Homework 6

Paul Lagarde

## March 15, 2021

Exercise 1

13.6.10a: to test H_0 : median healthy caloric intake = median bulimic intake against the alternate hypothesis that the two medians are not equal, we use a non-parametric method to adjust for the small sample size. Because our data is drawn from two independent populations and is not paired, we will use the Wilcoxon Rank Sum Test. Per R,

> *caloric.data <- read.table(file = "caloric.txt", header = TRUE)*
> *wilcox.test(caloric.data$Bulimic, caloric.data$Healthy, alternative="two.sided")*
> *Wilcoxon rank*
> *sum test with continuity correction*
> *W = 61.5, p-value = 0.0009651*
> *alternative hypothesis: true location shift is not equal to 0*

Thus, because $p < \alpha$ we reject H_0.

b) no. Bulima is by definition a condition that requires individuals to purge out calories before their body can digest them. That individuals with bulima have lower caloric intake than their healthy counterparts is in fact evidence that they are unhealthy, not that they can healthily live off that calory intake.

13.6.10a:

Beds per 1000, 1980

Beds per 1000, 1986
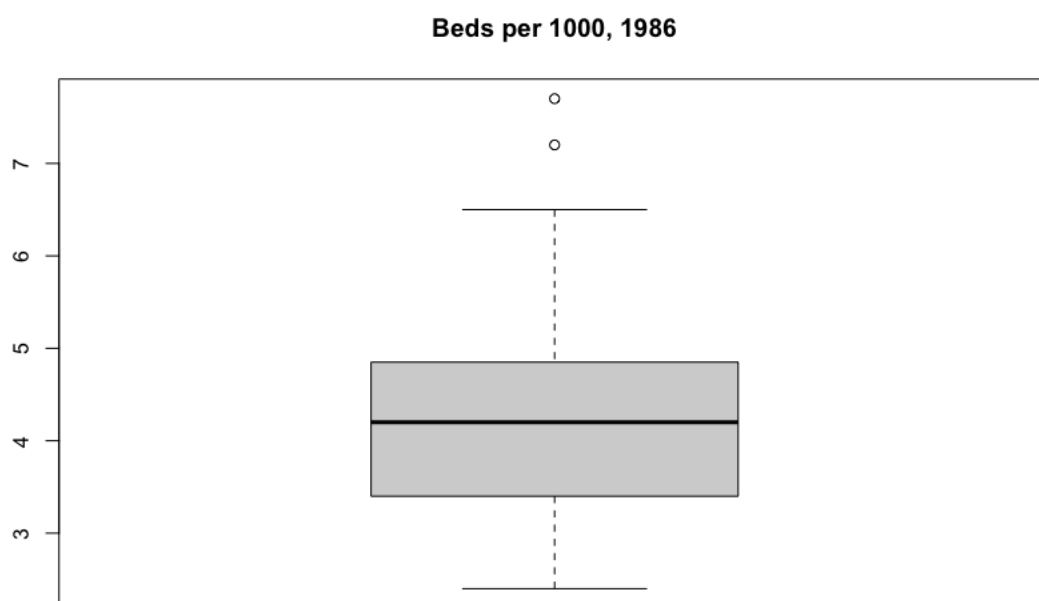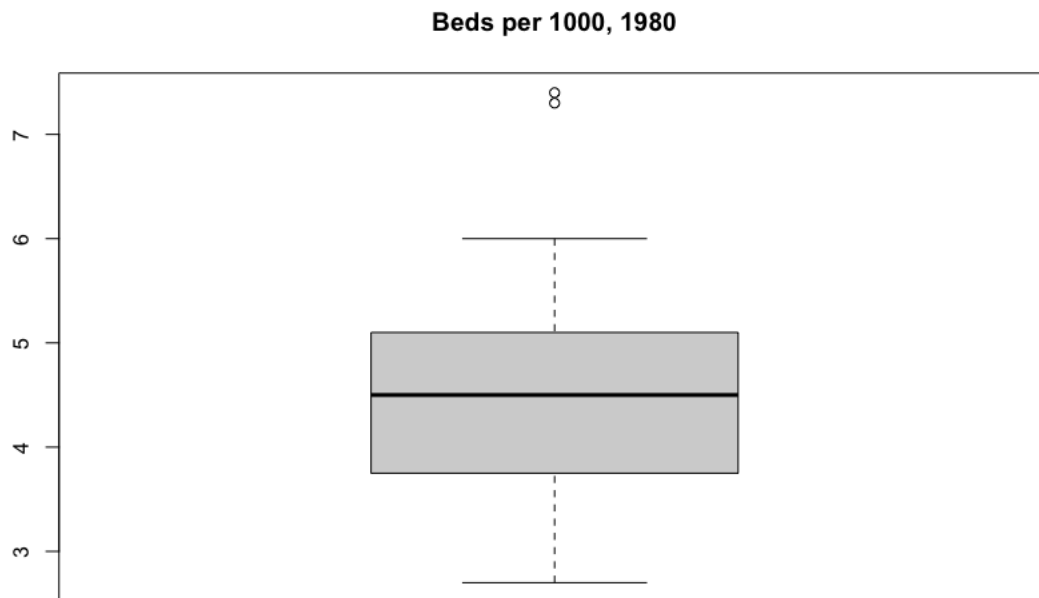
b: per R, we have:

> wilcox.test(bed.data$bed86, bed.data$bed80, paired = TRUE, alternative = "two.sided")

>Wilcoxon signed rank test with continuity correction

>data: bed.data$bed86 and bed.data$bed80

>V = 129.5, p-value = 5.683e-07

>alternative hypothesis: true location shift is not equal to 0

> reject h-naught

Thus, because $p < \alpha$, we reject H_0 under the Wilcoxon signed-rank test, concluding that the

number of beds/1000 people did change.

c) also per R
>*wilcox.test(bed.data$bed86, bed.data$bed80, alternative = "two.sided")*
>*Wilcoxon rank sum test with continuity correction*
>*data: bed.data$bed86 and bed.data$bed80*
>*W = 1025, p-value = 0.06544*
>*alternative hypothesis: true location shift is not equal to 0*
Thus, because $p > \alpha$, we fail to reject that the number of beds/1000 did not change under the Wilcoxon rank sum test.

d) in this case, the correct test to use is pretty clearly the Wilcoxon signed-rank test, which is what we use to evaluate differences among pairs. The RST is useful for independent data sets.

e) as we saw in Chapter 11, we would reject the null hypothesis under a t–test of similar sensitivity as well:

> *diff.beds <- bed.data$bed80 - bed.data$bed86*
> *t.test(diff.beds)*
>*One Sample t-test*
>*data: diff.beds*
>*t = 6.8721, df = 50, p-value = 9.513e-09*

13.6.16a: no. A two-sample t–test requires data to be normally distributed, eg a continuous distribution. With a set of outcomes $O = \{1, 2, 3, 4, 5\}$ we are dealing with a discrete distribution, which disqualifies thte t–test.

b) per R, we have:
> *insure.data <- read.table(file = "insure.txt", header = TRUE)*
> *unpaired data*
> *wilcox.test(insure.data$stage, insure.data$group, paired = FALSE)*
>*Wilcoxon rank sum test with continuity correction*
>*data: insure.data$stage and insure.data$group*
>*W = 1092260, p-value < 2.2e-16*
>*alternative hypothesis: true location shift is not equal to 0*

Thus, under $\alpha = .05$ and $p = 2.2e - 16$, we have that $p < \alpha$ so we reject, concluding that the median stage at which cancer is detected is different for those who are or are not insured.

c) surprisingly, the data indicates the opposite. $describeBy(insure.data\$stage, insure.data\$group)$ gives us that uninsured women (group 0) have a mean and median detection stages of 2.8 and 3, while those who are uninsured have 2.26 and 2. An un-intuitive result.

Exercise 2

Let A be a population from which a sample $m = 10$ is drawn, from B draw $n = 9$, with A and B independent. Per table A.7, for $m = 10$, $n = 9$, we see that the value of W with the lowest probability is $P(W_0 = 36) = 0.00$; the smallest non-zero is $P(W_0 = 38) = .0001$.

For a permutation, we see that we have $(n + m)! = 19! = 1.217e^{17}$ possible equally likely outcomes, hence each with a likelihood of $\frac{1}{19!} = 8.22e^{-18}$.