Unit 1 – Review of BIOSTATS 540
Practice Problems
SOLUTIONS - *Art of Stat*

Dear BIOSTATS 640 Fall 2023,
See syllabus for the answers to the questions about the syllabus!

#1. (Reviews BIOSTATS 540 Unit 1).
The following table lists length of stay in hospital (days) for a sample of 25 patients.

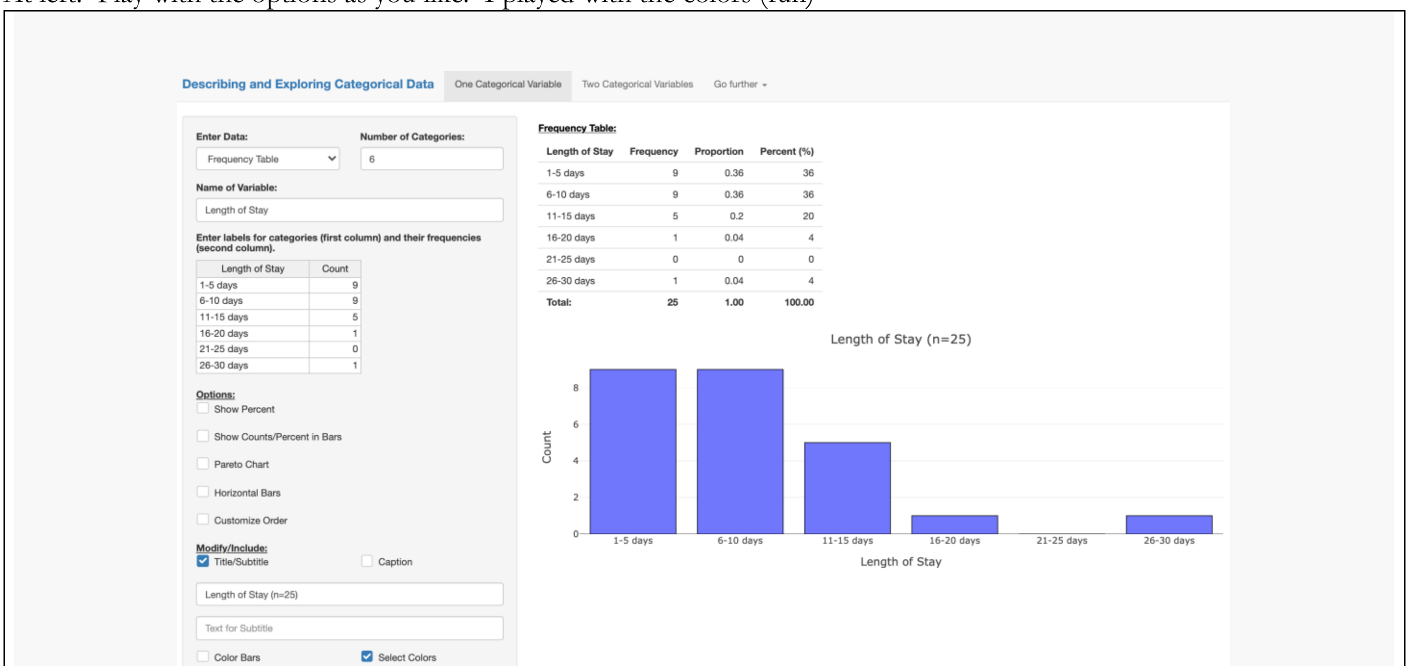
5	10	6	11	5	14	30	11	17	3
9	3	8	8	5	5	7	4	3	7
9	11	11	9	4					

Construct a frequency/relative frequency table for these data using 5-day class intervals. Include columns for the frequency counts and relative frequencies.

Solution:
Step 1 (by hand): Create 5-day intervals and obtain number of observations (frequency) in each 5-day interval

Interval	Observations	Frequency (simple count!)
1-5 days	5 3 5 5 4 5 4 3 3	9
6-10 days	9 9 10 6 8 8 9 7 7	9
11-15 days	11 11 11 14 11	5
16-20 days	17	1
21-25 days		0
25-30 days	30	1

Step 2: Launch www.artofstat.com
Online Web Apps > Explore Categorical Data > at top: One Categorical Variable
At left: enter data: frequency table, number of categories: 6, name of variable: Length of Stay
At left: Play with the options as you like! I played with the colors (fun)



#2. (Reviews BIOSTATS 540 Unit 2).
The following table lists fasting cholesterol levels (mg/dl) for two groups of men.

Group 1:

233	291	312	250	246	197	268	224	239	239
254	276	234	181	248	252	202	218	212	325

Group 2:

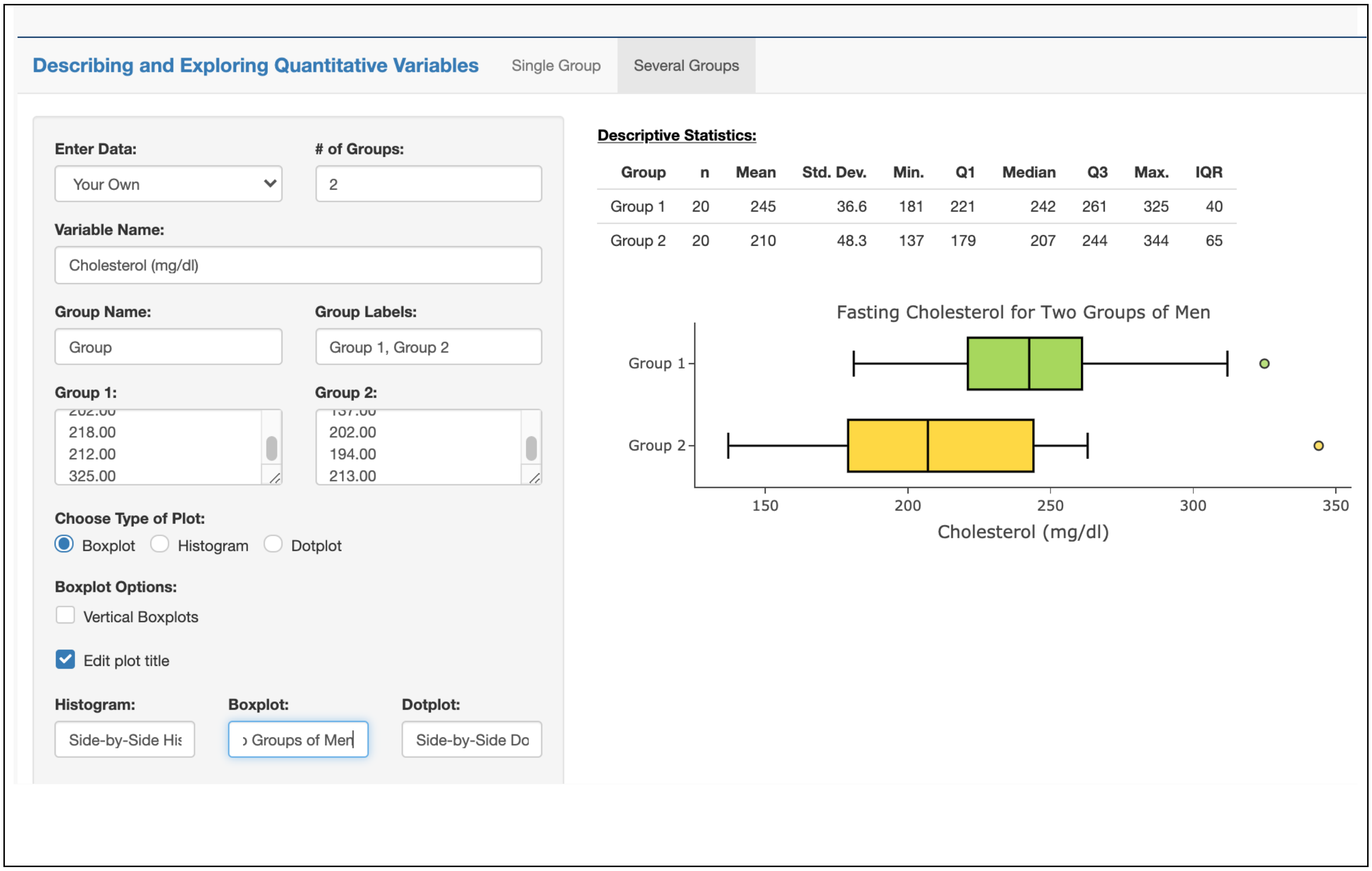
344	185	263	246	224	212	188	250	148	169
226	175	242	252	153	183	137	202	194	213

