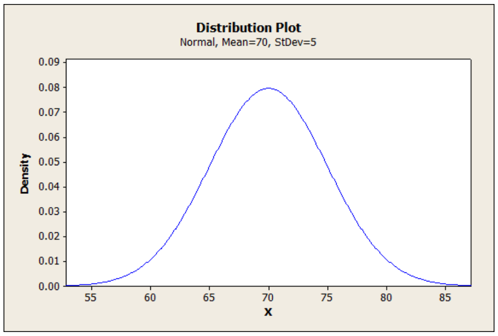
## STAT 350 Exam 1 Review Problems

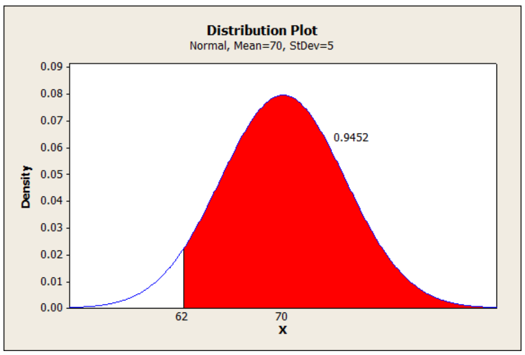
**1.** The weight of the eggs produced by a certain breed of hen is Normally distributed with mean 70 grams(g) and standard deviation 5 g.

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

Sketches are not required on the exam, but you can draw them if it will help you solve the problem.




b) What is the probability that a randomly selected egg is more than 62 grams?



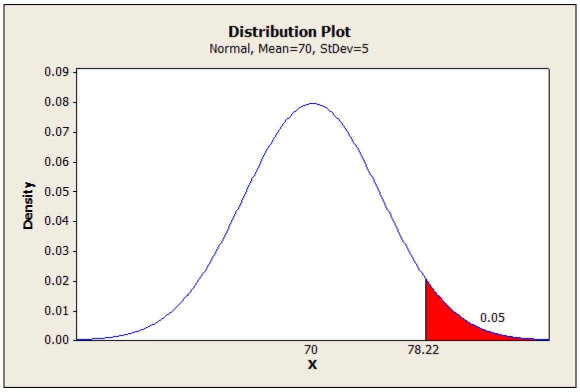

$$P(X > 62) = P\left(Z > \frac{62-70}{5}\right) = P(Z > -1.6)$$

Method 1
 = P(Z < 1.6) = 0.9452
Method 2
 = 1 – P(Z < -1.6) = 1 – 0.0548 = 0.9452

c) What is the probability that a randomly selected egg is between 62 and 69.4 grams?

$$P(62 < X < 69) = P\left(\frac{62-70}{5} < Z < \frac{69.4-70}{5}\right) = P(-1.6 < Z < -0.12) = P(Z < -0.12) - P(Z < -1.6)$$
$$= 0.4522 - 0.0548 = 0.3974$$

d) How much does an egg need to weigh in order for it to be in the largest 5% of weights of all of the eggs?

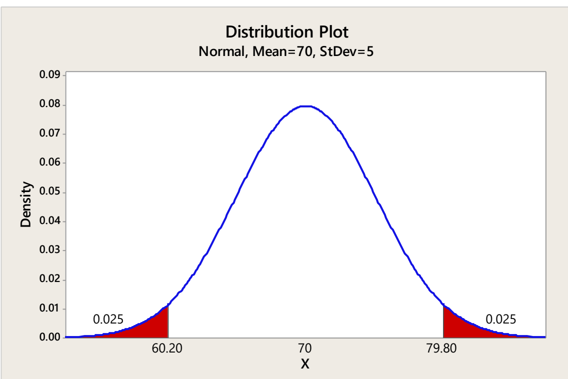



Note: we try to make sure that all of the percentages are exact on the exam to reduce confusion.
P(Z > z) = 0.05 ==> P(Z < z) = 1 – 0.05 = 0.95 ==> z = 1.645

$$z = \frac{x-\mu}{\sigma} \Rightarrow 1.645 = \frac{x-70}{5} \Rightarrow x = 70 + 1.645 \cdot 5 = 78.225$$

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

e) Find a symmetric interval around the mean such that 95% of all of the weights of the eggs lie in this interval.




