Review for Exam 2

1. Ground level ozone is a health hazard because it may cause asthma and heart attacks, etc. The World Health Organization (WHO) recommends ozone levels of more than 51 ppb (parts per billion)[1]. In 2008, the US Environmental Protection Agency (EPA) recommended no more than 75 ppb; this is now under review until 2013.[2]

   In Dallas, Texas, the ozone level is reported daily.[3] Assume the values are normally distributed with mean 49 ppb and standard deviation 14 ppb.

   (a) What is the probability that the daily reading is above the WHO limit? The EPA limit? Before calculating, decide which probability will be larger.

   The WHO probability will be larger because the cut-off is lower.
   For the WHO, the z-value is
   $$z = \frac{51 - 49}{14} = 0.14.$$
   From the table, $z = 0.14$ corresponds to a proportion of $0.5398 = 53.98$ %. Thus about 54% of the daily values meet WHO standards, and about 46% of the readings exceed them.
   For the EPA, the z-value is
   $$z = \frac{75 - 49}{14} = 1.86.$$
   Since $z = 1.86$ corresponds to $0.9713 = 97.13\%$, about 97% of the daily values meet EPA standards; 3% exceed them.

   (b) How often is the weekly average above the WHO limit? Above the EPA limit? (Think of the seven weekly readings as random sample.)
   By the Central Limit Theorem, the mean of seven readings is normally distributed with mean 49 ppb and standard deviation $14/\sqrt{7} = 5.29$ ppb.

   For WHO, we have $z = \frac{51-49}{5.29} = 0.38$.
   For EPA, we have $z = \frac{75-49}{5.29} = 4.9$
   From the table, $z = 0.38$ corresponds to $0.6554 = 65.54\%$ and $z = 4.9$ is off the chart. So about $100 - 65.5 = 34.5\%$ of the weeks are above the WHO level and almost none above the EPA level.

   (c) How often is the monthly average above the WHO limit? Above the EPA limit? (Take a 31-day month.)

---

[1] In fact, 100 μg/m3 8-hour mean http://whqlibdoc.who.int/hq/2006/WHO_SDE_PHE_OEH_06.02_eng.pdf, which converted to ppb gives about
[2] 2008 standard at http://www.epa.gov/glo/pdfs/fs20100106std.pdf;. Now suspended till 2013
http://www.economist.com/blogs/democracyinamerica/2011/09/sorry-air
[3] Texas Commission on the Environment http://www.tceq.state.tx.us/nav/data/ozone_data.html. Estimates of mean, SD from data.

For the WHO, we have $z = \frac{51-49}{14/\sqrt{31}} = 0.80$. From the table, this corresponds to a proportion of 0.7881 = 78.81%, so 21.2% of the months are above the WHO level.

Since the monthly average is less variable than the weekly average, the monthly average will almost never be above the EPA level.

(d) Estimate the Estimate the mean and standard deviation of a single year-long sample of daily readings.

The sample is large enough that we expect its mean to be near the population mean, 49 ppb. The standard deviation of the sample will be close to the standard deviation of the population, 14 ppb.

(e) Consider the means of all year-long samples of daily readings. (Take a 365 day year.)

(i) What is the distribution of these means? Give the name of the distribution and its mean and standard deviation.

(ii) Using the "Rule of Thumb" for the normal distribution, find the range of values where the sample mean is likely to lie. (That is, 95% of the time).

(i) By the Central Limit Theorem, the sampling distribution of means is normal with mean 49 ppb and standard error (SD)

$$SE = \frac{14}{\sqrt{365}} = 0.73 \text{ ppb.}$$

(ii) By the Rule of Thumb, about 95% of the sample means are within 2 SDs of the mean, that is within $2(0.73) = 1.46$ ppb. Thus, they are between
$$49 - 1.46 = 47.54 \text{ and } 49 + 1.46 = 50.46 \text{ ppb.}$$
Thus we expect that in 95% of the years, the mean will be between 47.54 ppb and 50.46 ppb.

2. The weights of figure skaters are normally distributed. For male skaters, the mean is 170 lbs with a standard deviation of 10 lbs. For female skaters, the mean is 110 lbs with a standard deviation of 5 lbs. Let X be the weight of a randomly selected male skater and Y be the weight of a randomly selected female skater.

