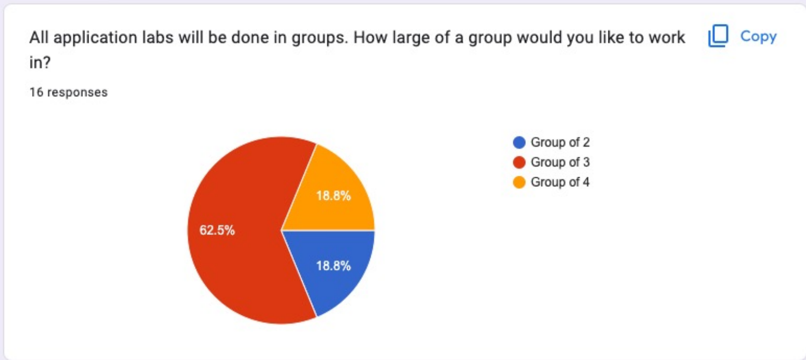
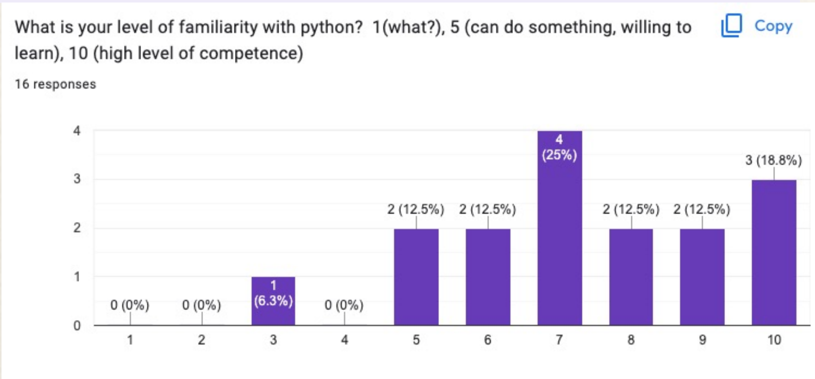
ANNOUNCEMENTS

1) Thanks for completing pre-class survey. I'm so looking forward to working together with you.. Some thoughts… (*See Slack for more*)





2) Are you ready to "learn by doing" and collaborate using jupyter notebooks and python? ***Starting next class, I will expect you to check that you can run the code prior to class. Class is not the time to update/install python. Let me know of any issues via Slack.***

Today’s plan…. Barnes 1.3-1.5

1) Statistical significance testing

- Normal Distribution, including “standardizing"

- Applying the Z-statistic

- Central Limit Theorem

- Applying the T-statistic (for small N<30 normally distributed samples)