P(70 – b < X < 70 + b) = 0.95

$$0.95 + \frac{1-0.95}{2} = 0.975$$

$$P(X < 70 + b) = P\left(Z < \frac{70+b-70}{5}\right) = P\left(Z < \frac{b}{5}\right) = 0.975$$

$$\frac{b}{5} = 1.96$$

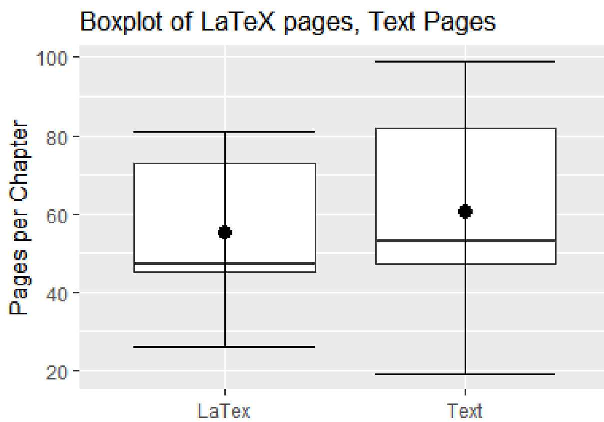
b = (5)(1.96) = 9.8
P(70 – 9.8 < X < 70 + 9.8) = 0.95 or (60.2, 79.8)

## STAT 350 Exam 1 Review Problems

**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. The data and the R output follows:

| | 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 |

| | | | | |
|---|---|---|---|---|
| **LaTeX** | `> mean(LaTex)`<br>`[1] 55.30769` | `> sd(LaTex)`<br>`[1] 18.59832` | `> var(LaTex)`<br>`[1] 345.8974` | `> quantile(LaTex)`<br>`0% 25% 50% 75% 100%`<br>`26 45 47 73 81` |
| **Text** | `> mean(Text)`<br>`[1] 60.61538` | `> sd(Text)`<br>`[1] 23.27208` | `> var(Text)`<br>`[1] 541.5897` | `> quantile(Text)`<br>`0% 25% 50% 75% 100%`<br>`19 47 53 82 99` |




a) What are the mean and the standard deviation for LaTex pages and Text pages?

| | LaTeX | Text |
|---|---|---|
| $\bar{x}$ mean() | 55.30769 | 60.61538 |
| s sd() | 18.59832 | 23.27208 |

b) What are the five number summaries for LaTex pages and Text pages?

Code: `quantile()`

Variable Minimum Q1 Median Q3 Maximum
LaTeXPages 26 45 47 73 81
TextPages 19 47 53 82 99

Note that you have to label these in your answer, not just put down the numbers.

c) Are there any outliers in this data using the 1.5 IQR rule? If there are any outliers, please state what they are. Justify your answer.

LaTeX

$$IQR = Q_3 - Q_1 = 73 - 45 = 28$$

$$1.5\ IQR = 1.5 \times 28 = 42$$

$$IF_L = Q_1 - 1.5\ IQR = 45 - 42 = 3$$
$$IF_H = Q_3 + 1.5\ IQR = 73 + 28 = 101$$

There are no outliers.

Text

$$IQR = Q_3 - Q_1 = 82 - 47 = 35$$
$$1.5\ IQR = 1.5 \times 35 = 52.5$$
$$IF_L = Q_1 - 1.5\ IQR = 47 - 52.5 = -5.5$$
$$IF_H = Q_3 + 1.5\ IQR = 82 + 52.5 = 134.5$$

There are no outliers.

There are no problems with the fact that the value of the fence (-5.5) is not a possible data point. That is the lower fence. This just means that there can't be any outliers.

These calculations are consistent with the boxplots.

d) Using the side-by-side boxplots shown above, i) Which one is more precise? ii) Which one has more pages? iii) What are their shapes?

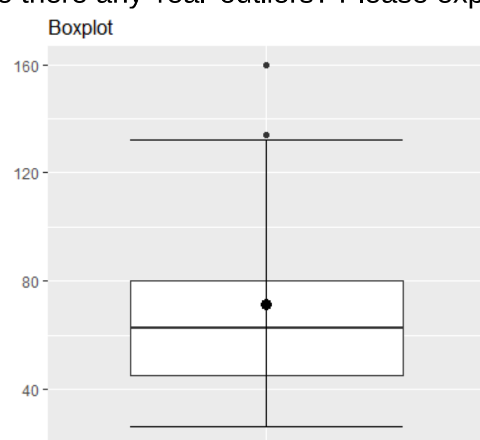
i) Which one is more precise?

The number of pages in the LaTeX chapters are more precise because the boxplot is smaller.

ii) Which one has more pages?

I would say that they have about the same number of pages on average because the boxes are in approximately the same location. The range of Text is such that it has a larger maximum and smaller minimum.