Preliminary:
I created a little excel file with the data that I could use in an EDIT/COPY/PASTE into Art of Stat. It looks like this (partial showing)

	A	B	C
1	group	ychol	
2	1.00	233.00	
3	1.00	291.00	
4	1.00	312.00	
5	1.00	250.00	
6	1.00	246.00	
7	1.00	197.00	
8	1.00	268.00	
9	1.00	224.00	
10	1.00	239.00	
11	1.00	239.00	
12	1.00	254.00	
13	1.00	276.00	
14	1.00	234.00	
15	1.00	181.00	
16	1.00	248.00	
17	1.00	252.00	
18	1.00	202.00	
19	1.00	218.00	
20	1.00	212.00	
21	1.00	325.00	
22	2.00	344.00	
23	2.00	185.00	

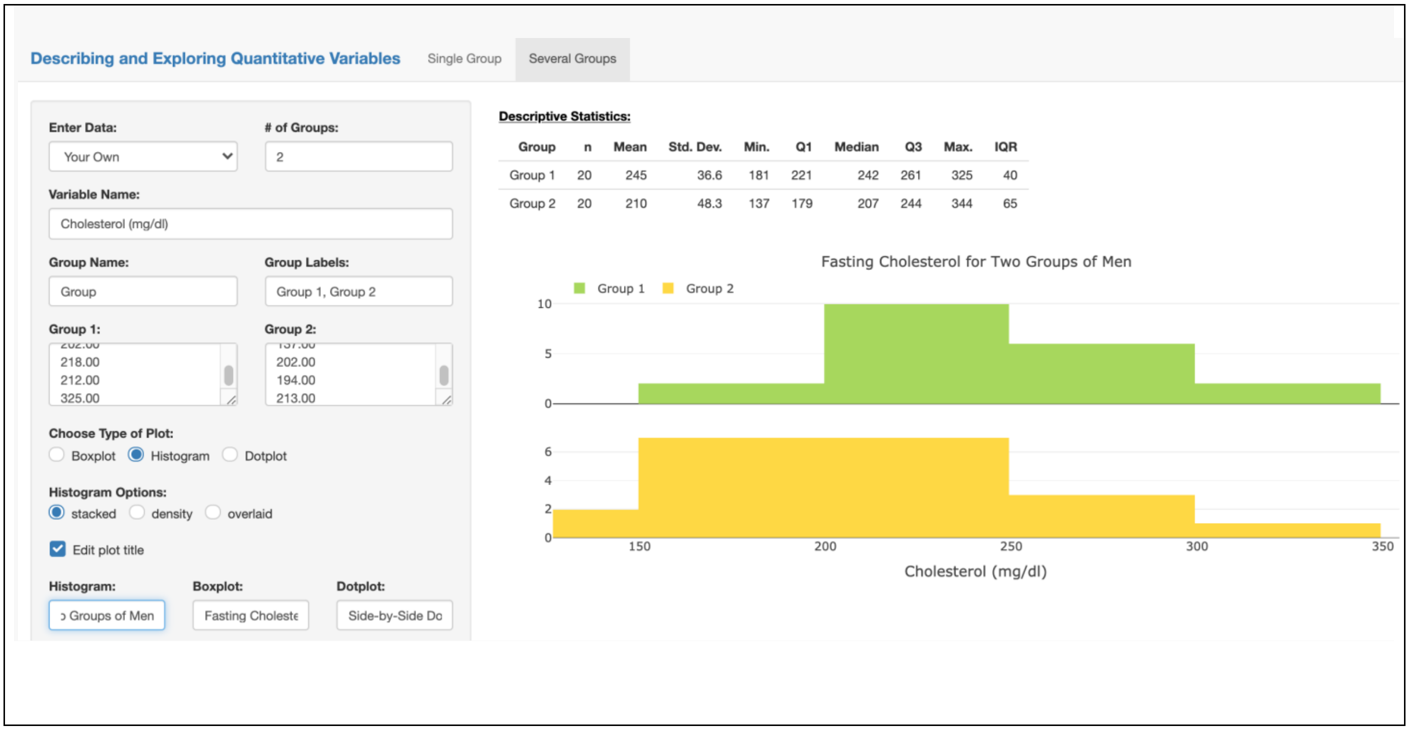
2a. Side-by-side Box Plot:
Launch www.artofstat.com

Online Web Apps > Explore Quantitative Data > at top: Several Groups
At left: enter data: your own, # Groups: 2, Variable Name: Cholesterol (mg/dl)
At left: Group Name: Group, Group Labels: Group 1, Group 2
Paste your data in from excel, *separately* for each group. Tip – Paste ONLY the data, not the variable names in row 1.
Choose Type of Plot: Boxplot



2a. Side-by-side Histogram:
Launch www.artofstat.com

Online Web Apps > Explore Quantitative Data > at top: Several Groups
At left: enter data: your own, # Groups: 2, Variable Name: Cholesterol (mg/dl)
At left: Group Name: Group, Group Labels: Group 1, Group 2
Paste your data in from excel, separately for each group (note: You may have already done this).
Choose Type of Plot: Histogram



In 1-2 sentences, compare the two distributions. What conclusions do you draw?

Comparison of the distributions: Men in Group 1 tend to have higher fasting cholesterol values, as reflected in the upward shift in location of data points. The variation in fasting cholesterol is slightly greater for men in Group 2; this is most easily seen in the side-by-side box plot, which shows a larger interquartile range (the size of the box itself) and a more distant outlier.

#3. (Reviews BIOSTATS 540 Unit 6 – Bernoulli and Binomial).

