## 3 or more samples ⇒ *F*-test for 1-way ANOVA

✓ Three or more independent samples (groups)

✓ **Hypotheses:** $H_0$ : all the underlying population means are the same i.e. $\mu_1 = \mu_2 = \mu_3 = \ldots = \mu_k$

vs $H_1$ : at least 2 of the underlying population means are not the same

Note: SHOULD NOT be written using symbols

✓ **Assumptions** for 1-way ANOVA

1. Observations *within* the samples are **independent** – *CRITICAL!*
2. The samples/groups are **independent**, i.e., observations *between* samples/groups are independent of each other – *CRITICAL!*
3. **The Normality Assumption**: The population or underlying distributions are Normal. No clusters or multi-modes allowed.
   Check by looking at plots of each sample and consider sample sizes. Plots should be unimodal and not too skewed for the $n_{tot}$ you have.
4. **Equality of Standard Deviations**
   The standard deviations of the underlying distributions (or populations) are equal.
   Check by using the ratio (fraction) $\dfrac{\textbf{largest sd}}{\textbf{smallest sd}} < 2$ as a guide

✓ Data not Normal and/or standard deviations not equal? Don't use the *F*-test!

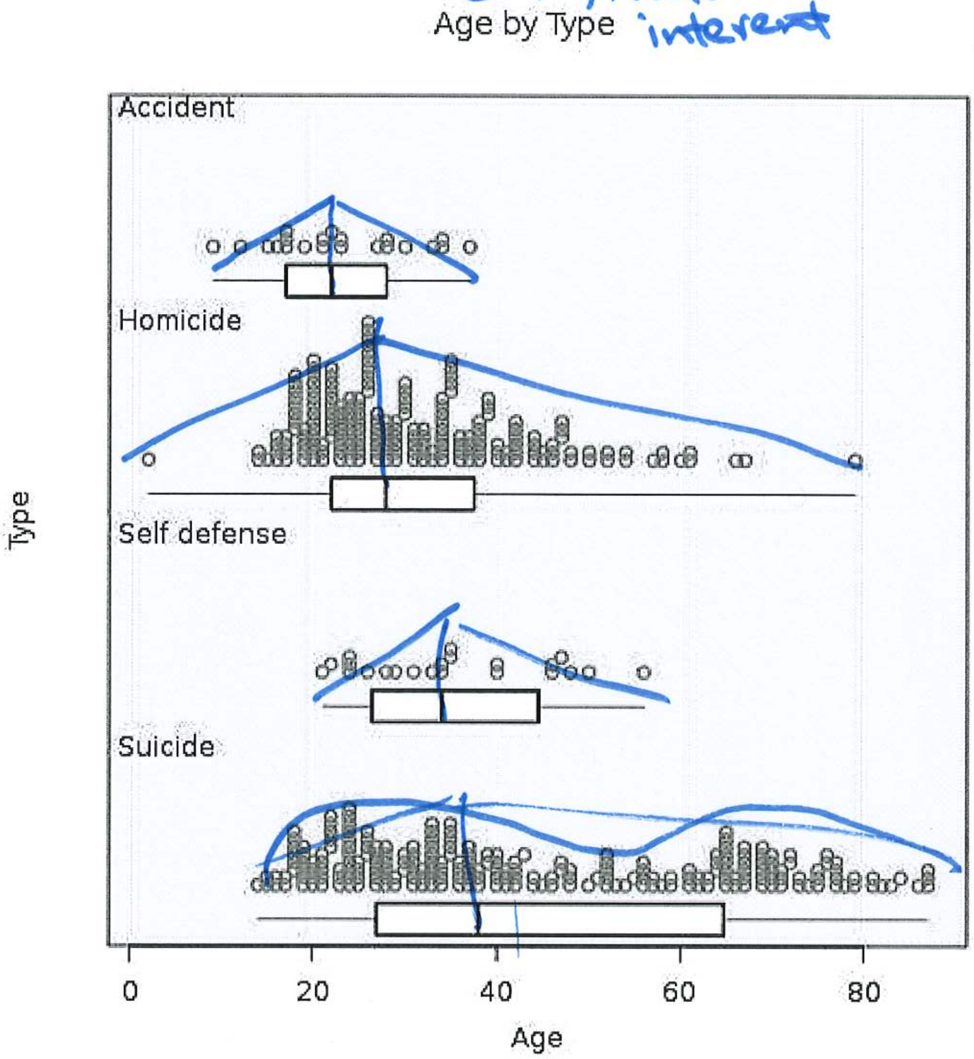
✓ **One-way ANOVA Table**
**(SPSS will have the numbers laid out in this way)**

