A Simulation Study on the Comparison of Hypothesis Testing Procedures

Autumn Biggie, Ruben Sowah, Jingjing Li

4/22/2022

Report Section

As we see in part I, to compare two tests we usually have to make assumptions like normal distribution and equal or unequal variances which may not be true for the population. Furthermore, some parameters are inaccessible or unrealistic to obtain at all, which undoubtedly impedes the efficiency and understanding of the statistic tests. On the other hand, what kinds of test methods are better are often ambiguous. An efficient strategy to address these issues is simulation study.

Simulation study allows researchers to specify the parameters of the population with desirable distribution which can provide the investigator with prior knowledge of the truth or falsehood of a null hypothesis. Using simulation method, we can easily generate the data which conform with the requirements of the forementioned two tests and compare the efficiency of the two tests under different parameter via extensive sampling procedure from the populations with specified parameters. Anyway, the numerical simulation technique is an effective method to make statistical inference because it offers the researcher control over specified parameters and can easily produce many samples without vast amounts of time and resources.

For simulation study, we intend to generate two data sets of random samples with predetermined parameters including sample size, mean difference and variance using normal distribution function, `rnorm()` via bootstrap approach. Basically, we created nested for-loops to generate 100 samples for all possible combinations of the above parameters. Next, in order to compare the means of two samples, under the null hypothesis (H0), $\mu_1 - \mu_2 = 0$, and alternative hypothesis (HA) $\mu_1 - \mu_2 != 0$, we compute t values using two tests (the Pooled Variances Test and the Unequal Variances Test), count the rejection times and get the rejection ratio which is can be regarded as type I error (when H0 is true) or power (when HA is true). The Pooled Variances Test assumes that the populations have equal variance, σ^2, while the Unequal Variances Test assumes the variances of the populations, σ_1^2 and σ_2^2, are unequal. To compare two test methods, we change the variance, sample size and mean difference and get a series of rejection ratios. The possible values for true mean difference ($\mu_1 - \mu_2$) were -5, -1, 0, 1, and 5. The possible values for sample size were $n_1 = 10, 30, 70$ and $n_2 = 10, 30, 70$. The value of σ_2^2 was held constant at 1, while σ_1^2 took on possible values of 1, 4, and 9. The resulting data set, `rez`, contains the values for the parameters specified, as well as variables `alpha1` and `alpha2`. It is important to note that `alpha1` and `alpha2` do not always refer to the Type I Error Rate, α, but only when the null hypothesis $\mu_1 - \mu_2 = 0$ is true. Otherwise, `alpha1` and `alpha2` refer to the power $1 - \beta$ of each test, where `alpha1` corresponds to the **Pooled Variances Test** and `alpha2` corresponds to the **Unequal Variances Test**.

The results, displayed graphically, reveal clear differences between the performance of the tests under different conditions. We first discuss our observations from Display Group 1:

When mean difference is -5 or 5 which is far from 0, the power values of both tests are all equal to 1 regardless of variance and sample size. When difference is -1 or 1, the power value is significantly influenced by sample size. When sample size of two samples are equal,the power values of both tests increase with sample size. Under the same sample size, the power values of both tests decrease with rising variance. When var1 (1) is

equal to var2 (1), the power values of the tests are equal or very similar, which are less influenced by sample size. In the case when two sample sizes are equal, the power values of the two tests are equal or very close (the blue dots and red dots overlap). When the size of sample 2 is larger than that of sample 1, the power value of test one (with pooled variance) is greater than that of test two (unequal variances test). On the other hand, When the size of sample 2 is smaller than that of sample 1, the power values of test two (with unequal variance) are correspondingly greater than those of test one (pooled test).

The figures in Display Groups 2 and 3 group the variable combinations by differences in value. For example, instead of displaying the values of `var_1` and `var_2`, observations are grouped by the absolute value of the difference between `var_1` and `var_2`. The purpose is to gain insight into the probability of rejection for each test when at different "distances" between means, variances, and sample sizes. This format also improves graph readability.

