Chapter 3. Sample Theory

In inferential Statistics, we will have the following situation: we are interested in studying a characteristic (a random variable) X, relative to a population P of (known or unknown) size N. The difficulty or even the impossibility of studying the entire population, as well as the merits of choosing and studying a random sample from which to make inferences about the population of interest, have already been discussed in the previous chapter. Now, we want to give a more rigorous and precise definition of a random sample, in the framework of random variables, one that can then employ probability theory techniques for making inferences.

1 Random Samples and Sample Functions

We choose n objects from the population and actually study $X_i,\ i = \overline{1,n}$, the characteristic of interest *for the i^{th} object selected*. Since the n objects were randomly selected, it makes sense that for $i = \overline{1,n}$, X_i is a random variable, one that has *the same* distribution (pdf) as X, the characteristic relative to the entire population. Furthermore, these random variables are independent, since the value assumed by one of them has no effect on the values assumed by the others. Once the n objects have been selected, we will have n numerical values available, $x_1, \ldots, x_n$, the observed values of $X_1, \ldots, X_n$.

Definition 1.1. *A **random sample of size n** from the distribution of X, a characteristic relative to a population P, is a collection of n independent random variables $X_1, \ldots, X_n$, having the same distribution as X. The variables $X_1, \ldots, X_n$, are called **sample variables** and their observed values $x_1, \ldots, x_n$, are called **sample data**.*

Remark 1.2. The term *random sample* may refer to the objects selected, to the sample variables, or to the sample data. It is usually clear from the context which meaning is intended. In general, we use capital letters to denote sample variables and corresponding lowercase letters for their observed values, the sample data.

We are able now to define sample functions, or statistics, in the more precise context of random variables.

Definition 1.3. *A **sample function** or **statistic** is a random variable*

$$Y_n = h_n(X_1, \ldots, X_n),$$

where $h_n : \mathbb{R}^n \to \mathbb{R}$ *is a measurable function. The value of the sample function* Y_n *is* $y_n = h_n(x_1, \ldots, x_n)$.

We will revisit now some sample numerical characteristics discussed in the previous chapter and define them as sample functions. That means they will have a pdf, a cdf, a mean value, variance, standard deviation, etc. A sample function will, in general, be an approximation for the corresponding population characteristic. In that context, the standard deviation of the sample function is usually referred to as the **standard error**.

In what follows, $\{X_1, \ldots, X_n\}$ denotes a sample of size n drawn from the distribution of some population characteristic X.

2 Sample Mean

Definition 2.1. *The* ***sample mean*** *is the sample function defined by*

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{2.1}$$

and its value is $\overline{x}_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i$.

Now that the sample mean is defined as a random variable, we can discuss its distribution and its numerical characteristics.

Proposition 2.2. *Let* X *be a characteristic with* $E(X) = \mu$ *and* $V(X) = \sigma^2$. *Then*

$$E\left(\overline{X}\right) = \mu \ \text{ and } \ V\left(\overline{X}\right) = \frac{\sigma^2}{n}. \tag{2.2}$$

Moreover, if $X \in N(\mu, \sigma)$, *then* $\overline{X} \in N\left(\mu, \dfrac{\sigma}{\sqrt{n}}\right)$.

Proof. Since $X_1, \ldots, X_n$ are identically distributed, with the same distribution as X, $E(X_i) = E(X) = \mu$ and $V(X_i) = V(X) = \sigma^2$, $\forall i = \overline{1, n}$. Then, by the usual properties of expectation, we have

$$E\left(\overline{X}\right) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\, n\mu = \mu.$$

Further, since $X_1, \ldots, X_n$ are also independent, by the properties of variance, it follows that

$$V\left(\overline{X}\right) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) = \frac{1}{n^2}\, n\sigma^2 = \frac{\sigma^2}{n}.$$

The last part follows from the fact that $\overline{X}$ is a linear combination of independent, normally distributed random variables. □

Remark 2.3. As a consequence, the standard deviation of $\overline{X}$ is

$$\mathrm{Std}(\overline{X}) \;=\; \sqrt{V(\overline{X})} \;=\; \frac{\sigma}{\sqrt{n}}.$$

So, when estimating the population mean μ from a sample of size n by the sample mean $\overline{X}$, the *standard error* of the estimate is $\sigma/\sqrt{n}$, which oftentimes is estimated by $s/\sqrt{n}$. Either way, notice that as n increases and tends to ∞, the standard error decreases and approaches 0. That means that the larger the sample on which we base our estimate, the more accurate the approximation.