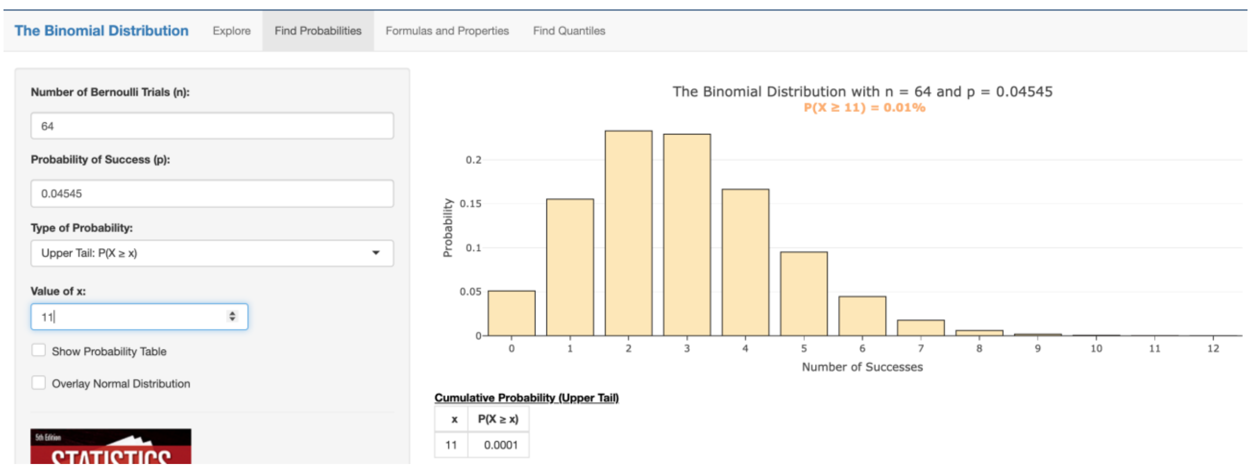
Consider the following setting.
Seventy-nine firefighters were exposed to burning polyvinyl chloride (PVC) in a warehouse fire in Plainfield, New Jersey on March 20, 1985. A study was conducted in an attempt to determine whether or not there were short- and long-term respiratory effects of the PVC. At the long term follow-up visit at 22 months after the exposure, 64 firefighters who had

been exposed during the fire and 22 firefighters who where not exposed reported on the presence of various respiratory conditions. Eleven of the PVC exposed firefighters had moderate to severe shortness of breath compared to only 1 of the non-exposed firefighters.

Calculate the probability of finding 11 or more of the 64 exposed firefighters reporting moderate to severe shortness of breath if the rate of moderate to severe shortness of breath is 1 case per 22 persons. **Show your work**.

Answer: .0001 or .01%
X ~ Binomial (n=64, p = 1/22 = 0.04545)
Want Pr [X ≥ 11]

Launch www.artofstat.com
Online Web Apps > Binomial Distribution > at top: Find Probabilities
At left: Number of Trials, n: 64, Probability of Success, p: 0.04545
At left: Type of Probability: Upper Tail, Value of x: 11



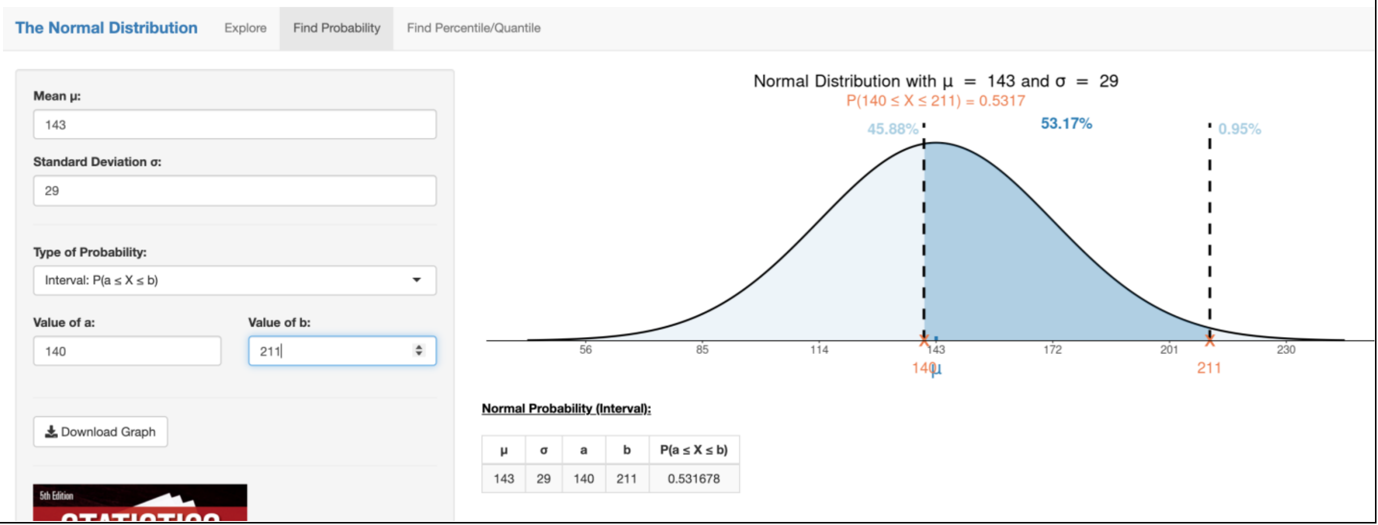
#4. (Reviews BIOSTATS 540 Unit 7- Normal Distribution).

The Air Force uses ACES-II ejection seats that are designed for men who weigh between 140 lb and 211 lb. Suppose it is known that women's weights are distributed Normal with mean 143 lb and standard deviation 29 lb.

4a. What proportion of women have weights that are ***outside*** the ACES-II ejection seat acceptable range?

Answer: .4683 or 48.63%
X ~ Normal (mean = 143, sd=29)
Want Pr [X ≤ 140] + Pr [X ≥ 211]
Art of Stat doesn't offer us this option. The best we can do is Pr [140 ≤ X ≤ 211]
However!
Pr [X ≤ 140] + Pr [X ≥ 211] = 1 - Pr [140 ≤ X ≤ 211]
= 1 - .5317
= .4863

Launch www.artofstat.com
Online Web Apps > Normal Distribution > at top: Find Probability
At left: Mean: 143, Standard Deviation: 29
At left: Type of Probability: Interval, Value of a: 140, Value of b: 211

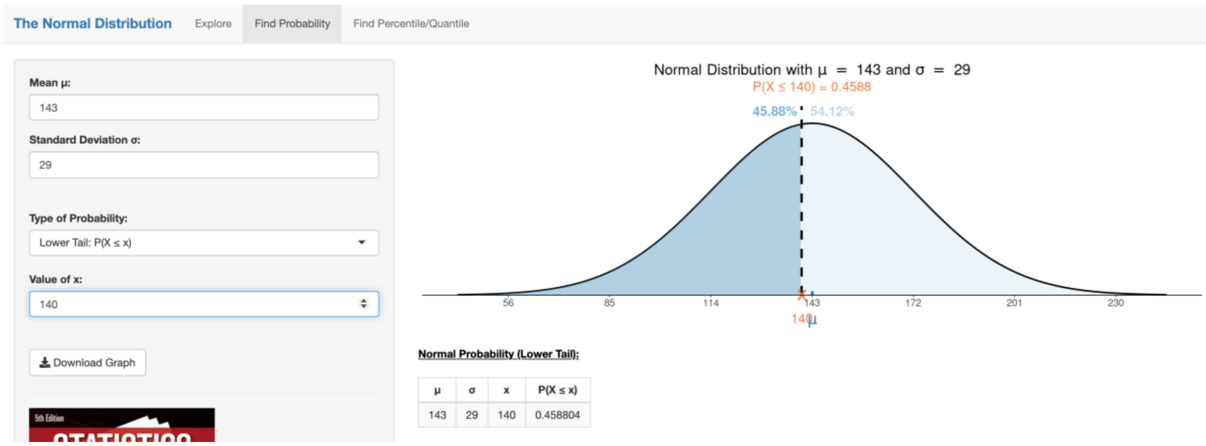


4b. In a sample of 1000 women, how many are expected to have weights below the 140 lb threshold?

Answer: 459
X ~ Normal (mean = 143, sd=29)
Expected # women w weight below threshold = [total # women] * Pr [that womean is below threshold]
= [1000 women] * Pr [X ≤ 140]
= [1000] * [.4588]
= 458.8 or 459 women

Launch www.artofstat.com

Online Web Apps > Normal Distribution > at top: Find Probability
At left: Mean: 143, Standard Deviaion: 29
At left: Type of Probability: Lower Tail, Value of x: 140



#5. (Reviews BIOSTATS 540 Units 8 and 9).

