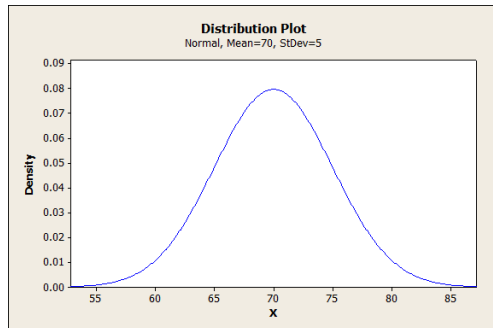


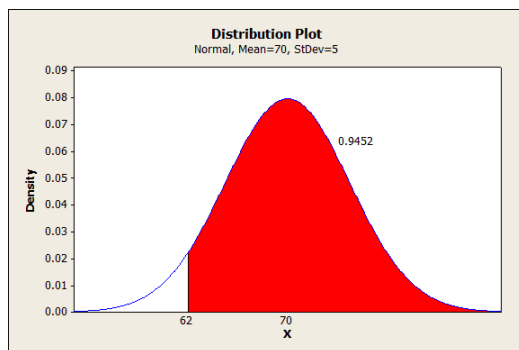
Solution to STAT 350 Exam 1 Review Problems Spring 2015

1. The weight of the eggs produced by a certain breed of hen is Normally distributed with mean 70 grams(g) and standard deviation 5g.

a) Sketch the graph of the normal distribution that corresponds to the weight of the eggs.



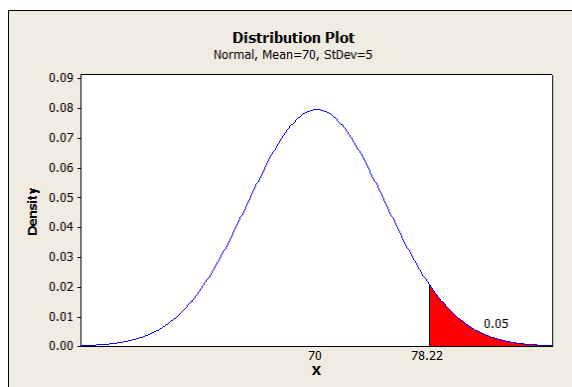
b) What proportion of the eggs are more than 62 grams?



$$z = \frac{62-70}{5} = -1.6 ,$$

$$P(Z > -1.6) = 1 - P(Z < -1.6) = 1 - 0.0548 = 0.9452$$

c) How much does an egg need to weight in order for it to be the largest 5%?



The z-score corresponding to the top 5% is $z = 1.645$

$$x = 70 + 1.645 \cdot 5 = 78.225$$

An egg need to weight 78.225 grams in order for it to be the largest 5%.

Solution to STAT 350 Exam 1 Review Problems Spring 2015

2. The editor of a statistics text would like to plan for the next edition. A key variable is the number of pages that will be in the final version. Text files are prepared by the authors using a word processor called LaTeX, and separate files contain figures and tables. For the previous edition of the text, the number of pages in the LaTeX files can easily be determined, as well as the number of pages in the final version of the text. Here are the data:

	Chapter												
	1	2	3	4	5	6	7	8	9	10	11	12	13
LaTeX pages	77	73	59	80	45	66	81	45	47	43	31	46	26
Text pages	99	89	61	82	47	68	87	45	53	50	36	52	19

