assignment-4

Gage Clawson

11/13/2018

# read in the lobster traps dataset  
lob\_traps <- as.data.frame(read\_csv("lobster\_traps.csv"))

## Parsed with column specification:  
## cols(  
## YEAR = col\_integer(),  
## MONTH = col\_integer(),  
## DATE = col\_character(),  
## FISHING\_SEASON = col\_character(),  
## SITE = col\_character(),  
## SWATH\_START = col\_character(),  
## SWATH\_END = col\_character(),  
## TRAPS = col\_integer(),  
## OBSERVER = col\_integer(),  
## NOTES = col\_character()  
## )

# Read in the lobster size data set and make it tidy  
  
lob\_size <- read\_csv("lobster\_size\_abundance.csv")

## Parsed with column specification:  
## cols(  
## YEAR = col\_integer(),  
## MONTH = col\_integer(),  
## DATE = col\_character(),  
## SITE = col\_character(),  
## SBC\_LTER\_TRANSECT = col\_integer(),  
## LOBSTER\_TRANSECT = col\_character(),  
## SIZE = col\_integer(),  
## COUNT = col\_integer()  
## )

lob\_size1 <- as.data.frame(lob\_size)  
  
lob\_size2 <- expand.dft(lob\_size1, freq = "COUNT") ##expand.dft filters out all count = 0 rows and makes data tidy (each observation has it's own row)

Yes boss

lob\_size3 <- mutate(lob\_size2, DATE = as.Date(DATE, format = '%d-%b-%y')) # make date formats same for both datasets yyyy-mm-dd  
  
three\_sites <- c("AQUE", "CARP", "MOHK")  
  
lob\_traps1 <- mutate(lob\_traps, DATE = as.Date(DATE, format = '%m/%d/%y')) %>%  
 filter(SITE %in% three\_sites, TRAPS != 0) %>%  
 expand.dft(freq = "TRAPS")   
   
  
  
# make date formats the same and filter for only the 5 sites

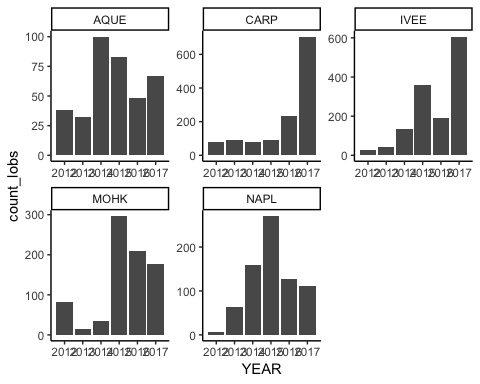
1. Lobster abundance and fishing pressure (2012 - 2017) Describe trends in lobster abundance (counts) and fishing pressure (trap buoys) at the five locations from 2012 - 2017. Ignore transect information - we are only interested in evaluating abundance and pressure on the order of SITE. Note: you are not expected to use regression here - just think of ways to clearly describe annual totals visually and in text, noting important trends, events and differences.

# find the counts of lobsters per each site per each year  
lob\_counts <- lob\_size3 %>%  
 group\_by(SITE, YEAR) %>%  
 dplyr::summarise(count\_lobs = length(SIZE))  
  
  
# find the counts of traps per each site per each year  
trap\_counts <- lob\_traps1 %>%  
 group\_by(SITE, YEAR) %>%  
 dplyr::summarise(count\_traps = length(OBSERVER))  
  
  
  
new\_lob\_traps <- left\_join(lob\_counts, trap\_counts, by = c("YEAR", "SITE"))

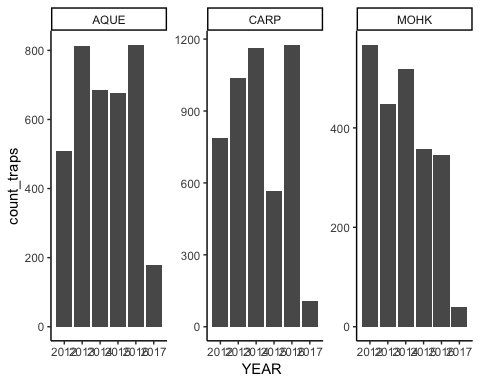
## Warning: Column `SITE` joining factors with different levels, coercing to  
## character vector

new\_lob\_traps[is.na(new\_lob\_traps)] <- 0 ## new\_lob\_traps is the final table for counts per each site per each year for number of lobsters and traps