| | Sum of Squares (SS) | *df* | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Between Groups (BG) | BGSS | $df_1 = k - 1$ | $s_B^2$ | $f_0 = \dfrac{s_B^2}{s_W^2}$ | $\Pr(F \geq f_0)$ |
| Within Groups (WG) | WGSS | $df_2 = n_{tot} - k$ | $s_W^2$ | | |
| Total (T) | TSS | $df_1 + df_2$ | | | |

where F is the *F*-statistic, $f_0 = \dfrac{s_B^2}{s_W^2}$ where $s_B^2$ = the **between** group variation

and $s_W^2$ = the **within** group variation

**Example:** In 1989, 464 people were killed by a gun in the United States in a single week in May. These deaths have been grouped into four classes: **Accident**; **Homicide**; **Self defense**; and; **Suicide**. The age was also recorded for each person.




The age for suicides is centred highest while the age for accidents is centred lowest. The age for accidents is the least variable. The accident ages are reasonably symmetric while the self defense group is `slightly / moderately negatively (left) / positively (right) skewed` while the other two groups (homicide and suicide) are both `slightly / moderately / strongly negatively (left) skewed` / `positively (right) skewed`.

**Assumptions**

- The observations within each type of death are independent, that is, the samples are random.

- The samples of age for each type of death are independent of each other.

- The underlying distributions (populations) of age for each type of death are Normally distributed. That is, the data for each type of death come from a Normal distribution.
- The standard deviations of the underlying distributions (populations) of age for each type of death are equal.

There `are / are no` concerns about the validity of the *F*-test.

### Descriptives

Age (in years)

| | N | Mean | Std. Deviation | Std. Error | 95% Confidence Interval for Mean Lower Bound | 95% Confidence Interval for Mean Upper Bound | Minimum | Maximum |
|---|---|---|---|---|---|---|---|---|
| Accident | 21 | 23.00 | 7.797 | 1.702 | 19.45 | 26.55 | 9 | 37 |
| Homicide | 195 | 30.86 | 11.879 | .851 | 29.18 | 32.54 | 2 | 79 |
| Self defense | 22 | 35.14 | 10.204 | 2.176 | 30.61 | 39.66 | 21 | 56 |
| Suicide | 222 | 44.25 | 20.418 | 1.370 | 41.55 | 46.95 | 14 | 87 |
| Total | 460 | 37.17 | 17.841 | .832 | 35.53 | 38.80 | 2 | 87 |

Let $\mu_1$ be the underlying mean age for accidents, and similarly define, $\mu_2$, $\mu_3$ and $\mu_4$ for homicide, self defense and suicide respectively.

$H_0$ : all the underlying population mean ages are the same

i.e. $\mu_1 = \mu_2 = \mu_3 = \mu_4$

vs $H_1$ : the underlying population mean age is different for at least two of the groups (at least two types of death)

### ANOVA

Age (in years)

| | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Between Groups | 23188.885 | 3 | 7729.628 | 28.676 | .000 |
| Within Groups | 122915.226 | 456 | 269.551 | | |
| Total | 146104.111 | 459 | | | |

We have `no / weak / some / strong / very strong` evidence (*P-value* = .000) against there being no difference in the underlying mean age for the four groups, that is, we have `no / weak / some / strong / very strong` evidence that the underlying mean age is different for at least two of the groups (at least two types of death).

