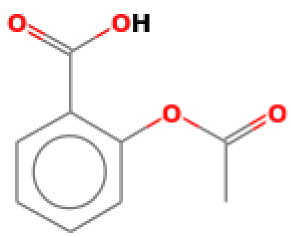
Exam II Solutions
Administered: Monday, October 17, 2022
22 points

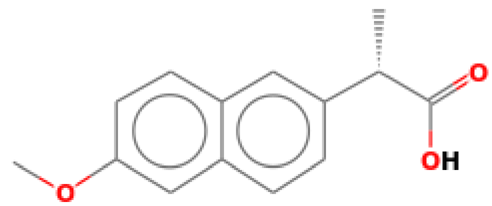
For each problem part: 0 points if not attempted or no work shown,
1 point for partial credit, if work is shown,
2 points for correct numerical value of solution

Problem 1. (16 points) Consider the following data for the enthalpy of fusion for two biochemicals.

acetylsalicylic acid (aspirin)
$C_9H_8O_4$



naproxen (Aleve)
$C_{14}H_{14}O_3$



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

Enthalpy of fusion of aspirin

$\Delta_{fus}H$ (kJ/mol)	Temperature (K)	Method	Reference
29.17	409.2	DSC	Xu, Sun, et al., 2004
31.01	412.7	DSC	Perlovich and Bauer-Brandl, 2001
29.8	414.	N/A	Kirklin, 2000

Enthalpy of fusion of naproxen

$\Delta_{fus}H$ (kJ/mol)	Temperature (K)	Method	Reference
34.2	428.8	DSC	Wassvik, Holmén, et al., 2006
31.5	428.5	N/A	Neau, Bhandarkar, et al., 1997
29.41	439.2	N/A	Claramonte, Vialard, et al., 1993

Perform the following tasks.
(a) Determine the sample mean of the enthalpy of fusion of aspirin.
(b) Determine the sample mean of the enthalpy of fusion of naproxen.
(c) Determine the sample variance of the enthalpy of fusion of aspirin.
(d) Determine the sample variance of the enthalpy of fusion of naproxen.
(e) Identify the appropriate distribution to describe the difference of means in this case?
(f) Determine the lower limit of a 90% confidence interval on the difference of means of the enthalpy of fusion.
(g) Determine the upper limit of a 90% confidence interval on the difference of means of the enthalpy of fusion.
(h) Explain your findings in language a non-statistician can understand.

Solution:

(a) Determine the sample mean of the enthalpy of fusion of aspirin.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 29.993 \ \frac{kJ}{mol}$$

Based on this data set sample mean of the enthalpy of fusion of aspirin is 29.9 kJ/mol.

(b) Determine the sample mean of the enthalpy of fusion of naproxen.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 31.703 \ \frac{kJ}{mol}$$

Based on this data set sample mean of the enthalpy of fusion of naproxen is 31.7 kJ/mol.

(c) Determine the sample variance of the enthalpy of fusion of aspirin.

$$s^2 = \frac{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)} = 0.8744 \left(\frac{kJ}{mol}\right)^2$$

Based on this data set the sample variance of the enthalpy of fusion of aspirin is 0.87 $(\text{kJ/mol})^2$.

(d) Determine the sample variance of the enthalpy of fusion of naproxen.

$$s^2 = \frac{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)} = 5.7670 \left(\frac{kJ}{mol}\right)^2$$

Based on this data set the sample variance of the enthalpy of fusion of naproxen is 5.77 $(\text{kJ/mol})^2$.

(e) Identify the appropriate distribution to describe the mean of the enthalpy of fusion 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 90% confidence interval on the difference of means of the enthalpy of fusion.
(g) Determine the upper limit of a 90% confidence interval on the difference of means of the enthalpy of fusion.

$$C.I. = 1 - 2\alpha = 0.90$$

$$\alpha = \frac{1 - C.I.}{2} = \frac{1 - 0.90}{2} = 0.05$$

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

$$v = 2.59 \sim 3$$

The limits on the t-distribution are given by

```
>> tlo = icdf('t',0.05,3)
```

```
tlo = -2.353363434801824
```

and for the upper limit

```
>> thi = icdf('t',0.95,3)

t =  2.353363434801823
```

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

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

$$P[-5.212 < (\mu_1 - \mu_2) < 1.792] = 0.90$$

(h) Translate your conclusions from (d) and (e) into a sentence that a non-statistician can understand.

We are 90% confidence that the difference between the enthalpy of fusion of aspirin and naproxen lies within the range from -5.212 to 1.792 kJ/mol. While the sample mean of the enthalpy of fusion is larger for aspirin than naproxen, this confidence interval includes both positive and negative numbers, indicating a lack of confidence at this level that the enthalpy of fusion of aspirin is actually greater than that of naproxen.

Problem 2. (6 points)
Consider a battery with a lifetime that follows the normal distribution with a mean lifetime of 18 months and a standard deviation of 2 months.
(a) What is the probability that a battery lasts at least 15 months?
(b) If you want to be 99.9% sure that the battery doesn't die, when should you replace the battery?
(c) If you have a system with a back-up battery, what is the probability that at least 1 battery continues to work in the time period determined in part (b)?

Solution:

(a) What is the probability that a battery lasts at least 15 months?

$$P(x \geq 15) = 1 - P(x \leq 15.0)$$

We use the cdf function in MATLAB:

```
>> p = 1 - cdf('normal',15.0,18.0,2.0)

p =   0.933192798731142
```

There is a 93.3% chance that a battery lasts at least 15.0 months.

(b) If you want to be 99.9% sure that the battery doesn't die, when should you replace the battery?

$$1 - P(x \leq x_{hi}) = 0.999$$

$$P(x \leq x_{hi}) = 0.001$$

```
>> xhi = icdf('normal',0.001,18.0,2.0)

xhi =  11.819535387664374
```

There is a 0.1% probability that a battery dies before 11.82 months. Therefore, if you change the battery at 11.82 months, there is 99.9% that you never lost power.

(c) If you have a system with a back-up battery, what is the probability that at least 1 battery continues to work in the time period determined in part (b)?

This problem requires the binomial distribution. If we define the random variable x as the number of working batteries, then $x \geq 1$, n=2 and p, the probability of a single battery still functioning until 11.82 months, given in part (b) as 0.999. We use the cdf function in Matlab because the binomial distribution is discrete and a discrete pdf gives the probability that x is less than or equal to a given value.

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

```
>> p = 1 - cdf('binomial',0,2,0.999)

p =   0.999999000000000
```

If you have a back-up battery, then you are 99.9999% sure that you don't lose power in 11.82 months. There is a one in million likelihood that you lose power.