Corollary 2.4. *Let X be a characteristic with $E(X) = \mu$ and $V(X) = \sigma^2$ and for $n \in \mathbb{N}$ let*

$$Z_n = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}.$$

Then the variable Z_n converges in distribution to a Standard Normal variable, as $n \to \infty$, i.e. $F_{Z_n} \overset{n\to\infty}{\longrightarrow} F_Z = \Phi$.
Moreover, if $X \in N(\mu, \sigma)$, then the statement is true for every $n \in \mathbb{N}$.

Proof. This is a direct consequence of the Central Limit Theorem (CLT).

□

3 Sample Moments and Sample Variance

Definition 3.1. *The statistic*

$$\overline{\nu}_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \tag{3.3}$$

is called the ***sample moment of order k*** *and its value is* $\frac{1}{n}\sum_{i=1}^{n} x_i^k$.

The statistic

$$\overline{\mu}_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^k \tag{3.4}$$

is called the ***sample central moment of order k*** *and its value is* $\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^k$.

Remark 3.2. Just like for theoretical (population) moments, we have

$$\begin{aligned} \overline{\nu}_1 &= \overline{X}, \\ \overline{\mu}_1 &= 0, \\ \overline{\mu}_2 &= \overline{\nu}_2 - \overline{\nu}_1^2. \end{aligned}$$

Next we discuss the distributions and characteristics of these new sample functions.

Proposition 3.3. *Let X be a characteristic with the property that for $k \in \mathbb{N}$, the theoretical moment $\nu_{2k} = \nu_{2k}(X) = E\left(X^{2k}\right)$ exists. Then*

$$E\left(\overline{\nu}_k\right) = \nu_k \text{ and } V\left(\overline{\nu}_k\right) = \frac{1}{n}\left(\nu_{2k} - \nu_k^2\right). \tag{3.5}$$

Proof. First off, the condition that ν_{2k} exists for X ensures the fact that all theoretical moments of X of order up to k also exist. The rest follows as before. We have

$$E\left(\overline{\nu}_k\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i^k) = \frac{1}{n}\sum_{i=1}^{n} E(X^k) = \frac{1}{n}\, n\nu_k = \nu_k$$

and

$$\begin{aligned} V\left(\overline{\nu}_k\right) &= \frac{1}{n^2}\sum_{i=1}^{n} V(X_i^k) &&= \frac{1}{n^2}\sum_{i=1}^{n} V(X^k) \\ &= \frac{1}{n^2}\, n\left(\nu_{2k} - \nu_k^2\right) &&= \frac{1}{n}\left(\nu_{2k} - \nu_k^2\right). \end{aligned}$$

□

Corollary 3.4. *Let X be a characteristic satisfying the hypothesis of Proposition 3.3 and for $n \in \mathbb{N}$*

let

$$Z_n = \frac{\overline{\nu}_k - \nu_k}{\sqrt{\frac{\nu_{2k} - \nu_k^2}{n}}}.$$

Then $Z_n \xrightarrow{d} Z$, *as* $n \to \infty$.

Proposition 3.5. *Let* X *be a characteristic with* $V(X) = \mu_2 = \sigma^2$ *and for which the theoretical moment* $\nu_4 = E(X^4)$ *exists. Then*

$$\begin{aligned} E(\overline{\mu}_2) &= \frac{n-1}{n}\,\sigma^2, \\ V(\overline{\mu}_2) &= \frac{n-1}{n^3}\Big[(n-1)\mu_4 - (n-3)\sigma^4\Big], \\ \operatorname{cov}(\overline{X}, \overline{\mu}_2) &= \frac{n-1}{n^2}\,\mu_3. \end{aligned} \tag{3.6}$$

Proof. We only prove the first assertion, as it is the most important and oftenly used property of $\overline{\mu}_2$. Using Proposition 3.3, Remark 3.2, the properties of expectation and the fact that $X_1, \ldots, X_n$ are independent and identically distributed, we have

$$\begin{aligned} E(\overline{\mu}_2) &= E(\overline{\nu}_2) - E\left(\overline{\nu}_1^2\right) = \nu_2 - E\left(\left(\frac{1}{n}\sum_{i=1}^n X_i\right)^2\right) \\ &= \nu_2 - \frac{1}{n^2} E\Big(\sum_{i=1}^n X_i^2 + 2\sum_{i<j} X_i X_j\Big) \\ &= \nu_2 - \frac{1}{n^2}\left[\sum_{i=1}^n E\left(X_i^2\right) + 2\sum_{i<j} E(X_i)E(X_j)\right] \\ &= \nu_2 - \frac{1}{n^2}\left[n\nu_2 + 2\,\frac{n(n-1)}{2}\,\nu_1^2\right] = \nu_2 - \frac{1}{n}\nu_2 - \frac{n-1}{n}\nu_1^2 \\ &= \frac{n-1}{n}\left(\nu_2 - \nu_1^2\right) = \frac{n-1}{n}\,\sigma^2. \end{aligned}$$

□

Remark 3.6.