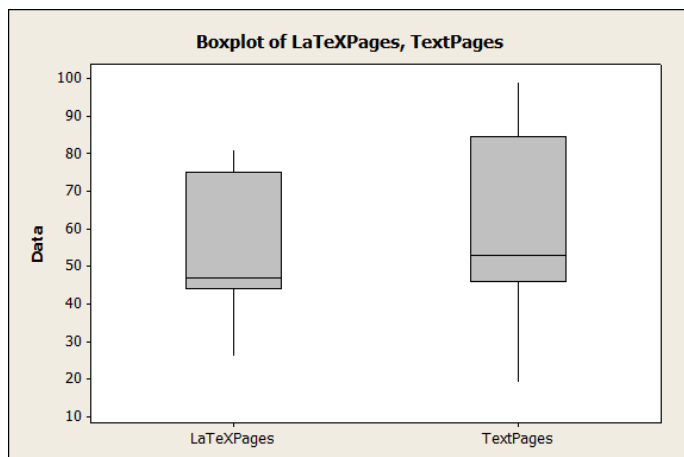
a) Find the mean and the standard deviation for LaTeX pages and Text pages.

Variable	Mean	StDev
LaTeXPages	55.31	18.60
TextPages	60.62	23.27

b) Find the five number summaries for LaTeX pages and Text pages.

Variable	Minimum	Q1	Median	Q3	Maximum
LaTeXPages	26.00	44.00	47.00	75.00	81.00
TextPages	19.00	46.00	53.00	84.50	99.00

c) Draw a side-by-side boxplot for LaTeX pages and Text pages.



Solution to STAT 350 Exam 1 Review Problems Spring 2015

3. You see a news report of an experiment that claims to show that a meditation technique increased job satisfaction of employees. The experimenter interviewed the employees and assessed their levels of job satisfaction. The subjects then learned how to meditate and did so regularly for a month. The experimenter re-interviewed them at the end of the month and assessed their job satisfaction levels again.

(a) There was no control group in this experiment. Why is this a blunder? What lurking variables might be confounded with the effect of meditation?

Without a control group, there is no reason to believe that the difference seen is due to the variable being studied. Other lurking variables might also be responsible for the change. For example, changes in job conditions, time away from work to meditate, and the act of doing something for yourself may also play a role in the levels of job satisfaction.

(b) The experimenter who diagnosed the effect of the treatment knew that the subjects had been meditating. Explain how this knowledge could bias the experimental conclusions.

The experimenter's knowledge may make him or her expect to see results from the meditation, and therefore they may see results that aren't really present.

(c) Briefly discuss a proper experimental design, with controls and blind diagnosis, to assess the effect of meditation on job satisfaction.

In a proper experiment, a control group would be set up. This group could also have consideration from the company in terms of their health, such as a time set aside for daily exercise or personal time. A test or survey should be standardized to determine job satisfactions, and this should be given both before and after the experiment. The researcher giving the survey should not know which subjects were in which groups.

4. According to a book review published in the Wall Street Journal on Sep 26, 2012, 1% of 40-year-old women have breast cancer. 80% of these women who actually have breast cancer will have a positive mammogram. 10% of 40-year-old women who do not have breast cancer will also have a positive mammogram. If a 40-year-old woman has positive mammogram, what is the probability that she has breast cancer?

Solution: We want to find $P(\text{Cancer} \mid \text{Test} +)$. [a tree diagram will be helpful]

We know: $P(\text{Cancer}) = 0.01$; $\Rightarrow P(\text{No Cancer}) = 0.99$

$$P(\text{Test} + \mid \text{Cancer}) = 0.8$$

$$P(\text{Test} + \mid \text{No Cancer}) = 0.1$$

$$\begin{aligned} \text{So, } P(\text{Test} +) &= P(\text{Test} + \text{ and Cancer}) + P(\text{Test} + \text{ and NoCancer}) \\ &= 0.01 * 0.8 + 0.99 * 0.1 = 0.107 \end{aligned}$$

$$P(\text{Cancer} \mid \text{Test} +) = P(\text{Cancer and Test} +) / P(\text{Cancer}) = 0.01 * 0.8 / 0.107 = 0.075$$

Solution to STAT 350 Exam 1 Review Problems Spring 2015

5. If children are given more choices within a class of products, will they tend to prefer that product to a competing product that offers fewer choices? Marketers want to know. An experiment prepared three sets of beverages. Set 1 contained two milk drinks and two fruit drinks. Set 2 had two fruit drinks and four milk drinks. Set 3 contained four fruit drinks but only two milk drinks. The researchers divided 210 children aged 4 to 12 years into three groups at random. They offered each group one of the sets. As each child chose a beverage to drink from the set presented, the researchers noted whether the choice was a milk drink or a fruit drink.