(a) What is $P(X < 150)$? 0.0228

(b) Approximately 90% of the male figure skaters weigh more than how many pounds? (Ans: 157 lb)

(c) The weight of a pair of figure skaters (a male and a female) can be thought of as a new random variable. Let the random variable $W = X + Y$. What is the mean of W? (Ans 280 lb)

(d) Suppose we consider the weights of the male partner and the female partner to be independent. What is the standard deviation of the random variable W? (Ans 11.18 lb)

(e) It does not seem likely that the weights of the male partner and the female partner would be independent. If the correlation r between X and Y equals 0.77, what is the standard deviation of the random variable W? (Ans 14.21 lb)

3. A politician is urging tighter restrictions on drivers' licenses issued to teens. He claims that "in one of every five auto accidents a teenager is behind the wheel."

(a) If the politician is correct, what is the probability that you would find no teen drivers in a random sample of 5 accidents? One or more? All five?
$$\text{None} = (0.8)^5 = 0.328,$$
$$\text{One or more} = 1 - 0.328 = 0.672,$$
$$\text{All five} = (0.2)^5 = 0.0003.$$

(b) You decide to do a little research and find 67 records of car accidents over the past few months. Only nine of the accidents had a teenager behind the wheel. Assume that we can consider these 67 accidents as a random sample of all accidents. If the politician is correct, what is the chance that you would observe 9 or fewer accidents with a teenager behind the wheel?
Use the CLT. We have
$$\hat{p} = \frac{9}{67} = 0.1343$$
The CLT tells us that Mean $= 0.2$ and
$$SE = \sqrt{\frac{0.2(1-0.2)}{67}} = 0.04887.$$
Since CLT tells us the distribution of sample proportions, $\hat{p}$, is normal, we find the z-value
$$z = \frac{0.1343 - 0.2}{0.04887} = -1.344.$$
From the table, we have probability $= 0.0901$

(c) How would your answer change if you had observed a larger number of accidents with teen drivers out of the 67? Probability would increase, because 9 out of 67 is less than what the politician said.

(d) How would your answer change if you had observed 9 accidents with teen drivers out of a random sample of more than 67 accidents? Smaller probability, because the proportion will decrease and be farther then the politicians estimate.

4. A New Zealand study on women's body image and ethnicity surveyed a group of women who were all slightly lighter than average. A question about what they thought about their own weight gave the following results.

	Body Image		
	Underweight	Right weight	Overweight
Asian	3	16	31
European	3	14	83

Let A be the event that a randomly selected woman in the study is Asian, E that she is European, U that she thinks she is underweight, R that she thinks she is the right weight, and O that she thinks she is overweight. Using these letters, express the following probabilities and find the values:

(a) The probability that a woman is Asian. $P(A)$,
(b) The probability that a woman things she is the right weight. $P(R)$
(c) The probability of an Asian woman who thinks she is the right weight. $P(A \text{ and } R)$.
(d) The probability that an Asian woman thinks she is overweight. $P(O|A)$
(e) The probability that a European woman thinks she is overweight $P(O|E)$.
(f) The probability that a woman who thinks she is overweight is Asian.
(g) Are ethnicity and body image independent?
Solution: Finding row and column totals gives

	Underweight	Right	Overweight	Total
Asian	3	16	31	50
European	3	14	83	100
Total	6	30	114	150

There are 150 women in the study.

(a) $P(A) = 50/150 = 0.333 = 33.3\%$

(b) $P(R) = 30/150 = 0.2 = 20\%$

(c) $P(A \text{ and } R) = 16/150 = 0.107 = 10.7\%$.

(d) $P(O|A) = 31/50 = 0.62 = 62\%$

(e) $P(O|E) = 83/100 = 0.83 = 83\%$

(f) $P(O|E) = 83/100 = 0.83 = 83\%$

(g) $P(A|O) = 31/114 = 0.272 = 27.2\%$.

(h) Not independent as $P(A)$ and $P(A|O)$ have different values.