2) Hypothesis Testing

- 5-steps

- type I and type II errors; a priori vs. a posteriori

3) Resampling and Monte Carlo techniques

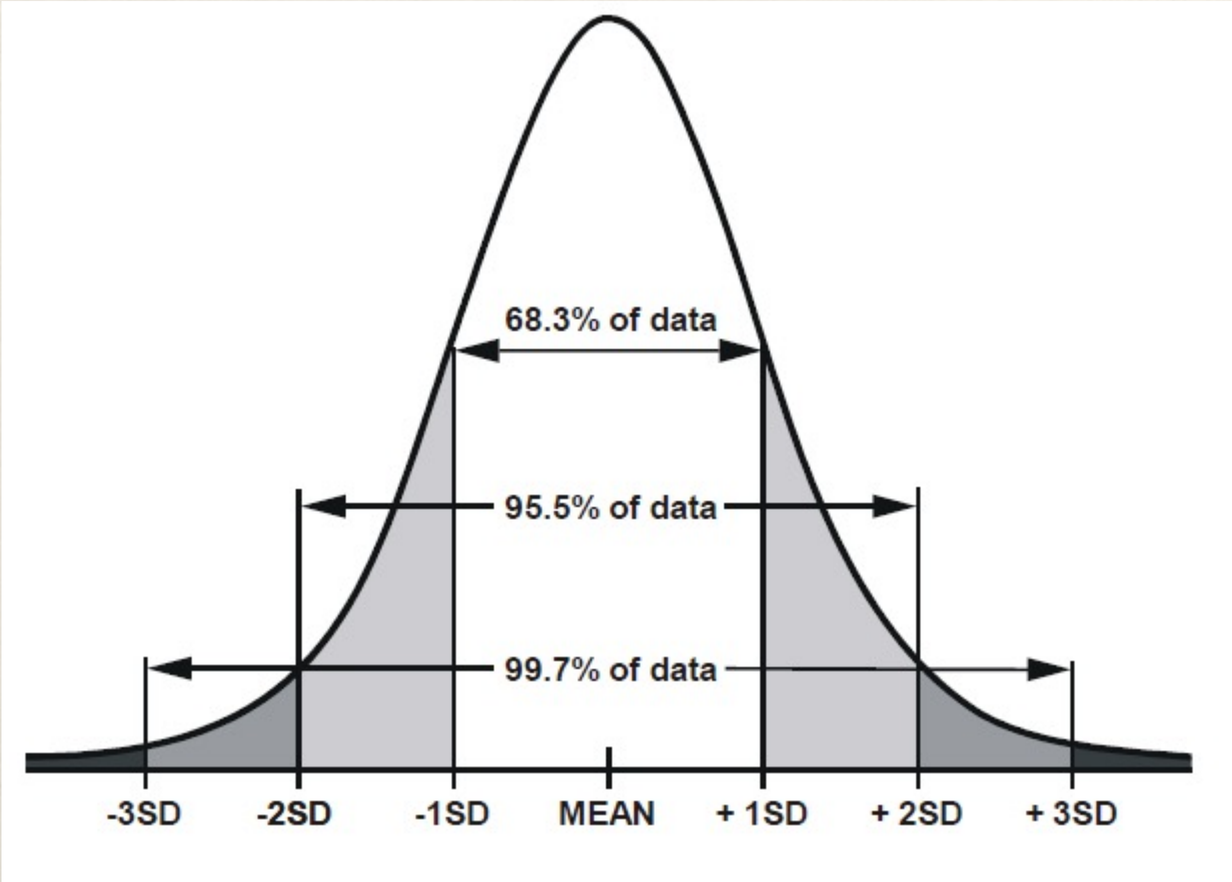
- Resampling: Bootstrap, Jacknife

- Montecarlo

Normal Distribution

The probability density function for a variable x that is normally distributed about its mean is given by:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)} \tag{68}$$



When we "standardize" a normal distribution - the mean is equal to 0 and the standard deviation equal to 1:

Often, one standardizes the normal distribution by defining a variable z:

$$z = \frac{x - \mu}{\sigma} \tag{71}$$

Note that in this case, the mean of $z = 0$ and the standard deviation of $z = 1$. This is very useful for discussing properties of the normal distribution with others who may have variables with different means and standard deviations. For example, recall in the discussion of skewness and kurtosis, the standard normal has a skewness of 0 and a kurtosis of 3.

If one standardizes to z, then

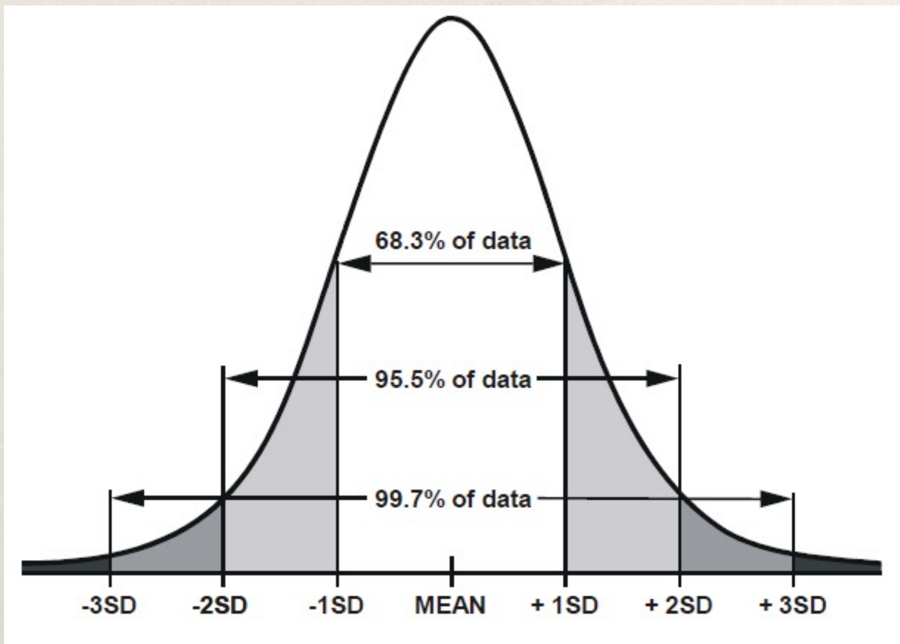
$$f(z) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-z^2}{2}\right) \tag{72}$$

Use a standard normal distribution to answer the following:
What is the probability of getting a given value of z?

$$Pr(-1 \leqslant z \leqslant 1) = 68.27\% \tag{73}$$
$$Pr(-2 \leqslant z \leqslant 2) = 95.45\% \tag{74}$$
$$Pr(-3 \leqslant z \leqslant 3) = 99.73\% \tag{75}$$

Recall that z is a standardized normal variable, and so, $z = 2$ is $z = 2$ standard deviation.

Thus, there is a 4.55% probability that z will fall outside of 2 standard deviations of its mean (two-tailed probability) and a 2.275% change it will exceed +2 standard deviations (one-tailed probability).

