50000,20000)=24,369 hours

    b. Is it unusual for an LED light bulb to last longer than 85,000 hours?

    No. z-score for 85K is 1.75<2

    c. If 20 LED light bulbs are tested, what is the probability that the sample mean is greater than 60,000 hours?

    (central limit theorem) $\frac{\sigma}{\sqrt{n}} = 4472$, z-score=2.235 or 1.267%

12. The average IQ on all IQ tests is set at 100, but IQ tests differ by their standard deviations. Use your knowledge of a normal distribution to determine which of two students received a better score on their IQ test if the first student received a score of 135 on a test with a standard deviation of 15, and the second student received a score of 145 on a test with a standard deviation of 20.

    35/15=2.333 compared to 45/20=2.25, 135 corresponds to the higher z-score and so is the higher IQ

13. The random number generator in your calculator generates numbers between 0 and 1 that are uniformly distributed. What is the probability that when it produces a random number that the value will be between 0.25 and 0.60?

    35%

14. Polling firm has conducted a survey and found that the Green Party candidate is receiving 52% of the vote and the Libertarian candidate is receiving 48% of the vote.
    a. If 1500 people were surveyed, calculate a 95% confidence interval for the Green Party candidate.

       (49.47,54.53) (1-PropZInt)

    b. Based on this information, what is a statistically accurate claim about who is winning the election (if anyone) according to the poll?

       The election is too close to call.

    c. If the polling firm wanted to relax their confidence to 90%, what would happen to the interval and explain why.

       The interval will shorten since the z-value used to calculate it is smaller

15. The average breaking length from 60 mph for domestic automobiles is given below. Use this data to answer the following questions.
    131, 136, 129, 127, 146, 146, 155, 139, 133, 131, 131, 140, 137, 144, 133, 122, 133, 143, 144, 133, 134, 143, 129, 128, 140, 144, 146, 130, 139, 140, 135, 136
    a. Find the mean of the sample.

       136.78 (1-VarStats)

    b. Find the standard deviation of the sample.

       7.161

c. Use this data to construct a 95% confidence interval for the mean breaking distance of domestic automobiles.

(135.815,137.745) TInterval

d. If you were running a highway safety program, would a 95% confidence interval be enough to base safe following distance on? What would happen to the interval if you increased the confidence level to 99.9%?

This is probably not safe enough for a recommendation since 5% of accidents is a lot. The interval would get wider to allow for a greater margin of safety and fewer accidents (t-value increases).

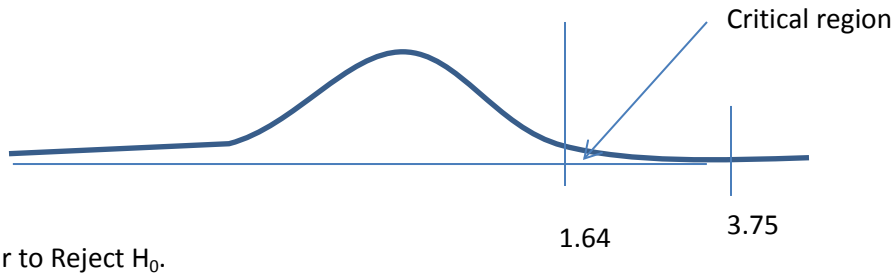
16. A TV station is thinking about developing a program for television. In order to guarantee enough audience for the program, the station needs to show their backers they will have enough viewers to justify the cost of producing the pilot. The backers set a threshold of 20% of the surveyed audience being interested in seeing the program before they will give the money for the pilot. A survey of 100 randomly selected viewers of the station determines that 35% of the population says they would be interested in watching the show based on a plot synopsis. Conduct a hypothesis test to determine if this is enough to earn the backing of investors.

a. State the null and the alternative hypothesis.
H_0: $p \leq .2$, H_1: $p > .2$

b. Test statistic

Z=3.75

c. Sketch the sampling distribution, clearly labeling the critical values and the rejection region.



d. Reject or Fail or to Reject H_0.

e. State your conclusion in the context of the problem.

There is definitely more than the threshold of 20% needed to support the pilot.

17. For a school science project, a student wishes to determine if the flight times of shuttle missions changed over time. She has chosen ten flight times from the beginning of the shuttle program and ten flight times from the end of the shuttle program. The data is shown in the chart below.

Early	54	54	192	169	122	146	145	247	191	167
Late	285	283	262	259	332	259	330	382	333	306

Conduct a hypothesis test at the 20% level of significance to determine if mean flight times of shuttle missions have increased over time.

a. State the null and alternative hypothesis.