5. The Department of Animal Regulations released information on pet ownership for the population consisting of all households in a particular county. Let the random variable X be the number of licensed dogs per household. The distribution for the random variable X is given below:

Value of X	0	1	2	3	4	5
Probability	0.52	0.22	0.13		0.03	0.01

a. The probability for $X = 3$ is missing. What is it? 0.09

b. What is the probability that a randomly selected household from this community owns at least one licensed dog? 0.48

c. Find the cdf.

| Value of X	0	1	2	3	4	5
Probability	0.52	0.22	0.13	0.09	0.03	0.01
cdf	0.52	0.74	0.87	0.96	0.99	1

d. What is the average number of licensed dogs per household in this county ? 092

6. You take samples from a population containing 62% students. Find the means and standard errors for the proportion of students in samples of $n = 100$ and of $n = 400$. What do you observe?
Mean is 62% for each sample.

For $n = 100$: Keeping more decimal places: Standard error $= \sqrt{\frac{0.62(1-0.62)}{100}} = 0.048 = 4.8\%$.

For $n = 400$: Standard error $= \sqrt{\frac{0.62(1-0.62)}{400}} = 0.024 = 2.4\%$.

Notice that the standard error for $n = 100$ is twice that for $n = 400$.
To **halve** the standard error, you have to **quadruple** the sample size.

7. The distribution for a population of test scores is displayed at the ***top left***.   The other five graphs, A
     to E, represent possible distributions of sample means for random samples drawn from the
     population.

