Final Policies

You should have the following items on your desk (and nothing else) during the final:

- Pen and/or pencil and/or eraser - make sure you have enough writing utensils and they are functional. You should not try to open your bag to get another pen. Have everything ready prior to the midterm.
- Scientific calculator.
- UCI Student ID card - we will come around to check your ID during the midterm.

The following items are **not** permitted during the midterm

- Scrap paper
- Watches - remaining time will be displayed on the screen.
- Cell phones - turn your cell phone off completely and put it in your bag not in your pocket.
- Laptops / Ipads/Tablets
- Graphing calculator
- Notes of any kind, books etc.
- Any note taking device or any device with internet connectivity

Note that if you use an eraser (for pen or pencil) and do not erase well, you may not receive any points if previous answer and the latest answer are visible simultaneously.

Multiple choice questions have one correct answer. Only choose one of the choices. Write your answers in the boxes provided. You should only write a single letter in each of these boxes.

For questions that are **not** multiple choice, show your work. Write your final answers in the boxes provided. You should only write numbers in these boxes. For instance if your answer is $P(A) = 0.12$ only write 0.12 inside the box.

You will have 115 minutes. You may leave the classroom if you finish early. Make sure to review your answers before leaving.

You may not start the final until you are prompted to start. Just write your name and student ID number and wait for the prompt to start. Once you begin make sure to write your name and Student ID on top of every page.

Row:	Seat:

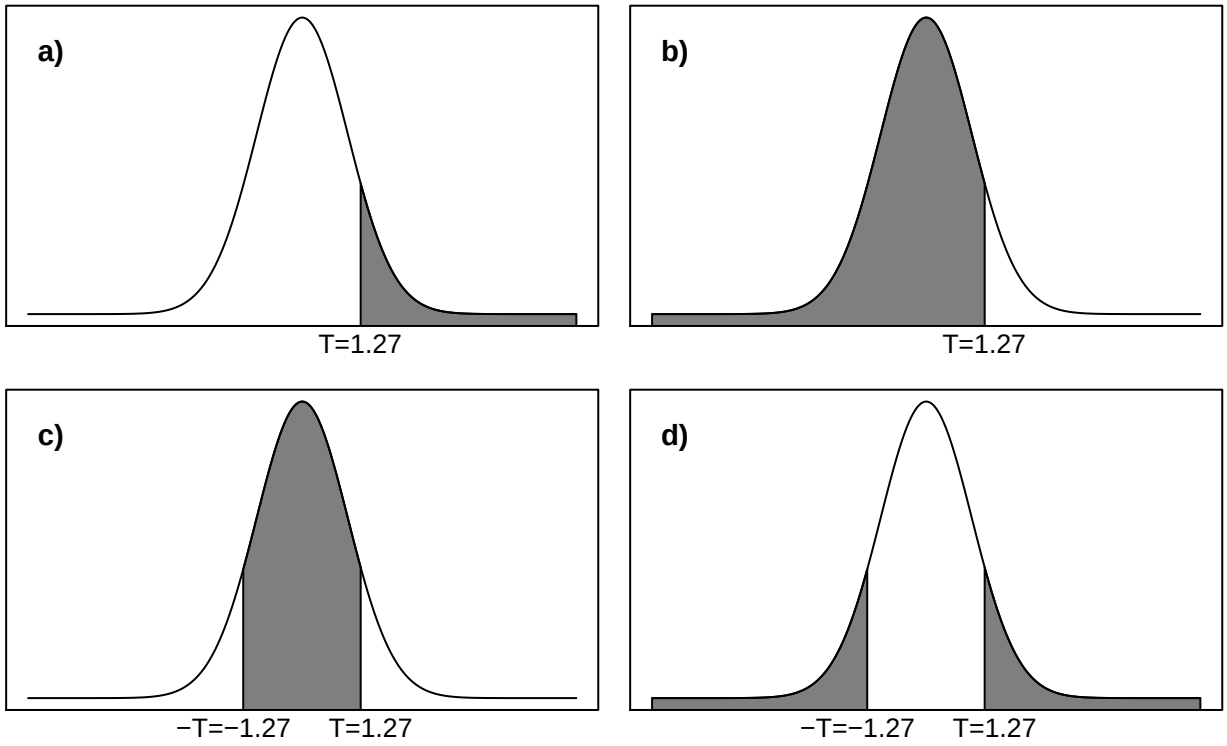
1. For a normally distributed random variable, 80% of the observations are contained within _____ standard deviations of the mean. Make sure to write the final numeric answer in the box. Round to two decimal places.