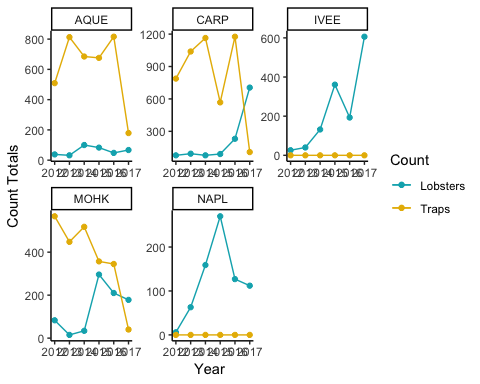
# graphs displaying trends  
  
ggplot(new\_lob\_traps, aes(x = YEAR, y = count\_lobs)) +  
 geom\_col() +   
 facet\_wrap(~SITE, scales = "free") +  
 theme\_classic() +  
 scale\_x\_continuous(breaks = seq(2012,2017, by = 1))



ggplot(trap\_counts, aes(x = YEAR, y = count\_traps)) +  
 geom\_col() +   
 facet\_wrap(~SITE, scales = "free") +   
 theme\_classic() +  
 scale\_x\_continuous(breaks = seq(2012,2017, by = 1))



## line graph of lobster vs trap counts per each year per each site  
lob\_vs\_trap\_counts <- ggplot(new\_lob\_traps, aes(x = YEAR)) +  
 geom\_line(aes(y = count\_lobs, color = "blue")) +  
 geom\_point(aes(y = count\_lobs, color = "blue")) +  
 geom\_line(aes(y = count\_traps, color = "red")) +  
 geom\_point(aes(y = count\_traps, color = "red")) +   
 facet\_wrap(~SITE, scales = "free") +  
 theme\_classic() +  
 scale\_color\_manual(values = c("#00AFBB", "#E7B800"), name = "Count", labels = c("Lobsters", "Traps")) +  
 labs(x = "Year", y = "Count Totals")  
  
lob\_vs\_trap\_counts



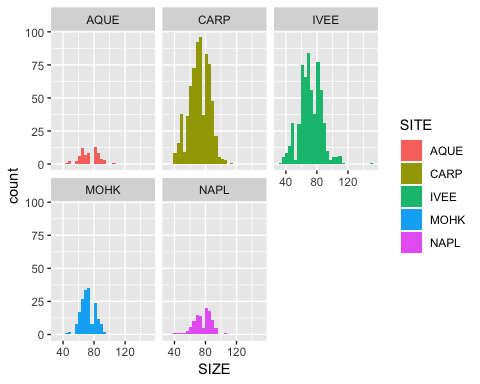
## add caption to graph to explain what is contained within it.

1. Compare mean lobster size by site in 2017 Compare mean lobster sizes (carapace length (mm)) across the five sites for lobster observations collected in
2. Warning: the size data are not in tidy format. There are rows that contain size information for multiple lobsters observed (e.g., if the researcher saw 3 lobsters all with carapace length ~ 60 mm, then they will have a single row where COUNT = 3 and SIZE = 60). You’ll want to get this into case format - where each lobster has its own row - before doing statistical analyses. There are many ways to do this. One hint: function expand.dft in the vcdExtra package (it doesn’t like tibbles, so you might need to coerce to data.frame first).

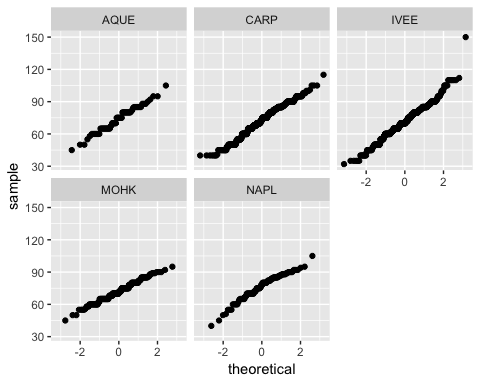
lob\_mean <- lob\_size3 %>%  
 filter(YEAR == 2017) %>%  
 group\_by(SITE) %>%  
 dplyr::summarise(mean\_size = mean(SIZE), sample\_size = length(SIZE), sd\_size = sd(SIZE))  
  
lob\_size\_2017 <- lob\_size3 %>%  
 filter(YEAR == 2017) %>%  
 select(YEAR, SITE, SIZE)

