

DATA557 Homework 3

Will Wright

January 26, 2019

Instructions

Submit your solutions in pdf format to the dropbox on the canvas page by 5:00PM, Wednesday January 30. You do not need to include your R code with your solutions for this assignment.

For question 1, you are to work in groups. For the other questions, you may work together to help each other solve problems, but you should do all the work, create your own solutions, and hand in your own work without copying others' work.

Question 1. (This is Q3 of Exercise 3.)

Suppose that a new experiment is being designed to determine the effect on output of temperatures higher than 100. In particular, the aim of the new experiment is to test the null hypothesis that the mean output is the same for temperature 100 and temperature 120. The researcher would like to have at least 90% power to detect a difference between these conditions in mean output equal to 75. Your job is to determine the sample sizes for each group and to decide which test statistic will be used to test the null hypothesis. Justify your answers.

For this question, you should continue to work with your group on the answer that you developed in class on Jan 23. The group member who did the original posting will receive feedback from me on canvas by Saturday morning and should relay it to the other group members. Work together either in person or electronically to develop your final solution. The group member who did the original posting should include the group's solution in their HW 3 submission, including the names of all group members. The grade assigned for the solution to this question will be applied to the grade for the HW for all group members.

Question 1 Findings:

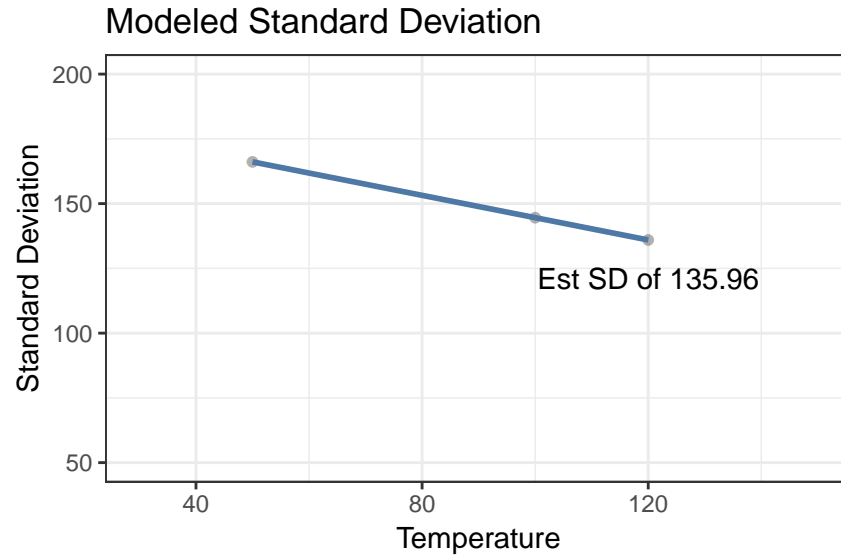
Team: Tara Wilson, Lauren Heintz, Ben Bordeur Mathieu, and Will Wright

Summary: We started by confirming the relative normality of the residual distributions. Without knowing the variance of the temp = 120 condition, we tested several possibilities. Ultimately, we recommend the most conservative option of $n = 92$ via a Welch's t-test simulation with unequal variance where we assume the variance of the temp = 120 condition is equal to the higher variance of the temp = 50 condition. The justification for this choice is that the Welch t-test allows for different variances and sample sizes between the two samples. We assumed a variety of variances for testing, but the real variance is still an unknown. Without having this information, we took a conservative estimate with our recommendation. The Welch model also provides some flexibility in the event there are sampling errors and we need to discard any data points.

Details: In our approach to this question, we first inspected the known experimental parameters to help determine which statistical test to use. In particular, we were interested in what we could learn about the distributions from the two samples of $n=30$ from the prior experiment for temperatures of 50 and 100.