You test the following hypotheses:

$H_0 : \mu = 5$

$H_A : \mu > 5$

with a sample of n = 397, and obtain a T-statistic of 1.27. Answer the next three questions based on this information.



2. How many degrees of freedom does the test-statistic T=1.27 have?
3. Assuming you had already assigned your degrees of freedom to a variable named `degFree`, which line of code would find the appropriate T-critical value for this test?
4. Determine distribution Continuous Uniform

5. Determine distribution Exponential
6. Determine distribution Binom/Geom/Poisson
7. Give a ggplot with best line of fit. Ask them to fill in the blank for geom_smooth()
8. Type I Type II error
9. Which of the following is false?
   (a) The sample mean is an unbiased estimator of the population mean.
   (b) As n increases the t-distribution approaches a normal distribution.
   (c) As number of samples increases for large samples, the sampling distribution approaches a normal distribution
   (d) The standard deviation can be negative.

We have information on daily air quality measurements for New York between May and September 1973. We would like to know if there is a relationship between temperature (degrees F) and ozone (ppb). Below is a summary of the linear model regression temperature (variable is called `Temp`) on ozone (variable is called `Ozone`). Answer the next four questions based on this information.

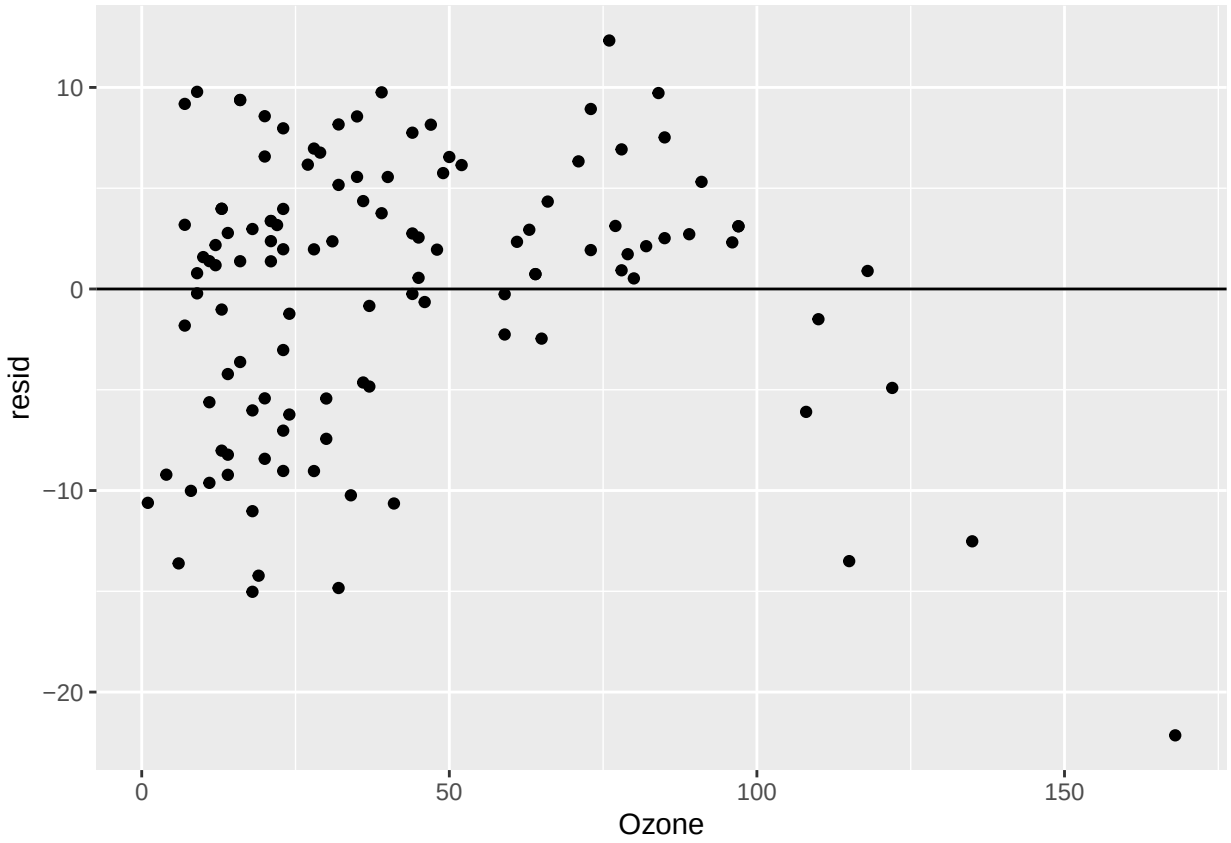
```
##
## Call:
## "lm(formula = _____________, data = airquality)"
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -22.147  -4.858   1.828   4.342  12.328
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 69.41072    1.02971   67.41   <2e-16 ***
## Ozone        0.20081    0.01928   10.42   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.819 on 114 degrees of freedom
##   (37 observations deleted due to missingness)
## Multiple R-squared:  0.4877, Adjusted R-squared:  0.4832
## F-statistic: 108.5 on 1 and 114 DF,  p-value: < 2.2e-16
```

