Testing for Significant Differences Between Distributions

A Crib Sheet

Dr. Andrew Hamilton-Wright

School of Computer Science
University of Guelph

2021-04-03

The overall question

If you have done an experiment and collected some sets of measures from two or more experimental setups, a common question is:

> *Q: Is there a significant difference between the distributions represented by the points making up my data sets?*

This set of slides gives you some advice on how to proceed to answer this question.

More detailed answers can be found in excellent online sources. My two favourites are these – they both give excellent explanations:

McDonald, J.H. (2014): Handbook of Biological Statistics (3rd ed.). Sparky House Publishing, Baltimore, Maryland. `http://www.biostathandbook.com/index.html`

NIST/SEMATECH (2013): e-Handbook of Statistical Methods, `http://www.itl.nist.gov/div898/handbook/`

Some questions will depend on the answers to one or both of these questions, as applied to each of your sets of points:

- $\mathcal{Q}$: Is this set of points normally distributed?

---

*Razali & Wah (2011): Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests Journal of Statistical Modelling and Analytics 2(1):21–33. ISBN: 978-967-363-157-5
Link via ResearchGate (Apr 2021)

Preliminary Tools

Some questions will depend on the answers to one or both of these questions, as applied to each of your sets of points:

- $\mathcal{Q}$: Is this set of points normally distributed?
  - **Shapiro-Wilk** (the most sensitive)*
  - **Anderson-Darling**,
  - **Kolmogorov-Smirnov**
  - a **Q-Q plot** (quartile-quartile plot)

*Razali & Wah (2011): Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests Journal of Statistical Modelling and Analytics 2(1):21–33. ISBN: 978-967-363-157-5
Link via ResearchGate (Apr 2021)

Some questions will depend on the answers to one or both of these questions, as applied to each of your sets of points:

- $\mathcal{Q}$: Is this set of points normally distributed?
  - **Shapiro-Wilk** (the most sensitive)*
  - **Anderson-Darling**,
  - **Kolmogorov-Smirnov**
  - a **Q-Q plot** (quartile-quartile plot)

- $\mathcal{Q}$: Are the variances between two of my sets of points different? (*a.k.a.* $\mathcal{Q}$: Is my data homoscedastic?)

*Razali & Wah (2011): Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests Journal of Statistical Modelling and Analytics 2(1):21–33. ISBN: 978-967-363-157-5
Link via ResearchGate (Apr 2021)

Preliminary Tools

Some questions will depend on the answers to one or both of these questions, as applied to each of your sets of points:

- $\mathcal{Q}$: Is this set of points normally distributed?
  - **Shapiro-Wilk** (the most sensitive)*
  - **Anderson-Darling**,
  - **Kolmogorov-Smirnov**
  - a **Q-Q plot** (quartile-quartile plot)

- $\mathcal{Q}$: Are the variances between two of my sets of points different? (*a.k.a.* $\mathcal{Q}$: Is my data homoscedastic?)
  - ***F*-test (of equality of variances)**

*Razali & Wah (2011): Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests Journal of Statistical Modelling and Analytics 2(1):21–33. ISBN: 978-967-363-157-5
Link via ResearchGate (Apr 2021)

$\mathcal{Q}$: Is there a significant difference between the distributions represented by the points making up my data sets?

If all of the following are true:

- you have **exactly two** distributions of data (sets of points),

$\mathcal{Q}$: Is there a significant difference between the distributions represented by the points making up my data sets?

If all of the following are true:

- you have **exactly two** distributions of data (sets of points),
- and they are **normally distributed**

Statistics:

$\mathcal{Q}$: Is there a significant difference between the distributions represented by the points making up my data sets?

If all of the following are true:

- you have **exactly two** distributions of data (sets of points),
- and they are **normally distributed**
- and your observations are **independent**

Statistics:

$\mathcal{Q}$: Is there a significant difference between the distributions represented by the points making up my data sets?

If all of the following are true:

- you have **exactly two** distributions of data (sets of points),
- and they are **normally distributed**
- and your observations are **independent**
- and your variances are equal (*a.k.a.* **homoscedasticity**)

Statistics:

$\mathcal{Q}$: Is there a significant difference between the distributions represented by the points making up my data sets?