Consider the setting of a single sample of n=16 data values that are a random sample from a normal distribution. Suppose it is of interest to perform a type I error $\alpha = 0.01$ statistical hypothesis test of $H_O: \mu \geq 100$ versus $H_A: \mu < 100$, $\alpha = 0.01$. Suppose further that σ is unknown.

5a. State the appropriate test statistic

Answer: Student t test with degrees of freedom = 15.

Solution:
This is a one-sample setting of normally distributed data where the population variance is not known and interest is in the mean. Because the sample size is 16, the degrees of freedom is (16-1)=15.

5b. Determine the critical region for values of the sample mean $\bar{X}$.

Answer: $\bar{X} \leq (S/4)(-2.602) + 100$

Solution:

Given: $n=16, \ \alpha=0.01, \text{one sided(left)}, \ \mu_O=100 \ \rightarrow$

(1) Solution for $\hat{SE} = \dfrac{S}{\sqrt{16}} = \dfrac{S}{4}$

(2) Solution for $t_{CRITICAL} = t_{.01;df=15} = -2.602$

(3) Solution for $\bar{X}_{CRITICAL}$ is obtained by its solution in the following expression:

$$t_{OBSERVED} \leq t_{CRITICAL} \rightarrow$$

$$\frac{\bar{X} - \mu_{NULL}}{\hat{SE}} \leq -2.602 \rightarrow$$

$$\bar{X} - \mu_{NULL} \leq (\hat{SE})(-2.602) \rightarrow$$

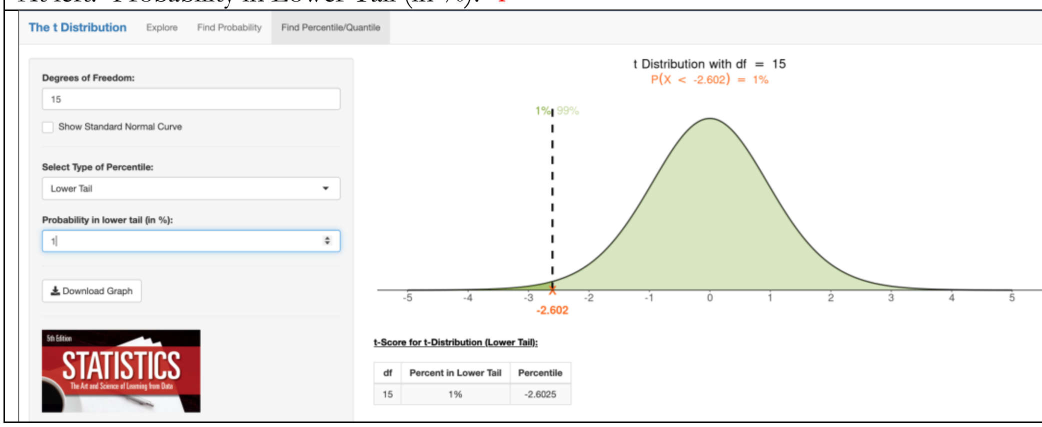
$$\bar{X} \leq (\hat{SE})(-2.602) + \mu_{NULL} \rightarrow$$

$$\bar{X} \leq (S/4)(-2.602) + \mu_{NULL} \rightarrow$$

$$\bar{X} \leq (S/4)(-2.602) + 100$$

Solution for $t_{CRITICAL}$ using Art of Stat

Launch www.artofstat.com
Online Web Apps > t Distribution > at top: Find Percentile/Quantile
At left: Degrees of Freedom: 15, Type of Percentile: Lower Tail
At left: Probability in Lower Tail (in %): 1



#6. (Reviews BIOSTATS 540 Unit 9).

An investigator is interested in the mean cholesterol level μ of patients with myocardial infarction. S/he drew a simple random sample of n=50 patients and from these data constructed a 95% confidence interval for the mean μ. In these calculations, it was assumed that the data are a simple random sample from a normal distribution with known variance. The resulting width of the confidence interval was 10 mg/dl.

How large a sample size would have been required if the investigator wished to obtain a confidence interval width equal to 5 mg/dl?

Answer: 200

Solution:

(Step 1) Solution for value of confidence coefficient:
95% CI and σ known → Desired confidence coefficient is 97.5^{th} percentile of Normal(0,1) = 1.96

(Step 2) Expression for CI width:

$$\text{width} = (\text{upper limit}) - (\text{lower limit})$$

$$= \left(\bar{X} + \frac{1.96\sigma}{\sqrt{n}}\right) - \left(\bar{X} - \frac{1.96\sigma}{\sqrt{n}}\right)$$

$$= \frac{1.96\sigma}{\sqrt{n}} - \left(-\frac{1.96\sigma}{\sqrt{n}}\right)$$

$$= \frac{(2)(1.96)\sigma}{\sqrt{n}}$$

(Step 3) Using known width=10 and known n=50, obtain σ = 18.0384

$$10 = \frac{(2)(1.96)\sigma}{\sqrt{n}} \rightarrow$$

$$\frac{(10)(\sqrt{n})}{(2)(1.96)} = \sigma \rightarrow$$

$$\frac{(10)(\sqrt{50})}{(2)(1.96)} = \sigma \rightarrow$$

$$\sigma = 18.0384$$

(Step 4) Using known width=5 and σ = 18.0384 known, obtain n=200

$$5 = \frac{(2)(1.96)\sigma}{\sqrt{n}} \rightarrow$$

$$\sqrt{n} = \frac{(2)(1.96)\sigma}{5} \rightarrow$$

$$\sqrt{n} = \frac{(2)(1.96)(18.0384)}{5} \rightarrow$$

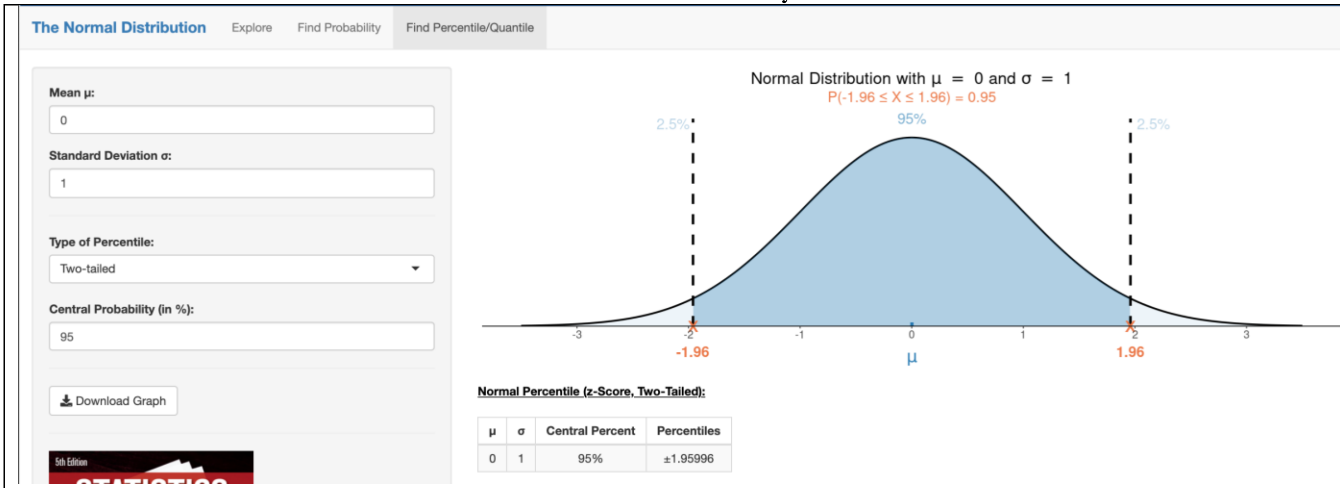
$$\sqrt{n} = 14.1421 \rightarrow$$

$$n = 199.99$$

Or 200, by rounding up.