Both figures in Display Group 2 reveal that Type 1 Error Rate increases as difference in variance increases. However, this error rate, although increasing, remains more stable for the Unequal Variances Test than the Pooled Variances Test. In addition, the power of both tests decreases as variance increases, but only when the difference in means is 1 or -1. When the difference in means is 5 or -5, the power is 1, presumably because the falsehood of the null hypothesis is so apparent that it would be very difficult for both tests to fail to reject. Although the power of both tests decreases as variance increases, as stated previously, the power values remain slightly more constant for the Unequal Variances test, although the difference is minute. The figures in Display Group 3 allow us to examine this difference more clearly. When the absolute mean difference is 1, the power values of both tests decrease as a function on increasing difference in variance, but there are more low power values for the Pooled Variances Test.

Lastly, the the figures in Display Group 3 clearly show the difference in Type I Error rate discussed previously. When the mean absolute difference is 0 (the null hypothesis is true) and the variances are equal, there is not much difference between the Type I Error Rates of the two tests. However, when variances are unequal, the Type I Error rates of the Pooled Variances Test climb much higher than those of the Unequal Variances Test. The Figures in Display Group 3 plot sample size differences in different colors, but because of the interaction between sample size and test type as they relate to power discussed previously, the output is not very interpretable. Identification of this interaction is one of the advantages of the figure in Display Group 1.

Based on our investigation, we made some conclusions shown below. First, the difference of two samples' means play critical roles in power value. When the two means are far different, the power will reach the maximum, 1 no matter what test method is used. Second, when the sample size and variance of two samples are consistent, the power values of two tests are equal or very close, pooled method may be intuitively adopted. Third, sample size increase led to improved power, while increased variance values reduce the power value. When sample size and variance is different,we have look into the details. When the sample size of the sample with lower variance is greater than that with higher variance, the unequal test method preferred. Otherwise, the pooled method is recommended.

Corresponding R Code

Objective: Conduct a simulation study to compare the Type I Error Rate and Power of two hypothesis testing procedures under various conditions. The procedures to be tested include: * Pooled Variance Test * Unequal Variance Test

1) Define variables

- `var1` and `var2` are the true respective variances for nonsmokers and smokers populations.

- `n1` and `n2` are the sample sizes for nonsmokers and smokers populations.

- `mu1` and `mu2` are the respective true means for nonsmokers and smokers populations.

- `rez` is the dataset generated using simulation combinations (135 in our case).

- `yval1` is a set of 9 random numbers based on sample n1 and yval2 is as set of 9 random numbers based on sample n2.

Assign variables

The following variables will be used to set conditions for the simulation. We'll be adjusting the values of variance for each group, as well as sample size and mean.

Generate Simulation

The following loop is used to generate random samples from the Normal distribution using all combinations of variance, sample size, and mean defined above. We also compute variables named `alpha1` and `alpha2`. It is important to note that `alpha1` and `alpha2` do not always refer to the Type I Error Rate, α, but only when the null hypothesis $\mu_1 - \mu_2 = 0$ is true. Otherwise, `alpha1` and `alpha2` refer to the power $1 - \beta$ of each test, where `alpha1` corresponds to the **Pooled Variance Test** and `alpha2` corresponds to the **Unequal Variances Test**.

```
##   var_1 var_2 n_1 n_2 mu_diff alpha1 alpha2
## 1     1     1  10  10      -5   1.00   1.00
## 2     1     1  10  10      -1   0.54   0.54
## 3     1     1  10  10       0   0.03   0.03
## 4     1     1  10  10       1   0.55   0.53
## 5     1     1  10  10       5   1.00   1.00
## 6     1     1  10  30      -5   1.00   1.00
```

Graphical Displays

We use the following graphical displays to represent the data from our simulation study.

First, new variables are created to allow for better visualizations.

- `abs_mu_diff` is the absolute difference of the means from each group

- `abs_var_diff` is the absolute difference of the variances from each group

- `abs_size_diff` is the absolute difference in sample sizes from each group

These are added to the `rez` data set.

Display Group 1

This graph shows the probability of rejection for the **Pooled Variances Test** in red and the **Unequal Variances Test** in blue as a function of difference in means, faceted by all possible combinations of `n_1`, `n_2`, and `var_1`. The results are discussed in the report above.