If all of the following are true:

- you have **exactly two** distributions of data (sets of points),
- and they are **normally distributed**
- and your observations are **independent**
- and your variances are equal (*a.k.a.* **homoscedasticity**)

then

Statistics: Independent two-sample t-test

$\mathcal{Q}$: Is there a significant difference between the distributions represented by the points making up my data sets?

If all of the following are true:

- you have **exactly two** distributions of data (sets of points),
- and they are **normally distributed**
- and your observations are **independent**
- and your variances are equal (*a.k.a.* **homoscedasticity**)

then

- you can use the **(independent two-sample) t-test** to answer your question.

Statistics:

$\mathcal{Q}$: Is there a significant difference between the distributions represented by the points making up my data sets?

If all of the following are true:

- you have **more than two** distributions of data (sets of points),
- and they are **normally distributed**
- and your observations are **independent**
- and your variances are equal (*a.k.a.* **homoscedasticity**)
- and your question is "$\mathcal{Q}$: do these samples come from distributions with the same mean?"

then

Statistics: ANOVA

$\mathcal{Q}$: Is there a significant difference between the distributions represented by the points making up my data sets?

If all of the following are true:

- you have **more than two** distributions of data (sets of points),
- and they are **normally distributed**
- and your observations are **independent**
- and your variances are equal (*a.k.a.* **homoscedasticity**)
- and your question is "$\mathcal{Q}$: do these samples come from distributions with the same mean?"

then

- you can use **ANOVA** (Analysis Of VARiance)

but

Statistics: ANOVA

Q: Is there a significant difference between the distributions represented by the points making up my data sets?

If all of the following are true:

- you have **more than two** distributions of data (sets of points),
- and they are **normally distributed**
- and your observations are **independent**
- and your variances are equal (*a.k.a.* **homoscedasticity**)
- and your question is "*Q*: do these samples come from distributions with the same mean?"

then

- you can use **ANOVA** (Analysis Of VARiance)

but

- this only tells you that one or more of the means of your distributions is different from the others, and doesn't tell you which one.

Statistics: ANOVA

$\mathcal{Q}$: Is there a significant difference between the distributions represented by the points making up my data sets?

If all of the following are true:

- you have **more than two** distributions of data (sets of points),
- and they are **normally distributed**
- and your observations are **independent**
- and your variances are equal (*a.k.a.* **homoscedasticity**)
- and your question is "$\mathcal{Q}$: do these samples come from distributions with the same mean?"

then

- you can use **ANOVA** (Analysis Of VARiance)

but

- this only tells you that one or more of the means of your distributions is different from the others, and doesn't tell you which one.
- You run a **post hoc test** to determine that.

Statistics:

$\mathcal{Q}$: Is there a significant difference between the distributions represented by the points making up my data sets?

If all of the following are true:

- you collected your points in **paired observations** from a set of subjects or entities, so that you have **two measures** from the **same subject** under two conditions, and you want to know whether the change in condition produced a change in measurement between the **paired measures**
- and **these differences** are **normally distributed**

Statistics:

$\mathcal{Q}$: Is there a significant difference between the distributions represented by the points making up my data sets?

If all of the following are true:

- you collected your points in **paired observations** from a set of subjects or entities, so that you have **two measures** from the **same subject** under two conditions, and you want to know whether the change in condition produced a change in measurement between the **paired measures**
- and **these differences** are **normally distributed**
- and your observations **between subjects** are **independent**

Statistics:

Q: Is there a significant difference between the distributions represented by the points making up my data sets?

If all of the following are true:

- you collected your points in **paired observations** from a set of subjects or entities, so that you have **two measures** from the **same subject** under two conditions, and you want to know whether the change in condition produced a change in measurement between the **paired measures**
- and **these differences** are **normally distributed**
- and your observations **between subjects** are **independent**
- and your question is "*Q*: do the differences described by these pairs have a non-zero mean?" (*i.e.*; "*Q*: Is there a difference between these paired values?)

then

Statistics:

$\mathcal{Q}$: Is there a significant difference between the distributions represented by the points making up my data sets?