Solution for confidence coefficient in Normal(0,1) using Art of Stat

Launch www.artofstat.com
Online Web Apps > Normal Distribution > at top: Find Percentile/Quantile
At left: Mean: 0, Standard Deviation: 1
At left: Type of Percentile: Two Tailed
At Left Central Probability (in %): 95



#7. (Reviews BIOSTATS 540 Units 9 and 10).

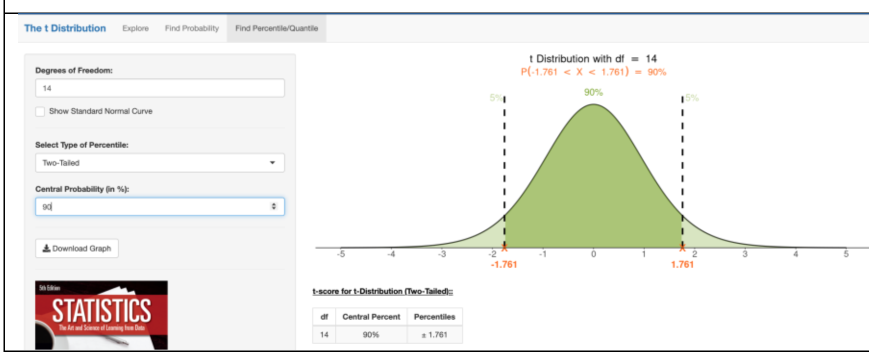
In (a) – (d) below, you may assume that the data are a simple random sample (or samples) from a normal distribution (or distributions). Each setting is a different setting of confidence interval estimation. In each, state the values of the confidence coefficients *(recall – these will be the values of specific percentiles from the appropriate probability distribution)*.

7a.

For a single sample size of n=15 and the estimation of the population mean μ when the variance is unknown using a 90% confidence interval, what are the values of the confidence coefficients?

Answer: -1.761 and + 1.761

Launch www.artofstat.com
Online Web Apps > t Distribution > at top: Find Percentile/Quantile
At left: Degrees of Freedom: 14, Type of Percentile: Two Tailed
At left: Central Probability (in %): 90



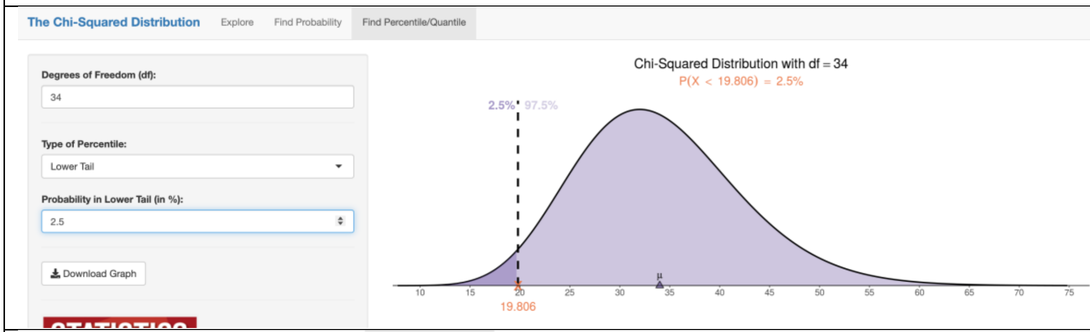
7b.

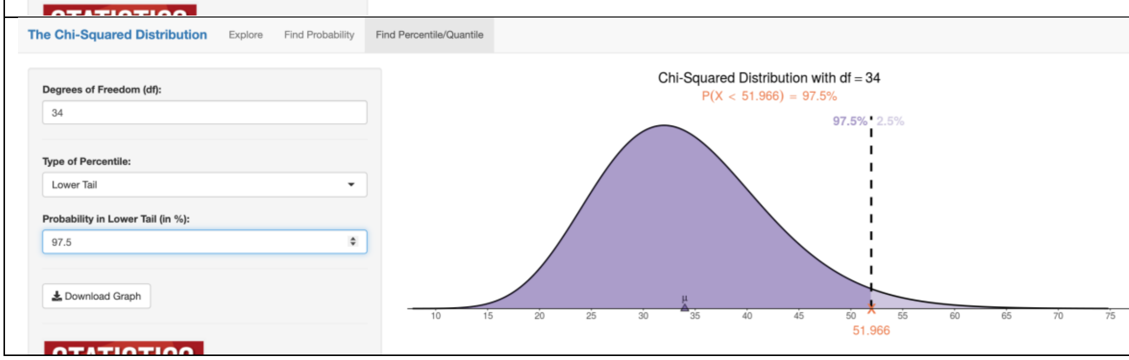
For a single sample size n=35 and the estimation of a variance parameter σ^2 using a 95% confidence interval, what are the values of the confidence coefficients?

Answer: 19.81 and 51.97

Launch www.artofstat.com

Online Web Apps > chi square Distribution > at top: Find Percentile/Quantile
At left: Degrees of Freedom: 34, Type of Percentile: Lower Tail (you have no choice here)
So you need to do TWO calculations:
1st: At left: Probability in Lower Tail (in %): 2.5
2nd : At left: Probability in Lower Tail (in %): 97.5

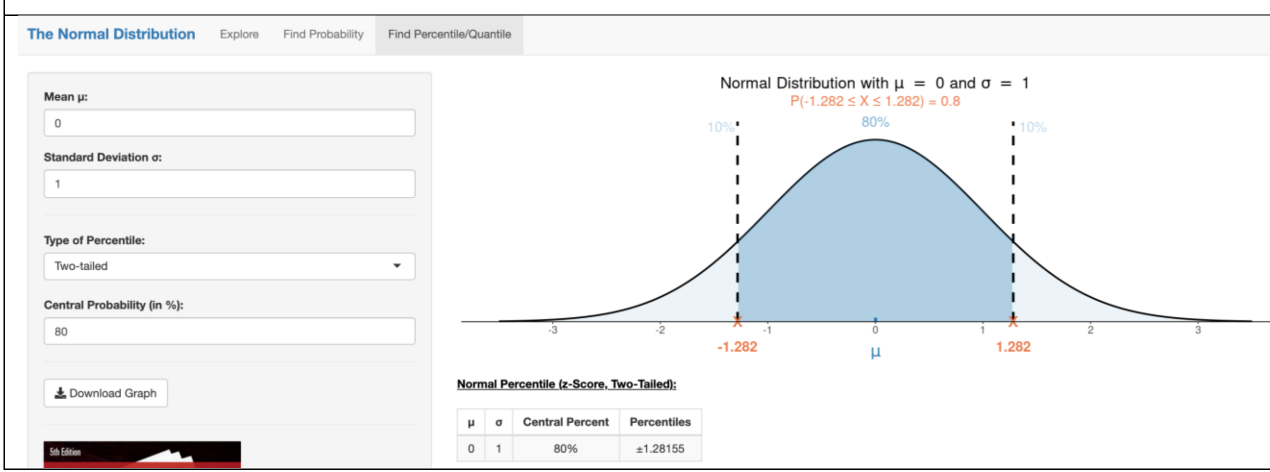




7c.
For a single sample size of n=25 and the estimation of the population mean μ when the variance is known using a 80% confidence interval, what are the values of the confidence coefficients?