$H_0: \mu=\mu_0$, $H_1:\mu\neq\mu_0$

b. Test statistic t=-6.7038 (2-SampTTest)

c. P-Value $5.379x10^{-6}$

d. Reject or Fail to Reject H_0.

e. State your conclusion in the context of the problem.

Mean flight times for later shuttle flights are longer than mean flight times for earlier shuttle flights.

18. Ohio State University has a 6-year graduation rate of 78.1%. DeVry University, on the other hand, has a 33.6% graduation rate after 6 years. Suppose that each is based on a sample of 1000 students.
    a. Construct a 98% confidence interval for both schools to determine the proportion that graduates after 6 years.

    OSU (0.75058,0.81142) 1-PropZInt
    DeVry (0.30125,0.37075)

    b. State the appropriate null and alternative hypothesis if Ohio State University wishes to claim that they have a higher graduation rate than DeVry University.

    $H_0: \mu\leq\mu_0$, $H_1:\mu>\mu_0$

c. Reject or Fail to Reject H_0.

P=1.369x10^{-89}

d. State the conclusion in the context of the problem.

OSU has a higher graduation rate than DeVry.

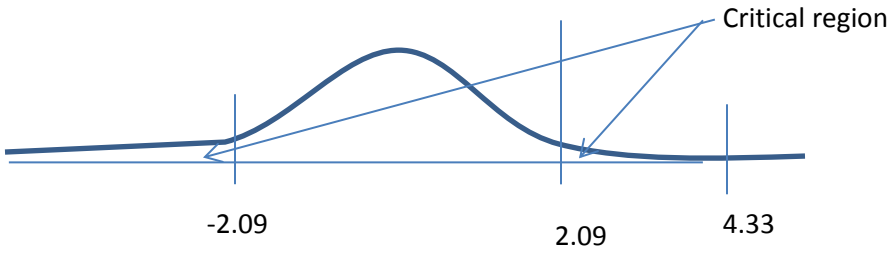
19. The data below gives budgets and gross (both in millions) for 20 PG-13 movies. Determine if there is a correlation between money spent on a movie and money earned.

116	70	75	52	120	125	150	30	225	70
103	66	121	116	101	213	290	100	322	19
80	40	70	300	113	68	72	160	29	132
117	48	228	373	380	118	114	120	120	234

a. State the null and alternative hypothesis to test for a significant correlation.

b. H_0: ρ=0, H_1:ρ≠0

c. Sketch the rejection region for a 5% level of significance and label the critical value.



d. Use the test statistic to determine whether there is a correlation between money spent and money generated on PG-13 movies.

There is a correlation between money spent and money earned on PG-13 movies.

e. State the regression equation that most closely models this data.

$$\hat{y} = 1.13x + 46.56$$

f. Use your equation to predict the likely amount of money to be earned at the box office if $100 million is spent on a new PG-13 movie.

$159.56 million

Stat 1450 Formulas

Unit 2

Class Width = $\frac{\max - \min}{\#classes}$

Class Midpoint = $\frac{lower + upper}{2}$

Unit 3

Range = maximum – minimum

Coefficient of Variation = $\frac{S}{\bar{x}} \cdot 100\%$

Chebyshev's Theorem:

At least 1-$\frac{1}{k^2}$ of the data lie within *k* standard deviations of the mean.

IQR = $Q_3 - Q_1$

z-score = $\frac{x - mean}{stdev} = \frac{x - \mu}{\sigma}$

Unit 4

P(A or B) = P(A) + P(B) – P(A and B)

P(A and B) = P(A) * P(B) if A and B are independent events

P(A and B) = P(A) * P(B|A) if A and B are dependent events

P(complement of E) = 1-P(E)

P(at least 1) = 1-P(none)

P(A|B) = $\frac{P(AandB)}{P(B)}$

Unit 5

Binomial: μ = np σ = $\sqrt{npq}$ $q = 1 - p$

Poisson: $\sigma = \sqrt{\mu}$

Unit 6

$z = \frac{x - \mu}{\sigma}$

$z = \frac{\left(\bar{x} - \mu\right)}{\left(\frac{\sigma}{\sqrt{n}}\right)}$

Unit 7

Proportions:

$CI = \hat{p} \pm z \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}}$ or $\hat{p} - E \le p \le \hat{p} + E$

$n = \frac{z^2 \hat{p}\hat{q}}{E^2} = \hat{p}\hat{q}\left(\frac{z}{E}\right)^2$

Means:

$CI = \bar{x} \pm z \cdot \frac{\sigma}{\sqrt{n}}$ or $\bar{X} - E \le \mu \le \bar{X} + E$

n = $\left(\frac{z * \sigma}{E}\right)^2$

Unit 8

Hypothesis Testing:

Proportions $z = \frac{\left(\hat{p} - p\right)}{\sqrt{pq/n}}$

Means $z = \frac{\left(\bar{x} - \mu\right)}{\left(\frac{\sigma}{\sqrt{n}}\right)}$

Unit 9

Two proportions:

$$\bar{p} = \frac{r_1 + r_2}{n_1 + n_2}$$

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\bar{p}\bar{q}}{n_1} + \frac{\bar{p}\bar{q}}{n_2}}}$$

$$E = z_c \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

$$(\hat{p}_1 - \hat{p}_2) - E < (p_1 - p_2) < (\hat{p}_1 - \hat{p}_2) + E$$

Two means:

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}}}$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}}}$$

$$t = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{n}}}$$

Unit 10

$$E = \frac{Total\ of\ row\ of\ interest \cdot total\ of\ column\ of\ interest}{n}$$

Unit 11

Correlation $t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$

TABLE A-3 t Distribution: Critical t Values

Degrees of Freedom	Area in One Tail: 0.005	0.01	0.025	0.05	0.10
	Area in Two Tails: 0.01	0.02	0.05	0.10	0.20
1	63.657	31.821	12.706	6.314	3.078
2	9.925	6.965	4.303	2.920	1.886
3	5.841	4.541	3.182	2.353	1.638
4	4.604	3.747	2.776	2.132	1.533
5	4.032	3.365	2.571	2.015	1.476
6	3.707	3.143	2.447	1.943	1.440
7	3.499	2.998	2.365	1.895	1.415
8	3.355	2.896	2.306	1.860	1.397
9	3.250	2.821	2.262	1.833	1.383
10	3.169	2.764	2.228	1.812	1.372
11	3.106	2.718	2.201	1.796	1.363
12	3.055	2.681	2.179	1.782	1.356
13	3.012	2.650	2.160	1.771	1.350
14	2.977	2.624	2.145	1.761	1.345
15	2.947	2.602	2.131	1.753	1.341
16	2.921	2.583	2.120	1.746	1.337
17	2.898	2.567	2.110	1.740	1.333
18	2.878	2.552	2.101	1.734	1.330
19	2.861	2.539	2.093	1.729	1.328
20	2.845	2.528	2.086	1.725	1.325
21	2.831	2.518	2.080	1.721	1.323
22	2.819	2.508	2.074	1.717	1.321
23	2.807	2.500	2.069	1.714	1.319
24	2.797	2.492	2.064	1.711	1.318
25	2.787	2.485	2.060	1.708	1.316
26	2.779	2.479	2.056	1.706	1.315
27	2.771	2.473	2.052	1.703	1.314
28	2.763	2.467	2.048	1.701	1.313
29	2.756	2.462	2.045	1.699	1.311
30	2.750	2.457	2.042	1.697	1.310
31	2.744	2.453	2.040	1.696	1.309
32	2.738	2.449	2.037	1.694	1.309
33	2.733	2.445	2.035	1.692	1.308
34	2.728	2.441	2.032	1.691	1.307
35	2.724	2.438	2.030	1.690	1.306
36	2.719	2.434	2.028	1.688	1.306
37	2.715	2.431	2.026	1.687	1.305
38	2.712	2.429	2.024	1.686	1.304
39	2.708	2.426	2.023	1.685	1.304
40	2.704	2.423	2.021	1.684	1.303
45	2.690	2.412	2.014	1.679	1.301
50	2.678	2.403	2.009	1.676	1.299
60	2.660	2.390	2.000	1.671	1.296
70	2.648	2.381	1.994	1.667	1.294
80	2.639	2.374	1.990	1.664	1.292
90	2.632	2.368	1.987	1.662	1.291
100	2.626	2.364	1.984	1.660	1.290
200	2.601	2.345	1.972	1.653	1.286
300	2.592	2.339	1.968	1.650	1.284
400	2.588	2.336	1.966	1.649	1.284
500	2.586	2.334	1.965	1.648	1.283
1000	2.581	2.330	1.962	1.646	1.282
2000	2.578	2.328	1.961	1.646	1.282
Large	2.576	2.326	1.960	1.645	1.282