##exploratory graphs  
  
hists <- ggplot(lob\_size\_2017, aes(x = SIZE)) +  
 geom\_histogram(aes(fill = SITE)) +   
 facet\_wrap(~SITE)  
  
hists

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



qq <- ggplot(lob\_size\_2017, aes(sample = SIZE)) +  
 geom\_qq() +  
 facet\_wrap(~SITE)  
  
qq



## histogram and qqplots indicate normality  
  
# Levene's test for equal variances   
# We'll use the function leveneTest from the 'car' package  
  
# H0: Variances are equal  
# HA: Variances are unequal  
  
lobster\_levene <- leveneTest(SIZE ~ SITE, data = lob\_size\_2017)  
lobster\_levene

## Levene's Test for Homogeneity of Variance (center = median)  
## Df F value Pr(>F)   
## group 4 8.3893 1.065e-06 \*\*\*  
## 1663   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# we reject the null hypothesis of equal variances (p < 0.05)  
  
var\_table <- lob\_size\_2017 %>%  
 group\_by(SITE) %>%  
 dplyr::summarise(variance = var(SIZE))  
  
# since largest variance < 4X larger than the smallest variance, can still use ANOVA

lobster\_aov <- aov(SIZE ~ SITE, data = lob\_size\_2017)  
new\_lobster\_aov <- summary(lobster\_aov)  
  
  
# H0: Mean sizes across all sites are equal  
# HA: There is at least one significant difference in means between the 5 sites  
  
# reject the null  
  
# At least two samples were taken from populations with different means. Which ones are different? All three are different from eachother? Or something else?  
  
  
# Post hoc testing using Tukey's HSD  
  
lobster\_ph <- TukeyHSD(lobster\_aov)  
lobster\_ph

## Tukey multiple comparisons of means  
## 95% family-wise confidence level  
##   
## Fit: aov(formula = SIZE ~ SITE, data = lob\_size\_2017)  
##   
## $SITE  
## diff lwr upr p adj  
## CARP-AQUE -1.6657352 -6.24294710 2.911477 0.8582355  
## IVEE-AQUE -2.4433772 -7.05292315 2.166169 0.5968998  
## MOHK-AQUE -1.8955224 -7.02720717 3.236162 0.8514711  
## NAPL-AQUE 2.3366205 -3.19311600 7.866357 0.7775633  
## IVEE-CARP -0.7776420 -2.76097123 1.205687 0.8216104  
## MOHK-CARP -0.2297872 -3.23309697 2.773523 0.9995765  
## NAPL-CARP 4.0023556 0.36042398 7.644287 0.0228728  
## MOHK-IVEE 0.5478548 -2.50450730 3.600217 0.9882889  
## NAPL-IVEE 4.7799976 1.09751057 8.462485 0.0037001  
## NAPL-MOHK 4.2321429 -0.08607271 8.550358 0.0579286

tukey\_data <- as.data.frame(lobster\_ph$SITE)   
  
  
## only significant differences between Naples and Carp, and Naples and IV

The mean lobster size (mm) differed significantly in five Long-Term Ecological Research (LTER) Sites in the Santa Barbara Channel:Arroyo Quemado (n= 67 , Naples Reef (n= 112 ), Mohawk Reef (n= 178), Isla Vista (n= 606), Carpinteria(n= 705 ) studied (one-way ANOVA, F(4,1663) = r, *p* < 0.001, *alpha*= 0.5; Table …). Post-hoc analysis by Tukey’s HSD revealed that the mean lobster size in Naples Reef differed significantly from Carpinteria and Isla Vista (pairwise *p* < 0.001).