Answer: -1.282 and + 1.282

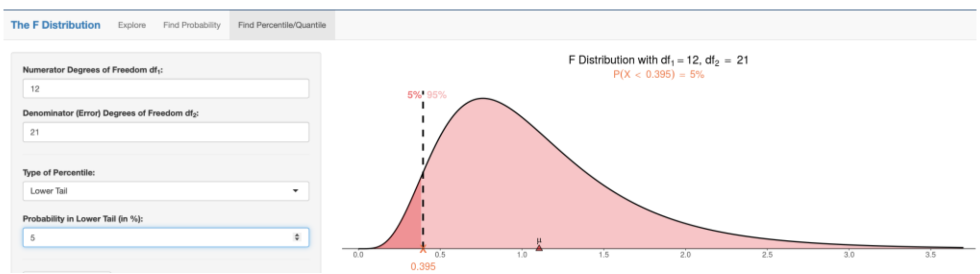
Launch www.artofstat.com
Online Web Apps > Normal Distribution > at top: Find Percentile/Quantile
At left: mean: 0, standard deviation: 1
At left: Central Probability (in %): 80

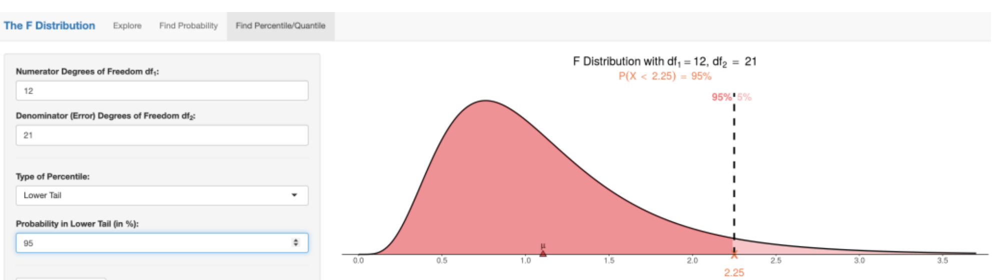


7d.
For the setting of two independent samples, one with sample size $n_1 = 13$ and the other with sample size $n_2 = 22$, it is of interest to construct a 90% confidence interval estimate of the ratio of the two population variances, $[\ \sigma_1^2/\sigma_2^2\ ]$. What are the values of the confidence coefficients?

Answer: 0.39 and 2.25

Launch www.artofstat.com
Online Web Apps > F Distribution > at top: Find Percentile/Quantile
At left: Numerator Degrees of Freedom: 12, Denominator Degrees of Freedom: 21
Type of Percentile: Lower Tail (like the chi square, here, you have no choice)
So you need to do TWO calculations:
1st: At left: Probability in Lower Tail (in %): 5
2nd : At left: Probability in Lower Tail (in %): 95





#8. **(Reviews BIOSTATS 540 Unit 10).**

A study was investigated of length of hospital stay associated with seat belt use among children hospitalized following motor vehicle crashes. The following are the observed sample mean and sample standard deviations for two groups of children: 290 children who were ***not*** wearing a seat belt at the time of the accident plus 123 children who ***were*** wearing a seat belt at the time of the accident.

Group	Sample size, n	Sample mean	Sample standard deviation
Seat belt = no	$n_{NO} = 290$	$\bar{X}_{NO} = 1.39$ days	$S_{NO} = 3.06$ days
Seat belt = yes	$n_{YES} = 123$	$\bar{X}_{YES} = 0.83$ days	$S_{YES} = 2.77$ days

You may assume normality. You may also assume that the unknown variances are equal. Construct a 95% confidence interval estimate of the difference between the two population means. In developing your answer, you may assume that the population variances are unknown but EQUAL.

8a.
What is the value of the point estimate?

Answer: 0.56

Solution:

$$\text{Point estimate of } [\mu_1 - \mu_2] = [\bar{X}_1 - \bar{X}_2] = [1.39 - 0.83] = 0.56$$

8b.
What is the value of the estimated standard error of the point estimate?

Answer: 0.32

Solution:

(1) Preliminary: Obtain S^2_{pool}

$$S^2_{pool} = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{(n_1-1)+(n_2-1)} = \frac{(289)(9.3636)+(122)(7.6729)}{(289)+(122)} = 8.8617377$$

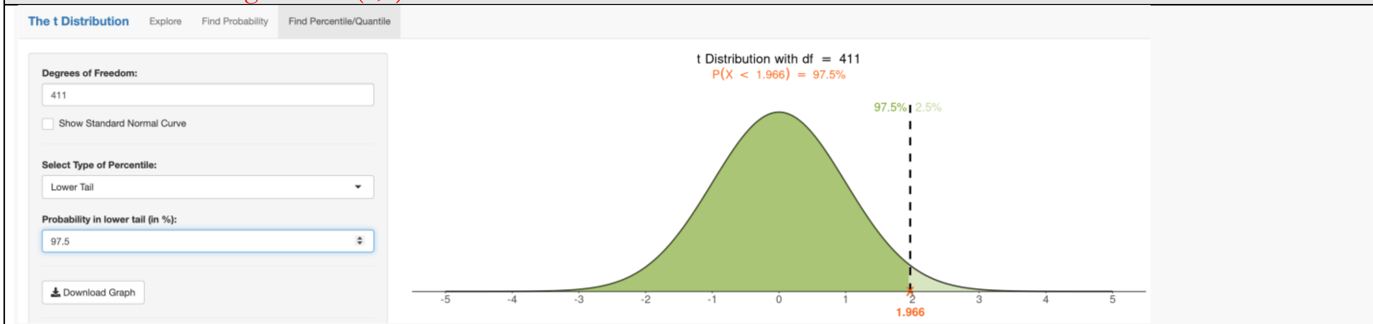
(2) Solution for SE of point estimate, $\hat{SE}[\bar{X}_1 - \bar{X}_2]$

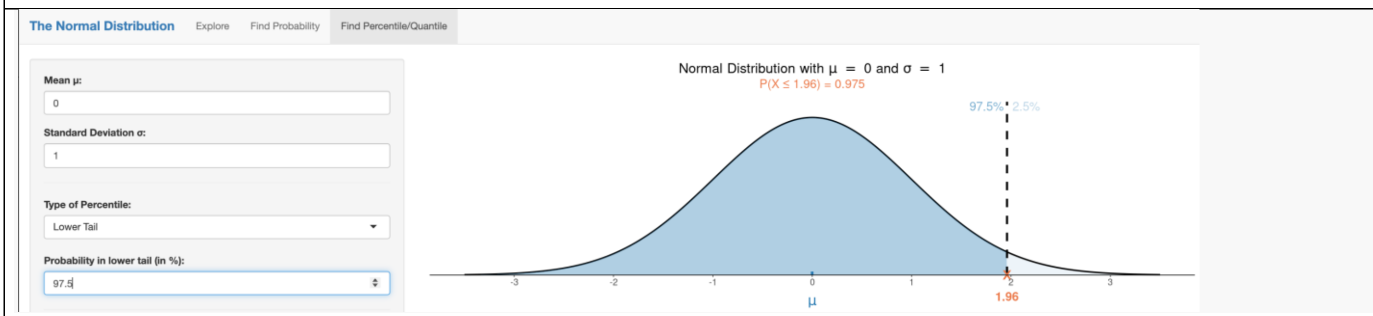
$$\hat{SE}[\bar{X}_1 - \bar{X}_2] = \sqrt{\frac{S^2_{pool}}{n_1} + \frac{S^2_{pool}}{n_2}} = \sqrt{\frac{8.8617377}{290} + \frac{8.8617377}{123}} = 0.320319$$

(3) Solution for degrees of freedom, df:
$df = (n_1 - 1) + (n_2 - 1) = (290\text{-}1) + (123\text{-}1) = 411$

8c.
What is the value of the confidence coefficient?