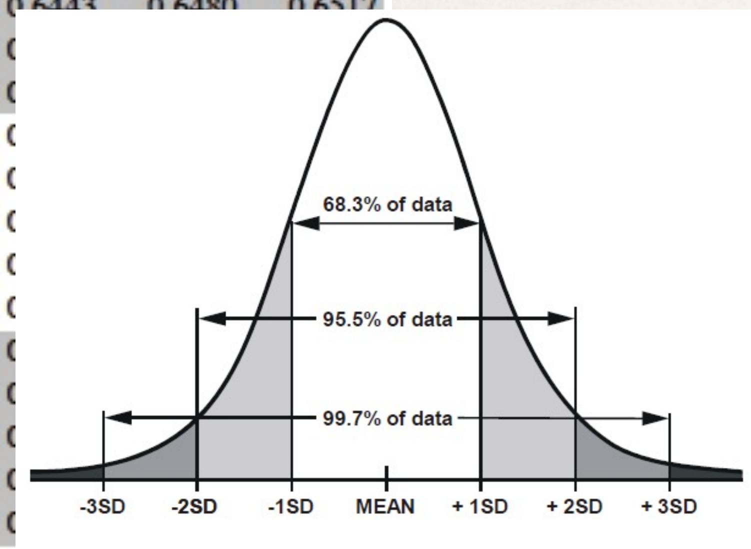


Let's use a z-statistic table to assess 1-sided probabilities

What is the probability of z<1 or z<2?

Table of the Cumulative Normal Probability distribution F(z).

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772			
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123			
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454			
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764			
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051			
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315			
1.0	$\boxed{0.8413}$	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554			
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770			
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962			
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131			
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279			
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406			
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	$\boxed{0.9772}$	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817



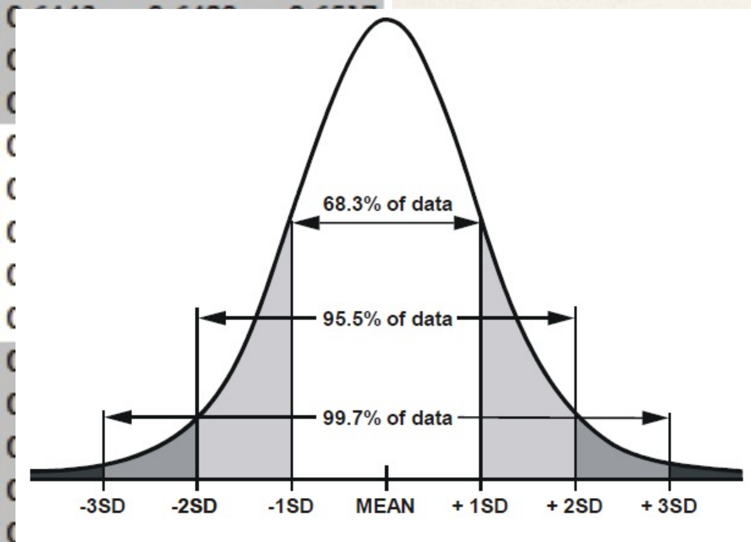
84.13% of the distribution has a value less than z=1;
97.72% of the distribution has a value less than z=2

Let's use the z-statistic table to assess 2-sided probabilities

What is the probability of that z lies within 1 standard deviation of the mean?

Table of the Cumulative Normal Probability distribution F(z).

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	[illegible]	[illegible]	[illegible]
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	[illegible]	[illegible]	[illegible]
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	[illegible]	[illegible]	[illegible]
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	[illegible]	[illegible]	[illegible]
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	[illegible]	[illegible]	[illegible]
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	[illegible]	[illegible]	[illegible]
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	[illegible]	[illegible]	[illegible]
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	[illegible]	[illegible]	[illegible]
1.1	[illegible]	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	[illegible]	[illegible]	[illegible]
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	[illegible]	[illegible]	[illegible]
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	[illegible]	[illegible]	[illegible]
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	[illegible]	[illegible]	[illegible]
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	[illegible]	[illegible]	[illegible]
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817



100-84.13%=15.87% of the distribution has values greater than z=1.
100-84.13%=15.87% of the distribution has values less than z=-1.
Therefore, 68.27% of the distribution has values between -1 < z < 1.

More often we are interested in the difference between a sample mean and its underlying population…

Question: Is this sample mean consistent with the population mean?

Step #1: Calculate the sample mean and sample standard error.

Let $X = [x_1, x_2, x_3, x_4, ..., x_N]$ be a random sample of size N drawn from a normal distribution with population mean μ and standard deviation σ.

For $N > 30$, $\overline{X}$ is normally distributed with,

$$\mu_{\overline{X}} = \mu \quad = \quad \text{mean} \tag{80}$$

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{N}} \quad = \quad \text{standard error of the mean} \tag{81}$$

In other words, the sample mean is an estimate of the population mean, and the error of your estimate (the variance of the distribution of $\overline{X}$) decreases as N increases (as it should).

$\sigma_{\overline{X}}$ is known as the standard error of the mean.

Question: Is this sample mean consistent with the population mean?

Step #2: Calculate the z-statistic. The z-statistic is now the number of standard errors that the sample mean deviates from the population mean.

So, how do we test whether the mean of the sample is different from the mean of the population? We can use the z-statistic!

Recall,

$$z = \frac{x - \mu}{\sigma} \tag{82}$$

Now, we want to test $\overline{X}$, not x. So, replace μ with the population mean of $\overline{X}$ and σ with the population standard deviation of $\overline{X}$.

Plugging things in leads to

$$z = \frac{\overline{X} - \mu_{\overline{X}}}{\sigma_{\overline{X}}} = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{N}}} \tag{83}$$

Example – Have ENSO dynamics changed?

1.3.3.1 Statistical significance of the mean

(a) What is the probability that December 2013 had an ENSO index of 0.50 or greater?