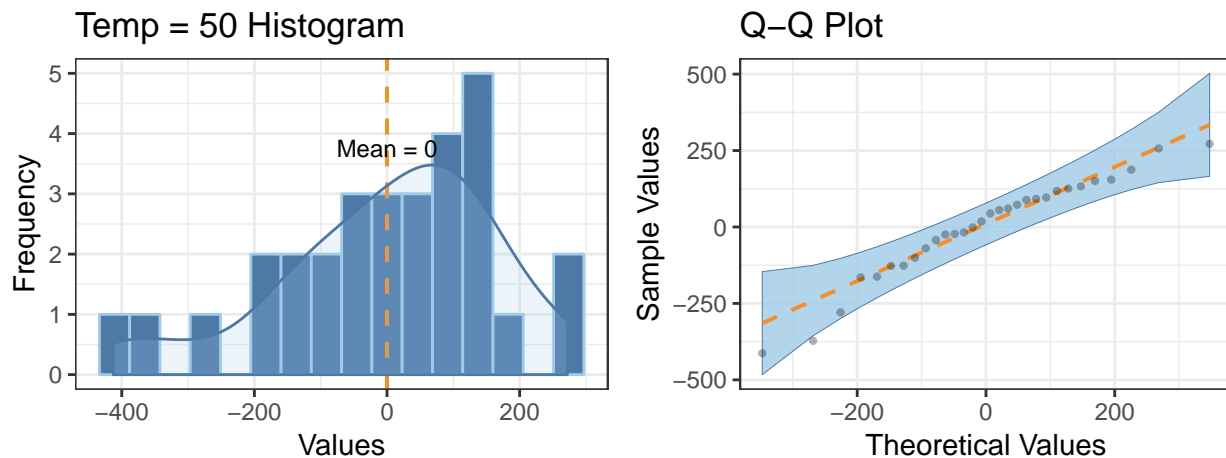
- The mean and standard deviation for the sample with a temperature of 50 are 899.82 and 166.11.
- The mean and standard deviation for the sample with a temperature of 100 are 1034.14 and 144.57.

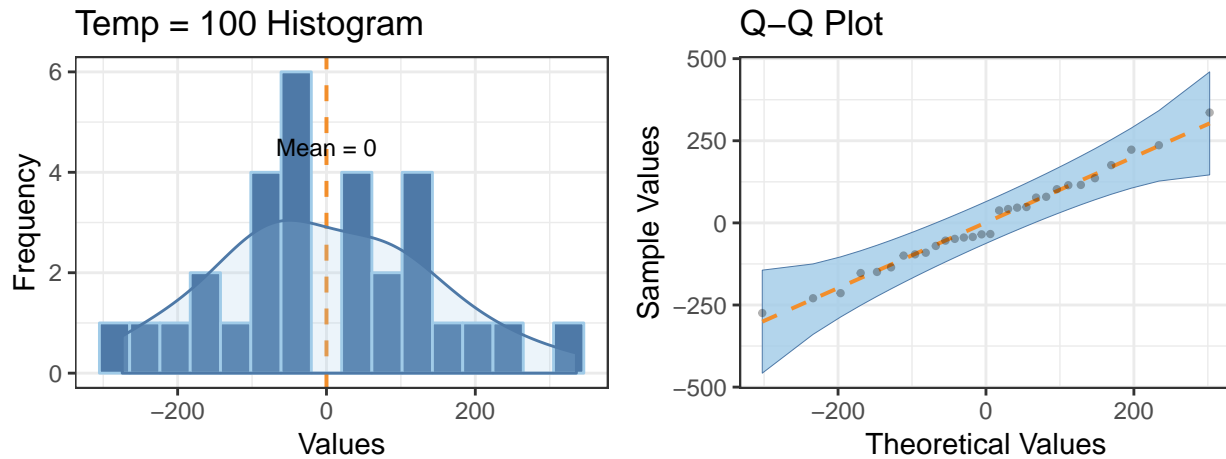
Next, in order to estimate the standard deviation for a new sample with a temperature of 120, we used a linear model to extrapolate. Given more data or context about the machine where measurements are taken, we could likely make better assumptions about what standard deviation is reasonable. For instance, if we know that failures happen more frequently above 100 degrees, a linear model would not be appropriate. Given our lack of knowledge of the true distribution of variance, this is an optimistic approximation.



Our model shows an estimated standard deviation of 135.96.