**3.** In the following boxplot, are there any 'real' outliers? Please explain your answer.




The point at around 160 is a real outlier as there is a sufficient gap between the upper whisker and the point indicating that this is probably more than simple tail behavior. The other explicit point is not a real outlier as it is not far from the whisker.

## STAT 350 Exam 1 Review Problems

Suppose that 8% of tires manufactured by a certain company are defective. Assume that we have a random sample of 16 of these tires (enough for 4 cars). Note: This is a binomial distribution. What is the probability that more than one of the tires are defective? Hint: Write down the complete formula that is to be used.

n = 16, p = 0.08

$$P(X > 1) = 1 - P(X \le 1) = 1 - P(X = 0) - P(X = 1) = 1 - \binom{16}{0} 0.08^0 (1 - 0.08)^{16} - \binom{16}{1} 0.08^1 (1 - 0.08)^{15}$$
$$= 1 - 0.92^{16} - (16)(0.08)(0.92)^{15} = 1 - 0.263 - 0.366 = 0.371$$

**5**. Cars pass your house at an average of 2.5 cars per day (you live in the country). Let X be the number of cars that pass by your house. Assume that X is a Poisson random variable.

a) What is the probability that the at least 2 cars pass by your house in 1.5 days?

$\lambda' = 1.5\lambda = (1.5)(2.5) = 3.75$

$$P(X \ge 2) = 1 - P(X < 2) = 1 - P(X = 0) - P(X = 1) = 1 - \frac{e^{-3.75} 3.75^0}{0!} - \frac{e^{-3.75} 3.75^1}{1!} = 1 - 0.0235 - 0.088$$
$$= 0.888$$

b) In the next 2 months (60 days), what is the probability that the <u>average</u> number of cars that passes your house per day is at least 2?

Since it says 'average', this is a sampling distribution.
$\mu_{\bar{x}} = \lambda = 2.5$

$$\sigma_{\bar{X}} = \frac{\sqrt{\lambda}}{\sqrt{n}} = \frac{\sqrt{2.5}}{\sqrt{60}} = 0.204$$
$$P(\bar{X} \ge 2) = P\left(Z \ge \frac{2 - 2.5}{0.204}\right) = P(Z \ge -2.45) = 1 - P(Z < -2.45) = 1 - 0.0071 = 0.9929$$

c) If the number of cars that passes your friend's house, Y, has a Poisson distribution with an average of 25 cars per day, what is the standard deviation of 3Y – 5X?

$$\sigma_{3Y-5X} = \sqrt{3^2\sigma_Y^2 + 5^2\sigma_X^2} = \sqrt{(9)(25) + (25)(2.5)} = \sqrt{287.5} = 16.956$$

Note that the formula uses the variance, not the standard deviation.

d) If the vehicles that pass by your house are either trucks (T) or cars (C), what is the sample space of the next 2 vehicles that pass by your house?

S = {TT, TC, CT, CC}

e) Suppose that 55% of the vehicles are trucks, 25% of the vehicles are red and 17% are red trucks (they are trucks and red). In addition, the vehicles that are not trucks are cars (we are assuming that there are no motorcycles). Given that a vehicle is red, what is the probability that it is a truck?

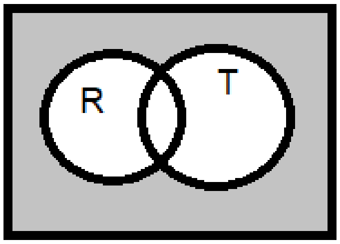
P(T) = 0.55     P(R) = 0.25     P(T ∩ R) = 0.17

$$P(T|R) = \frac{P(T \cap R)}{P(R)} = \frac{0.17}{0.25} = 0.68$$

f) Suppose that 55% of the vehicles are trucks, 25% of the vehicles are red and 17% are red trucks. In addition, the vehicles that are not trucks are cars (we are assuming that there are no motorcycles). What is the probability that the next vehicle is not red and is not a truck?