(b) What is the probability that the average 2011-2013 monthly ENSO index was 0.50 or greater assuming that ENSO dynamics have not changed?

(a) Assume the ENSO index is standard normal. $\mu = 0, \sigma = 1$. Look at table for $z = 0.50$. **1-.69 = 31.0%**. So, not very rare.

(b) Here, we are testing for the sample mean.

$$\overline{X} = 0.5 \tag{85}$$

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{N}} = \frac{1}{\sqrt{36}} \tag{86}$$

$$z = \frac{\overline{X} - \mu_{\overline{X}}}{\sigma_{\overline{X}}} = \frac{0.50 - 0}{.1667} = 3.0 \tag{87}$$

The $\boldsymbol{Pr}(z \geqslant 3.0) \approx 1 - 0.9987 = 0.1\%$.

Such a low probability implies that we had either a very rare event, or, that the dynamics of ENSO changed in 2011-2013 compared to climatological ENSO variability.

We can also use a z-statistic to test differences in two sample means:

Question: Is sample mean #1 different than sample mean #2?

We can manipulate the equation for the z-statistic to obtain an equation for the difference of two means (as opposed to the difference between a sample mean and the population), where σ_1 can be different from σ_2:

$$z = \frac{(\overline{x_1} - \overline{x_2}) - \Delta_{1,2}}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}} \tag{84}$$

$\Delta_{1,2}$ is the hypothesized difference between the two means, which is typically 0 in practice.

Example – Calculating Confidence Limits using Z-statistics

Question: What interval contains the true parameter 95% of the time?

1.3.4.1 Example: Calculating confidence limits

Say we have temperatures for 30 winters with a mean of 10° C and a standard deviation of 5° C. What is the 95% confidence interval on the true population mean? You may assume that the temperatures are normally distributed.

$$z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{N}}} \tag{88}$$

$$\mu = \overline{x} - z\frac{\sigma}{\sqrt{N}} \tag{89}$$

Since we are interested in the spread about the mean, and since the normal distribution is symmetric about the mean, this can be rewritten as,

$$\mu = \overline{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{N}} \tag{90}$$

Since 95% of the population lies within $z_{\alpha/2} = \pm 1.96$, it follows that the confidence limit for μ is

$$\mu = 10.0 \pm 1.96\frac{5.0}{\sqrt{30}} = 10.0^{\circ} \pm 1.7^{\circ} \tag{91}$$

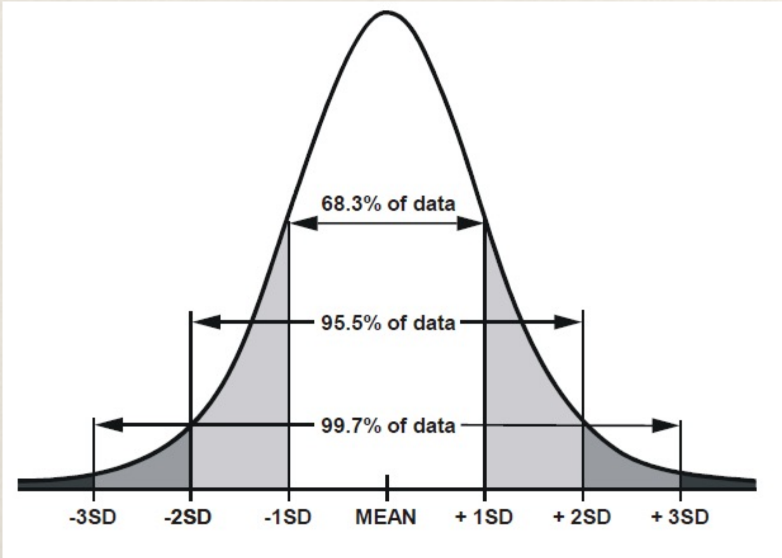
$$8.3^{\circ} \leqslant \mu \leqslant 11.7^{\circ} \tag{92}$$

Central Limit Theorem:
Large N (More Data) matters. Large N makes things Normal.

The arithmetic mean of a sufficiently large number of iterates of independent random variables, each with a well-defined expected value and well-defined variance, will be approximately normally distributed, with standard error of $\sigma/\sqrt{N}$, where N is the length of each sample.

This theorem is the basis of most of what we do in statistics, and most of what we do when assessing the significance of geophysical signals.

What the central limit theorem says is that if you have a sample that is large enough, you can use the normally distributed z-statistic to estimate probabilities of getting that mean - no matter the distribution of the underlying data.



Central Limit Theorem: It does not always apply!

