6.1

A. As seen in the lecture, if we want to test with a null hypothesis that the true distribution of F of $X_1, \ldots, X_n$ is equal to F_0 we can use the Kolmogorov-Smirnov test statistic D_n:

$$D_n = \sup_x \left|\hat{F}_n(x) - F_0(x)\right| .$$

And, as *theorem 3.1* in the syllabus states, this statistic is *nonparametric*. And if we look at the proof for Kolmogorov-Smirnov test statistic, a simular argument can be made for the adjusted version. The following equality holds:

$$\begin{aligned}\tilde{D}_n &= \sup_x \left|\hat{F}_n(x) - \Phi((x - \bar{X})/S)\right| \\ &= \max_{i=1,..,n} \left(\max\left\{\hat{F}_n(X_i) - \Phi((X_i - \bar{X})/S)\ ,\ \hat{F}_n(X_{i-1}) - \Phi((X_i - \bar{X})/S)\right\}\right) \\ &= \max_{i=1,..,n} \left(\max\left\{\frac{1}{n} - \Phi((X_{(i)} - \bar{X})/S)\ ,\ \frac{i-1}{n} - \Phi((X_{(i)} - \bar{X})/S)\right\}\right) .\end{aligned}$$

However, as it turns out, finding critical values for $\Phi(x)$ is not so simple. In theorem 3.1 F_0 has a uniform distribution under the null hypothesis, but now we are dealing with a normal distribution, which is not distribution free. What we can say, is that the statistic is independent of location and scale. Since the location and scale parameters for a normal distribution are it's expectation $\bar{X}$ and variance S^2 which are included in the statistic.

B. Using the adjusted KS statistic $\tilde{D}_n$, we can not reject the null hypothesis that the measurement errors of the 1879 data are normally distributed. The mean $\tilde{D}_n$ over 1000 bootstrap samples is 0.12, and the corresponding right-tailed p-value is $1 - L(0.12 \times \sqrt{n = 100}) = 0.13$, with L as in Smirnov's `https://projecteuclid.org/download/pdf_1/euclid.aoms/1177730256`. No α level was specified but $0.13 >$ most common ones.

For the 1882 data, we have $p = 1 - L(0.22 \times \sqrt{n = 23}) = 9.6\text{e-}5$. This p-value is probably $< \alpha$, implying that we can reject the null hypothesis (that the 1882 measurement errors are normally distributed).

C. The p-values found with a standard KS-test over the 1879 and 1882 data are respectively 0.49 and 0.72. These values are very different from those found in the previous part. As the syllabus states for the Kolmogorov-Smirnov test-statistic: *"the postulated null-distribution F_0 occurs in the definition of the test statistic D_n."* So the normal distribution is too complex (i.e., distribution **un**free) for the Kolmogorov-Smirnov test. However, in our adjusted test, we use the sample mean and variance so as to undo some of that 'parametricity.'

6.2

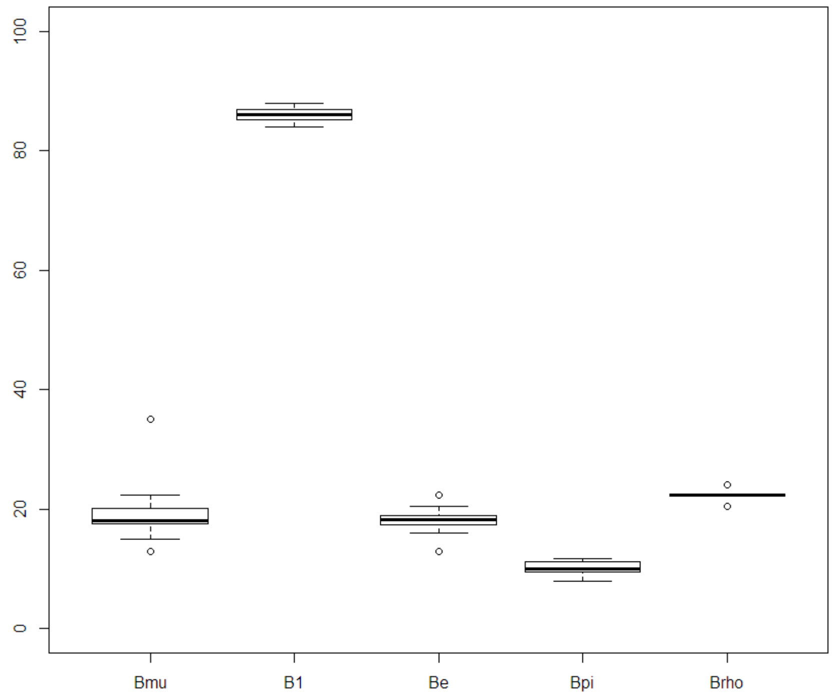
A. Individual datapoints contribute unequally to a weighted mean. For example, outliers affect the inverse-variance weighted mean less than they would affect an arithmetic mean.