Probability of Rejection Vs. True Difference in Means Over All Possible Cor

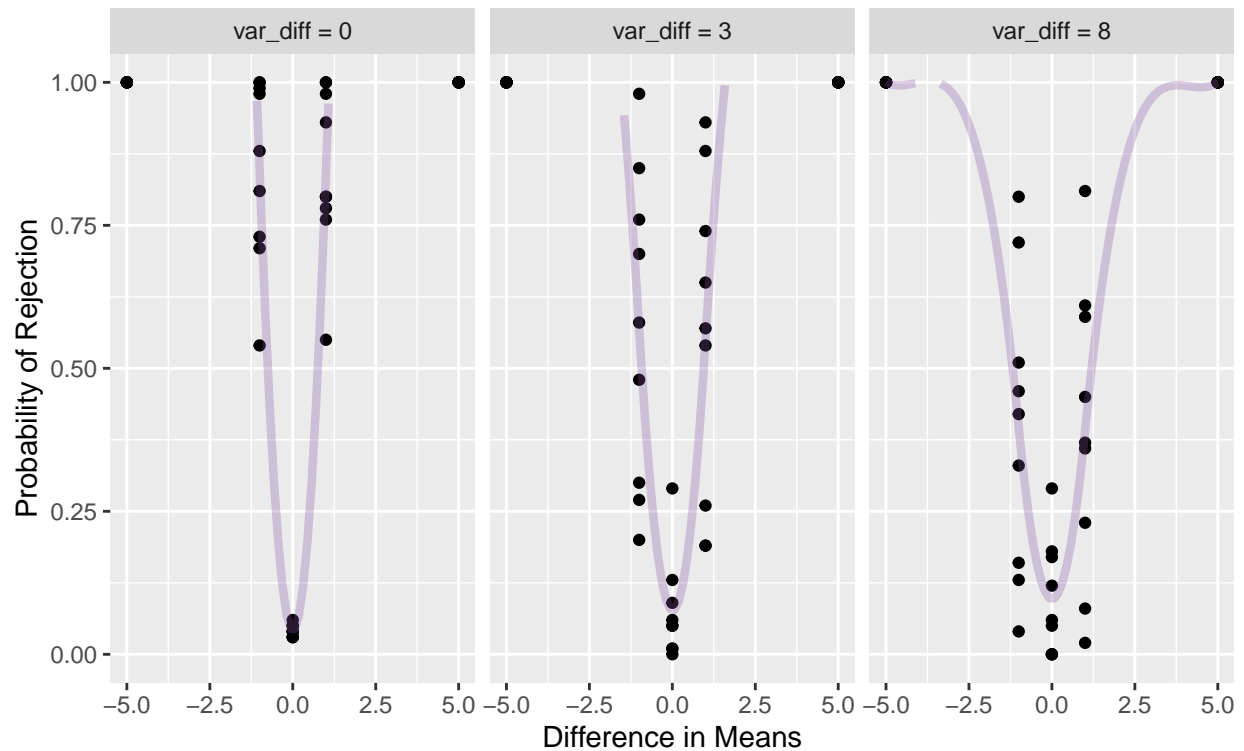## Pooled Variance Test in red, Unequal Variance Test in blue
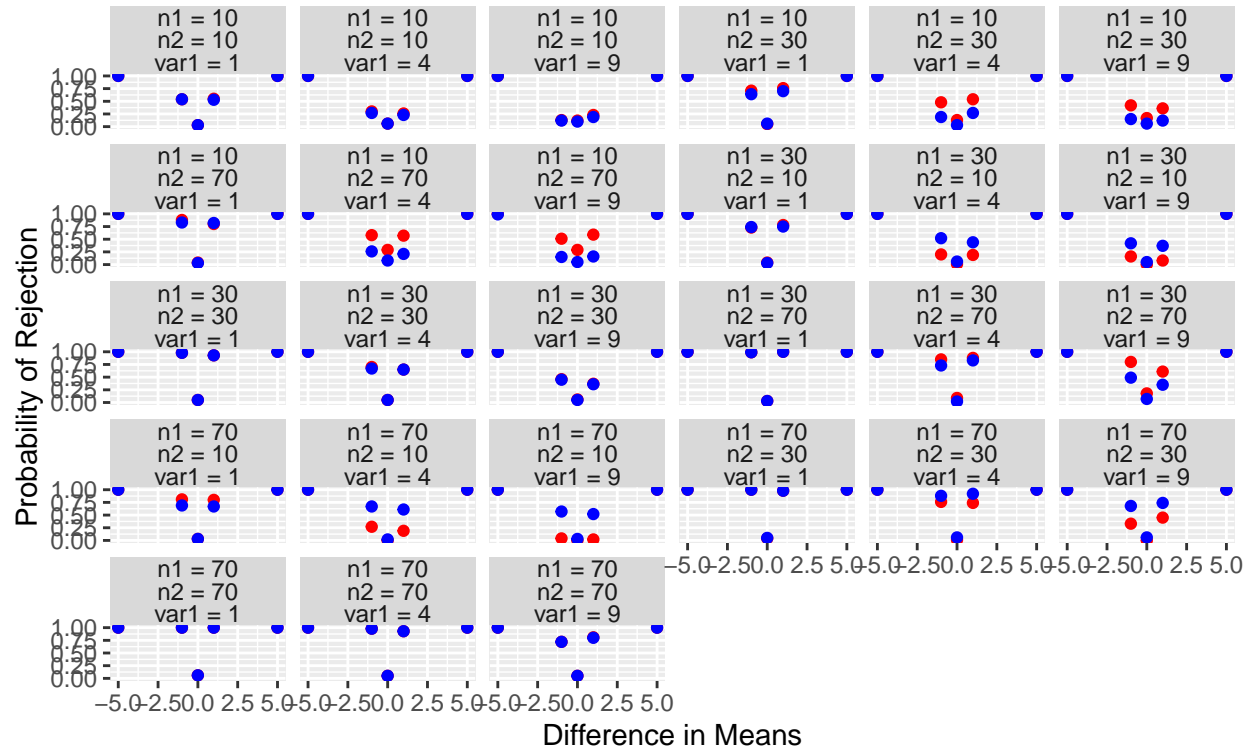
Display Group 2

