Boring but important disclaimers:

- If you are not getting this from the GitHub repository or the associated Canvas page (e.g. CourseHero, Chegg etc.), you are probably getting the substandard version of these slides Don't pay money for those, because you can get the most updated version for free at

  `https://github.com/julianmak/academic-notes`

  The repository principally contains the compiled products rather than the source for size reasons.
- Associated Python code (as Jupyter notebooks mostly) will be held on the same repository. The source data however might be big, so I am going to be naughty and possibly just refer you to where you might get the data if that is the case (e.g. JRA-55 data). I know I should make properly reproducible binders etc., but I didn't...
- I do not claim the compiled products and/or code are completely mistake free (e.g. I know I don't write Pythonic code). Use the material however you like, but use it at your own risk.
- As said on the repository, I have tried to honestly use content that is self made, open source or explicitly open for fair use, and citations should be there. If however you are the copyright holder and you want the material taken down, please flag up the issue accordingly and I will happily try and swap out the relevant material.

OCES 3301 :

basic **Data Analysis** in ocean sciences

Session 6: statistical tests

January 22, 2025

Outline

(Just overview here; for actual content see Jupyter notebooks)

- recall hypothesis testing (mostly focus on *analysis*)
  - $\rightarrow$ null hypothesis etc.
  - $\rightarrow$ Z-test
  - $\rightarrow$ (Student's) t-test
  - $\rightarrow$ χ^2
  - $\rightarrow$ ANOVA + F-test
- some others

Recall: sea cucumber

Figure: Moldy sea cucumber.

e.g. say from samples,

$$\{\mu_1 = 3.00, \quad \sigma_1 = 0.5\}$$

$$\{\mu_2 = 3.20, \quad \sigma_2 = 0.5\}$$

- mean is different, so has effect?

  $\rightarrow$ but could just be a fluke?

- hypothesis testing as a tool to say whether differences are statistically significant

Recall: Hypothesis testing

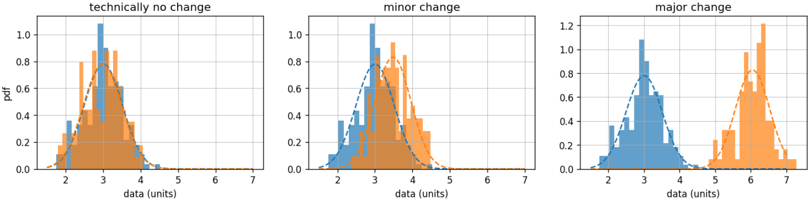


Figure: Control and varied sample distributions and associated Gaussian pdf.

- ▶ we are basically dealing with samples (going to assume CLT holds here)

Q. variability in data always exist, so how to distinguish change from noise?

? non-overlapping confidence intervals?

$\rightarrow$ quite strict (and **under-powered**; see later)

Recall: Hypothesis testing

Start with a null hypothesis H_0 (opposite to what you want to show usually)

- assume H_0
- decide test and significance level (depends on the thing you want to show)
- compute test statistics (depends on test)
- if associated probability of computed test statistic is low, then it is either:
  1. a really surprising result
  2. or H_0 is incompatible with data
- if latter, reject H_0, and there is statistical evidence in support for *not* H_0 (which is the thing you wanted anyway)

Recall: Hypothesis testing

e.g. sea cucumber, want to know if diet has any effect on weight

H_0 : diet has **NO** bearing on weight

test : large enough samples, assume Gaussian statistics, do two-tailed **Z-test**

α : choose $\alpha = 0.05$

$\rightarrow$ how far you are in tails of the pdf

$\rightarrow$ for Gaussian pdf, corresponds to Z-score of around 2, because 95% CI is around $(-2\sigma, +2\sigma)$

compute : compute Z-statistics (see notebook)

conclude : if Z-statistic large or corresponding p-value small, then reject H_0

(see actual code syntax in notebook)

Recall: Type I and II errors

Type I errors (false-positives)

- rejecting H_0 when H_0 is true

  $\rightarrow$ related to choice of significance α

Type II errors (false-negatives)

- fail to reject H_0 when H_0 is false

	H_0 true	H_0 false
reject H_0	Type I	✓
fail to reject H_0	✓	Type II

or, with H_0 = someone is innocent,

	innocent	murderer
found guilty	wrongful conviction	✓
found not guilty	✓	fail to prosecute

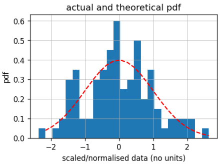
Recall: Z-test

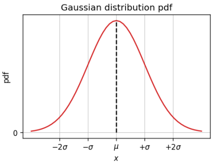
- Z-test calculates Z-statistic with sample mean $\overline{x}$

$$Z = \frac{\overline{x} - \mu}{\sigma / \sqrt{N}}$$

$\rightarrow$ if Z big enough (so the associated p-value is small) then evidence to reject H_0

