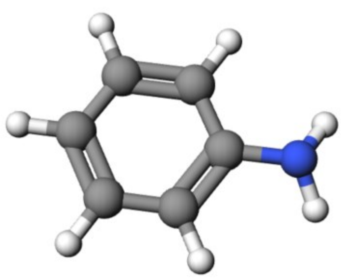
Exam II Solutions
Administered: Monday, October 11, 2024
24 points

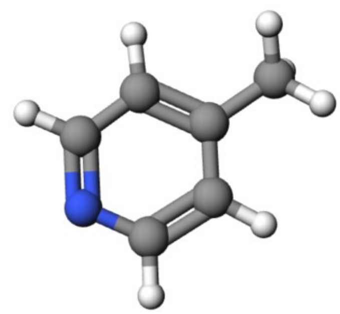
For each problem part: 0 points if not attempted or no work shown,
1 point for partial credit, if work is shown,
2 points for correct numerical value of solution

Problem 1. (16 points) Consider the following data for the standard enthalpy of formation for two organic chemicals with the same chemical stoichiometry.

Aniline
C_6H_7N



4-methyl-Pyridine
C_6H_7N



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

Standard Enthalpy of Formation of Aniline

$\Delta_f H°_{gas}$ (kJ/mol)	Reference
87.03	Hatton, Hildenbrand, et al., 1962
82.4	Vriens and Hill, 1952
83.2	Cole and Gilbert, 1951
81.0	Anderson and Gilbert, 1942
85.4	Lemoult, 1907

Standard Enthalpy of Formation of 4-methyl-Pyridine

$\Delta_f H°_{gas}$ (kJ/mol)	Reference
59.20	Good, 1972
56.78	Cox, Challoner, et al., 1954
48.07	Constam and White, 1903

Perform the following tasks.
(a) Determine the sample mean of the standard enthalpy of formation of aniline.
(b) Determine the sample mean of the standard enthalpy of formation of 4-methyl-pyridine.
(c) Determine the sample variance of the standard enthalpy of formation of aniline.
(d) Determine the sample variance of the standard enthalpy of formation of 4-methyl-pyridine.
(e) Identify the appropriate distribution to describe the difference of means in this case?

(f) Determine the lower limit of a 95% confidence interval on the difference of means of the standard enthalpy of formation.
(g) Determine the upper limit of a 95% confidence interval on the difference of means of the standard enthalpy of formation.
(h) Explain your findings in language a non-statistician can understand.

Solution:

(a) Determine the sample mean of the standard enthalpy of formation of aniline.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 83.8\ \frac{kJ}{mol}$$

Based on this data set sample mean of the standard enthalpy of formation of aniline is 83.8 $\frac{kJ}{mol}$.

(b) Determine the sample mean of the standard enthalpy of formation of 4-methyl-pyridine.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 54.7\ \frac{kJ}{mol}$$

Based on this data set sample mean of the standard enthalpy of formation of 4-methyl-pyridine is 54.7 $\frac{kJ}{mol}$.

(c) Determine the sample variance of the standard enthalpy of formation of aniline.

$$s^2 = \frac{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)} = 5.79\ \frac{kJ^2}{mol^2}$$

Based on this data set the sample variance of the standard enthalpy of formation of aniline is 5.79 $\frac{kJ^2}{mol^2}$.

(d) Determine the sample variance of the standard enthalpy of formation of 4-methyl-pyridine.

$$s^2 = \frac{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)} = 34.3\ \frac{kJ^2}{mol^2}$$

Based on this data set the sample variance of the standard enthalpy of formation of 4-methyl-pyridine is 34.3 $\frac{kJ^2}{mol^2}$.

(e) Identify the appropriate distribution to describe the difference of the mean of the standard enthalpy of formation 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 95% confidence interval on the difference of means of the standard enthalpy of formation.
(g) Determine the upper limit of a 95% confidence interval on the difference of means of the standard enthalpy of formation.

$$C.I. = 1 - 2\alpha = 0.95$$

$$\alpha = \frac{1 - C.I.}{2} = \frac{1 - 0.95}{2} = 0.025$$

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

$$v = 2.41 \sim 2$$

The limits on the t-distribution are given by

```
>> tlo = icdf('t',0.025,2)

tlo = -4.3027
```

and for the upper limit

```
>> thi = icdf('t',0.975,2)

thi =  4.3027
```

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

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

$$P[13.86 < (\mu_1 - \mu_2) < 44.38] = 0.95$$

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

We are 95% confidence that the difference between the standard enthalpy of formation of aniline and 4-methyl-pyridine lies within the range from 13.9 to 44.4 kJ/mol. The sample mean of the standard enthalpy of formation is larger for aniline than for 4-methyl-pyridine. This confidence interval on the difference includes strictly positive numbers, indicating that all differences within the 95% confidence level show a lower standard enthalpy of formation for 4-methyl-pyridine relative to aniline.