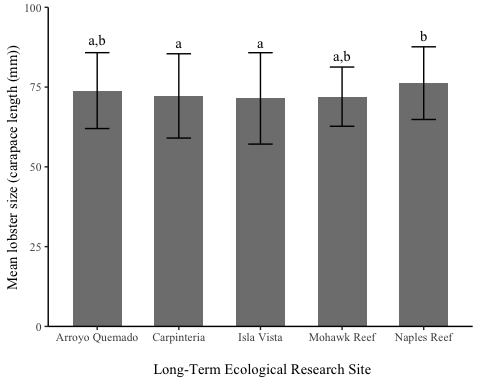
Ilayda: needs referencing for F value.

Lobster size ANOVA results summary.

Ilayda: I tried to make a table as in the example but it doesn’t work. Maybe you can fix this Gage.

Gage: Figured it out. If we knit to pdf in the file “xtable\_code.Rmd”, we can get the table. I also took a screenshot of the table and put it in our project file.

ggplot(lob\_mean, aes(x = SITE, y = mean\_size)) +  
 geom\_col(colour = NA, fill = "gray50", width = 0.6) +  
 geom\_errorbar(aes(ymin =mean\_size - sd\_size, ymax = mean\_size + sd\_size), color = "gray0", width = .3) +  
 scale\_y\_continuous(expand = c(0,0), limits = c(0,100)) +  
 scale\_x\_discrete(labels = c("Arroyo Quemado","Carpinteria","Isla Vista","Mohawk Reef","Naples Reef")) +  
 annotate("text", x = 1, y = 90, label = "a,b", family = "Times New Roman") +  
 annotate("text", x = 2, y = 89, label = "a", family = "Times New Roman") +  
 annotate("text", x = 3, y = 89, label = "a", family = "Times New Roman") +  
 annotate("text", x = 4, y = 85, label = "a,b", family = "Times New Roman") +  
 annotate("text", x = 5, y = 91, label = "b", family = "Times New Roman") +  
 theme(panel.grid.major = element\_blank(), panel.grid.minor = element\_blank(),  
 panel.background = element\_blank(), axis.line = element\_line(colour = "black"),text = element\_text(family = "Times New Roman"))+  
 xlab("\n Long-Term Ecological Research Site")+  
 ylab("Mean lobster size (carapace length (mm))") #we can change this title



**Figure 2. Lobster size in five Long-Term Ecological Research Sites.** Mean lobster sizes (mm) for sites Arroyo Quemado, Carpinteria,Isla Vista,Mohawk Reef,Naples Reef in the Santa Barbara Channel. Error bars indicate +/- 1 standard deviation. Like letters indicate values that do not differ significantly (by one-way ANOVA with Tukey’s HSD; F(4,1663) = 3.424, *p* < 0.001), with = 0.05 for all post-hoc pairwise comparisons).

1. Changes in lobster size at MPA and non-MPA sites (comparing only 2012 and 2017 sizes) From the data description (<http://sbc.lternet.edu/cgi-bin/showDataset.cgi?docid=knb-lter-sbc.77>): “Data on abundance, size and fishing pressure of California spiny lobster (Panulirus interruptus) are collected along the mainland coast of the Santa Barbara Channel. Spiny lobsters are an important predator in giant kelp forests off southern California. Two SBC LTER study reefs are located in or near the California Fish and Game Network of Marine Protected Areas (MPA), Naples and Isla Vista, both established as MPAs on 2012-01-01. MPAs provide a unique opportunity to investigate the effects of fishing on kelp forest community dynamics. Sampling began in 2012 and is ongoing.” At Isla Vista and Naples Reef, the two protected MPA sites (with zero fishing pressure), how do lobster sizes in 2012 and 2017 compare? At the non-MPA sites?

mpa <- c("IVEE", "NAPL")   
  
'%!in%' <- function(x,y)!('%in%'(x,y))  
  
## filter for mpa sites and 2017 or 2012  
  
mpa\_site <- lob\_size3 %>%  
 filter(SITE %in% mpa, YEAR == 2012|YEAR == 2017) %>%  
 select(SITE,YEAR,SIZE)  
  
## filter for non mpa sites and 2017 or 2012  
non\_mpa\_site <- lob\_size3 %>%  
 filter(SITE %!in% mpa, YEAR==2012|YEAR==2017) %>%  
 select(SITE,YEAR,SIZE)