P(T) = 0.55     P(R) = 0.25     P(T ∩ R) = 0.17

# STAT 350 Exam 1 Review Problems

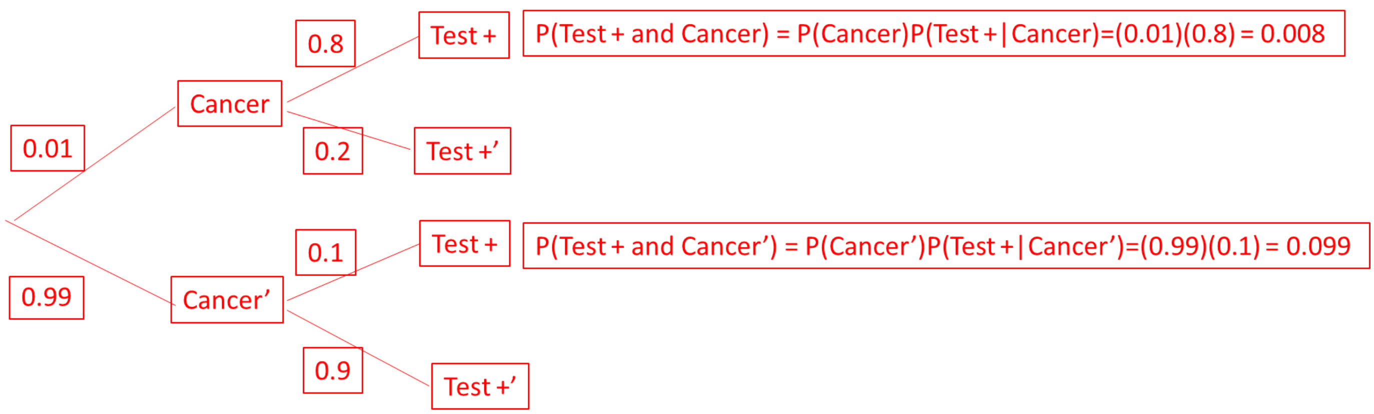



P(T' ∩ R') = [P(T U R)]' = 1 – [P(T) + P(R) - P(T ∩ R)] = 1 – [0.55 + 0.25 – 0.17] = 1 – 0.63 = 0.37
Note that these are NOT independent so you cannot multiply P(T') and P(R') together to get the answer.

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(Cancer | Test +).
We know: P(Cancer) = 0.01 ⇒ P (Cancer') = 1 – 0.01 = 0.99
P(Test + | Cancer ) = 0.8
P(Test + | Cancer') = 0.1

Method 1




So, P(Test + ) = P (Test + ∩ Cancer) + P(Test + ∩ Cancer')
=0.008 + 0.099 = 0.107

$$P(Cancer|Test+) = \frac{P(Cancer \cap Test+)}{P(Test+)} = \frac{0.008}{0.107} = 0.0748$$

Method 2

$$P(Cancer|Test+) = \frac{P(Cancer)P(Test+|Cancer)}{P(Cancer)P(Test+|Cancer) + P(Cancer')P(Test+|Cancer')}$$
$$= \frac{(0.01)(0.8)}{(0.01)(0.8) + (0.99)(0.1)} = 0.0748$$

b) If the 40-year-old woman has a positive mammogram, should she be worried that she has breast cancer?

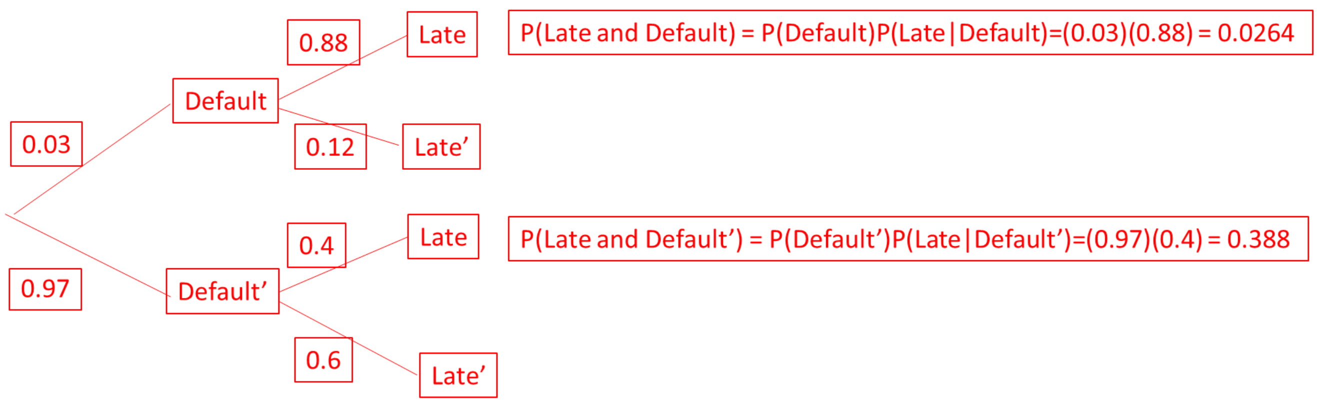
Since she only has a 7% chance of having breast cancer, she should get some further tests to be sure since this is cancer. However, she should not really be worried that she has cancer unless breast cancer runs in her family.

