# 2 Calculations Exercises

## General instructions.

There are 5 exercises below, and you will be required to provide five solutions, each worth ten points. You have three options for completing the exercises.

1. In the past, I've required that R and SAS be used for at least one solution each. If you wish to delevop skill in both languages, you can use this option. You can provide both R and SAS solutions for the same exercises (e.g. R and SAS code for exercise 1), and then provide three solutions by choosing among the remaining four exercises (e.g. 2,4,5) in the language of your choice.
2. We have effectively two cohorts in these classes (600, 601, 602). One cohort is expected to learn SAS, one will not be using SAS in their program. Thus, for this summer, I will not require solutions in both languages. You can use either language to complete all the exercises. You will be *allowed* to submit homework as in past summers (point 1), but not *required*.
3. Last summer I allowed that one solution could be implemented in Python. I will continue that practice this summer. You can embed Python code in R Markdown, using the syntax `{python}` instead of `{r}`. If you choose this, you will be expected to comment on the differences or similarities between R/SAS and Python, and I would prefer your Python solution to be included in R Markdown. I won't be teaching Python this summer, but if you're familiar with Python, this may help understand the inner workings of R or SAS.

## Exercise 1

Cohen gives a formula for effect size, $d$, for the difference between two means $m_1$ and $m_2$, as

$$d = \frac{|m_1 - m_2|}{s_{pooled}}$$

where $s_{pooled}$ is a pooled standard deviation. Use the formula

$$s_{pooled} = \sqrt{(s_1^2 + s_2^2)/2}$$

Calculate the effect size $d$ for the differences among calories per serving,

- 1936 versus 2006
- 1936 versus 1997
- 1997 versus 2006

Use the values from Wansink, Table 1 as given in Homework 1 or in the course outline.

### Answer

Enter the R code in the chunks below. If you choose SAS for this exercise, use the marked portion in the SAS homework template.

#### 1936 versus 2006

```
# code for 1936 versus 2006 here
```

#### 1936 versus 1997

```
# code for 1936 versus 1997 here
```

#### 1997 versus 2006

```
# code for 1997 versus 2006 here
```

Cohen recommends that $d = 0.2$ be considered a small effect, $d = 0.5$ a medium effect and $d = 0.8$ a large effect. Should any of the differences be considered *large*?

## Exercise 2.

Suppose you are planning an experiment and you want to determine how many observations you should make for each experimental condition. One simple formula (see Kuehl, "Design of Experiments : Statistical Principles of Research Design and Analysis") for the required replicates $n$ is given by

$$n \geq 2 \times \left( \frac{CV}{\%Diff} \right)^2 \times \left( z_{\alpha/2} + z_\beta \right)^2$$

where

$$\begin{aligned} \%Diff &= \frac{m_1 - m_2}{(m_1 + m_2)/2} \\ CV &= \frac{s_{pooled}}{(m_1 + m_2)/2} \end{aligned} \tag{1}$$

and $z$ are quantiles from the normal distribution with $\mu = 0$ and $\sigma^2 = 1$.

Use this formula to calculate the number of replicates required to detect differences between calories per serving,

- 1936 versus 2006
- 1936 versus 1997
- 1997 versus 2006

You will need to research how to use the normal distribution functions (`*norm` in R, ). Use $\alpha = 0.05$ and $\beta = 0.2$ for probabilities, and let `mean = 0` and `sd = 1` (both $z$ should be positive).

Since $n$ must be an integer, you will need to round *up*. Look up the built in functions for this.

### Answer

Enter the R code in the chunks below. If you choose SAS for this exercise, use the marked portion in the SAS homework template.

#### 1936 versus 2006

**1936 versus 1997**

**1997 versus 2006**

To check your work, use the rule of thumb suggested by van Belle ("Statistical Rules of Thumb"), where

$$n = \frac{16}{\Delta^2}$$

with

$$\Delta = \frac{\mu_1 - \mu_2}{\sigma}$$

.

How does this compare with your results? Why does this rule of thumb work? How good is this rule of thumb?

A comment about the notation. When planning for experiments, we can assume known parameters (i.e. $\sigma^2$), but when we plan for experiments using the results of past experiments, we can use the corresponding estimates (i.e. $s^2$).

# Exercise 3

The probablity of an observation $x$, when taken from a normal population with mean $\mu$ and variance $\sigma^2$ is calculated by

$$L(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

For values of $x = \{-0.1, 0.0, 0.1\}$, write code to calculate $L(x; \mu = 0, \sigma = 1)$.

## Answer

Enter the R code in the chunks below. If you choose SAS for this exercise, use the marked portion in the SAS homework template.

$x = -0.1$

$x = 0.0$

$x = 0.1$

You can confirm your results using the built in normal distribution function. Look up `dnorm` in R help and use the same values for `x`, `mean` and `sigma` as above. You should get matching results to at least 12 decimal places.

```
#compare with dnorm
```

# Exercise 4

## Part a

Write code to compute

$$7 - 1 \times 0 + 3 \div 3$$

Type this in verbatim, using only numbers, `+`,`-`,`*` and `/`, with no parenthesis. Do you agree with the result? Explain why, one or two sentences.

#### Answer

## Part b

According to "Why Did 74% of Facebook Users Get This Wrong?" (https://profpete.com/blog/2012/11/04/why-did-74-of-facebook-users-get-this-wrong/), most people would compute the result as 1. Use parenthesis `( )` to produce this result.

#### Answer

## Part c

Several respondents to the survey cited in Part 2 gave the answer 6. Add *one* set of parenthesis to produce this result.

#### Answer

# Exercise 5.

### Part a

Quoting from Wansink and Payne

> Because of changes in ingredients, the mean average calories in a recipe increased by 928.1 (from 2123.8 calories … to 3051.9 calories … ), representing a 43.7% increase.

Show how 43.7% is calculated from 2123.8 and 3051.9, and confirm W&P result.

#### Answer

> The resulting increase of 168.8 calories (from 268.1 calories … to 436.9 calories … ) represents a 63.0% increase … in calories per serving.

### Part b

Repeat the calculations from above and confirm the reported 63.0% increase in calories per serving. Why is there such a difference between the change in calories per recipe and in calories per serving?

**Answer**

**Part c**

Calculate an `average_calories_per_serving` by dividing `average_calories_per_recipe` by `average_servings_per_reci`
for years 1936 and 2006, then calculate a percent increase. Which of the two reported increases (a or b) are consistent with this result?

**Answer**