**Multiple Comparisons**

Dependent Variable: Age (in years)

Tukey HSD

| (I) Type of death | (J) Type of death | Mean Difference (I-J) | Std. Error | Sig. | 95% Confidence Interval Lower Bound | 95% Confidence Interval Upper Bound |
|---|---|---|---|---|---|---|
| Accident | Homicide | -7.862 | 3.771 | .160 | -17.58 | 1.86 |
| | Self defense | -12.136 | 5.009 | .074 | -25.05 | .78 |
| | Suicide | -21.248* | 3.748 | .000 | -30.91 | -11.58 |
| Homicide | Accident | 7.862 | 3.771 | .160 | -1.86 | 17.58 |
| | Self defense | -4.275 | 3.693 | .654 | -13.80 | 5.25 |
| | Suicide | -13.386* | 1.611 | .000 | -17.54 | -9.23 |
| Self defense | Accident | 12.136 | 5.009 | .074 | -.78 | 25.05 |
| | Homicide | 4.275 | 3.693 | .654 | -5.25 | 13.80 |
| | Suicide | -9.111 | 3.670 | .064 | -18.57 | .35 |
| Suicide | Accident | 21.248* | 3.748 | .000 | 11.58 | 30.91 |
| | Homicide | 13.386* | 1.611 | .000 | 9.23 | 17.54 |
| | Self defense | 9.111 | 3.670 | .064 | -.35 | 18.57 |

*. The mean difference is significant at the 0.05 level.

p-val

$\mu_4 > \mu_1$

$\mu_4 > \mu_2$

$\mu_4 = \mu_3$

**Assuming the Tukey pairwise comparisons are valid:**

We have very strong evidence (P-value = 0.000) that the underlying mean age for suicides is less than / greater than that for accidents and homicides.

With 95% confidence, we estimate, that the underlying mean age for suicides is somewhere between 11.6 and 30.1 years less than / greater than that for accidents.

With 95% confidence, we estimate, that the underlying mean age for suicides is somewhere between 9.2 and 17.5 years less than / greater than that for homicides.

We cannot say that suicides have the lowest / highest age on average because there is not a significant difference between suicide and self defense, that is, with 95% confidence, we estimate, that the underlying mean age for suicides is somewhere between 0.4 years less than / greater than and 18.6 years less than / greater than that for self defense.

## Appendix D: Fruitfly Data

**Questions X** to **BB** refer to the information in this appendix.

It had already been established that increased sexual activity decreases the number of days for which female fruitflies live. Researchers Hanley and Shapiro (1994) designed a study to see if the same were true for male fruitflies. The sexual activity of male fruitflies was manipulated by supplying individual male fruitflies with either one or eight receptive virgin females per day. The lifespan of these males (the number of days they lived for) was compared with the lifespan of males that were supplied daily with one or eight newly inseminated females. Newly inseminated females are not receptive because they will not re-mate for at least two days. There was also a group of males kept with no females.

Thus there were five groups in total and 125 male fruitflies were randomly assigned to one of these five groups. This meant that there were 25 male fruitflies in each group.

The five groups were:

> **GP1:** Male kept alone
>
> **GP2:** Male supplied daily with 1 newly inseminated female unwilling to mate
>
> **GP3:** Male supplied daily with 1 receptive virgin female willing to mate
>
> **GP4:** Male supplied daily with 8 newly inseminated females unwilling to mate
>
> **GP5:** Male supplied daily with 8 receptive virgin females willing to mate

The *underlying* mean lifespan for **GP1** ($\mu_{GP1}$) is defined to be the mean lifespan (in days) if all of the 125 male fruitflies used in the study had been kept alone. The *underlying* mean lifespans for **GP2**, **GP3**, **GP4** and **GP5** ($\mu_{GP2}$, $\mu_{GP3}$, $\mu_{GP4}$, and $\mu_{GP5}$) are defined in a similar manner.

An *F*-test for one-way analysis of variance was conducted to compare the lifespan of the fruitflies in the different groups.

Dot/box plots of the lifespan for each group are shown in Figure 9, page 36. Output from the *F*-test is shown in Table 6, page 36.