Then, we inspected the distribution of residuals for the $\text{temp} = 50$ and $\text{temp} = 100$ cases to see if normality is a reasonable assumption:



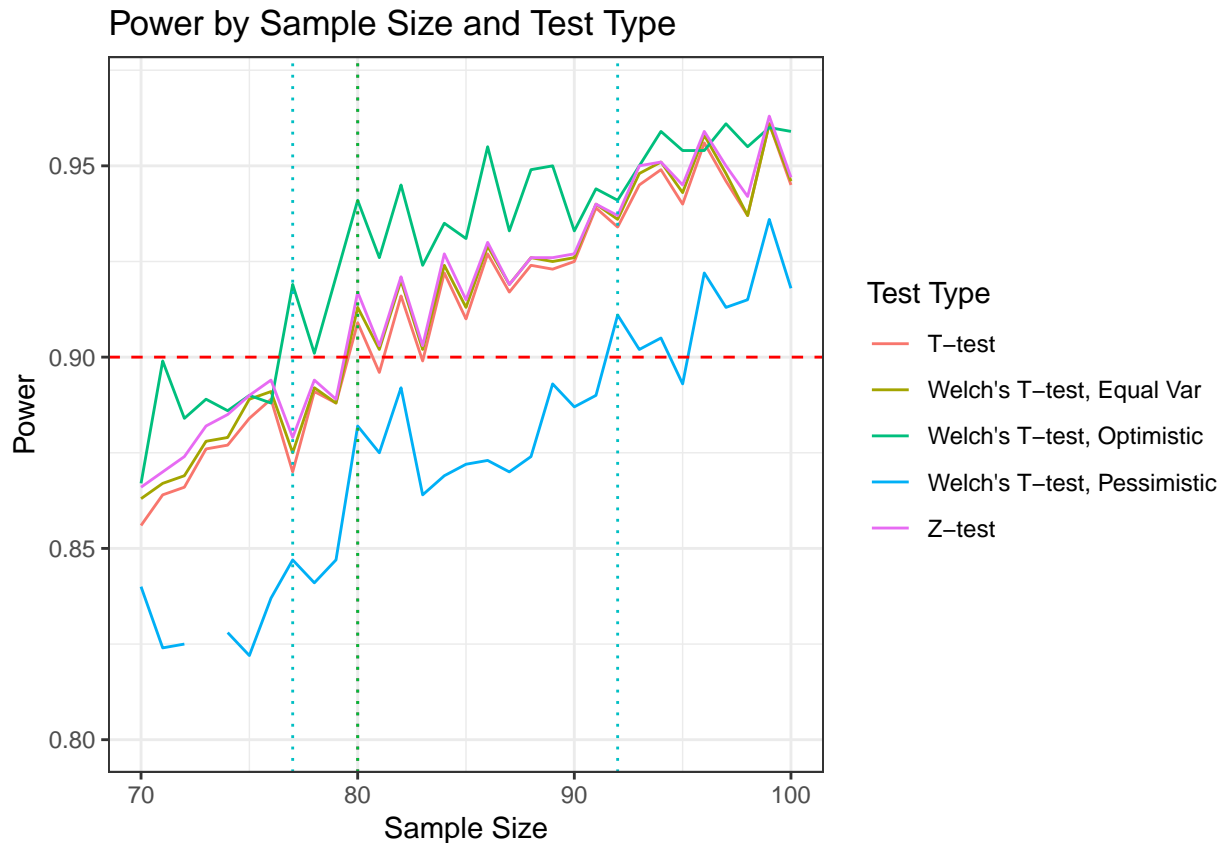


Based on these results, normality seems plausible, which satisfies the required assumption to move forward with the following simulations:

To examine the relationship between sample sizes and power, we investigated 5 conditions:

1. Assuming a completely normal distribution with equal variance using a Z-statistic
2. Assuming a t-distribution with equal variance using a T-statistic
3. Assuming a t-distribution with unequal variance using a T-statistic (Welch's T-test) (Optimistic with our linear approximation)
4. Assuming a t-distribution with unequal variance using a T-statistic (Welch's T-test) (Pessimistic with our highest known sd for temp=50)
5. Assuming a t-distribution with variance equal to the temp = 100 condition using a T-statistic (Welch's T-test)

Below, we simulated 1000 samples in each condition to get a mean power per each sample size from 1 to 100 and graphed the result between 70 and 100 for clarity:



The results of this power analysis show that the needed sample size is 80, 80, 80, 77, or 92 depending on if we wanted to go with a Z-test, T-test, Welch's T-test with equal variance, Welch's T-test with the modeled optimistic variance, or Welch's T-test with the pessimistic temp = 50 variance, respectively.