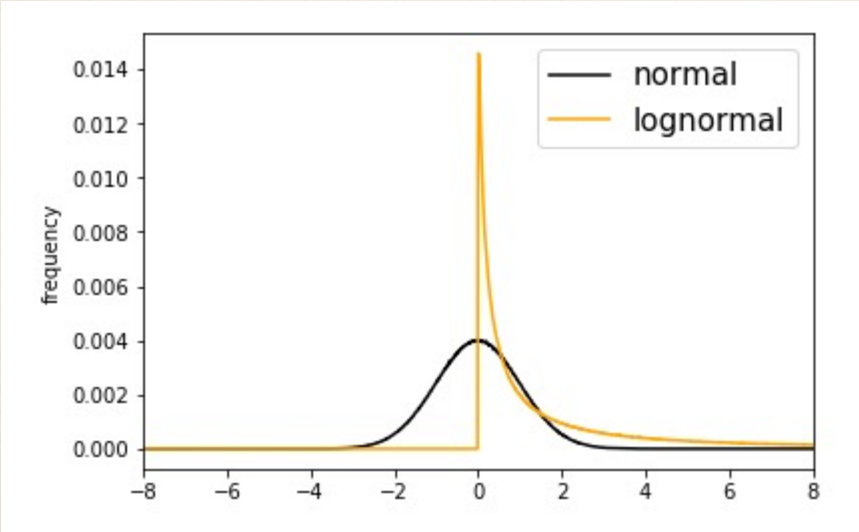
Question: Does it work for rain?

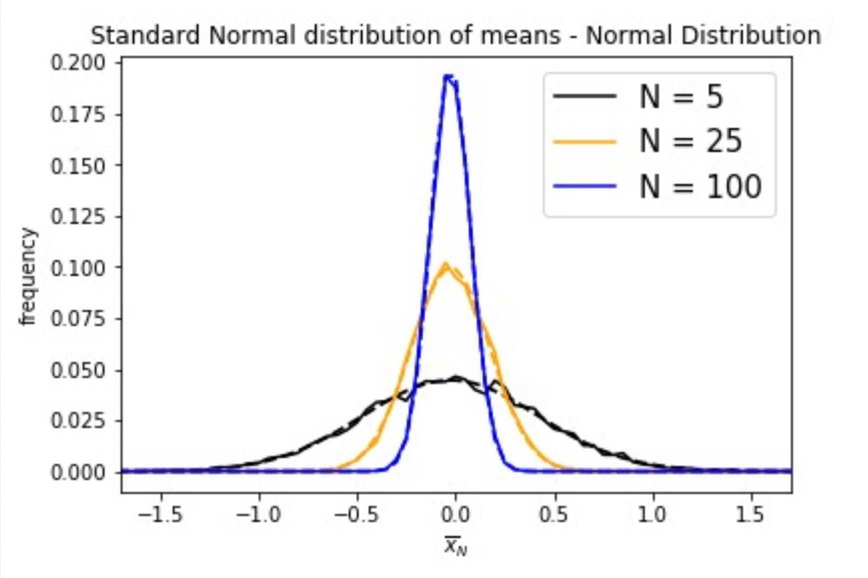
1.3.5.1 Example: Rain rate

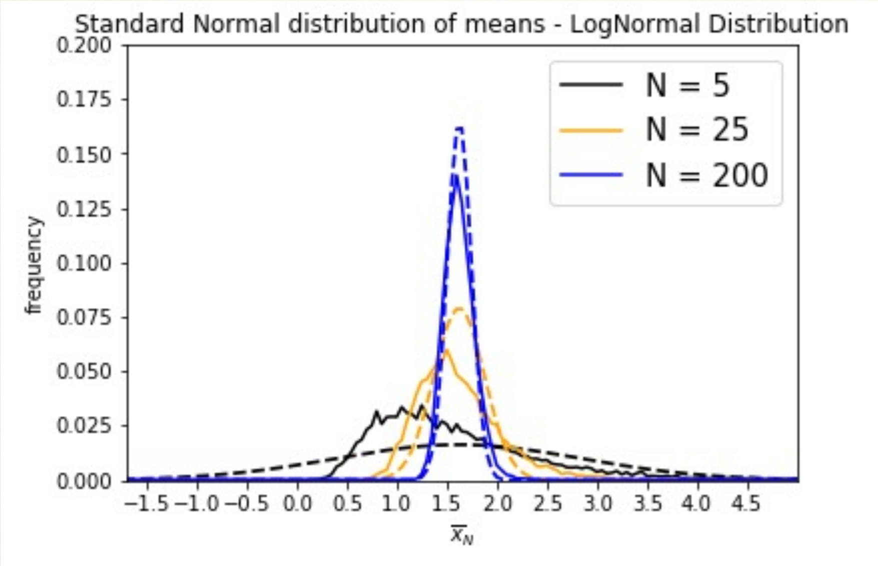
Rain rates measured at minute-intervals are lognormal. Thus, you *cannot* use the z-statistic to determine, say, the probability of getting a rain rate at any given time of 2 mm/sec. or higher.

However, if you want to know the probability of having a monthly average rain-rate of 2 mm/sec. or higher, you can use the z-statistic!

Central Limit Theorem Example in Python







Try the python code – central_limit_theorem.ipynb

Introducing the t-statistic….

Use it when the sample is small (N<30).

Consider the z-statistic for the sample mean:

$$z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{N}}} \qquad (93)$$

To apply this formula, we need to know the σ of the underlying distribution. But often, we don't know what the underlying distribution is. As we showed in the previous example, if we have a large enough sample, the sample standard deviation s is a good approximation of the true population σ.

However, if $N < 30$, this is not the case, and the z-statistic does not apply!

In this case, one must use the Student's t-statistic, introduced by William Sealy Gosset in 1908 to monitor Guinness quality at the brewery in Dublin, Ireland. At the time, he was not allowed to publish anything that he developed while at Guinness, so he used the pseudonym "Student".