If all of the following are true:

- you collected your points in **paired observations** from a set of subjects or entities, so that you have **two measures** from the **same subject** under two conditions, and you want to know whether the change in condition produced a change in measurement between the **paired measures**
- and **these differences** are **normally distributed**
- and your observations **between subjects** are **independent**
- and your question is "$\mathcal{Q}$: do the differences described by these pairs have a non-zero mean?" (*i.e.*; "$\mathcal{Q}$: Is there a difference between these paired values?)

then

- you can use the **(paired)** t**-test** to answer your question.

Statistics: Paired sample t-test

$\mathcal{Q}$: Is there a significant difference between the distributions represented by the points making up my data sets?

If all of the following are true:

- you collected your points in **paired observations** from a set of subjects or entities, so that you have **two measures** from the **same subject** under two conditions, and you want to know whether the change in condition produced a change in measurement between the **paired measures**
- and **these differences** are **normally distributed**
- and your observations **between subjects** are **independent**
- and your question is "$\mathcal{Q}$: do the differences described by these pairs have a non-zero mean?" (*i.e.*; "$\mathcal{Q}$: Is there a difference between these paired values?)

then

- you can use the **(paired)** t**-test** to answer your question.

Statistics: Paired sample t-test

Here is something that people frequently get mixed up for the paired t-test ...

The Normality requirement only applies **to the differences**

Note that only the **differences** calculated between the pairs need to be normally distributed – the distributions of the unpaired points <u>does not matter at all</u>.
This is because the paired t-test **converts the set of differences into a single distribution**, and uses a t-test to determine if **this distribution** is non-zero.

Bonferroni correction

If you have point sets $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ (or more), and want to make **multiple comparisons**, such as $\mathcal{A}$:$\mathcal{B}$ and $\mathcal{B}$:$\mathcal{C}$, then use a correction such as the **Bonferroni correction**:

- used to correct the estimation done when multiple inferences are made on the same data (*a.k.a.* the "**multiple comparisons** problem")

Bonferroni correction

If you have point sets $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ (or more), and want to make **multiple comparisons**, such as $\mathcal{A}$:$\mathcal{B}$ and $\mathcal{B}$:$\mathcal{C}$, then use a correction such as the **Bonferroni correction**:

- used to correct the estimation done when multiple inferences are made on the same data (*a.k.a.* the "**multiple comparisons** problem")
- built into most statistical packages, but math is trivial
- we just:
  - divide our α by the number of tests we are doing (making it lower), or equivalently

Bonferroni correction

If you have point sets $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ (or more), and want to make **multiple comparisons**, such as $\mathcal{A}$:$\mathcal{B}$ and $\mathcal{B}$:$\mathcal{C}$, then use a correction such as the **Bonferroni correction**:

- used to correct the estimation done when multiple inferences are made on the same data (*a.k.a.* the "**multiple comparisons** problem")
- built into most statistical packages, but math is trivial
- we just:
  - divide our α by the number of tests we are doing (making it lower), or equivalently
  - multiply our p by the number of tests we are doing (making them higher).

Bonferroni correction

If you have point sets $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ (or more), and want to make **multiple comparisons**, such as $\mathcal{A}$:$\mathcal{B}$ and $\mathcal{B}$:$\mathcal{C}$, then use a correction such as the **Bonferroni correction**:

- used to correct the estimation done when multiple inferences are made on the same data (*a.k.a.* the "**multiple comparisons** problem")
- built into most statistical packages, but math is trivial
- we just:
  - divide our α by the number of tests we are doing (making it lower), or equivalently
  - multiply our p by the number of tests we are doing (making them higher).

Why do we do this?

This decreases our likelihood of saying that any test has produced a statistically significant result — because we are attempting to control for the likelihood of finding such a result by random chance.

Statistics: Non-parametric tests

Q: Is there a significant difference between the distributions represented by the points making up my data sets?

- If your data are **NOT normally distributed**
    - or your variances do not conform to your constraints
- but your observations are **independent**

then you can use:

---

†

Statistics: Non-parametric tests

$\mathcal{Q}$: Is there a significant difference between the distributions represented by the points making up my data sets?

- If your data are **NOT normally distributed**
  - or your variances do not conform to your constraints
- but your observations are **independent**

then you can use:

- **Mann-Whitney-Wilcoxon** (MWW)† if there are **two** distributions (analogue of the t-test, but with no model),

---

†

$\mathcal{Q}$: Is there a significant difference between the distributions represented by the points making up my data sets?