**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(Late|Default) = 0.88 P(Default) = 0.03 P(Late|Default$^c$) = 0.4
P(Default$^c$) = 1 – P(Default) = 1 – 0.03 = 0.97

Method 1




$$P(Default|Late) = \frac{P(Default \cap Late)}{P(Late)} = \frac{0.0264}{0.0.0264 + 0.388} = 0.0637$$

Method 2

$$P(Default|Late) = \frac{P(Default)P(Late|Default)}{P(Defaul)P(Late|Default) + P(Default')P(Late|Default')}$$
$$= \frac{(0.88)(0.03)}{(0.88)(0.03) + (0.4)(0.97)} = 0.0637$$

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?

P(Default) = 0.03 P(Default$^C$) = 0.97 P(Late|Default) = 0.88 P(Late|Default$^c$) = 0.4

Method 1: see tree diagram above

$$P(Default'|Late) = \frac{P(Default \cap Late)}{P(Late)} = \frac{0.388}{0.0.0264 + 0.388} = 0.936$$

$$P(Default'|Late) = \frac{P(Default')P(Late|Default')}{P(Default')P(Late|Default') + P(Default)P(Late|Default)}$$
$$= \frac{(0.4)(0.97)}{(0.4)(0.97) + (0.88)(0.03)} = 0.936$$

Method 2:

P(Default'|Late) = 1 – P(Default|Late).= 1 – 0.0637 = 0.936

(0.936)(100) = 93.6 therefore between 93 and 94 would NOT default.

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

No, the policy is not reasonable.
Only 3% of the customers defaulted. 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 though it does make the probability higher.

**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.

$$\mu = \sum_{i=0}^{5} x_i * p_i = 0 * 0.3 + 1 * 0.1 + 2 * 0.1 + \ldots + 5 * 0.1 = 2.2$$

$$\sigma^2 = E(X^2) - \left(E(X)\right)^2 = \left[\sum_{i=0}^{5} x_i{}^2 * p_i\right] - (\mu^2)$$
$$= [0 * 0.3 + 1 * 0.1 + 4 * 0.1 + 9 * 0.2 + 16 * 0.2 + 25 * 0.1] - 2.2^2 = 8 - 2.2^2 = 3.16$$
$$\sigma = \sqrt{3.16} = 1.78$$

Notice that you calculate the variance first and then take the square root to obtain the standard deviation.

**9.**

$$f_X(x) = \begin{cases} kx(2-x) & 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

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

$$\int_0^2 kx(2-x)dx = k\int_0^2 (2x - x^2)dx = k\left[\frac{2x^2}{2} - \frac{x^3}{3}\right]_0^2 = k\left(4 - \frac{8}{3}\right) = \frac{4}{3}k = 1 \Rightarrow k = \frac{3}{4}$$

b) Find P( 1 < X < 3)

$$P(1 < X < 3) = P(1 < X < 2) = \int_1^2 \frac{3}{4}x(2-x)dx = \frac{3}{4}\int_1^2 (2x - x^2)dx = \frac{3}{4}\left[\frac{2x^2}{2} - \frac{x^3}{3}\right]_1^2 = \frac{3}{4}\left(\frac{4}{3} - \frac{2}{3}\right) = \frac{3}{4}\left(\frac{2}{3}\right)$$
$$= 0.5$$

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

$$E(X) = \int_0^2 x\frac{3}{4}x(2-x)dx = \frac{3}{4}\int_0^2 (2x^2 - x^3)dx = \frac{3}{4}\left[\frac{2x^3}{3} - \frac{x^4}{4}\right]_0^2 = \frac{3}{4}\left(\frac{16}{3} - 4\right) = 1$$

$$E(X^2) = \int_0^2 x^2\frac{3}{4}x(2-x)dx = \frac{3}{4}\int_0^2 (2x^3 - x^4)dx = \frac{3}{4}\left[\frac{2x^4}{4} - \frac{x^5}{5}\right]_0^2 = \frac{3}{4}(8 - 6.4) = 1.2$$

$\sigma^2 = E(X^2) - (E(X))^2 = 1.2 - 1^2 = 0.2$

$$\sigma = \sqrt{\sigma^2} = \sqrt{0.2} = 0.447$$

d) Find the 80th percentile of X.