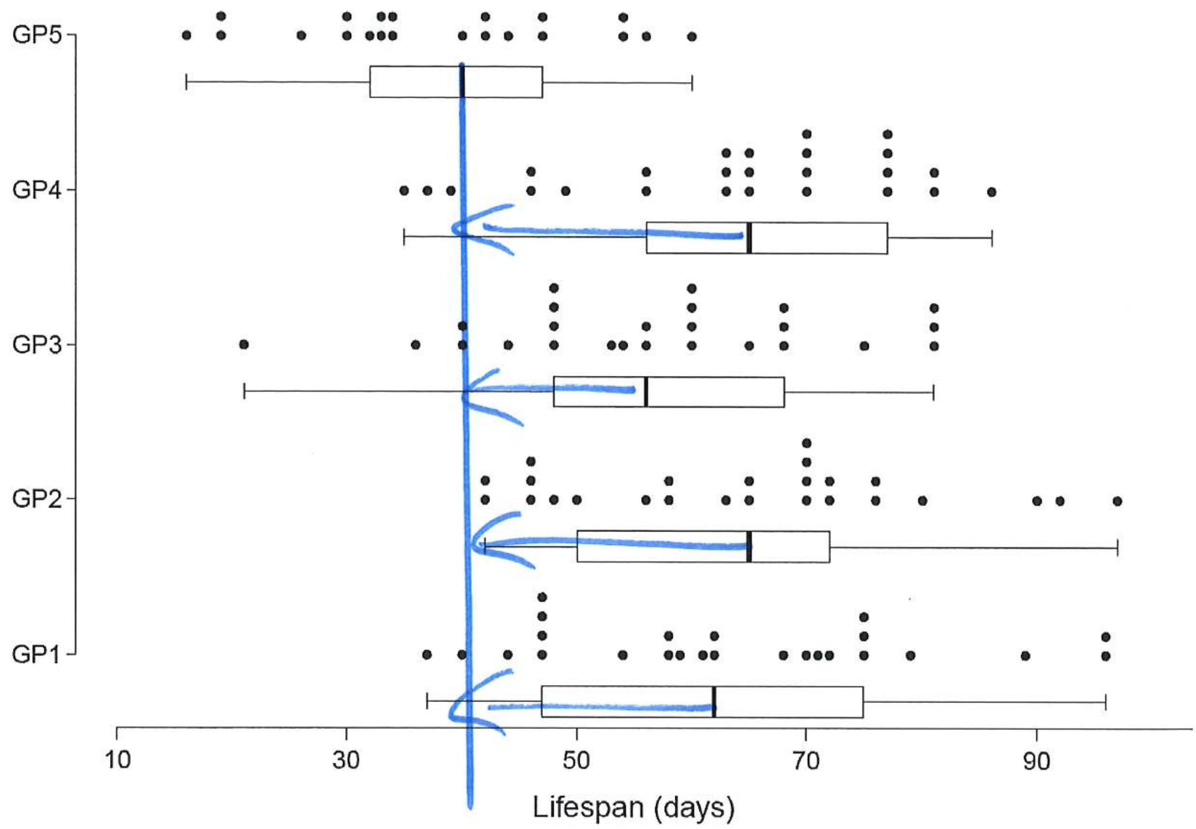



**Figure 9:** Lifespan of male fruitflies by group

**ANOVA**

Lifespan

| | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Between Groups | 11939.280 | ++ | 2984.820 | ++ | .000 |
| Within Groups | 26313.520 | ++ | 219.279 | | |
| Total | 38252.800 | ++ | | | |

Note: Some values have been replaced by ++.

**Descriptives**

Lifespan

| | N | Mean | Std. Deviation | Std. Error | 95% Confidence Interval for Mean | | Minimum | Maximum |
|---|---|---|---|---|---|---|---|---|
| | | | | | Lower Bound | Upper Bound | | |
| GP1 | 25 | 63.56 | 16.452 | 3.290 | 56.77 | 70.35 | 37 | 96 |
| GP2 | 25 | 64.80 | 15.652 | 3.130 | 58.34 | 71.26 | 42 | 97 |
| GP3 | 25 | 56.76 | 14.928 | 2.986 | 50.60 | 62.92 | 21 | 81 |
| GP4 | 25 | 63.36 | 14.540 | 2.908 | 57.36 | 69.36 | 35 | 86 |
| GP5 | 25 | 38.72 | 12.102 | 2.420 | 33.72 | 43.72 | 16 | 60 |
| Total | 125 | 57.44 | 17.564 | 1.571 | 54.33 | 60.55 | 16 | 97 |

**Table 6:** *F*-test output

**Questions X** to **AB** refer to the information in **Appendix D**, pages 35 and 36.

X. Which **one** of the following statements is **false**?

(1) The response variable is lifespan (in days).