Problem 2. (8 points)
A couple in a doomsday cult has decided that the collapse of civilization is imminent and that humanity will not return to its current level of technology for a thousand years. They wish to record their thoughts for posterity and have chosen archival "M-Disc" optical drives as the most reliable means of information storage for a millenium. Manufacturers of the M-Disc state that the drive can last "up to a thousand years". Because no one has tested the technology for a thousand years, lifetime projections are based on theory. Suppose that M-discs have a lifetime that follows the normal distribution with a mean lifetime of 900 years and a standard deviation of 100 years. They make ten copies of their most important data.
(a) What is the probability that a single M-Disc remains readable after 1,000 years?
(b) What is the probability that at least one of the ten M-Discs is still readable after 1,000 years?
(c) Recalculate the answer to part (b) if the couple makes twenty copies of their most important data. What is the probability that at least one of the twenty M-Discs is still readable after 1,000 years?
(d) If the couple wants a 99% probability that at least one of their drives is readable after 1,000 years, how many copies should they make?

Solution:

(a) What is the probability that a single M-Disc remains readable after 1,000 years?

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

We use the cdf function in MATLAB:

```
>> p = 1 - cdf('normal',1000,900,100)

p =  0.1587
```

There is a 15.87% chance that a single M-Disc remains readable after 1,000 years.

(b) What is the probability that at least one of the ten M-Discs is still readable after 1,000 years?

This problem requires the binomial distribution. If we define the random variable x as the number of working drives, then $x \geq 1$, n=10 and p, the probability of a single drives still functioning at 1,000 years, is given in part (a) as 0.1587. We use the cdf function in Matlab.

$$P(x \geq 1) = 1 - P(x < 1) = 1 - P(x \leq 0)$$

```
>> p = 1 - cdf('binomial',0,10,0.1587)

p = 0.822373037532886
```

There is a 82.24% probability that at least one of the ten M-Discs is still readable after 1,000 years.

(c) Recalculate the answer to part (b) if the couple makes twenty copies of their most important data. What is the probability that at least one of the twenty M-Discs is still readable after 1,000 years?

We just rework part (b) with n = 20.

$$P(x \geq 1) = 1 - P(x < 1) = 1 - P(x \leq 0)$$

```
>> p = 1 - cdf('binomial',0,20,0.1587)

p = 0.9684
```

There is a 96.84% probability that at least one of the twenty M-Discs is still readable after 1,000 years.

(d) If the couple wants a 99% probability that at least one of their drives is readable after 1,000 years, how many copies should they make?

To answer this question, we can simply repeat part (b) increasing the number of trials until the probability is greater than 0.99. To do this, I wrote a little script in Matlab that looped from 1 to 40 copies and printed the probability of each at the end.

```
>> for i = 1:1:40; p(i) = 1 - cdf('binomial',0,i,0.1587); end; p'
```

Here is the output. I highlighted the first entry greater than 99%.

```
>> p =    0.158700000000000
   0.292214310000000
   0.404539899003000
```

```
0.499039417031224
0.578541861548369
0.645427268120643
0.701697960669897
0.749038494311584
0.788866085264336
0.822373037532886
0.850562436476417
0.874278177807609
0.894230230989542
0.911015893331502
0.925137671059792
0.937018322662603
0.947013514856048
0.955422470048393
0.962496924051713
0.968448662204706
0.973455859512819
0.977668414608135
0.981212437209824
0.984194023424625
0.986702431907137
0.988812755963474
0.990588171592071
0.992081828760409
0.993338442536132
0.994395631705648
0.995285044953962
0.996033308319768
0.996662822289421
0.997192432392090
0.997637993371465
0.998012843823414
0.998328205508638
0.998593519294417
0.998816727782393
0.999004513083327
```

At twenty seven copies, the probability first exceeds 99%. We can explicitly check this using the same command as in part (b).

```
>> p = 1 - cdf('binomial',0,27,0.1587)

p =   0.990588171592071
```

If the couple makes 27 copies of their data, there is a 99.06% probability that at least one will be readable in a thousand years.