1. For large samples, i.e. when $n \to \infty$, $\overline{X}$ and $\overline{\mu}_2$ are uncorrelated.
2. If X has a symmetric distribution, then $\mu_3 = 0$ and, hence, $\overline{X}$ and $\overline{\mu}_2$ are uncorrelated for every $n \in \mathbb{N}$.

3. As before, one can show that under the assumptions of Proposition 3.5, the sequence

$$Z_n = \frac{\overline{\mu}_2 - \sigma^2}{\sqrt{\frac{\mu_4 - \sigma^4}{n}}}$$

converges in distribution to a Standard Normal variable, as $n \to \infty$.
4. Notice that the sample central moment of order 2 is the first statistic whose expected value *is not* the corresponding population function, in this case the theoretical variance. This is the motivation for the next definition.

Definition 3.7. *The statistic*

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2 \tag{3.7}$$

*is called the **sample variance** and its value is* $\dfrac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2$.
The statistic $s = \sqrt{s^2}$ *is called the **sample standard deviation**.*

Remark 3.8. Notice that the sample central moment of order 2 is no longer equal to the sample variance, as we are used. In fact, we have

$$s^2 = \frac{n}{n-1}\, \overline{\mu}_2.$$

Then, by Proposition 3.5, we have for the sample variance

$$\begin{aligned} E\left(s^2\right) &= \mu_2 = \sigma^2, \\ V\left(s^2\right) &= \frac{1}{n(n-1)} \Big[(n-1)\mu_4 - (n-3)\sigma^4 \Big], \\ \operatorname{cov}(\overline{X}, s^2) &= \frac{1}{n}\, \mu_3. \end{aligned} \tag{3.8}$$

4 Sample Proportions

Definition 4.1. *Assume a subpopulation A of a population consists of items that have a certain attribute. The **population proportion** is then the probability*

$$p = P(i \in A), \tag{4.1}$$

i.e. the probability for a randomly selected item i to have this attribute.
*The **sample proportion** is*

$$\overline{p} \;=\; \frac{\text{number of sampled items from } A}{n} \tag{4.2}$$

Proposition 4.2. *Let p be a population proportion. Then*

$$E(\overline{p}) = p, \;\; V(\overline{p}) = \frac{p(1-p)}{n} = \frac{pq}{n} \;\; \textit{and} \;\; \sigma(\overline{p}) = \sqrt{\frac{pq}{n}}, \tag{4.3}$$

where $q = 1 - p$.

Proof. We use the indicator random variable

$$X_i \;=\; \begin{cases} 1, & i \in A \\ 0, & i \notin A \end{cases} .$$

Then $X_i \in Bern(p)$ and, so, we know that $E(X_i) = p$ and $V(X_i) = pq$, for every $i = 1, \ldots, n$. But notice that $\overline{p} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$, i.e. the sample mean of the sample $X_1, \ldots, X_n$. Thus, by Proposition 2.2,

$$\begin{aligned} E(\overline{p}) &= p, \\ V(\overline{p}) &= \frac{pq}{n}, \\ \sigma(\overline{p}) &= \sqrt{\frac{pq}{n}}. \end{aligned}$$

□

5 Sample Distribution Function

Thus far, we have been able to define sample functions that mimicked their theoretical correspondents (mean, moments, variance, proportion) and, hopefully, will provide good inferential estimates for the entire population. The ultimate goal of Statistics is to derive the probability distribution that generated a sample from the sample itself, i.e. to define a sample function that gives some information of the distribution function of a characteristic, relative to the entire population. The idea is suggested by the shape of the cumulative distribution function of discrete random variables.

Definition 5.1. *Let X be a characteristic and $X_1, \ldots, X_n$ sample variables for a random sample of size n. The* ***sample distribution function*** *(sdf) or* ***empirical distribution function*** *is the sample function $\overline{F}_n : \mathbb{R} \to \mathbb{R}$ defined by*

$$\overline{F}_n(x) = \frac{1}{n} I(X_i \leq x) = \frac{\text{card}\{X_i \mid X_i \leq x\}}{n}, \tag{5.1}$$

where $I(A)$ is the indicator of event A, and its value is $\dfrac{\text{card}\{x_i \mid x_i \leq x\}}{n}$.

Remark 5.2.