$\rightarrow$ Gaussian, large samples, known σ (could approximate with s)

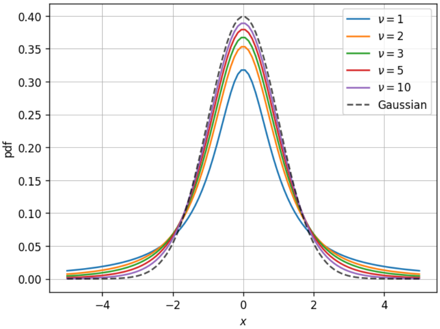




(Student's) t-test

- Gaussian data, **small** samples, unknown σ
- t-statistic pdf is

$$p(x) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\pi\nu}\,\Gamma(\nu/2)} \left(1 + \frac{x^2}{\nu}\right)^{-(\nu+1)/2},$$



$\rightarrow$ ν ('nu') is called the degree of freedom (d.o.f.)

$\rightarrow$ Γ ('gamma') is the Gamma function

(Student's) *t*-test

- Z-test statistic is so easy to compute there is no inbuilt function (?)...

(Student's) *t*-test

- Z-test statistic is so easy to compute there is no inbuilt function (?)...
- ...but *t*-tests does exist in e.g. scipy (see notebook)

  $\rightarrow$ d.o.f., related to sample size, is computed automatically

  $\rightarrow$ with H_0 that mean is the **same**

  $\rightarrow$ gives *t*-statistic and *p*-values

(Student's) t-test

- Z-test statistic is so easy to compute there is no inbuilt function (?)...
- ...but t-tests does exist in e.g. scipy (see notebook)
  $\rightarrow$ d.o.f., related to sample size, is computed automatically
  $\rightarrow$ with H_0 that mean is the **same**
  $\rightarrow$ gives t-statistic and p-values

- a few types
  $\rightarrow$ one sample (test against population mean)
  $\rightarrow$ two sample paired (test against each other's means)
  $\rightarrow$ two sample independence / unpaired (as above)

χ^2-test

- categorical data with discrete probabilities (e.g. dice, population distributions)
- χ^2-test ('kai squared') pdf

$$p(x) = \frac{x^{\nu/2-1} e^{-x/2}}{2^{\nu/2}\Gamma(\nu/2)} \qquad \text{if } x > 0,$$

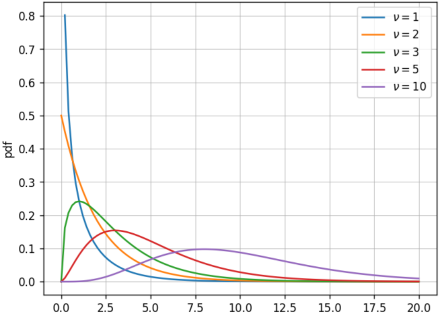


Figure: χ^2 pdf.

χ^2 test

- scipy has this (see notebook)

  $\rightarrow$ computes d.o.f.

  $\rightarrow$ with H_0 that means are the **same**

  $\rightarrow$ gives χ^2-statistic and p-values

- sample applications

  $\rightarrow$ how to tell if a die (singular of dice) is 'fair'? (see notebook)

  $\rightarrow$ sea cucumber activity affected by seasonality

  $\rightarrow$ some ocean applications you can make up

- robust but "weak"

  $\rightarrow$ tells you there are deviations of distribution, but doesn't tell you where

ANOVA and F-test

- to probe whether the **means** in groups of data are different to each other, by analysing their **variances**

  $\rightarrow$ ANOVA = ANalysis Of **VAriance**
- idea: if variance <u>between</u> groups dominates variance <u>within</u> individual group, then the means **between** groups are significant

  $\rightarrow$ high F regime

- F-statistic pdf looks like χ^2

  $\rightarrow$ one-way F-test works on H_0 = all means are the same

  $\rightarrow$ large F-statistic is p value small

ANOVA and F-test

- use this to avoid **multiple testing** between pairs of data
  $\rightarrow$ reduce false-positives / Type I errors

- robust but “weak”
  $\rightarrow$ tells you there is/are pair(s) with different means, but doesn’t tell you where

- not the only use, but not going to go through the others

- beware of assumptions going into test
  $\rightarrow$ e.g. normality, homogeneity, …

Other pre/post-analysis tests

- plot out the data!
- Shapiro–Wilk's test

  → normality of data
- Levene and/or Bartlett's test

  → homogeneity
- Tukey's Honest Significance Test

  → post-analysis for F-test

  → sometimes Tukey's range test, Tukey HSD, etc.

Jupyter notebook

go to 06 Jupyter notebook to get some code practise

- extended example on the Shannon diversity index / Shannon entropy there also (doesn't quite fit anywhere...)

  $\rightarrow$ feature importance, uses in ecology, machine learning...
- all banana skins previously apply here

Statistics is just a <u>tool</u>, no more and no less

- YOU are the user and the onus is on YOU to know enough about to tool to not abuse it