$$0.8=\int_0^y \frac{3}{4}x(2-x)dx=\frac{3}{4}\int_0^y (2x-x^2)dx=\frac{3}{4}\left[\frac{2x^2}{2}-\frac{x^3}{3}\right]_0^y=\frac{3}{4}\left(y^2-\frac{y^3}{3}\right)=0.75y^2-0.25y^3$$

The integration finds the CDF.
Therefore, the equation that you need to find the roots for is $0.75y^2 - 0.25y^3 - 0.8 = 0$
I prefer to remove the fractions by multiplying the equation by 4 (this is not required) to obtain
$3y^2 - y^3 - 3.2$ -= 0

This is a cubic (this will NOT be on the exam) so you can solve it via a graphing calculator or Wolfram alpha (which is what I use)

The three roots are -0.91, 1.43, 2.48. The only one that is in the range of (0,2) is **1.43** so this is the answer.

**10.**

$$f_X(x)=\begin{cases} k(e^{-x}+e^{-4x}) & x>0 \\ 0 & \text{otherwise}\end{cases}$$

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

$$\int_0^\infty k(e^{-x}+e^{-4x})dx=k\left(\int_0^\infty (e^{-x}+e^{-4x})dx\right)=k\left(\left[-e^{-x}+\left[-\frac{1}{4}e^{-4x}\right]\right]_0^\infty\right)=k\left[\left(-0-\frac{1}{4}(0)\right)-\left(-1-\frac{1}{4}\right)\right]$$
$$=\frac{5}{4}k\Rightarrow k=\frac{4}{5}$$

b) Find P(X > 2).

$$P(X>2)=\int_2^\infty \frac{4}{5}(e^{-x}+e^{-4x})dx=\frac{4}{5}\left(\int_2^\infty (e^{-x}+e^{-4x})dx\right)=\frac{4}{5}\left(\left[-e^{-x}+\left[-\frac{1}{4}e^{-4x}\right]\right]_2^\infty\right)$$
$$=\frac{4}{5}\left[\left(-0-\frac{1}{4}(0)\right)-\left(-e^{-2x}-\frac{1}{4}e^{-8}\right)\right]=\frac{4}{5}(0.135)=0.11$$

c) Find the mean for X.

You could do this problem from scratch but that would involve integrating by parts. So, I am using a different method.

$$E[X]=\frac{4}{5}\int_0^\infty (xe^{-x}+xe^{-4x})dx$$
$$=\frac{4}{5}\int_0^\infty xe^{-x}dx+\frac{1}{5}\int_0^\infty 4xe^{-4x}dx$$

Note, that this is a weighted sum of two expected values for exponential distributions.
Let $\boldsymbol{Y\sim Exp(\lambda_1=1)}$ **and** $\boldsymbol{Z\sim Exp(\lambda_2=4)}$
Then it follows that $\boldsymbol{E[X]=\frac{4}{5}E[Y]+\frac{1}{5}E[Z]=\frac{4}{5}\times 1+\frac{1}{5}\times\frac{1}{4}=\frac{17}{20}=0.85}$.

d) Find the cdf of this function.

$$F(s)=\int_0^s \frac{4}{5}(e^{-x}+e^{-4x})dx=\frac{4}{5}\left(\int_0^s (e^{-x}+e^{-4x})dx\right)=\frac{4}{5}\left(\left[-e^{-x}+\left[-\frac{1}{4}e^{-4x}\right]\right]_0^s\right)$$
$$=\frac{4}{5}\left[\left(-e^{-s}-\frac{1}{4}e^{-4s}\right)-\left(-1-\frac{1}{4}\right)\right]=\frac{1}{5}(-4e^{-s}-e^{-4s}+4+1)=\frac{1}{5}(-4e^{-s}-e^{-4s}+5)$$

**11.**

$$f_X(x) = \begin{cases} kx^{-2} & -2 < x < -1 \\ 0 & \text{otherwise} \end{cases}$$

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

$$\int_{-2}^{-1} kx^{-2}dx = k\left[\frac{-1}{x}\right]_{-2}^{-1} = k\left(\frac{-1}{-1} - \frac{-1}{-2}\right) = \frac{1}{2}k = 1 \Rightarrow k = 2$$

b) Find P(X > 0).

P(X > 0) = 0 because the pdf is 0 in that range.

c) Find the cdf of this function.