1. So, the sample distribution function at a value x is given by

$$\overline{F}_n(x) = \frac{\text{number of sample elements } x_i \leq x}{n}.$$

Defining it this way makes it an excellent tool for measuring how faithful the sample is to the probability distribution: where observations are densely packed, this function grows rapidly, which is exactly what is expected from the true distribution function (for where the distribution function grows rapidly, the probability density−its derivative, is large, which is propitious to a high concentration of observations), while observations that are few and far between happen in regions of low probability density.

2. Assuming the sample data $x_1, \ldots, x_n$ are sorted in increasing order, a more explicit computational formula for the sample distribution function is

$$\overline{F}_n(x) = \begin{cases} 0, & \text{if } x < x_1 \\ \dfrac{i}{n}, & \text{if } x_i \leq x < x_{i+1},\ i = \overline{1, n-1} \\ 1, & \text{if } x \geq x_n. \end{cases} .$$

Thus $\overline{F}_n$ presents similar properties to those of a cumulative distribution function of a discrete random variable:

- it is a step function;
- it monotonically increases from 0 to 1;
- it is constant on semi-open intervals $[x_i, x_{i+1})$;
- its limits at $\pm\infty$ are 1 and 0, respectively.

In addition, here the height of each "step" is $\frac{1}{n}$.

3. The sample distribution function can also be viewed as a random variable (since it *is* a sample function). If F denotes the cumulative distribution function of the characteristic X, then for each $x \in \mathbb{R}$, $\overline{F}_n(x)$ is a discrete random variable with pdf

$$\overline{F}_n(x) \begin{pmatrix} i/n \\ C_n^i \left(F(x)\right)^i \left(1 - F(x)\right)^{n-i} \end{pmatrix}_{i=\overline{0,n}}.$$

Now that we have seen the similarities between a cdf and a sdf, the question that naturally arises is how much does the latter resemble the former, how well and in what sense, does it approximate it. The answer is given below.

Theorem 5.3 (Glivenko-Cantelli)**.** *Let X be a characteristic with cdf F and $X_1, \ldots, X_n$ be sample variables for a random sample of size n, with sdf $\overline{F}_n$. Then*

$$P\left(\lim_{n\to\infty}\left(\sup_{x\in\mathbb{R}} |\overline{F}_n(x) - F(x)|\right) = 0\right) = 1, \tag{5.2}$$

i.e. the sample distribution function converges almost surely to the cumulative distribution function.

6 Sample Functions for Comparing Two Populations

It will be necessary sometimes to compare characteristics of two populations. For that, we will need results on sample functions referring to both collections. Assume we have two characteristics $X_{(1)}$ and $X_{(2)}$, relative to two populations. We draw from both populations independent random samples of sizes n_1 and n_2, respectively. Denote the two sets of random variables by

$$X_{11}, \ldots, X_{1n_1} \text{ and } X_{21}, \ldots, X_{2n_2}.$$

Then we have two sample means and two sample variances, given by

$$\overline{X}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} X_{1i}, \quad \overline{X}_2 = \frac{1}{n_2}\sum_{j=1}^{n_2} X_{2j}$$

and

$$s_1^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}\left(X_{1i} - \overline{X}_1\right)^2, \quad s_2^2 = \frac{1}{n_2 - 1}\sum_{j=1}^{n_2}\left(X_{2j} - \overline{X}_2\right)^2,$$

respectively. In addition, denote by

$$s_p^2 = \frac{\sum_{i=1}^{n_1} \left(X_{1i} - \overline{X}_1\right)^2 + \sum_{j=1}^{n_2} \left(X_{2j} - \overline{X}_2\right)^2}{n_1 + n_2 - 2} = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

the **pooled variance** of the two samples, i.e. a variance that considers the sample data from both samples.

In inferential Statistics, when comparing the means of two populations, we will look at their difference and try to estimate it. Regarding that, we have the following result.

Proposition 6.1. *Let $X_{(1)}, X_{(2)}$ be two population characteristics with means $E(X_{(i)}) = \mu_i$ and variances $V(X_{(i)}) = \sigma_i^2, i = 1, 2$. Then*

$$\begin{aligned} E\left(\overline{X}_1 - \overline{X}_2\right) &= \mu_1 - \mu_2, \\ V\left(\overline{X}_1 - \overline{X}_2\right) &= \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}. \end{aligned} \tag{6.1}$$

The proof is straightforward computational, similar to the proof of Proposition 2.2 and we skip it.

In a similar fashion, we can compare two population proportions. Again, the random variable of interest is their difference.