Answer: 1.966 using Student-t.
1.96 using Normal(0,1). Close!





8d.
What are values of the lower and upper limits of the confidence interval?

Answer: [-0.07, + 1.19]

Solution:

$$CI = [\bar{X}_1 - \bar{X}_2] \pm \left(t_{.975;df=411}\right)\hat{SE}(\bar{X}_1 - \bar{X}_2)$$

$$= 0.56 \pm (1.966)(0.32)$$

$$= [\ -0.06912\ ,\ +1.18912\ ]$$

8e.
Write a clear interpretation of the confidence interval.

With 95% confidence, from these data, it is estimated that the difference in average length of stay (non-seat belt wearers minus seat belt wearers) is between -0.07 days and +1.19 days. Since this interval includes 0, these data do not provide statistically significant evidence that the length of hospital stay for children in motor vehicle crashes who were not wearing seat belts is different than the length of hospital stay for children in motor vehicle crashes who were wearing seat belts.

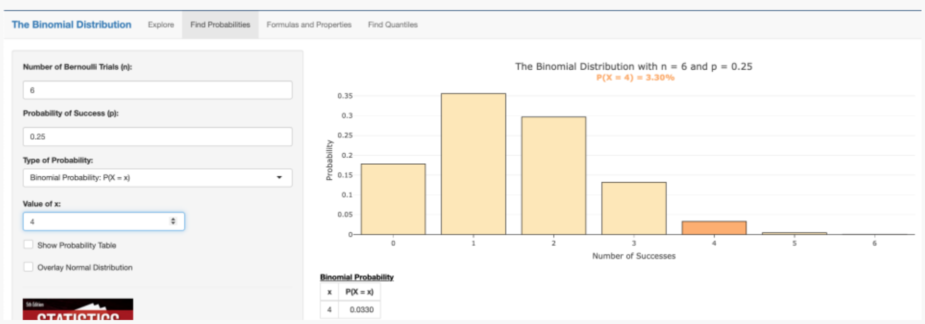
#9. (Reviews BIOSTATS 540 Unit 6).
A test consists of multiple-choice questions, each having four possible answers, one of which is correct. What is the probability of getting exactly four correct answers when six guesses are made?

Answer: .03

This is a binomial probability calculation.
N=6 π=.25 Want Pr[X = 4]

$$\Pr[X=4] = \binom{n}{x} \pi^x (1-\pi)^{n-x} = \binom{6}{4} .25^4 (.75)^2 = .032959$$

Launch www.artofstat.com
Online Web Apps > Binomial Distribution > at top: Find Probabilities
At left: Number of Trials, n: 6, Probability of Success, p: 0.24
At left: Type of Probability: Binomial Probability: P(X=x), Value of x: 4



#10. (Reviews BIOSTATS 540 Unit 6).
After being rejected for employment, woman "A" learns that company "X" has hired only 2 women among the last 20 new employees. She also learns that the pool of applicants is very large, with an approximately equal number of qualified men and women. Help her address the charge of gender discrimination by finding the probability of getting 2 or fewer women when 20 people hired under the assumption that there is no discrimination based on gender. Does the resulting probability really support such a charge?

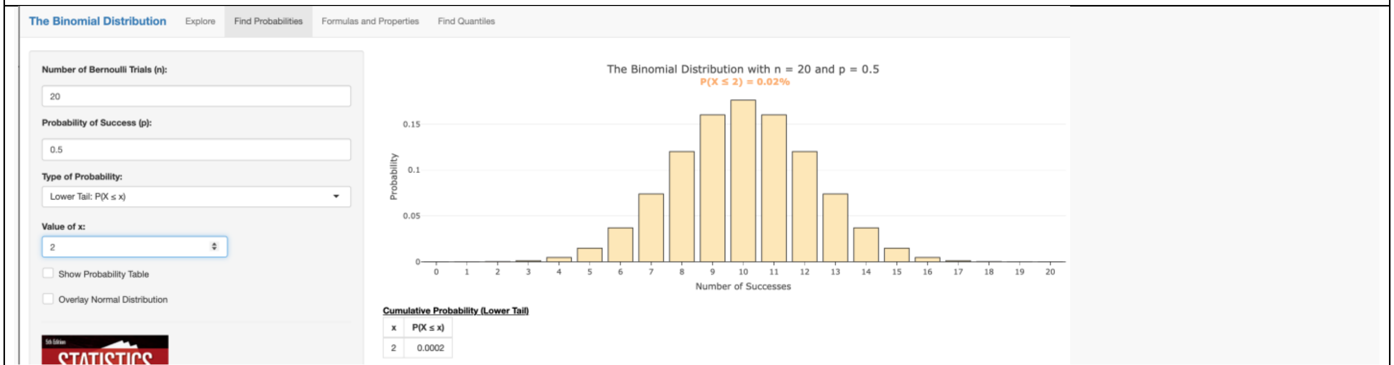
Answer: .0002
This is a very small probability. As such, it would support a charge of gender discrimination but only under the circumstances where, for each position filled, there were an equal number of men and women applicants.

Solution:
This is also a binomial probability calculation.
Here, n=20 π=.50 Want Pr[$X \le 2$]

$$\Pr[X \le 2] = \sum_{x=0}^{2} \binom{n}{x} \pi^x (1-\pi)^{n-x} = \sum_{x=0}^{2} \binom{20}{x} .50^x (.50)^{20-x} = .00020122$$

Launch www.artofstat.com

Online Web Apps > Binomial Distribution > at top: Find Probabilities
At left: Number of Trials, n: 20, Probability of Success, p: 0.50
At left: Type of Probability: Binomial Probability: P(X ≤ x), Value of x: 2



#11. **(Reviews BIOSTATS 540 Unit 7).**
Suppose the length of newborn infants is distributed normal with mean 52.5 cm and standard deviation 4.5 cm. What is the probability that the mean of a sample of size 15 is greater than 56 cm?

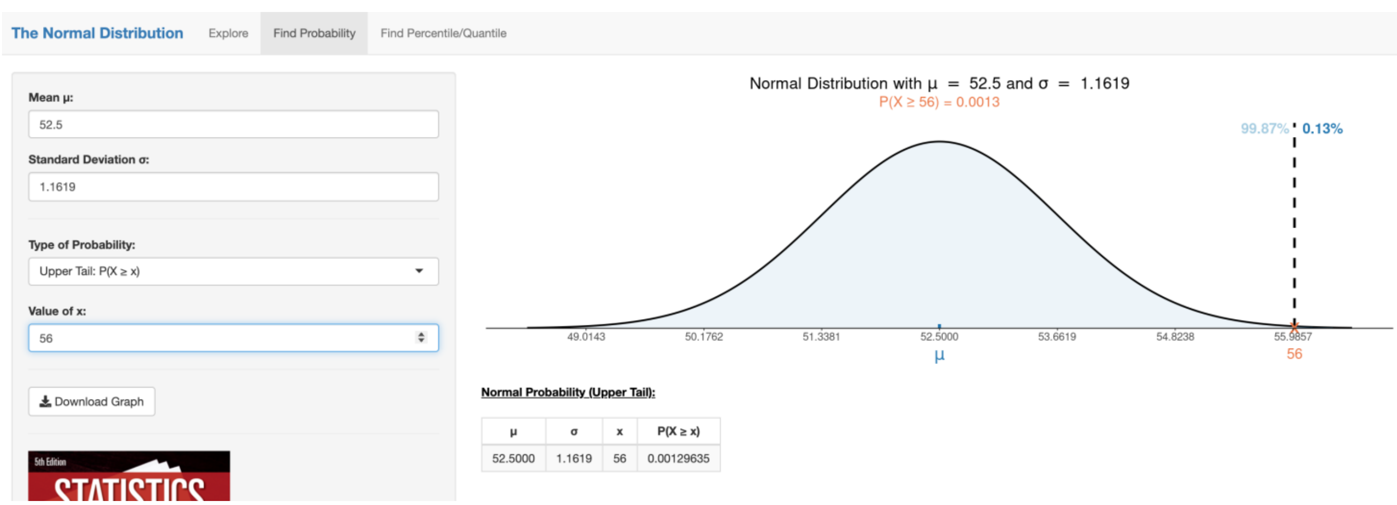
Answer: .0013

Solution:
This is a normal distribution probability calculation.