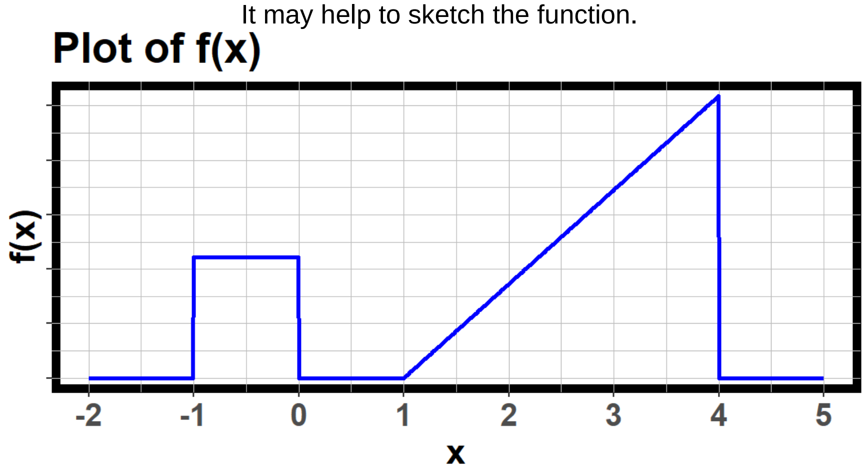
$$\int_{-2}^{x} 2s^{-2}ds = 2\left[\frac{-1}{s}\right]_{-2}^{x} = 2\left(\frac{-1}{x} - \frac{-1}{-2}\right) = -\left(\frac{2}{x} + 1\right)$$

$$F_X(x) = \begin{cases} 0 & x < -2 \\ -\frac{2}{x} - 1 & -2 \le x < -1 \\ 1 & x \ge -1 \end{cases}$$

**12.**

$$f_X(x) = \begin{cases} \frac{2}{9} & -1 < x < 0 \\ k(x-1) & 1 < x < 4 \\ 0 & \text{otherwise} \end{cases}$$

It may help to sketch the function.




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

**Using geometry:**

Notice that $f_X(x) \ge 0$ as long as $k > 0$. Next, the area of the rectangle is $\frac{2}{9}$, that implies that we need $k$ such that the area of the triangle is $\frac{7}{9}$. Area of triangle is $\frac{1}{2}\text{base} \times \text{height}$ the base has length 3 and the height is unknown but we know the area that is required $\frac{1}{2} \times 3 \times h = \frac{7}{9}$, which yields $h = \frac{14}{27}$. Knowing the base and height means we know two pairs of points $(1,0), \left(4, \frac{14}{27}\right)$, allowing us to obtain the line $y = mx + b$ which yield $y = \frac{14}{81}x - \frac{14}{81} = \frac{14}{81}(x-1)$ and thus $k = \frac{14}{81}$.

**Using Integration:**

$$\int_{-1}^{0} \frac{2}{9}dx + k\int_{1}^{4}(x-1)dx = 1$$

$$\frac{2}{9} + k\left[\frac{x^2}{2} - x\right]_1^4 = 1$$

$$\frac{2}{9} + k\left(8 - 4 - \left(\frac{1}{2} - 1\right)\right) = 1$$

$$\frac{2}{9} + \frac{9k}{2} = 1$$

$$k = \frac{14}{81}$$

b) Find the standard deviation for $X$.

We know that → $\sigma_X = \sqrt{\mathrm{Var}(X)} = \sqrt{\mathrm{E}[X^2] - (\mathrm{E}[X])^2}$
Therefore, we need to find $\mathrm{E}[X]$ and $\mathrm{E}[X^2]$.

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x f_X(x) = \frac{2}{9}\int_{-1}^{0} x dx + \frac{14}{81}\int_{1}^{4}(x^2 - x)dx$$

$$= \frac{1}{9}x^2|_{-1}^{0} + \frac{14}{81}\left[\frac{1}{3}x^3 - \frac{1}{2}x^2\right]_1^4 = \frac{2}{18} + \frac{14}{81}\left(\frac{64}{3} - 8 - \left(\frac{1}{3} - \frac{1}{2}\right)\right)$$

$$= -\frac{1}{9} + \frac{14}{\cancel{81}} \times \frac{\cancel{81}}{6} = \frac{20}{9} = 2.2222$$

$$\mathrm{E}[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x) = \frac{2}{9}\int_{-1}^{0} x^2 dx + \frac{14}{81}\int_{1}^{4}(x^3 - x^2)dx$$

$$= \frac{2}{27}x^3|_{-1}^{0} + \frac{14}{81}\left[\frac{1}{4}x^4 - \frac{1}{3}x^3\right]_1^4 = \frac{2}{27} + \frac{14}{81}\left(64 - \frac{64}{3} - \left(\frac{1}{4} - \frac{1}{3}\right)\right)$$