- If your data are **NOT normally distributed**
    - or your variances do not conform to your constraints
- but your observations are **independent**

then you can use:

- **Mann-Whitney-Wilcoxon** (MWW)† if there are **two** distributions (analogue of the *t*-test, but with no model), or

†This is also called the "**Mann-Whitney *U* test**" or "**Wilcoxon-rank-sum**" — but note that the "~~**Wilcoxon signed-rank**~~" test is a different beastie (more on next slide).

Statistics: Non-parametric tests

$\mathcal{Q}$: Is there a significant difference between the distributions represented by the points making up my data sets?

- If your data are **NOT normally distributed**
  - or your variances do not conform to your constraints
- but your observations are **independent**

then you can use:

- **Mann-Whitney-Wilcoxon** (MWW)† if there are **two** distributions (analogue of the t-test, but with no model), or
- Kruskal-Wallis (KW) if there are **more than two**.

†This is also called the "**Mann-Whitney U test**" or "**Wilcoxon-rank-sum**" — but note that the "~~Wilcoxon signed-rank~~" test is a different beastie (more on next slide).

Statistics: Non-parametric tests

Q: Is there a significant difference between the distributions represented by the points making up my data sets?

- If your data are **NOT normally distributed**
  - or your variances do not conform to your constraints
- but your observations are **independent**

then you can use:

- **Mann-Whitney-Wilcoxon** (MWW)† if there are **two** distributions (analogue of the t-test, but with no model), or
- Kruskal-Wallis (KW) if there are **more than two**.

Note that KW test functions like a **one-way ANOVA**, and to determine which pairs differ, you can use MWW without the **Bonferroni correction**.

†This is also called the "**Mann-Whitney *U* test**" or "**Wilcoxon-rank-sum**" — but note that the "~~**Wilcoxon signed-rank**~~" test is a different beastie (more on next slide).

Non-independent tests

Q: Is there a significant difference between the distributions represented by the points making up my data sets?

If your data are **NOT independent** then

- if you have **two** distributions (and ∴ would have considered **Mann-Whitney-Wilcoxon**) then

Non-independent tests

$\mathcal{Q}$: Is there a significant difference between the distributions represented by the points making up my data sets?

If your data are **NOT independent** then

- if you have **two** distributions (and $\therefore$ would have considered **Mann-Whitney-Wilcoxon**) then use the **Wilcoxon signed-rank** test and

Non-independent tests

$\mathcal{Q}$: Is there a significant difference between the distributions represented by the points making up my data sets?

If your data are **NOT independent** then

- if you have **two** distributions (and $\therefore$ would have considered **Mann-Whitney-Wilcoxon**) then use the **Wilcoxon signed-rank** test and
- if you have **more than two** distributions (and would have considered **Kruskal-Wallis**), then

Non-independent tests

$\mathcal{Q}$: Is there a significant difference between the distributions represented by the points making up my data sets?

If your data are **NOT independent** then

- if you have **two** distributions (and $\therefore$ would have considered **Mann-Whitney-Wilcoxon**) then use the **Wilcoxon signed-rank** test and
- if you have **more than two** distributions (and would have considered **Kruskal-Wallis**), then use the **Friedman** test instead.

Summary: Which test to use?

$\mathcal{Q}$: Is there a significant difference between the distributions represented by the points making up my data sets?

Independence	*Normality*	*N dist.*	$\rightarrow$	*Use*
Yes	Yes	2		
		3 or more		
	No	2		‡
		3 or more		

‡

◁Of differences only. See note on page 22.

Summary: Which test to use?

$\mathcal{Q}$: Is there a significant difference between the distributions represented by the points making up my data sets?

Independence	*Normality*	*N dist.*	$\rightarrow$	*Use*
Yes	Yes	2		***t*-test**
		3 or more		
	No	2		‡
		3 or more		

‡

◁Of differences only. See note on page 22.

Summary: Which test to use?

$\mathcal{Q}$: Is there a significant difference between the distributions represented by the points making up my data sets?

Independence	*Normality*	*N dist.*	$\rightarrow$	*Use*
Yes	Yes	2		***t*-test**
		3 or more		**ANOVA**
	No	2		‡
		3 or more		

‡

◁Of differences only. See note on page 22.