Recommendation: Given that we don't truly know what variance to expect for the temp = 120 condition, the best we can do is speculate. While a linear model for variance may be reasonable, it could just as easily be the case that the relationship with temperature is non-linear. If we had data from a variety of temperatures, we could have more confidence in selecting an appropriate variance. However, lacking this data, we're taking the most conservative approach and assuming that the variance increases to the higher temp = 50 condition at temp = 120. This results in an ideal n of 92.

Question 2

Data set: 'fev.csv'

The data come from a study to examine, among other things, the association between smoking and lung function as measured by forced expiratory volume (FEV) in children. Variables in the data set are listed below.

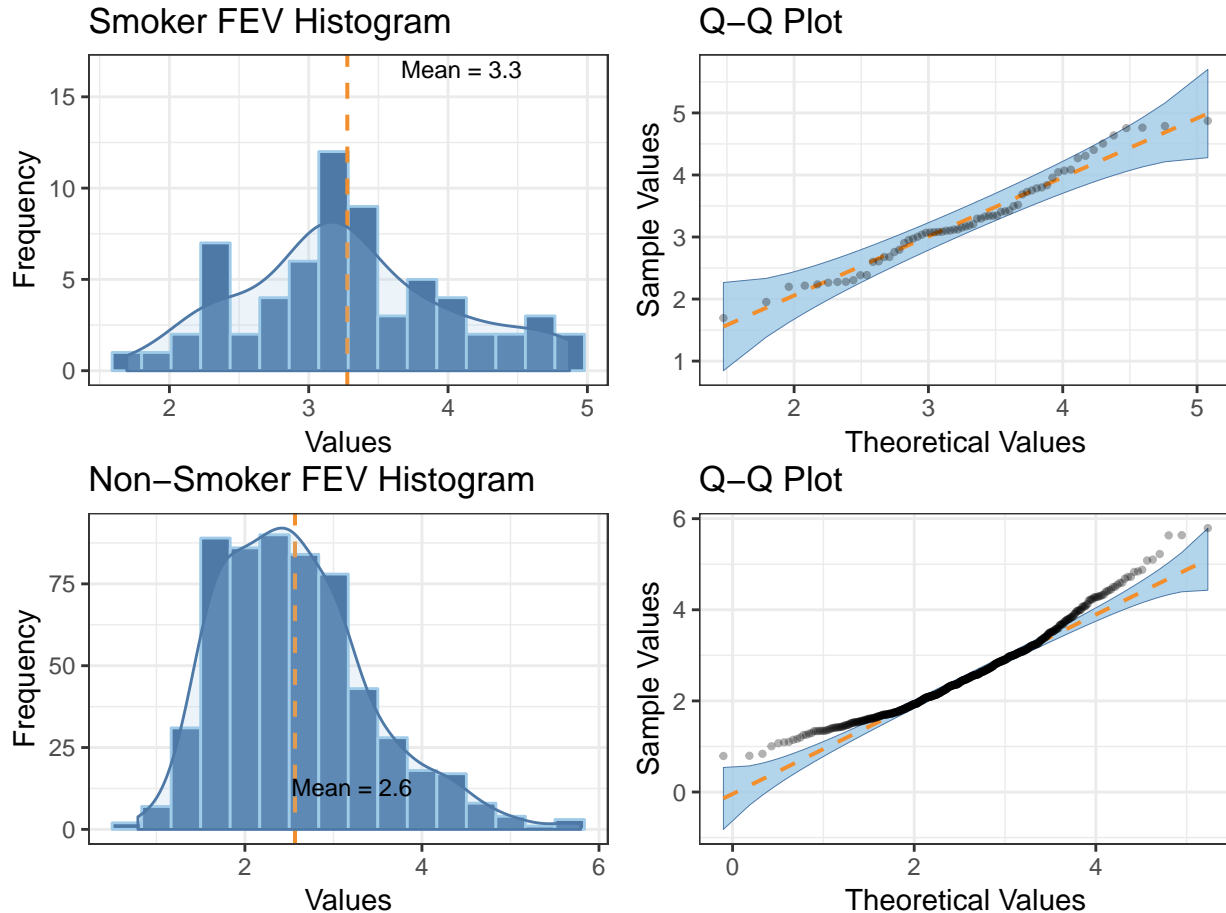
id: ID number age: Age (yrs) fev: FEV (liters) ht: Height (inches) male: Sex (0=female, 1=male) smoke: Smoking Status (0=non-current smoker, 1=current smoker)