This graph shows the probability of rejection (`alpha1`) for the **Pooled Variances Test** as a function of the difference in means, faceted by difference in variance. The results are discussed in the report above.

```
## 'geom_smooth()' using method = 'loess' and formula 'y ~ x'
```

Pooled Variances Test:
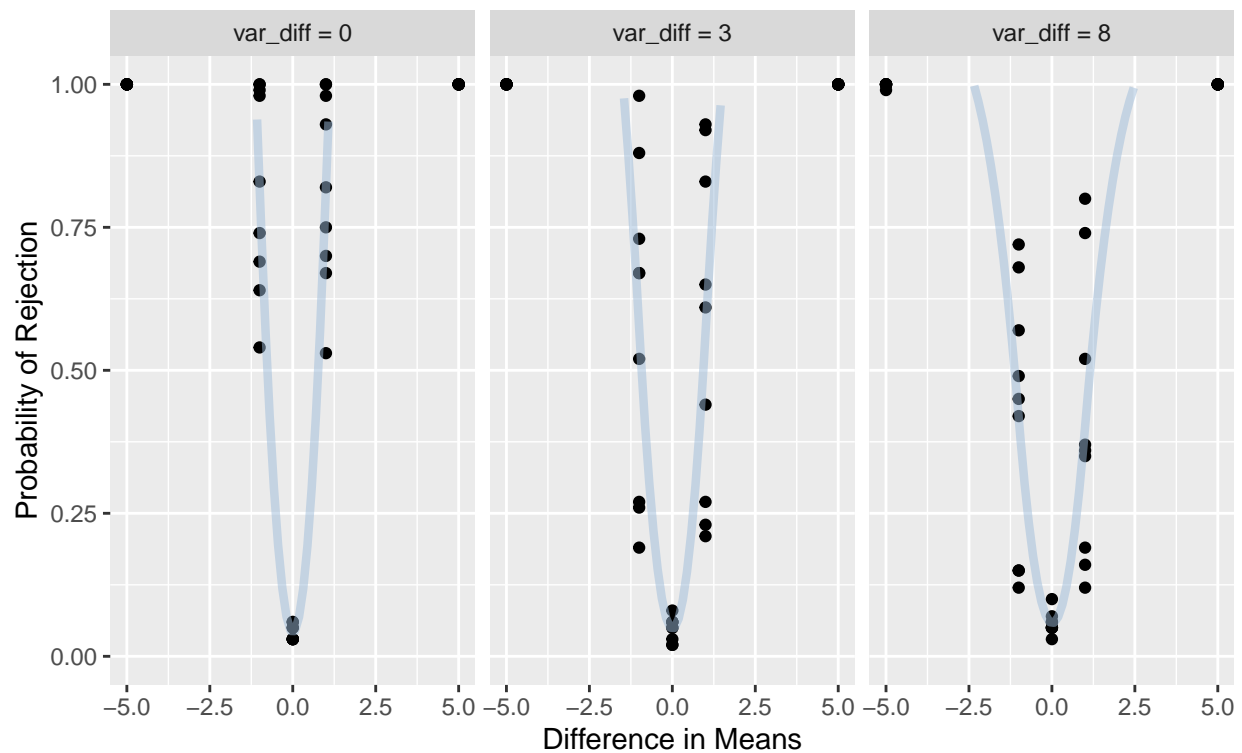a Comparison of Probability of Rejection over Difference in Variance