(2) The groups GP1, GP2 and GP4 can be viewed as control groups.

(3) This study is an experiment with 5 different 'treatment' levels.

(4) The units in this study are 125 male fruitflies.

(5) There is blocking in this study design with the female fruitflies blocked as either 'newly inseminated' or 'receptive virgin'.

Y. Which **one** of following is a **correct** set of hypotheses for this *F*-test?

(1) $H_0$: All five observed group means have the same value.
$H_1$: At least two observed group means have different values.

(2) $H_0$: The underlying means are the same for all five groups.
$H_1$: The underlying means are not all the same for the five groups.

(3) $H_0$: The underlying means are the same for all five groups.
$H_1$: Each of the five groups has a different underlying mean.

(4) $H_0$: The underlying means are not all the same for the five groups.
$H_1$: The underlying means are the same for all five groups.

(5) $H_0$: No two groups have the same underlying mean.
$H_1$: The underlying means are the same for all five groups.

Z. Which **one** of the following statements about this *F*-test is **true**?

(1) There is no concern about the validity of the *F*-test with regards to the Normality assumption because the data does not suggest clusters nor does it appear strongly skewed.

(2) There is concern about the validity of the *F*-test with regards to the equal underlying standard deviations assumption.

(3) We should be wary about using an *F*-test on these data because the observations within each group are not independent.

(4) It is not appropriate to use an *F*-test on these data because the number of male fruitflies in each group is not greater than or equal to 30.

(5) We should be wary about using an *F*-test on these data because the groups are not independent.

**Questions AA** to **DD** assume that the use of the *F*-test is appropriate. (Note that this may not be true.)

Refer to information given in Table 6, page 36, to answer these questions.

AA. The value of the *F*-test statistic, $f_0$, is approximately:

(1) 0.03
(2) 13.61
(3) 1.45
(4) 0.45
(5) 0.69

c.f. pg 36

BB. The degrees of freedom for this *F*-test are:

(1) $df_1 = 5$ $df_2 = 20$
(2) $df_1 = 4$ $df_2 = 121$
(3) $df_1 = 4$ $df_2 = 120$
(4) $df_1 = 4$ $df_2 = 20$
(5) $df_1 = 5$ $df_2 = 120$