2.1. It is desired to conduct a test of the null hypothesis that mean FEV is the same in smokers and non-smokers. Conduct an appropriate simulation study to determine an appropriate hypothesis test procedure

to use (consider only the 3 methods discussed in class). Provide a description of your simulation study in words and provide a summary of the results of your simulation study.

$H_0 : \mu_A = \mu_B$ where A are smokers and B are non-smokers $H_1 : \mu_A \neq \mu_B$

In order to determine an appropriate hypothesis test procedure, we must consider how large the sample size is and what the shapes of the distributions are:



The smokers have a relatively-normal distribution with $n = 65$, mean = 3.3, and $sd = 0.75$ while the non-smokers have a left-skewed distribution with $n = 589$, mean = 2.6, and $sd = 0.85$. Given the degree of skewness with the non-smoker data, it's questionable to use any of the 3 methods discussed in class, but we'll proceed with the best of the bunch, which is a Welch's t-test, given the uneven distribution. Below, a simulation is conducted to test the coverage of the proposed method as well as a standard t-test and z-test (just for comparison):

##	p.equal.var	p.Welch	p.Z
##	0.0276	0.0442	0.0190

As can be seen, Welch's t-test performed the closest to expectation so we'll proceed with that approach.

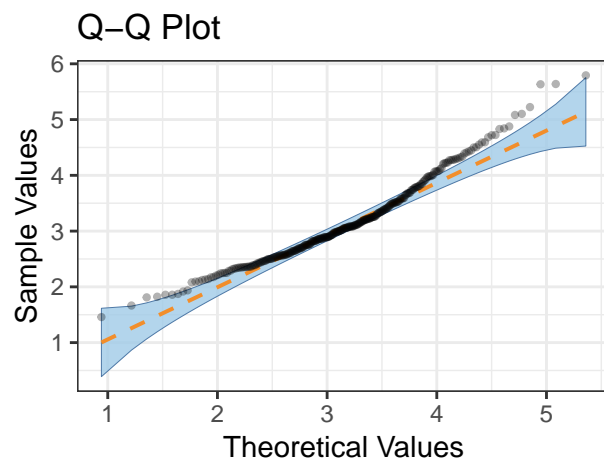
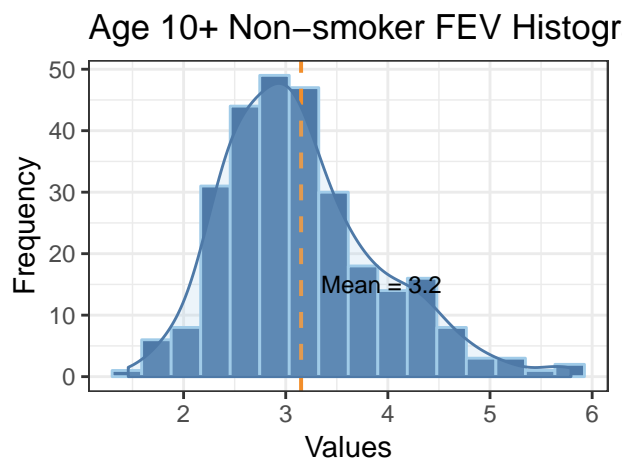
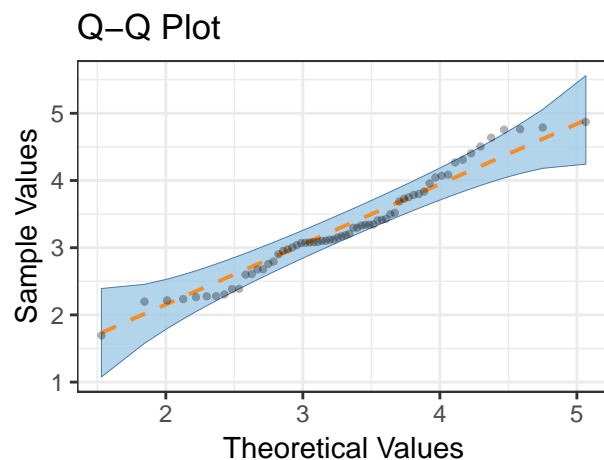
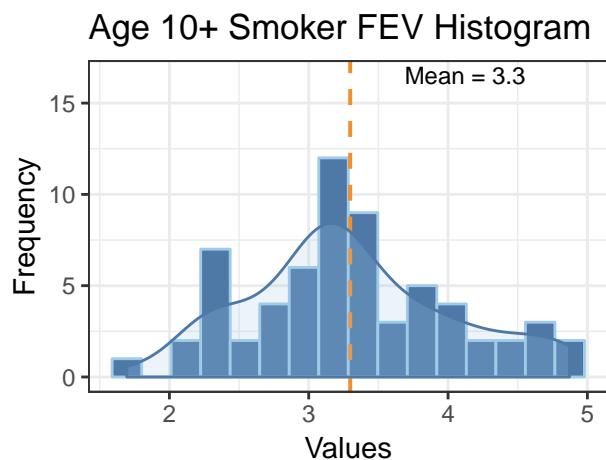
2.2. Using the method chosen in Q2.1, carry out the test of the null hypothesis. Report your results along with a confidence interval.

```
t.test(smokerDat$fev, nonSmokerDat$fev, alternative = "two.sided", var.equal = FALSE)
```

```
##
## Welch Two Sample t-test
##
## data: smokerDat$fev and nonSmokerDat$fev
## t = 7.1496, df = 83.273, p-value = 3.074e-10
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.5130126 0.9084253
## sample estimates:
## mean of x mean of y
##  3.276862  2.566143
```