The t-statistic is like the z-statistic but ...

The t-statistic is analogous to the z-score, except it also requires:

$$\begin{aligned} \nu &= \text{degrees of freedom} && (94) \\ \sqrt{N} &\rightarrow \sqrt{N-1} && (95) \end{aligned}$$

The t-statistic is defined as

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{N-1}}}, \tag{96}$$

where s is the sample standard deviation.

Recall from the Central Limit Theorem example that when $N \gtrsim 30$, the true standard deviation of the sample means is well approximated by $s/\sqrt{N}$, and so we can write σ.

However, in this case, we keep the nomenclature s since it is unknown what the true σ is.

Confidence intervals for the t-statistic work similarly to the z-statistic,

$$\mu = \bar{x} \pm t_c \frac{s}{\sqrt{N-1}} \tag{106}$$

t_c is the critical value for t. It depends on the sample size and the significance level desired. You can see values of this statistic in the t-table. Note that it is setup differently than the z-table since it requires a column for ν.

Degrees of Freedom in the t-statistic
Where did that N-1 come from?

ν is the number of degrees of freedom $= N - 1$. It is the number of independent samples (N) minus the number of parameters that must be estimated. In the case of the t-statistic, we calculate $\overline{x}$ and s but must estimate μ. Hence, $\nu = N - 1$. More on independent samples later.

Some notes on the t-statistic…

- Most often the t-distribution is applied when the sample mean is drawn from a normal distribution but with small N ($N<30$).
- The t-statistic depends on N through the degrees of freedom.
- Smaller values of N lead to longer tails for the t-statistic.
- As N increases, the t-statistic approaches the z-statistic.
- The KEY difference: t-statistic uses an estimate of the standard error based on the sample standard deviation. z-statistic uses the true standard error (population standard deviation).

Try the python code – z_t_pdfs.ipynb

You should find that the student-t probability density function is broader than the z probability density function… especially for small N.

DEMONSTRATION – Lecture #2 code

- **central_limit_theorem.ipynb**
- **z_t_pdfs.ipynb**

xterm

```
(base) jen@cu-biot-9-10 ~ % source activate culabenv2023clean
(culabenv2023clean) jen@cu-biot-9-10 ~ % jupyter notebook
```

TIPS for those programming in python (and those who are not!):

You can save jupyter notebook as .py code to be run at the command line.

1) **In Jupyter notebook – save out code as a .py file. Under File/Download as Python (.py),** *click accept at the bottom*
2) **Comment out anything with magic - recall # #get_ipython().magic('matplotlib inline')**
3) **Add plt.show() after plotting code to make plots show up on the computer screen when you run the python code.**
4) **To run the python code at the command line type "python X.py"**

You can also save jupyter notebook code as .html (show example).

Note: For homework, I will require you to submit your code in both .ipynb and .html format. More on homework format next Tuesday and on Slack…

Example – Compare the z-statistic and t-statistic

1.3.6.1 Example: Comparing the z-stat and t-stat

You have 5 years of monthly-mean temperature data derived from the MSU4 satellite. The mean temperature along 60N during January is ~ -60° C and the standard deviation is ~ 8° C. What are the 95% confidence limits on the true population mean?

---

z-statistic

The critical value $z_c = \pm 1.96$ for 95% confidence. Thus, the population mean μ is expected to lie within

$$-60 \pm 1.96\frac{8}{\sqrt{5}} = -60 \pm 7.0 \rightarrow -67.0 \leqslant \mu \leqslant -53.0 \qquad (104)$$

What about the t-statistic?

Example – Compare the z-statistic and t-statistic

1.3.6.1 Example: Comparing the z-stat and t-stat

You have 5 years of monthly-mean temperature data derived from the MSU4 satellite. The mean temperature along 60N during January is ~ -60^o C and the standard deviation is ~ 8^o C. What are the 95% confidence limits on the true population mean?

---

z-statistic

The critical value $z_c = \pm 1.96$ for 95% confidence. Thus, the population mean μ is expected to lie within

$$-60 \pm 1.96 \frac{8}{\sqrt{5}} = -60 \pm 7.0 \rightarrow -67.0 \leqslant \mu \leqslant -53.0 \qquad (104)$$

t-statistic

The critical value $t_c = \pm 2.78$ for $\nu = 5 - 1 = 4$. (Note: we want to use 0.025 since a two tailed test). Thus, the population mean μ is expected to lie within

$$-60 \pm 2.78 \frac{8}{\sqrt{4}} = -60 \pm 11.1 \rightarrow -71.1 \leqslant \mu \leqslant -48.9 \qquad (105)$$

Thus, using the t-statistic gives a wider confidence range than the z-statistic, reflecting the additional uncertainty associated with a small N. If we had erroneously used the z-stat instead of the t-stat, we would underestimate the 95% confidence bounds by 35%.

When to use the t-test?

If you are choosing between the z-test and t-test, use the t-statistic! It converges to the z-test for large N.

When applying the t-test, you are implicitly making a very strong assumption: *that the underlying distribution is normal.*

ALWAYS CHECK YOUR NORMALITY ASSUMPTION: plot your data, look at the skewness, kurtosis, etc. There are fancy tests to check for normality, e.g. the Kolmogorov-Smirnov test.

