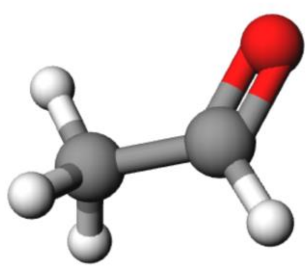
Exam II Solutions
Administered: Monday, October 16, 2023
24 points

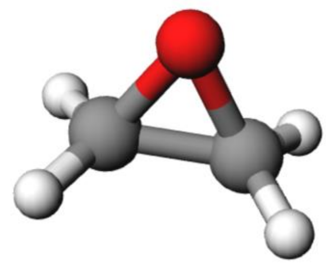
For each problem part: 0 points if not attempted or no work shown,
1 point for partial credit, if work is shown,
2 points for correct numerical value of solution

Problem 1. (16 points) Consider the following data for the critical temperature for two biochemicals.

Acetaldehyde
C_2H_4O



Ethylene Oxide
C_2H_4O



taken from the NIST Chemistry Webbook, http://webbook.nist.gov/chemistry/.

Critical Temperature of Acetaldehyde

Temperature (K)	Reference
466	Teja and Anselme, 1990
461	Hollmann, 1903
454.7	Van der Waals, 1881

Critical Temperature of Ethylene Oxide

Temperature (K)	Reference
468.9	Walters and Smith, 1952
469	Hess and Tilton, 1950
465.2	Maass and Boomer, 1922

Perform the following tasks.
(a) Determine the sample mean of the critical temperature of acetaldehyde.
(b) Determine the sample mean of the critical temperature of ethylene oxide.
(c) Determine the sample variance of the critical temperature of acetaldehyde.
(d) Determine the sample variance of the critical temperature of ethylene oxide.
(e) Identify the appropriate distribution to describe the difference of means in this case?
(f) Determine the lower limit of a 80% confidence interval on the difference of means of the critical temperature.
(g) Determine the upper limit of a 80% confidence interval on the difference of means of the critical temperature.
(h) Explain your findings in language a non-statistician can understand.

Solution:

(a) Determine the sample mean of the critical temperature of acetaldehyde.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 460.6\ K$$

Based on this data set sample mean of the critical temperature of acetaldehyde is 460.6 K.

(b) Determine the sample mean of the critical temperature of ethylene oxide.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 467.7\ K$$

Based on this data set sample mean of the critical temperature of ethylene oxide is 467.7 K.

(c) Determine the sample variance of the critical temperature of acetaldehyde.

$$s^2 = \frac{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)} = 32.06\ K^2$$

Based on this data set the sample variance of the critical temperature of acetaldehyde is 32.06 K^2.

(d) Determine the sample variance of the critical temperature of ethylene oxide.

$$s^2 = \frac{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)} = 4.69\ K^2$$

Based on this data set the sample variance of the critical temperature of ethylene oxide is 4.69 K^2.

(e) Identify the appropriate distribution to describe the difference of the mean of the critical temperature in this case.

In this case we do not know the true population variance so the appropriate distribution of the difference of sample means is the **t** distribution.

(f) Determine the lower limit of a 80% confidence interval on the difference of means of the critical temperature.
(g) Determine the upper limit of a 80% confidence interval on the difference of means of the critical temperature.

$$C.I. = 1 - 2\alpha = 0.80$$

$$\alpha = \frac{1 - C.I.}{2} = \frac{1 - 0.80}{2} = 0.10$$

$$v = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left[\left(\frac{s_1^2}{n_1}\right)^2 \Big/ (n_1 - 1)\right] + \left[\left(\frac{s_2^2}{n_2}\right)^2 \Big/ (n_2 - 1)\right]} \quad \text{if } \sigma_1 \neq \sigma_2$$

$$v = 2.57 \sim 3$$

The limits on the t-distribution are given by

```
>> tlo = icdf('t',0.1,3)

tlo =  -1.637744353696194
```

and for the upper limit

```
>> thi = icdf('t',0.9,3)

t =  1.637744353696194
```

We next insert all of these numbers into the equation for the confidence interval on the difference of means.

$$P\left[\left(\overline{X}_1-\overline{X}_2\right)+t_\alpha\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}<\left(\mu_1-\mu_2\right)<\left(\overline{X}_1-\overline{X}_2\right)+t_{1-\alpha}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}\right]=1-2\alpha$$

$$P[-12.8657 < (\mu_1 - \mu_2) < -1.40097] = 0.80$$

(h) Explain your findings in language a non-statistician can understand.

We are 80% confidence that the difference between the critical temperature of acetaldehyde and ethylene oxide lies within the range from -12.9 to -1.4 K. The sample mean of the critical temperature is smaller for acetaldehyde than for ethylene oxide. This confidence interval on the difference includes strictly negative numbers, indicating that all differences within the 80% confidence level show a higher critical temperature for ethylene oxide relative to acetaldehyde.

Problem 2. (8 points)
Consider a 6-cylinder automobile engine with six spark plugs, each with a lifetime that follows the normal distribution with a mean lifetime of 50,000 miles and a standard deviation of 2,000 miles.
(a) What is the probability that a sparkplug lasts at least 45,000 miles?
(b) What is the probability that all six sparkplugs last at least 45,000 miles?
(c) What is the probability that a sparkplug dies before 45,000 miles?
(d) What is the probability that no more than half the sparkplugs have died by 45,000 miles?

Solution:

(a) What is the probability that a sparkplug lasts at least 45,000 miles?

$$P(x \geq 45{,}000) = 1 - P(x \leq 45{,}000)$$

We use the cdf function in MATLAB:

```
>> p = 1 - cdf('normal',45000,50000,2000)

p = 0.993790334674224
```

There is a 99.38% chance that a sparkplug lasts at least 45,000 miles.

(b) What is the probability that all six sparkplugs last at least 45,000 miles?

This problem requires the binomial distribution. If we define the random variable x as the number of working sparkplugs, then x=6, n=6 and p, the probability of a single battery still functioning at 45,000 miles, is given in part (a) as 0.9938. We use the pdf function in Matlab because the binomial distribution is discrete and a discrete pdf gives the probability that x is equal to a given value.

$$P(x = 6)$$

```
>> p = pdf('binomial',6,6,0.9938)

p = 0.963371855549593
```

There is a 96.34% probability that all six sparkplugs are functioning at 45,000 miles.

(c) What is the probability that a sparkplug dies before 45,000 miles?

$$P(x \leq 45{,}000) = 1 - P(x \geq 45{,}000)$$

We calculated the probability of a lifetime beyond 45,000 miles in part (a). The probability of dying before 45,000 miles is one minus the answer in part (a). Alternatively, we can just use the cdf function.

```
>> p = cdf('normal',45000,50000,2000)

p =   0.006209665325776
```

There is a 0.628% chance that a sparkplug dies before 45,000 miles.

(d) What is the probability that no more than half the sparkplugs have died by 45,000 miles?

Half the six sparkplugs is three. No more than half means less than or equal to three. In this case, we can again use the binomial distribution but we can define the random variable to be the number of sparkplugs that have died.

$$P(x \leq 3)$$

```
>> p = cdf('binomial',3,6,0.0062)

p = 0.999999978054800
```

There is a very high probability that no more than three sparkplugs have died by 45,000 miles.