This graph shows the probability of rejection (`alpha2`) for the **Unequal Variances Test** as a function of the difference in means, faceted by difference in variance. The results are discussed in the report above.

```
## 'geom_smooth()' using method = 'loess' and formula 'y ~ x'
```

Unequal Variances Test:

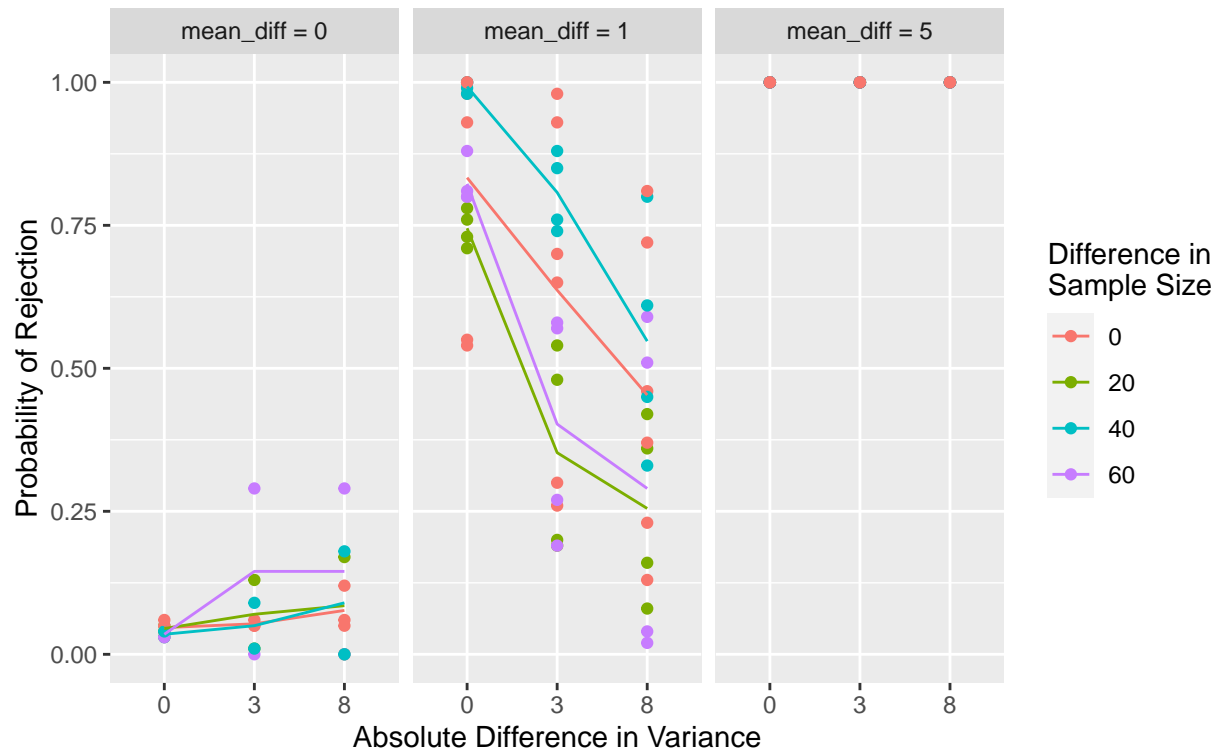a Comparison of Probability of Rejection over Difference in Variance

Display Group 3

This graph shows the probability of rejection (`alpha1`) for the **Pooled Variances Test** as a function of the difference in variances, faceted by difference in means. The results are discussed in the report above. Note that the points for `mean_diff = 5` are layered on top of one another, which is why most points in this facet are not visible.

```
## 'geom_smooth()' using method = 'loess' and formula 'y ~ x'
```