Proposition 6.2. *Assume we have two population proportions p_1 and p_2. From each population we draw independent samples of size n_1 and n_2, respectively, which yield the population proportions $\overline{p}_1$ and $\overline{p}_2$. Then*

$$\begin{aligned} E\left(\overline{p}_1 - \overline{p}_2\right) &= p_1 - p_2, \\ V\left(\overline{p}_1 - \overline{p}_2\right) &= \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}, \end{aligned} \tag{6.2}$$

with $q_i = 1 - p_1, i = 1, 2$.

Again, the proof is straightforward computational and we skip it.

7 Properties of Sample Functions

Before we state some properties that will be used later, let us review the notations of all the sample functions we discussed previously and their corresponding population characteristics.

So, as usually, let X be a characteristic of a population from which a random sample of size n is drawn and let $X_1, \ldots, X_n$ be the sample variables. We have the following correspondence:

Function	Population (theoretical)	Sample
Mean	$\mu = E(X)$	$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
Variance	$\sigma^2 = V(X)$	$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$
Standard deviation	$\sigma = \sqrt{V(X)}$	$s = \sqrt{s^2}$
Moment of order k	$\nu_k = E\left(X^k\right)$	$\overline{\nu}_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$
Central moment of order k	$\mu_k = E\left[(X - E(X))^k\right]$	$\overline{\mu}_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^k$
Proportion	$p = P(i \in A)$	$\overline{p} = \frac{\text{number of } X_i \text{ from } A}{n}$

Table 1: Notations

The following results will be used later on in Inferential Statistics. They are based either on properties of random variables or they are consequences of some Central Limit Theorem. We just state them without proof, so they can be referenced later.

Proposition 7.1. *Assume either that $X \in N(\mu, \sigma)$ or that the sample size is large enough and let*

$$Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}.$$

Then $Z \in N(0,1)$.

Proposition 7.2. *Assume either that* $X \in N(\mu, \sigma)$ *or that the sample size is large enough and let*

$$T = \frac{\overline{X} - \mu}{\frac{s}{\sqrt{n}}}.$$

Then $T \in T(n-1)$.

Proposition 7.3. *Assume that* $X \in N(\mu, \sigma)$ *and let*

$$V = \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \overline{X})^2 = \frac{(n-1)\, s^2}{\sigma^2}.$$

Then $V \in \chi^2(n-1)$.

For comparing two populations, assume we have two characteristics $X_{(1)}$ and $X_{(2)}$, relative to two populations. We draw from both populations independent random samples of sizes n_1 and n_2, respectively:

$$X_{11}, \ldots, X_{1n_1} \text{ and } X_{21}, \ldots, X_{2n_2}.$$

Recall the notations

$$\overline{X}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} X_{1i}, \quad \overline{X}_2 = \frac{1}{n_2} \sum_{j=1}^{n_2} X_{2j}$$

and

$$s_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} \left(X_{1i} - \overline{X}_1\right)^2, \;\; s_2^2 = \frac{1}{n_2 - 1} \sum_{j=1}^{n_2} \left(X_{2j} - \overline{X}_2\right)^2$$

and the **pooled variance** of the two samples

$$s_p^2 = \frac{\sum_{i=1}^{n_1} \left(X_{1i} - \overline{X}_1\right)^2 + \sum_{j=1}^{n_2} \left(X_{2j} - \overline{X}_2\right)^2}{n_1 + n_2 - 2} = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Proposition 7.4. *Assume either that* $X_{(1)} \in N(\mu_1, \sigma_1),\ X_{(2)} \in N(\mu_2, \sigma_2)$ *or that the sample sizes are large enough and let*

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \;\; \textit{and} \;\; T = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}.$$

Then $Z \in N(0,1)$ and $T \in T(n)$, where

$$\frac{1}{n} = \frac{c^2}{n_1 - 1} + \frac{(1-c)^2}{n_2 - 1} \quad \textit{and} \quad c = \frac{\dfrac{s_1^2}{n_1}}{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}.$$

Proposition 7.5. *Assume either that $X_{(1)} \in N(\mu_1, \sigma_1),\ X_{(2)} \in N(\mu_2, \sigma_2)$ or that the sample sizes are large enough and let*

$$T = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{s_p \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}.$$

Then $T \in T(n_1 + n_2 - 2)$.

Proposition 7.6. *Assume $X_{(1)} \in N(\mu_1, \sigma_1)$ and $X_{(2)} \in N(\mu_2, \sigma_2)$ and let*

$$F = \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2}.$$

Then $F \in F(n_1 - 1, n_2 - 1)$.

Remark 7.7. The elusive term "large enough" should be understood as follows: for one sample, $n > 30$ and for two samples, $n_1 + n_2 > 40$.