(a) What are the experimental subjects?

The 210 children in the study

(b) What is the factor and what are its levels? What is the response variable?

- **The factor is the set of choices that are presented to each subject.**
- **The levels correspond to the three sets of choices.**
- **The response variable is whether they chose a milk drink or a fruit drink.**

(c) Use a diagram to outline a completely randomized design for the study

TBD

Please refer to Figure 3.6, Figure 3.7, and Figure 3.9 for a completely randomized design diagram.

(d) Use a diagram to outline a blocked design for the study

TBD

Solution to STAT 350 Exam 1 Review Problems Spring 2015

6. At a self-service gas station, 40% of the customers pump regular gas, 35% pump mid-grade, and 25% pump premium gas. Of those who pump regular, 30% pay at least \$30. Of those who pump midgrade, 50% pay at least \$30. And of those who pump premium, 60% pay at least \$30.

a) What is the probability that the next customer pays at least \$30? (Draw a tree diagram to organize the information given.)

$$P(\text{Pay at least \$30}) = 0.445$$

b) What percent of customers who pay at least \$30 pump premium gas? (Write this as a conditional probability and use your result from the previous exercise.)

$$P(\text{premium} \mid \text{pay at least \$30}) = 0.337$$

7. The credit manager for a local department store discovers that 88% of all the store's credit card holders who defaulted on their payments were late (by a week or more) with two or more of their monthly payments before failing to pay entirely (defaulting). This prompts the manager to suggest that future credit be denied to any customer who is late with two monthly payments. Further study shows that 3% of all credit customers default on their payments and 40% of those who have not defaulted have had at least two late monthly payments in the past.

(a) What is the probability that a customer who has two or more late payments will default?

$$P(\text{Default} \mid \text{Late}) = 0.0637$$

(b) Under the credit manager's policy, in a group of 100 customers who have their future credit denied, how many would we expect *not* to default on their payments?

Between 93 to 94

(c) Does the credit manager's policy seem reasonable? Explain your response.

No, the policy is not reasonable.

Only 3% of the customers default. Of those who are late, only 6.37% default. Knowing that a customer is late on payments does not dramatically increase the chance that they will default on the payment.

8. An insurance company has the following information about drivers aged 16 to 18 years: 20% are involved in accidents each year; 10% in this age group are A students; among those involved in an accident, 5% are A students. find the percent of A students who are involved in accidents.

$$\text{Answer: } P(\text{Accident} \mid \text{A students}) = 0.1$$

Solution to STAT 350 Exam 1 Review Problems Spring 2015

9. Servings of fruits and vegetables. The following table gives the distribution of the number of servings of fruits and vegetables consumed per day in a population.

Number of servings X	0	1	2	3	4	5
Probability	0.3	0.1	0.1	0.2	0.2	0.1

Find the mean and the standard deviation for this random variable.

Answer:

$$\mu = \sum_{k=0}^5 k p(k) = 0 \cdot 0.3 + 1 \cdot 0.1 + 2 \cdot 0.1 + \dots + 5 \cdot 0.1 = 2.2$$

$$\sigma^2 = \left[\sum_{k=0}^5 k^2 p(k) \right] - (\mu^2) = [0 \cdot 0.3 + 1 \cdot 0.1 + 4 \cdot 0.1 + 9 \cdot 0.2 + 16 \cdot 0.2 + 25 \cdot 0.1] - 2.2^2 = 3.16$$

$$\sigma = \sqrt{3.16} = \mathbf{1.78}$$

10. $f(x) = kx(2-x)$, $0 < x < 2$

a) What is the constant k that makes the above function a valid density function?

$$k = 0.75$$

b) Find $P(0 < X < 1.5)$

$$0.84$$

c) Find the mean and the standard deviation for X .

