Mexico Stats Exam

Created by: Josh Stewart, Brian Stock, and Lynn Waterhouse

March 23, 2017

Statistics with Applications to Fisheries Exam

Please enter your answer after the question.

This is just to evaluate what you know at the start of the class. If you do not know how to answer a question or are not sure how to do it, try your best, and feel to write down I do not (or am not sure) know how to do this right now.

Please put all your answers into a word document and save it as "PreTest_FirstLastName.doc", so if Josh took the pre test it would be "PreTest_JoshStewart.doc". If asked for, please paste your R code and provide the output. When you are done, please email it to Josh (joshua.stewart6@gmail.com). Please email the test to Josh by midnight on March 24th.

This pre-test will not be graded, it is just so we can see where you all are and evaluate how well the course goes at the end.

Identifying Distributions

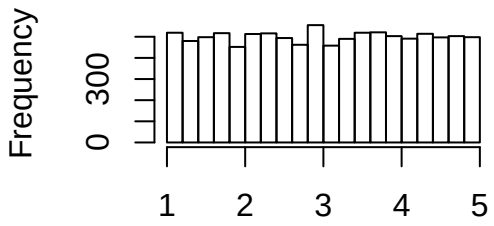
Problem 1.

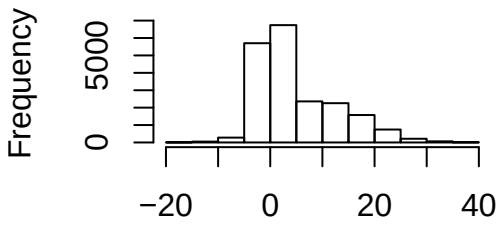
Label the distributions below which look like:

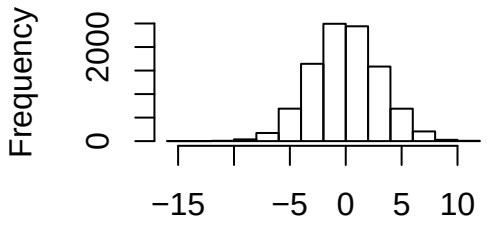
A. Normal

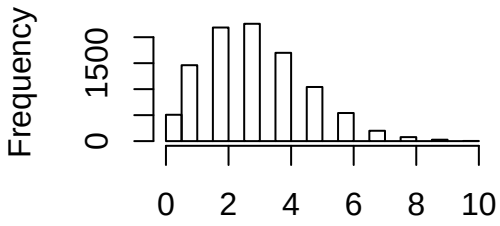
B. Poisson

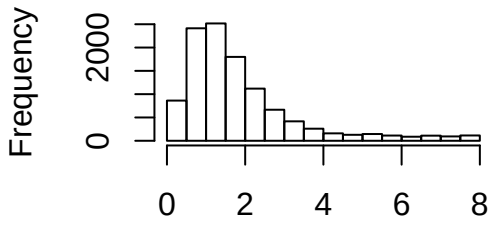
C. Uniform

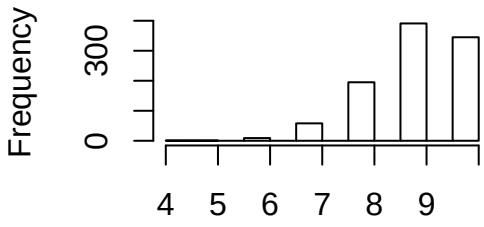












Problem 2.

Using R, read in the dataset "problem2.csv". Perform the appropriate test to evaluate if this data is normally distributed. Paste your R-code and output below.

Data Types.

For Problems 3 - 8. Identify the type of data discussed in the following examples. The type of data possible are: time series, ordinal, count, paired, continuous, categorical, and binomial. Note that some examples fall into multiple categories.

Problem 3.

At a fishing port, you track the number of every species of fish caught by the fishermen every day for 3 months.

Problem 4.

Sea state is classified using the Beaufort scale.

Beaufort.state	Wind.speed.km.hr	Wave.height.m	Description
0	< 1	0	mirror-like
1	1-5	0-0.2	ripples
2	6-11	0.2-0.5	wavelets
3	12-19	0.5-1	crests break
4	20-28	1-2	whitecaps
5	29-38	2-3	spray
6	39-49	3-4	foam crests

Problem 5.

For one fishing port in 2012 and 2013 you have the total number of fish caught. You want to see if there is a difference between 2012 and 2013.

Problem 6.

You study a population of grouper and collect length measurements of 500 fish.

Problem 7.

During a species' spawning season, you sample 200 females and record whether or not each female has eggs (sexual maturity).

Problem 8.

On a fishing boat, you record the species of each fish caught.

Hypothesis Testing, P-values, and Confidence Intervals

Problem 9.

Data has been collected on weights of Corvina caught in the southern edge of their range and in the northern edge of their range. We are curious if fish are different weights between the

two regions. We have 168 measurements of fish weights from each region. The mean and standard deviation for the southern region are 37 kg and 8 kg, respectively. The mean and standard deviation for the northern region are 56 kg and 12 kg respectively.

a. What is the appropriate test for this question?

b. What is the result of the test and what is the conclusion?

Problem 10.