$$\bar{X}_{n=15} \text{ is distributed Normal with } \mu_{\bar{X}}=52.5 \text{ and } se(\bar{X}_{n=15}) = \frac{\sigma}{\sqrt{15}} = \frac{4.5}{\sqrt{15}} = 1.1619$$

$$\text{Want } \Pr[\bar{X}_{n=15} \geq 56] = \Pr\left[\frac{\bar{X}_{n=15} - \mu_{\bar{X}}}{se(\bar{X}_{n=15})} \geq \frac{56 - 52.5}{1.1619}\right] = \Pr[\text{Z-score} \geq 3.0123] = .001296$$

Launch www.artofstat.com
Online Web Apps > Normal Distribution > at top: Find Probability
At left: mean: 52.5, standard deviation: 1.1619
At left: Type of Probability: Upper Tail P (X ≥ x)



#12. (Reviews BIOSTATS 540 Unit: 7).
Suppose that 25 year old males have a remaining life expectancy of an additional 55 years with a standard deviation of 6 years. Suppose further that this distribution of additional years life is normal. What proportion of 25 year-old males will live past 65 years of age?

Answer: 99.4%

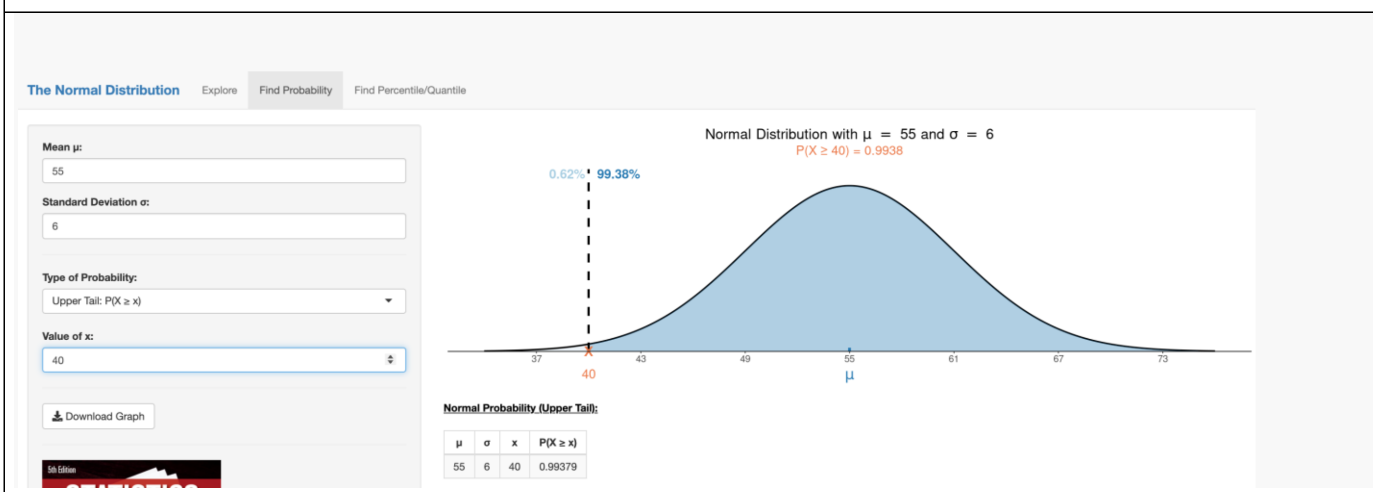
This is also a normal distribution probability calculation.

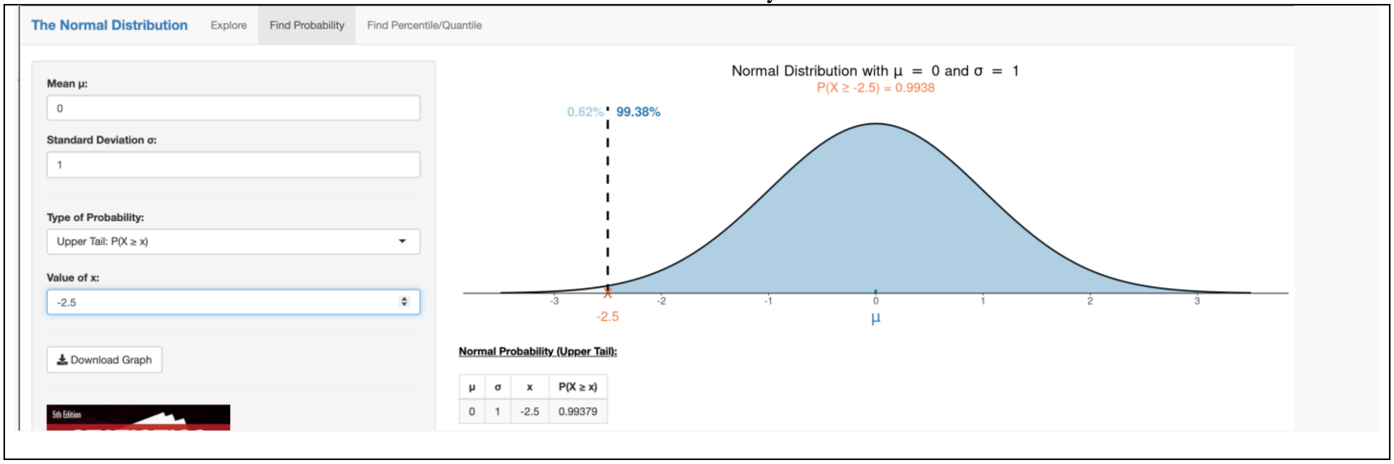
X is distributed Normal with μ = 55 and σ = 6

"*Living past 65 years of age*" corresponds to a remaining life expectancy of an *additional 40+ years*.
Thus, want:

$$\Pr[X \geq 40] = \Pr\left[\frac{X - m}{s} \geq \frac{40 - 55}{6}\right] = \Pr\left[\text{Z-score} \geq -2.5\right] = .9938$$

Launch www.artofstat.com
Online Web Apps > Normal Distribution > at top: Find Probability
At left: mean: 55, standard deviation: 6
At left: Type of Probability: Upper Tail P (X ≥ x)




The Normal Distribution
Explore
Find Probability
Find Percentile/Quantile
Mean μ:
0
Standard Deviation σ:
1
Type of Probability:
Upper Tail: P(X ≥ x)
Value of x:
-2.5
Download Graph
Normal Distribution with μ = 0 and σ = 1
P(X ≥ -2.5) = 0.9938
0.62%
99.38%
-3
-2
-1
0
1
2
3
-2.5
μ
Normal Probability (Upper Tail):
μ σ x P(X ≥ x)
0 1 -2.5 0.99379