(Jeremy) reflective symmetry about $x = 1$, bounded support ($0 < x < 2$) \Rightarrow mean = 1
variance = $\frac{1}{3} = 0.2 \Rightarrow$ SD = 0.45

d) Find the 80th percentile of X . (ie. find x so that $P(X < x) = 0.8$)

(Jeremy) $F(x) = 0.75x^2 - 0.25x^3 = \frac{3}{4}x^2 - \frac{1}{4}x^3$
 $0.8 \Rightarrow x^3 - 3x^2 + 3.2 = 0$
Cubic equation, use wolfram alpha.
80th percentile $x = 1.43$

11. $f(x) = k(e^{-x} + e^{-4x})$ for $x > 0$

a) What is the constant k that makes the above function a valid density function?

(Jeremy) $W \sim \text{Exp}(\lambda_W = 1)$ independent of $Y \sim \text{Exp}(\lambda_Y = 4)$
 $\Rightarrow X = k(W + 0.25Y) \Rightarrow k = 1/(1 + 0.25) = \frac{4}{5} = 0.8$

b) Find $P(X > 2)$

(Jeremy) $P(X > 2) = \frac{3}{4}P(W > 2) + \frac{1}{4}P(Y > 2)$
 $= \frac{3}{4}e^{-2} + \frac{1}{4}e^{-8}$
 $= 0.8e^{-2} + 0.2e^{-8} = 0.8e^{-2} + 0.2e^{-8} = 0.11$

c) Find the mean and the standard deviation for X .

(Jeremy) $E[X] = E[0.8W + 0.2Y]$
 $= 0.8E[W] + 0.2E[Y]$

Solution to STAT 350 Exam 1 Review Problems Spring 2015

$$=_{\text{\{Mean of Exponential Function\}}} 0.8/\lambda_W + 0.2/\lambda_Y$$

$$= 0.8/1 + 0.2/4 = 0.85$$

$$\text{Var}[X] = \text{Var}[0.8W + 0.2Y] =_{\text{\{W independent of Y\}}} \text{Var}[0.8W] + \text{Var}[0.2Y]$$

$$=_{\text{\{Variance Operator\}}} 0.8^2 \text{Var}[W] + 0.2^2 \text{Var}[Y]$$

$$=_{\text{\{Variance of Exponential Function\}}} 0.64/\lambda_W^2 + 0.04/\lambda_Y^2$$

$$= 0.64/1^2 + 0.04/4^2 = 0.6425$$

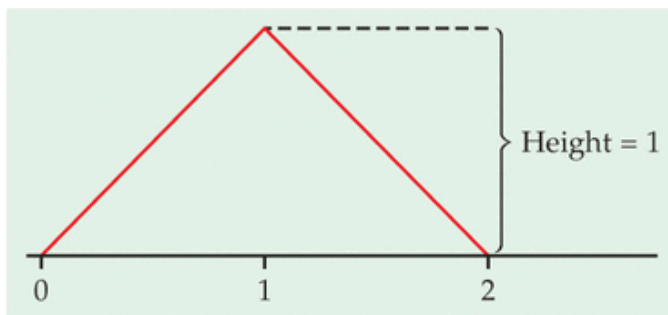
$$\Rightarrow \text{SD}[X] = 0.6425^{(1/2)} = 0.801561$$

Hard. Use Wolfram Alpha:

mean = 0.85, SD = 0.95

Random sample simulation in R provides strong evidence that TA made a boo-boo for SD.