Given the a p-val less than an alpha of 0.05, we reject the null hypothesis that there is no difference in mean FEV. The confidence interval is 0.51 to 0.91.

2.3. Conduct a test of the null hypothesis that mean FEV is the same in smokers and non-smokers, among children 10 years old or older. Compare your results to the previous results for all children and provide an interpretation of any differences in results. (Note: Children younger than 10 would be unlikely to smoke and had been included in the study for other purposes besides the analysis of the association between smoking and FEV.)



```
##
## Welch Two Sample t-test
##
## data: olderSmokerDat$fev and olderNonSmokerDat$fev
## t = 1.4333, df = 95.857, p-value = 0.155
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.05663529 0.35088918
## sample estimates:
## mean of x mean of y
## 3.297547 3.150420
```

In this case, with a p-val of 0.155, we fail to reject the null hypothesis that there is a difference in mean FEV. The cause for this difference is two-fold: the means are much closer together at 3.2 and 3.3 and the variance is higher due to a lower sample size with the age-subsetting.

Question 3

Read the paper “Why most published research findings are false.” (see Week 3 folder). Do you agree with the overall conclusion of the paper that most published research findings are false? Provide your answer with your rationale in a paragraph of 200 words or less. (No simulation studies required here!)

Answer: Based purely on this paper, I’m fine with accepting statement that most published research findings are false. That said, I disagree with the general sentiment it provokes-that we can’t trust scientific papers in general. My understanding is that, by volume, most research is victim to the 6 corollaries he defines (i.e. small studies, small effect sizes, flexible experimental design etc.). As evidence of this, you can find a scientific paper to cite as evidence for nearly any crazy idea (including that vaccines don’t work or that human-caused climate change is a myth). So it’s true that if I had to draw a random research finding from a hat, I’d bet against it being true. That said, research that has been reproducible and aggregated into meta-analyses with strong effect sizes are a subset of all research that I believe deserves our confidence. Ioannidis touches on this by stating that “Better powered evidence, e.g., large studies or low-bias meta-analyses, may help. . .” but he goes on to cast doubt regardless.