B. With B's being 0.25-trimmed means, and under the assumption that the observations come from a normal distribution, the 98% confidence interval for

$$D = B_1 - (B_\rho + B_\pi + B_\varepsilon + B_\mu)$$

is [15.41, 21.09].

C. The following figures shows all boxplots of the lepton data. We can conclude that B_1 is approximately the sum of all other B's and 17, and that some B's have a few outliers. The latter property is a reason to consider trimmed means.



D. The 98-% bootstrap confidence interval for the percentage of D particles with 20%-trimmed means is [14.05, 19.48].

E. The 98-% bootstrap confidence interval for the percentage of D particles with untrimmed means is [13.18, 18.02].

F. From the previous parts, we can conclude with some confidence that the lower bound for the percentage of D particles is > 12. This implies the existence of another, unknown, particle (according to the syllabus).

Appendix

```
rm(list=ls())
source("./light.txt")
source("./functions_Ch4.txt")
source("./lepton.txt")

### Exercise 6.1
## B
# Parametric 79
light79 = sort(light$`1879`)
m79 = mean(light79)
s79 = sd(light79)
B = 1000
n79 = length(light79)
EmpBS79= numeric(B);
for (i in 1:B)
{
  xstar = sample(light79, n79, replace=T)
  EmpBS79[i] = ks.test(xstar, pnorm, m79 , s79)$statistic
}
hist(EmpBS79)
d79 = mean(EmpBS79); d79 # 0.12
d79 * sqrt(n79) # 1.17
l79 = .870612 # https://projecteuclid.org/download/pdf_1/euclid.aoms/1177730256
p79 = 1 - l79; p79 # 0.13

# Parametric 82
light82 = sort(light$`1882`)
m82 = mean(light82)
s82 = sd(light82)
B = 1000
n82 = length(light82)
EmpBS82= numeric(B);
for (i in 1:B) {
  xstar = sample(light82, n82, replace = T)
  EmpBS82[i] = ks.test(xstar, pnorm, m82 , s82)$statistic
}
hist(EmpBS82)
d82 = mean(EmpBS82); d82 # 0.22
d82 * sqrt(n82) # 2.23
l82 = .999904 # https://projecteuclid.org/download/pdf_1/euclid.aoms/1177730256
p82 = 1 - l82; p82 # 9.6e-5

## C
ks.test(light79, pnorm, m79, s79) # 0.083, p = 0.49
ks.test(light82, pnorm, m82, s82) # 0.145, p = 0.72
```

```
### Exercise 6.2
## C
boxplot(lepton, ylim=c(0,100))

## D
t = .2
B1 = lepton$B1
Bmu = lepton$Bmu
Be = lepton$Be
Bpi = lepton$Bpi
Brho = lepton$Brho
Dut = mean(B1, trim=t) - (mean(Bmu, trim=t) + mean(Be, trim=t) +
      mean(Bpi, trim=t) + mean(Brho, trim=t))

meansBS = numeric(B)
for (i in 1:B) {
  meansBS[i] = mean(sample(lepton$B1, replace=TRUE), trim=t) -
               (mean(sample(lepton$Bmu, replace=TRUE), trim=t) +
                mean(sample(lepton$Be, replace=TRUE), trim=t) +
                mean(sample(lepton$Bpi, replace=TRUE), trim=t) +
                mean(sample(lepton$Brho, replace=TRUE), trim=t))
}
hist(meansBS)
z = meansBS-Dut
meansC = c(Dut-quantile(z, probs=1-alph/2), Dut-quantile(z, probs=alph/2))
meansC # [14.05, 19.48]

## E
Du = mean(B1) - (mean(Bmu) + mean(Be) + mean(Bpi) + mean(Brho))
meansBS = numeric(B)
for (i in 1:B) {
  meansBS[i] = mean(sample(lepton$B1, replace=TRUE), trim=t) -
    (mean(sample(lepton$Bmu, replace=TRUE), trim=t) +
       mean(sample(lepton$Be, replace=TRUE), trim=t) +
       mean(sample(lepton$Bpi, replace=TRUE), trim=t) +
       mean(sample(lepton$Brho, replace=TRUE), trim=t))
}
hist(meansBS)
z = meansBS-Du
meansC = c(Du-quantile(z, probs=1-alph/2), Du-quantile(z, probs=alph/2))
meansC # [13.18, 18.02]
```