Pooled Variances Test:

a Comparison of Probability of Rejection over Absolute Mean Difference

This graph shows the probability of rejection (`alpha2`) for the **Unequal Variances Test** as a function of the difference in variances, faceted by difference in means. The results are discussed in the report above. Note that the points for `mean_diff = 5` are layered on top of one another, which is why most points in this facet are not visible.

```
## 'geom_smooth()' using method = 'loess' and formula 'y ~ x'
```

Unequal Variances Test:

a Comparison of Probability of Rejection over Absolute Mean Difference
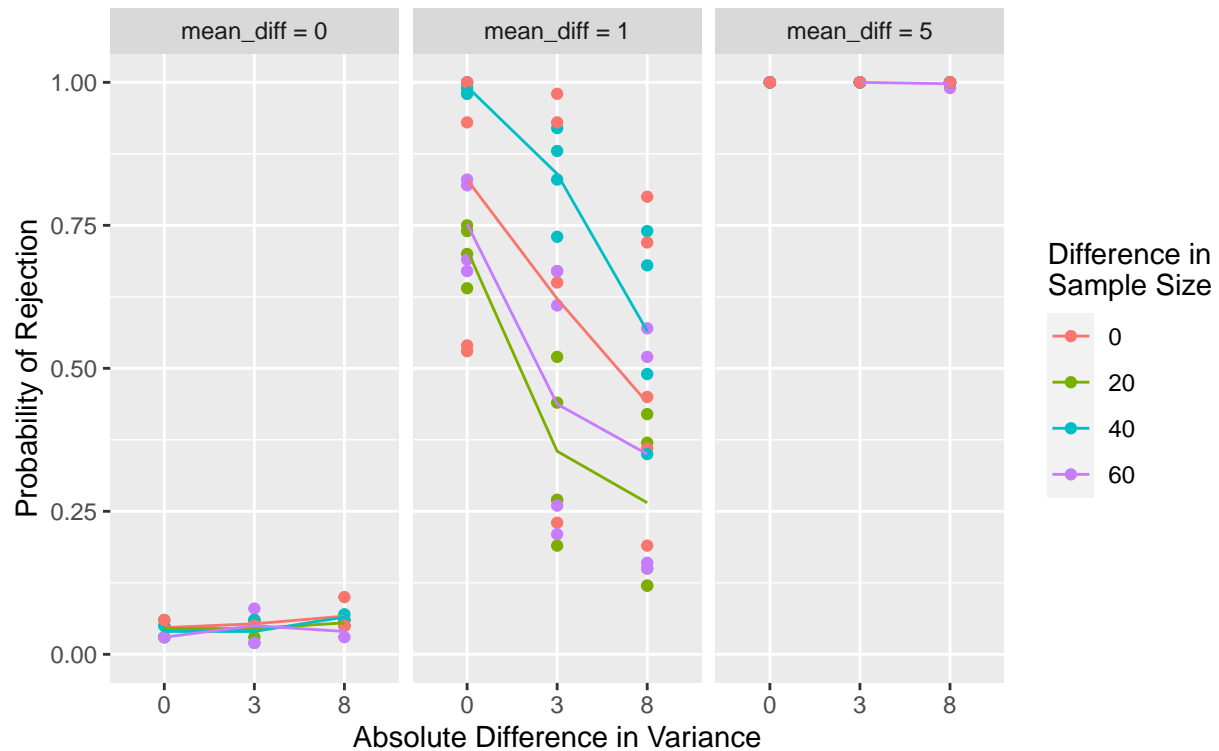
Member Contributions

Autumn Biggie initiated the idea of using nested loop to combine the parameters, built the framework of the data table, generated data sets, calculated t value of two tests calculated, wrote the code for display group 2 and 3, polished the code for group 1, constructed the original report. Ruben Sowah modified code and put forward to some suggestions. Jingjing Li introduced loop for make 100 samples at different parameters, computed p-value, rejection times and alpha values and prepared the draft code for graph1, modified the report.

References:

https://www.ime.usp.br/~abe/ICOTS7/Proceedings/PDFs/InvitedPapers/7G2_ERIC.pdf