Population Distribution

$\sigma = 28.19$

Median = 28 ⌐   ⌐$\mu = 40$

A

B

C

D

E
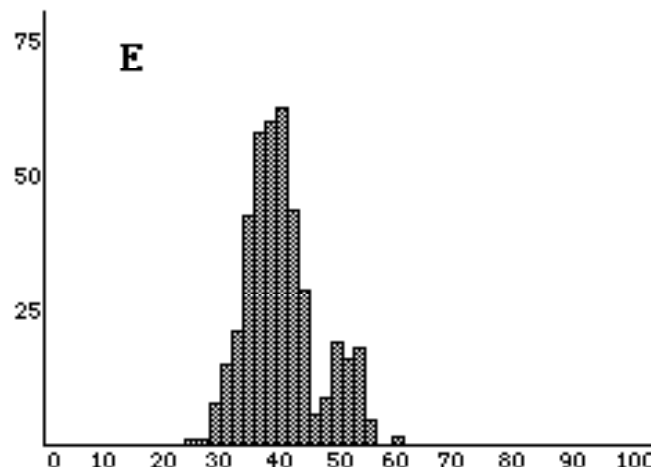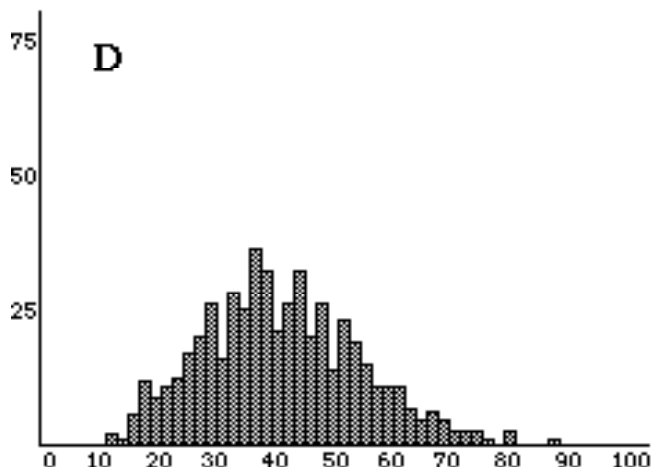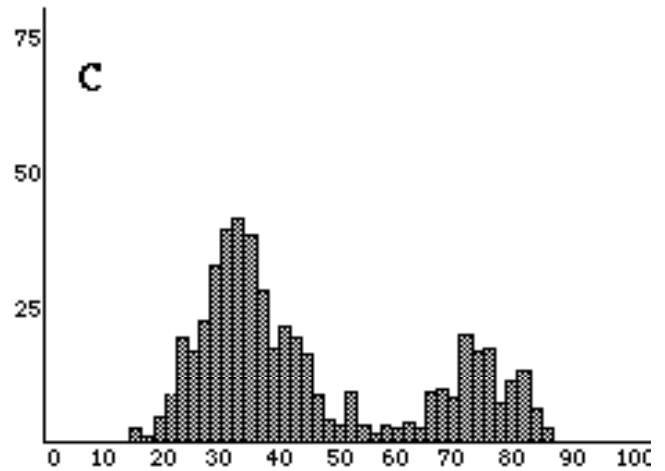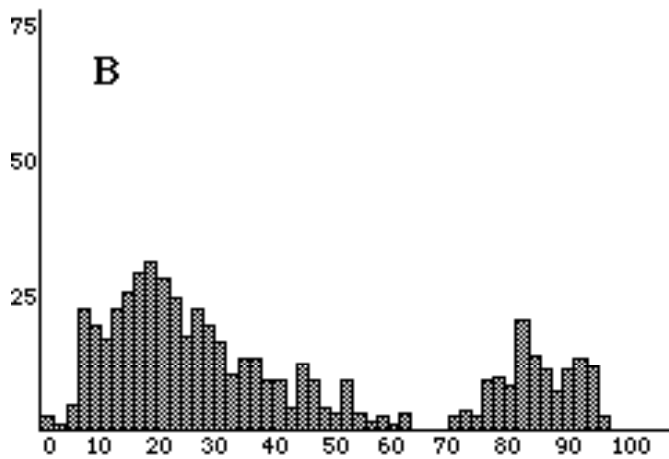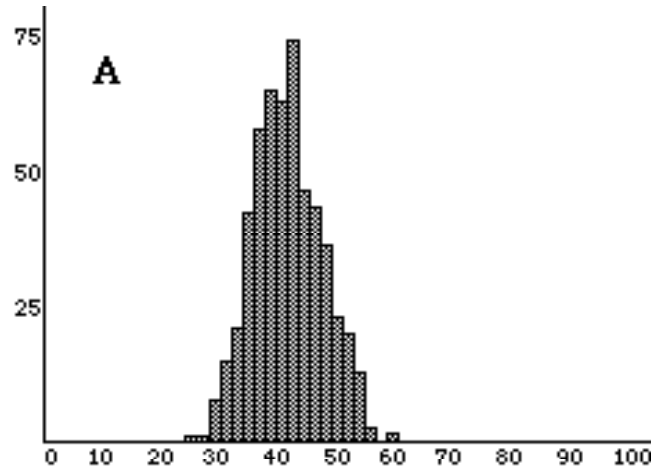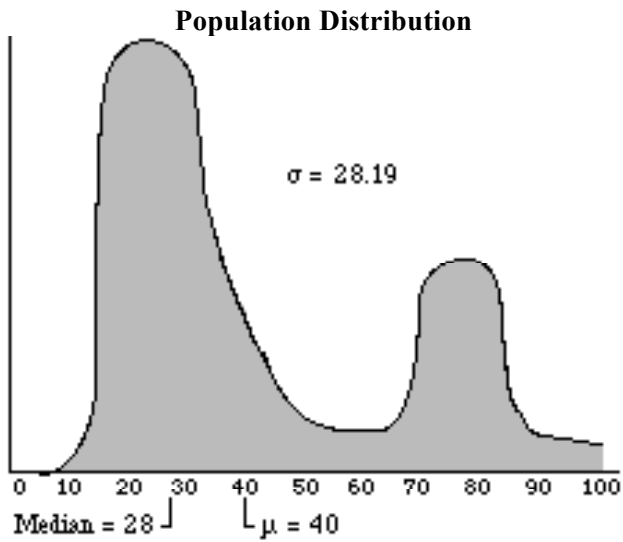
Which graph approximates the distribution of sample means for 500 samples of size **4**? (circle one)

    A   B   C   D   E

D: Sampling distribution is approximately normal, with SE $= \frac{28.19}{\sqrt{4}} = 14.1$.

The standard deviation of the sampling distribution you chose is (check one):

Smaller_____the same_____Larger_____

than the standard deviation of the population? Smaller

Which graph represents a distribution of sample means for 500 samples of size **25**? (circle one)

    A   B   C   D   E

A: Sampling distribution is approximately normal, with SE $= \frac{28.19}{\sqrt{25}} = 5.6$.

What do you expect for the shape of the sampling distribution you chose in?

     o   Shaped more like a normal distribution. Normal

     o   Shaped more like the population.

The standard deviation of the sampling distribution for n = 4 is Larger

      Smaller_____the same_____ Larger_____

than the standard deviation of the sampling distribution you chose for n = 25?