Now, we previously discussed the Central Limit Theorem, which tells us that for a "large enough" sample size, the distribution of sample means is normal. HOWEVER, note that the t-test applies for small N but for underlying *normal* distributions, whereas the Central Limit Theorem only applies for large N.

Thus, you cannot blindly apply the t-test to test differences in sample means if the underlying distributions are not already normal! This is a common mistake made in our field.

How many samples do you really have?
Are your samples independent?

More on how to assess that in unit 2…

In all of the cases thus far, it has been assumed that the N samples are *independent* samples. Often, N observations of a geophysical variable are not independent, for example, they exhibit either spatial or temporal correlations.

For example, the geopotential height is highly auto-correlated so that each day's value is not independent from the previous or following days. You cannot improve the your ability to know a 5-day wave by sampling every 3 hours instead of every 6. We will discuss this more later-on in the course.

Hypothesis Testing – Some Terminology

- significance/confidence level: α, typically 5% (0.05), often reported as $1-\alpha$
- critical value: t_c or z_c, the value that must be exceeded to reject the null hypothesis, one-sided t_α, two-sided $t_{\alpha/2}$
- p-value: probability of observing an effect given that the null hypothesis is true (probability of your t-score or z-score)

Hypothesis Testing – 5 basic steps – Write these Down!

1. State the significance level (α)
2. State the null hypothesis H_0 and the alternative H_1
3. State the statistic to be used, and the assumptions required to use it
4. State the critical region
5. Evaluate the statistic and state the conclusion

Hypothesis Testing – Example

Sam went skiing 10 times at Vail this winter. The average temperature on these 10 days was 35F, and the standard deviation of 10 daily temperatures is 5F. Sam knows that the climatological mean winter temperature for Vail is 32F. Is this a sign of climate change?

Let's suppose by a "sign of climate change", we mean "Is the 10-day average experience by Sam consistent with the null hypothesis that Vail temperatures have not warmed?" Let's walk through the 5-steps to answer this question.

Let μ be the true mean of the population from which the 10 days Sam went skiing were sampled.

1. let's use 95% confidence ($\alpha = 0.05$)

2. H_0: $\mu = 32^o$
   H_1: $\mu \neq 32^o$

3. For this problem, we will assume that the temperatures at Vail are normally distributed. Since we have $N = 10$ samples to estimate the true standard deviation (σ) from the sample standard deviation (s), we must use the t-test.

4. We will use a two-sided t-test (before he went skiing, Sam didn't know what to expect from the temperatures). Thus, to reject the null hypothesis we must have $t > t_{0.025} = 2.262$ (for $\nu = 9$)

5. $t = \frac{35-32}{\frac{5}{\sqrt{10-1}}} = 1.80$

   $t = 1.80 < t_{0.025}$, so we cannot reject the null hypothesis that the underlying population is different for the days Sam went skiing.

Hypothesis Testing – What to say/not say…!

Note, we should not say:

> "There was nothing different about the temperatures on the day Sam went skiing."
>
> *...nor...*
>
> "There has not been a change in temperatures at Vail over the past few decades."

We can only say

> "The data is consistent with the null hypothesis that the mean of the underlying population on the days Sam went skiing was the same as the climatology."
>
> *...or...*
>
> "The higher temperatures present when Sam went skiing are consistent with natural variability."

Errors in Hypothesis Testing

	H_0 is true	H_0 is false
Fail to Reject H_0	No Error	Type II Error (false negative)
Reject H_0	Type I Error (false positive)	No Error

The way typical hypothesis tests (frequentist approach) are setup, a 95% confidence level means you have a 5% chance of making a *Type I Error*, that is, you reject the null hypothesis (think you found something interesting) when you should not have. It is much more difficult to asses the Type II Error - the probability you "play it safe and fail to reject H_0 when something interesting was there". For typical hypothesis testing, the probability or a Type II error can be very large.

In engineering, you often care about the differences between Type I and Type II errors, and you design your statistics to reflect your judgements. For example, if H_0 is that the bridge will hold-up if 10 semi-trucks cross at the same time, and H_1 is that the bridge will not hold-up, you might be happier with a Type I Error, which requires that you redesign the bridge, rather than a Type II Error, where you think the bridge will be fine, and it won't be.

A priori (you have a reason to expect a relationship)
A posteriori (you don't have a reason)

If you have an "a priori" expectation of the sign of the result, you can use a 1-tailed test. Otherwise you should use a 2-tailed test.

An example of applying a posteriori statistics

You think that Arctic warming has caused blocking to increase in frequency between the 1980's and today. You look over the 4 seasons, two latitude bins (north of 45° N and south of 45°N), and 6 different longitude bins. You test for changes in mean blocking frequency at the 95% confidence level. How many "significant changes" in blocking should you expect by chance alone? How might you apply a posteriori statistics?

You have no a priori knowledge which season or longitude bin should exhibit changes in blocking due to sea ice - thus, you are giving the test $4 \times 6 \times 2 = 48$ chances to succeed. The number of changes you expect to show significant trends by chance alone is $0.05 \times 48 = 2.4$, so, you should find 2.4 "significant changes" by chance alone.