# sandro talked to allison and apparently we have to do 5 different t tests for EACH SITE, instead of only one ttest for MPA and non mpa   
  
mpa\_2012 <- lob\_size3 %>%  
 filter(SITE %in% mpa, YEAR == 2012) %>%  
 select(SITE,YEAR,SIZE) %>%  
 mutate( i = row\_number()) %>%  
 spread(SITE,SIZE) %>%  
 select(-i)  
  
mpa\_2017 <- lob\_size3 %>%  
 filter(SITE %in% mpa, YEAR == 2017) %>%  
 select(SITE,YEAR,SIZE) %>%  
 mutate( i = row\_number()) %>%  
 spread(SITE,SIZE) %>%  
 select(-i)  
  
non\_mpa\_2012 <- lob\_size3 %>%  
 filter(SITE %!in% mpa, YEAR == 2012) %>%  
 select(SITE,YEAR,SIZE) %>%  
 mutate( i = row\_number()) %>%  
 spread(SITE,SIZE) %>%  
 select(-i)  
  
non\_mpa\_2017 <- lob\_size3 %>%  
 filter(SITE %!in% mpa, YEAR == 2017) %>%  
 select(SITE,YEAR,SIZE) %>%  
 mutate( i = row\_number()) %>%  
 spread(SITE,SIZE) %>%  
 select(-i)

# F test for equal variances  
# H0: The variances are equal (ratio of variances = 1)  
# HA: The variances are not equal (ratio of variances != 1)  
  
mpa\_ftest <- mpa\_site %>%  
 var.test(SIZE ~ YEAR, data = .)  
  
mpa\_ftest

##   
## F test to compare two variances  
##   
## data: SIZE by YEAR  
## F = 0.75323, num df = 31, denom df = 717, p-value = 0.3346  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 95 percent confidence interval:  
## 0.477719 1.341900  
## sample estimates:  
## ratio of variances   
## 0.7532319

# retain the null hypothesis of equal variances  
  
# We can override the default setting in t.test() function of var.equal = FALSE, because the variances are actually likely equal.  
  
# H0: mean Lobster size at mpa sites in 2012 are equal to mean lobster size at mpa sites in 2017  
# HA: mean Lobster size at mpa sites in 2012 are NOT equal to mean lobster size at mpa sites in 2017  
  
mpa\_ttest <- mpa\_site %>%  
 t.test(SIZE ~ YEAR, data = ., var.equal = TRUE)  
  
mpa\_ttest # retain null they are equal

##   
## Two Sample t-test  
##   
## data: SIZE by YEAR  
## t = -1.9159, df = 748, p-value = 0.05576  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -9.7644724 0.1189292  
## sample estimates:  
## mean in group 2012 mean in group 2017   
## 67.37500 72.19777

mpa\_site\_2012 <- lob\_size3 %>%  
 filter(SITE %in% mpa, YEAR == 2012) %>%  
 select(SITE,YEAR,SIZE)  
  
mpa\_site\_2017 <- lob\_size3 %>%  
 filter(SITE %in% mpa, YEAR == 2017) %>%  
 select(SITE,YEAR,SIZE)  
  
#calculate the difference in means  
  
mpa\_mean\_2012 <- mean(mpa\_site\_2012$SIZE)  
  
mpa\_mean\_2017 <- mean(mpa\_site\_2017$SIZE)  
  
mpa\_mean\_2012 - mpa\_mean\_2017

## [1] -4.822772

# calculate effect size  
cohen\_d\_test <- cohen.d(mpa\_site\_2012$SIZE, mpa\_site\_2017$SIZE)  
cohen\_d\_test

##   
## Cohen's d  
##   
## d estimate: -0.3461506 (small)  
## 95 percent confidence interval:  
## inf sup   
## -0.701270908 0.008969759

## effect size is small which indicates that there likely is NOT a significant difference in mean lobster size at the two sites.

# F test for equal variances  
# H0: The variances are equal (ratio of variances = 1)  
# HA: The variances are not equal (ratio of variances != 1)  
  
non\_mpa\_ftest <- non\_mpa\_site %>%  
 var.test(SIZE ~ YEAR, data = .)  
  
non\_mpa\_ftest

##   
## F test to compare two variances  
##   
## data: SIZE by YEAR  
## F = 0.99085, num df = 198, denom df = 949, p-value = 0.953  
## alternative hypothesis: true ratio of variances is not equal to 1  
## 95 percent confidence interval:  
## 0.8037718 1.2406929  
## sample estimates:  
## ratio of variances   
## 0.9908519

# retain the null hypothesis of equal variances  
  
# We can override the default setting in t.test() function of var.equal = FALSE, because the variances are actually likely equal.  
  
