Homework 9
MTH 3220, Fall 2019
Can be handed in Tue., Dec. 3 (after Fall Break)

For any problems that involve computations, you must **show your work** to receive full credit.

Section in Book	**Problems**
14.1	2, 7, 11*
14.3	Problem 1 (below), 27, 31

* The data for **Problem 11** are in the file **ex_14_11.txt**. For **Part *a***, the R function **qnorm()** will return the normal distribution quantiles (percentiles) used to define the class intervals. For example, the values returned by the following commands will divide the N(0, 1) distribution into equi-probable intervals each having probability 1/6:

```
> qnorm(p = 1/6, mean = 0, sd = 1)
> qnorm(p = 2/6, mean = 0, sd = 1)
  ⋮
> qnorm(p = 5/6, mean = 0, sd = 1)
```

For **Part *b***, the quantiles are obtained from:

```
> qnorm(p = 1/6, mean = 0.5, sd = 0.002)
> qnorm(p = 2/6, mean = 0.5, sd = 0.002)
  ⋮
> qnorm(p = 5/6, mean = 0.5, sd = 0.002)
```

For **Part *c***, the function **sort()** will make it easier to tally up how many observations fall into each of the intervals from **Part *b***, for example:

```
> sort (my.data$x)
```

Additional Problem

1. A study from the University of Texas Medical Center examined whether the risk of **hepatitis C** was related to whether or not people had **tattoos**. A sample of ***n* = 626** patients being treated for non-blood-related disorders was tested for hepatitis C and asked whether they had any tattoos. The counts are shown below.*

		Hepatitis C		
		Yes	No	Total
Tattoo	Yes	25	88	113
	No	22	491	513
	Total	47	579	626

a) The ***estimated conditional probability*** that a patient will have hepatitis C, given that he or she has a tattoo, is the fraction of individuals that have hepatitis C among those that have tattoos, i.e.

$$\text{Estimated conditional probability} = \frac{n_{11}}{n_{1\cdot}}.$$

The ***estimated conditional probability*** that a patient will have hepatitis C, given that he or she *doesn't* have a tattoo, is the fraction of individuals that have hepatitis C among those who *don't* have tattoos, i.e.

$$\text{Estimated conditional probability} = \frac{n_{21}}{n_{2\cdot}}.$$

Compute these two **estimated conditional probabilities**. Which one is higher?

b) The ***estimated marginal column probabilities***, denoted $\hat{p}_{\cdot 1}$ and $\hat{p}_{\cdot 2}$, are defined as

$$\hat{p}_{\cdot 1} = \frac{n_{\cdot 1}}{n} \qquad \text{and} \qquad \hat{p}_{\cdot 2} = \frac{n_{\cdot 2}}{n}$$

Thus $\hat{p}_{\cdot 1}$ is the fraction of individuals that have hepatitis C in the entire sample, and $\hat{p}_{\cdot 2}$ is the fraction that *don't* have hepatitis C.

Compute $\hat{p}_{\cdot 1}$ and $\hat{p}_{\cdot 2}$. Which one is larger?

c) Carry out a **chi-squared test for independence** between **hepatitis C** status and **tattoo** presence. Use a level of significance $\alpha = 0.05$.

d) Verify that the sample size $n = 626$ is large enough to justify the use of the chi-squared test: The **estimated expected counts**, $\hat{e}_{11}$, $\hat{e}_{12}$, $\hat{e}_{21}$, and $\hat{e}_{22}$

$$\hat{e}_{ij} = n\hat{p}_{i\cdot}\hat{p}_{\cdot j} = \frac{n_{i\cdot}n_{\cdot j}}{n}$$

should all be **at least five**.

*Source: De Veaux, Velleman, and Bock