At an aquaculture facility they want to know if a certain type of feed makes the fish gain weight. They measure the weights of all the fish in pen #1 and then feed then feed "superfeed XX9" for three weeks, they weigh the fish again at the end of the three weeks. They have tagged the fish so they know every fish's starting and end weight.

a. What would be the appropriate test to use to determine if they fish gained weight by eating "superfeedXX9"?

b. Write out the null and alternative hypotheses in mathematical terms.

c. If the p-value calculated from this test is 0.067 and the significance level is 0.05, what conclusion would you draw?

Problem 11.

A new MPA is being designed. The government wants to establish it where the most turtles will be protected. They are fighting for site "Blanca" as they believe it has more turtles. The alternative site is called "Rojo". You are given the mean number of turtles seen during the year at "Blanca" and "Rojo" along with the standard deviation and sample size.

a. What would be the appropriate test to use to determine if "Blanca"" has more turtles than "Rojo"?

b. Write out the null and alternative hypotheses in mathematical terms.

c. If the p-value calculated from this test is 0.032 and the significance level is 0.05, what conclusion would you draw?

Problem 12.

The average weight of snapper caught each day in Port DeLeche over 40 days is given by the below table.

Day	Avg.Weight.Snapper.kg
1	38
2	37

Day	Avg.Weight.Snapper.kg
3	62
4	44
5	48
6	32
7	55
8	25
9	44
10	43
11	38
12	44
13	40
14	31
15	48
16	33
17	33
18	41
19	52
20	43
21	39
22	43
23	21
24	48
25	52
26	40
27	25
28	39
29	41
30	43
31	40
32	51
33	35
34	39
35	43
36	42
37	47
38	48
39	12
40	34

a. Calculate the mean and standard deviation.

b. Using the rule of thumb calculate the 95% confidence interval.

c. Calculate the exact 95% confidence interval using R.

Statistical Tests

Problem 13.

You are interested in what factors influence a population of sea bass and decide to monitor 3 variables in a marine park each year for 12 years: sea bass abundance (bass), sea surface temperature (sst), and chlorophyll *a* (chla).

a. Plot the chlorophyll time-series in "problem13.csv".

b. When you plot your chlorophyll data, you notice that one value is extremely high—biologically improbable. You realize that you made a data entry mistake, using μg per m^3 instead of mg per m^3. Using R,

- Identify which year the mistake is for ('which' function)

- Replace the value in μg with the value in mg (multiply by 1000), and

- Plot the corrected chlorophyll time-series

c. Is there evidence that temperature affects bass abundance in your park?

d. Is there evidence that chlorophyll *a* affects bass abundance in your park?

e. Is there evidence that temperature AND chlorophyll *a* affect bass abundance (fit a model with both sst and chla)?

f. Which of the 3 models is best supported by your data?

g. Plot the data and fitted line from the best model.

Problem 14.

You collected total catch data from 300 landing sites in 2010 and 2011. You would like to know if the catch rates across your 300 sites increased from 2010 to 2011.

a. Plot the data in "LandingsData.csv" and make any necessary transformations. If you transform the data, plot it as both transformed and untransformed.

b. Perform a statistical test to determine if the landings between 2010 and 2011 changed. Report your p-values and conclusions

Problem 15.

Josh and I are interested if mantas are different sizes in different locations in Mexico. We collect size data at 3 locations (Yelapa, Revillagigedos, Cabo Pulmo) over 5 years (2011-2015). Some of your length measurements in Yelapa are on dead mantas caught by fisheries :(.

a. Read the data in "problem15.csv" into R. Use R to create a summary table, showing for each location across all years: number of mantas (*n.mantas*), mean length (*mean*), variance (*var*), min length (*min*), max length (*max*), number dead (*n.dead*), and percent dead (*perc.dead*). Round all numbers to 2 decimal places.

b. Show the length data by location in a boxplot

c. You want to perform an ANOVA to test for differences in mean length by location. What are the assumptions of ANOVA?

d. Perform an ANOVA in R. Show the diagnostics plots. Are the assumptions met?

e. Print the ANOVA summary table. Are there differences in mean length by location?

f. Do a Tukey test for pairwise differences. Which locations have different mean lengths?

Problem 16

The fishing port of Saturna catches 5 species complexes each year: sharks, bottomfish, forage fish, pelagics, and coastal species. The scientific question of interest is if the species composition is changing from year to year. The catch (in thousands) for years 2015 and 2016 are shown in the following table:

Species	Count.2015	Count.2016
sharks	150	100
bottomfish	600	540
forage fish	480	300
pelagics	200	230
coastal	140	120

a. Select the appropriate analysis

b.Check if all assumptions are met

c.Calculate the degrees of freedom and a p-value for the analysis

d.Interpret p-value and write conclusion

Problem 17.

You collected data on blue whale presence or absence, and krill density at 500 different sites. You think that the probability of whales being present is related to how much krill is available for them to feed on. You previously normalized the krill data so it has a mean of 0 and is bounded by [-1,1].

a. Plot the presence/absence observations as a function of krill density from "Whale data.csv"

b. Use a statistical model to determine whether or not there is a relationship between krill density and whale abundance. Plot the results of the model along with the whale presence/absence observations and comment on the model fit.

R Skills

Problem 18.

You have collected data on crab carapace diameters in the town of Yelapa. You plan to examine differences in sizes between crabs collected near the river, in town, and at the edge of the forest. But first you must get your data in an appropriate format for analysis. Read in your data from "CrabData.csv" and prepare it for analysis by placing it in 'Long' format and removing NA values.