12. The sum of two uniform random numbers. Generate two random numbers between 0 and 1 and take Y to be their sum. Then Y is a continuous random variable that can take any value between 0 and 2. The density curve of Y is the triangle shown in Figure 14.



a) Find the density function.

$$f(y) = \begin{cases} y & \text{for } 0 < y < 1 \\ 2 - y & \text{for } 1 < y < 2 \end{cases}$$

b) Find the mean and the standard deviation of Y .

$$\text{(Jeremy)} E[Y] = E[W + X] = E[W] + E[X] = (0 + 1)/2 + (0 + 1)/2 = 1$$

$$\text{Var}[Y] = \text{Var}[W + X] =_{\text{\{W independent of X\}}} \text{Var}[W] + \text{Var}[X]$$

$$= (1 - 0)^2/12 + (1 - 0)^2/12 = \frac{1}{6}$$

$$\Rightarrow \text{SD}[Y] = \sqrt{\frac{1}{6}} = 0.41$$

c) Find the 90th percentile, that is the value x so that $P(Y < y) = 0.9$.

$$\text{(Jeremy)} \quad F_Y(y) = \begin{cases} 0 & \text{for } y \leq 0 \\ 0.5y^2 & \text{for } 0 < y < 1 \\ -0.5y^2 + 2y - 1 & \text{for } 1 < y < 2 \\ 1 & \text{for } y \geq 2 \end{cases}$$

$$\Rightarrow F_Y(y) = -0.5y^2 + 2y - 1 =_{\text{\{SET\}}} 0.9 \Rightarrow y^2 - 4y + 3.8 = 0$$

$$\Rightarrow_{\text{\{Quadratic Formula\}}} y = 2 - \sqrt{5}/5 = 1.553$$

13. Suppose the time until the earthquake in Lonely Mountain has the density function

Solution to STAT 350 Exam 1 Review Problems Spring 2015

$f(x) = 0.5 e^{-0.5x}$ for $x > 0$ and $f(x)=0$ elsewhere

- a) Find the probability that the next earthquake happens within two years.

$$P(X < 2) = \text{_}\{\text{CDF of Exponential Function}\} 1 - e^{(-2\lambda)} = 1 - e^{(-1)} = 0.63$$

- b) Find the median time until the next earthquake in Lonely Mountain.

$$S_X(x) = e^{(-\lambda x)} = 1 - p \Rightarrow x = -\ln(1 - p)/\lambda = 2\ln(2) = 1.39$$

14. The total sleep time per night among college students was approximately Normally distributed with mean $\mu = 6.78$ hours and standard deviation $\sigma = 1.24$ hours. You plan to take an SRS of size $n = 150$ and compute the average total sleep time.

- (a) What is the standard deviation for the average time?

$$\sigma_{\bar{X}} = \sigma/\sqrt{n} = 0.1012$$

- (b) What is the probability that your average will be below 6.9 hours?

0.8830

15. The emerald ash borer is a beetle that poses a serious threat to ash trees. Purple traps are often used to detect or monitor populations of this pest. In the counties of your state where the beetle is present, thousands of traps are used to monitor the population. These traps are checked periodically. The distribution of beetle counts per trap is discrete and strongly skewed. A majority of traps have no beetles, and only a few will have more than 1 beetle. For this exercise, assume that the mean number of beetles trapped is 0.3 with a standard deviation of 0.8.

- (a) Suppose that your state does not have the resources to check all the traps, and so it plans to check only an SRS of $n = 100$ traps. What are the mean and standard deviation of the average number of beetles \bar{X} in 100 traps?

$$\mu_{\bar{X}} = \mu = 0.3$$

$$\sigma_{\bar{X}} = \sigma/\sqrt{n} = 0.8/10 = 0.08$$

- (b) Use the central limit theorem to find the probability that the average number of beetles in 100 traps is greater than 0.5.

0.0062

- (c) Do you think it is appropriate in this situation to use the central limit theorem? Explain your answer.

Yes, it is appropriate. The sample size 100 is very large and the traps are independent.