Let H_0 : the mean in blocking frequency has not changed.

$$
\begin{aligned}
\boldsymbol{Pr}(\text{correctly not reject } H_0 \text{ when it is true for one test}) &= 0.95 && (113)\\
\boldsymbol{Pr}(\text{correctly not reject } H_0 \text{ when it is true for all 48 tests}) &= 0.95^{48} \approx 9\% && (114)
\end{aligned}
$$

Thus, your 95% confidence level is really a 9% confidence level!

By trial and error, we can calculate the significance level β for which $\beta^{48} \approx 0.95$ (our a posteriori statistic). In this case, $\beta \approx 0.999$. Thus, if we require the blocking changes for each chance to pass at the 99.9% confidence level, then the probability of correctly not rejecting the null hypothesis for all chances will be 95%.

Resampling and Monte Carlo techniques

Monte Carlo and resampling techniques haven't historically been discussed or used as often because of the computational constraints. These days, computer time is cheap and everyone has one, so, these techniques can be found more regularly in the peer reviewed literature.

Resampling techniques: best used when you have a long climatology or control run

Monte Carlo: useful when you think you know the underlying distribution or behaviour, but you don't have a long control or climatology - thus, you create many synthetic "realities" to determine the probabilities of certain events

1.5.1 Why use resampling and Monte Carlo techniques?

1. when spatial and temporal correlations are hard to estimate
2. it is unclear which statistical assumptions are appropriate (e.g. your underlying population is not Gaussian, and the central limit theorem does not apply) - this is when resampling techniques are useful
3. the statistic of interest does not have a straight-forward theory for estimation, but you know the underlying population's distribution - this is when Monte Carlo simulations are useful

Resampling techniques: Bootstrap and Jacknife

Bootstrapping involves constructing a number of resamples of the original dataset (and of equal size to the observed sample of interest) by random sampling with replacement from the original dataset. In this way, you never need to assume anything about the underlying distribution of the data since it is already built-in to the original dataset. In essence, you ask, by random chance, what is the probability that a particular event (or sample statistic) occurred?

This method is also useful when you are determining statistics other than the mean (e.g. extrema, median, skewness) when we don't have simple statistics for these variables.

The jackknife method predates the bootstrap method. It is a way of getting errors bars (or measuring the variance) of your sample. You systematically remove one value from your sample, and calculate the statistic, then put the value back into the sample and remove the next value, calculate the statistics...and on and on. Then, following a simple formula, you can estimate the variance of the parameter by using the mean of the jackknife estimates of your sample parameter of interest.

You will find the jackknife approach often used to estimate the slope of a line through multiple points, or the y-intercept, etc.

Bootstrap example

You believe that aerosols grow the most when you have high geopotential heights nearby. You composite the 500 geopotential height on the 20 August days when you have aerosol formation and growth over a site in Egbert, Canada, and you find that the average geopotential on these days is 5900 m. The mean at this station is 5886 m, so the heights are higher. Are these results significant? Or is this just random chance?

Run 2,500 experiments, within each experiment, randomly grab 20 days from the historical geopotential height data, and take the mean of the 20 days. After 2,500 iterations, you will have a distribution of the $\mathsf{N} = 20$ sample means under the null hypothesis of random chance. Now, you can look at this distribution and determine the 95% confidence bounds on the $\mathsf{N} = 20$ sample means - if the observed value of 5900 m outside of this range, you have reason to believe it may be more than random chance.

Try the python code – subsampling_example.ipynb

Monte Carlo

1.5.4 Monte Carlo

Monte Carlo simulations require that you make an assumption about the underlying distribution. This is useful when you don't have a large enough base population to perform the bootstrap approach.

The idea is that you create bunch of synthetic (i.e. fake) data according to your distribution - you create many "realities" based on the same underlying distribution, and then you perform your analysis on these realities to determine the distribution of the statistical parameters of interest.

Monte Carlo example

In January (31 days), the maximum daily temperature was 2.2 standard deviations from the climatological mean temperature. If we assume that the daily temperature is normally distributed, how rare is it to have a maximum of 2.2σ or greater in 31 daily samples?

We do not have a test for the maximum of a distribution - note, *this is not the mean.*

We can't use the bootstrap approach, since we don't have the population to resample from, we only have our 31 points.

However, since our null hypothesis is that the values come from a standard normal, we can create synthetic data to determine the confidence interval on the maximum in a sample of $N = 31$.

Try the python code – monte_carlo_example.ipynb

GROUP WORK
Assigned groups on Slack
~10 minutes

- **Tinker with**
  subsampling_example.ipynb
  monte_carlo_example.ipynb

For Thursday:

First Application Lab!! You will work in the same groups as today.

Please come prepared – i.e., ready to run Application Lab #1 code in a Jupyter notebook your laptop. Verify the code works BEFORE class. Everything you need is in the class google drive.

In Application Lab #1…

We will be analyzing the statistical significance of global warming in the CESM1 Large Ensemble.

And assessing the influence of ENSO on Colorado (Loveland Pass) snowpack using composite analysis.

I also recommend reviewing Barnes Chapter 1 notes, esp. 1.1-1.5.

See you then, if not before ☺