$$= \frac{2}{27} + \frac{14}{81} \times \frac{513}{12} = \frac{403}{54} = 7.4629$$

$$\mathrm{E}[X] = \frac{20}{9}$$

$$\mathrm{E}[X^2] = \frac{403}{54}$$

$$Var(X) = \frac{403}{54} - \left(\frac{20}{9}\right)^2 = \frac{1209}{162} - \frac{800}{162} = \frac{409}{162} = 2.5247$$

$$\sigma_X = \sqrt{\frac{403}{54}} = 1.58893$$

c) Find the cdf of this function.

The cdf will be piece-wise with 5 regions →

$$F_X(x) = \begin{cases} 0 & x < -1 \\ \frac{2}{9}(x+1) & -1 \le x < 0 \\ \frac{2}{9} & 0 \le x < 1 \\ \frac{7}{81}(x-1)^2 + \frac{2}{9} & 1 \le x < 4 \\ 1 & x \ge 4 \end{cases}$$

Region 1: $x < -1$ → No area has been accumulated by this point so the CDF will remain 0.

Region 2: $-1 \leq x < 0$ → Start to accumulate area but it depends on the value of x in this range, therefore we integrate from -1 up to some arbitrary $x$.

$$\frac{2}{9}\int_{-1}^{x} dt = \frac{2}{9}(x+1)$$

Region 3: $0 \leq x < 1$ → No new area is accumulated in the region but we have already accumulate the entire rectangle which has area 2/9 so the CDF remains 2/9 between 0 and 1.

Region 4: $1 \leq x < 4$ → Have already accumulated 2/9 of the area and now begin to accumulate more area depending on what value $x$ is in the range of 1 to 4.

Will use $u$ –substitution to make the function more clear but its not required.

$$\frac{2}{9}+\frac{14}{81}\int_{1}^{x}(t-1)dt = \frac{2}{9}+\frac{14}{81}$$

Let $u = t - 1$ then $du = dt$ and the limits of integration change from 1-x to 0-(x-1)

$$\frac{2}{9}+\frac{14}{81}\int_{0}^{x-1} udu = \frac{2}{9}+\frac{7}{81}(x-1)^2$$

Region 5: After 4 the entire area has been accumulated i.e., the CDF will be 1 for any $x \geq 4$.

d) Determine the 80th percentile of $X$.

Solving → $F_x(x^*) = 0.8$ for $x^*$. We note that before 1 the total area accumulated is only $\frac{2}{9} = 0.2222$.

Therefore the 80th percentile must occur within 1 to 4.

Solve → $\frac{2}{9}+\frac{7}{81}(x^*-1)^2 = 0.8$

$$(x^*-1)^2 = \frac{234}{35}$$

$$x^* = 1 \pm \sqrt{\frac{234}{35}}$$

$x^* = 1 + \sqrt{\frac{234}{35}} = 3.585675$ or $x^* = 1 - \sqrt{\frac{234}{35}} = -1.585675$

However, only one of these is in the support and therefore the 80th percentile is

$$x^* = 1 + \sqrt{\frac{234}{35}} = 3.585675$$

**13.** Suppose the time until an earthquake in Lonely Mountain has an exponential density function with an average of 2 years.

I strongly recommend that these types of problems be done **without** using integrals.

Since this is an exponential distribution with E(X) = 2, $\lambda = \frac{1}{2} = 0.5$
The distribution is f(x) 0.5 $e^{-x/2}$ for x > 0.

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

$$P(X < 2) = 0.5\int_0^2 e^{-x/2}dx = \left[-e^{-x/2}\right]_0^2 = (-e^{-1} + 1) = 0.632$$

or you can use the CDF
F(X) = 1 – $e^{-1}$ = 0.632

b) If you are living on Lonely Mountain, should you prepare for an earthquake to occur?

Yes, because there is a 63% chance of an earthquake occurring soon.

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

$$0.5 = 0.5\int_0^y e^{-x/2}dx = \left[-e^{-x/2}\right]_0^y = \left(-e^{-y/2} + 1\right) \Rightarrow e^{-\frac{y}{2}} = 0.5$$

Again, you can use this problem by using the CDF and not re-do the integration.

Take the ln (natural log) of each side

$$-\frac{y}{2} = \ln(0.5) \Rightarrow y = -2\ln(0.5) = 2ln2 = 1.386$$