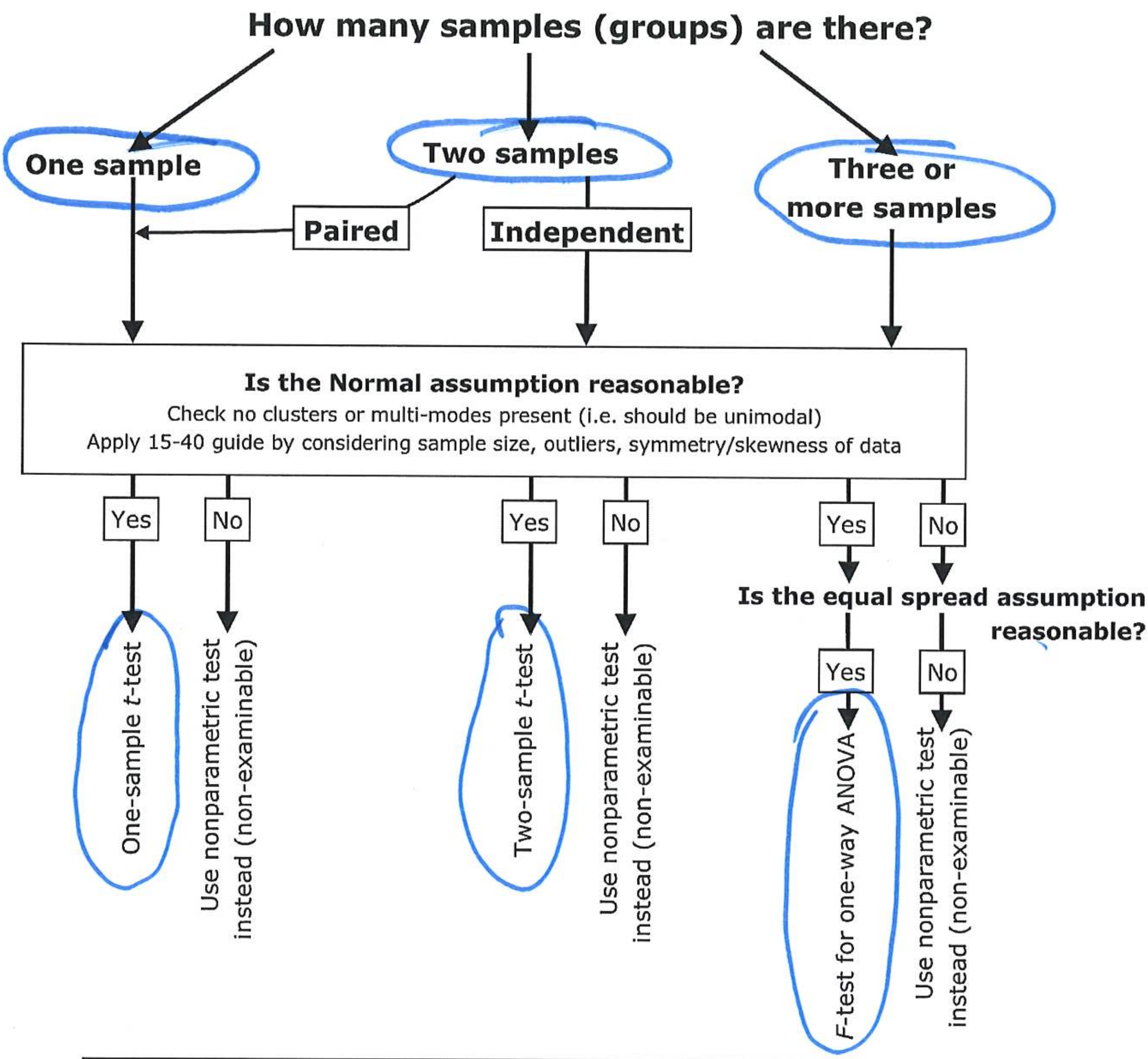
c.f. pg 36

CC. Analysis of variance (ANOVA) is (select **one** only):

(1) an overall test of no difference between sample variances.
(2) an *F*-test of no difference between population means.
(3) an overall test of no difference between population variances.
(4) an *F*-test for the equality of population variances.
(5) an *F*-test of no difference between sample means.

DD. Which **one** of the following statements about the one-way analysis of variance *F*-test is **false**?

T (1) The evidence of differences between the true group means comes from comparing the variability between group means with the variability within the groups. $f_0 = s_B^2 / s_W^2$
T (2) It should only be used when comparing independent samples.
T (3) It provides partial protection against multiple comparisons.
T (4) The null hypothesis is that all the true group means are the same.
F (5) It is not badly affected by the presence of only one or two outliers. may be




| **Assumptions** | **Checks** |
|---|---|
| **1. Independence — All tests**<br>- Single sample assumes indep. between observations.<br>- Paired data assumes indep. between pairs of observations.<br>- Two or more samples assumes indep. between observations *and* samples. | - The design of the experiment/study |
| **2. Normality — All tests**<br>- one-sample *t*-test<br>- two-sample *t*-test<br>- *F*-test for one-way ANOVA | - Plot the data<br>- Apply 15-40 guide |
| **3. Equal spread — *F*-test for one-way ANOVA only** | - Plot the data<br>- Check the standard deviation ratio: $\frac{\text{largest sd}}{\text{smallest sd}} < 2$ |