# H0: mean Lobster size at mpa sites in 2012 are equal to mean lobster size at mpa sites in 2017  
# HA: mean Lobster size at mpa sites in 2012 are NOT equal to mean lobster size at mpa sites in 2017  
  
non\_mpa\_ttest <- non\_mpa\_site %>%  
 t.test(SIZE ~ YEAR, data = ., var.equal = TRUE)  
  
non\_mpa\_ttest # reject null they are equal

##   
## Two Sample t-test  
##   
## data: SIZE by YEAR  
## t = 2.6973, df = 1147, p-value = 0.007093  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 0.7143078 4.5265173  
## sample estimates:  
## mean in group 2012 mean in group 2017   
## 74.92462 72.30421

non\_mpa\_site\_2012 <- lob\_size3 %>%  
 filter(SITE %!in% mpa, YEAR == 2012) %>%  
 select(SITE,YEAR,SIZE)  
  
non\_mpa\_site\_2017 <- lob\_size3 %>%  
 filter(SITE %!in% mpa, YEAR == 2017) %>%  
 select(SITE,YEAR,SIZE)  
  
#calculate the difference in means  
  
non\_mpa\_mean\_2012 <- mean(non\_mpa\_site\_2012$SIZE)  
  
non\_mpa\_mean\_2017 <- mean(non\_mpa\_site\_2017$SIZE)  
  
non\_mpa\_mean\_2012 - non\_mpa\_mean\_2017

## [1] 2.620413

# calculate effect size  
cohen\_d\_test\_non <- cohen.d(non\_mpa\_site\_2012$SIZE, non\_mpa\_site\_2017$SIZE)  
cohen\_d\_test\_non

##   
## Cohen's d  
##   
## d estimate: 0.2102816 (small)  
## 95 percent confidence interval:  
## inf sup   
## 0.05707948 0.36348365

## effect size is small which indicates that there likely is NOT a significant difference in mean lobster size at the two sites.  
  
## while there is a significance difference, the effect size is small, the absolute difference in means is small ....

1. Proportions of “legal” lobsters at the 5 sites in 2017 The legal minimum carapace size for lobster is 82.6 mm. What proportion of observed lobsters at each site are above the legal minimum? Does that proportion differ significantly across the 5 sites? Note: We’ll be doing chi-square in labs next week, or go ahead with maximum resourcefulness and check out the chisq.test() function on your own!

# filter for 2017  
# filter for above 82.6 mm  
  
  
legal\_min <- lob\_size3 %>%  
 filter(YEAR == 2017) %>%  
 mutate(above\_legal = case\_when(  
 SIZE >= 82.6 ~ 'yes',  
 SIZE < 82.6 ~ 'no'  
 )) %>%  
 count(SITE, above\_legal) %>%  
 spread(above\_legal, n) %>%  
 select(-SITE)  
   
rownames(legal\_min) <- c("AQUE", "CARP", "IVEE", "MOHK", "NAPL")

## Warning: Setting row names on a tibble is deprecated.

legal\_prop <- prop.table(as.matrix(legal\_min), 1)  
  
  
above\_x2 <- chisq.test(legal\_prop)

## Warning in chisq.test(legal\_prop): Chi-squared approximation may be  
## incorrect

above\_x2

##   
## Pearson's Chi-squared test  
##   
## data: legal\_prop  
## X-squared = 0.11095, df = 4, p-value = 0.9985

## The proportion does not differ significantly between each location.   
# there is no signficant association between legal lobster size and different test sites

Based on the obervations from five Long-Term Ecological Research (LTER) Sites in the Santa Barbara Channel:Arroyo Quemado (n= 67) , Naples Reef (n= 112 ), Mohawk Reef (n= 178), Isla Vista (n= 606), Carpinteria(n= 705 ),the proportion of observed lobsters that are above the legal minimum carapace size does not differ significantly by site.( (4)= 0.11095, *p*= 0.9985)