Summary: Which test to use?

$\mathcal{Q}$: Is there a significant difference between the distributions represented by the points making up my data sets?

Independence	*Normality*	*N dist.*	$\rightarrow$	*Use*
Yes	Yes	2		***t*-test**
		3 or more		**ANOVA**
	No	2		**Mann-Whitney-Wilcoxon**‡
		3 or more		

‡

◁Of differences only. See note on page 22.

Summary: Which test to use?

$\mathcal{Q}$: Is there a significant difference between the distributions represented by the points making up my data sets?

Independence	*Normality*	*N dist.*	$\rightarrow$	*Use*
Yes	Yes	2		***t*-test**
		3 or more		**ANOVA**
	No	2		**Mann-Whitney-Wilcoxon**[‡]
		3 or more		

‡ *a.k.a.* **Mann-Whitney *U* test** or **Wilcoxon rank sum** – Wilcoxon was a very busy person

◁ Of differences only. See note on page 22.

Summary: Which test to use?

$\mathcal{Q}$: Is there a significant difference between the distributions represented by the points making up my data sets?

Independence	*Normality*	*N dist.*	$\rightarrow$	*Use*
Yes	Yes	2		**t-test**
		3 or more		**ANOVA**
	No	2		**Mann-Whitney-Wilcoxon**‡
		3 or more		**Kruskal-Wallis**

‡ *a.k.a.* **Mann-Whitney U test** or **Wilcoxon rank sum** – Wilcoxon was a very busy person

◁ Of differences only. See note on page 22.

Summary: Which test to use?

Q: Is there a significant difference between the distributions represented by the points making up my data sets?

Independence	*Normality*	*N dist.*	$\rightarrow$	*Use*
Yes	Yes	2		***t*-test**
		3 or more		**ANOVA**
	No	2		**Mann-Whitney-Wilcoxon**‡
		3 or more		**Kruskal-Wallis**
No	n/a	2		
		3 or more		

‡ *a.k.a.* **Mann-Whitney *U* test** or **Wilcoxon rank sum** – Wilcoxon was a very busy person

◁ Of differences only. See note on page 22.

Summary: Which test to use?

$\mathcal{Q}$: Is there a significant difference between the distributions represented by the points making up my data sets?

Independence	*Normality*	*N dist.*	$\rightarrow$	*Use*
Yes	Yes	2		**t-test**
		3 or more		**ANOVA**
	No	2		**Mann-Whitney-Wilcoxon**[‡]
		3 or more		**Kruskal-Wallis**
No	n/a	2		
		3 or more		

[‡] *a.k.a.* **Mann-Whitney *U* test** or **Wilcoxon rank sum** – Wilcoxon was a very busy person

[◁] Of differences only. See note on page 22.

Summary: Which test to use?

$\mathcal{Q}$: Is there a significant difference between the distributions represented by the points making up my data sets?

Independence	*Normality*	*N dist.*	$\rightarrow$	*Use*
Yes	Yes	2		***t*-test**
		3 or more		**ANOVA**
	No	2		**Mann-Whitney-Wilcoxon**[‡]
		3 or more		**Kruskal-Wallis**
No	n/a	2		**Wilcoxon signed-rank**
		3 or more		**Friedman**

[‡] *a.k.a.* **Mann-Whitney *U* test** or **Wilcoxon rank sum** – Wilcoxon was a very busy person

◁Of differences only. See note on page 22.

Summary: Which test to use?

$\mathcal{Q}$: Is there a significant difference between the distributions represented by the points making up my data sets?

Independence	*Normality*	*N dist.*	$\rightarrow$	*Use*
Yes	Yes	2		**t-test**
		3 or more		**ANOVA**
	No	2		**Mann-Whitney-Wilcoxon**‡
		3 or more		**Kruskal-Wallis**
No	n/a	2		**Wilcoxon signed-rank**
		3 or more		**Friedman**
Yes	Yes◁	Pairs		**paired t-test**

‡ *a.k.a.* **Mann-Whitney *U* test** or **Wilcoxon rank sum** – Wilcoxon was a very busy person

◁ Of differences only. See note on page 22.