10. The model summary has a blank. Fill in the blank with appropriate code.
11. Interpret the coefficients (slope and intercept) in the context of air quality.
12. Below are a residual plot, a QQ-plot, and a histogram of residuals for the model. Check whether the conditions for a linear regression are met.

```
airquality <- airquality %>%
   add_predictions(model1) %>%
   add_residuals(model1)
```

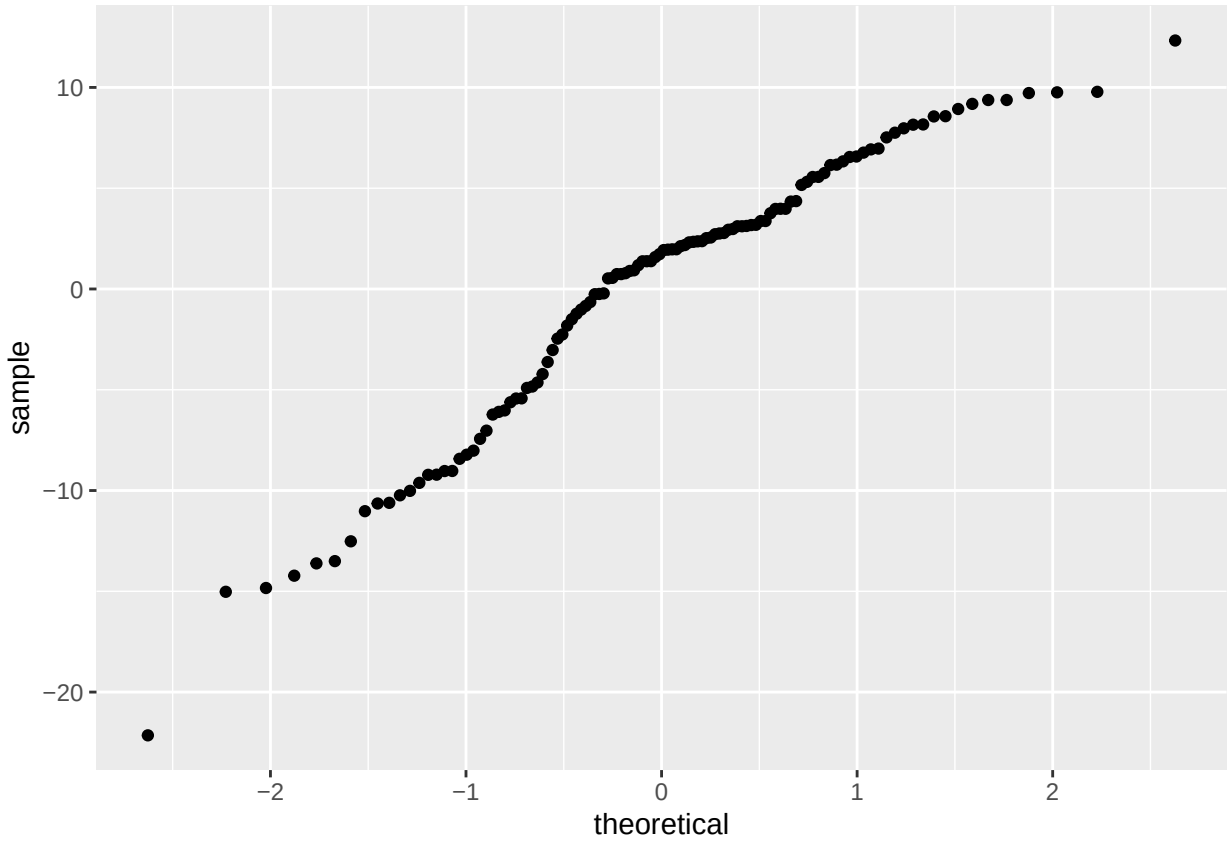
```
airquality %>%
   ggplot(aes(x=Ozone,y=resid))+
   geom_point()+
   geom_hline(yintercept=0)
```

```
## Warning: Removed 37 rows containing missing values (geom_point).
```




```
airquality %>%
   ggplot(aes(sample=resid)) +
   geom_qq()
```

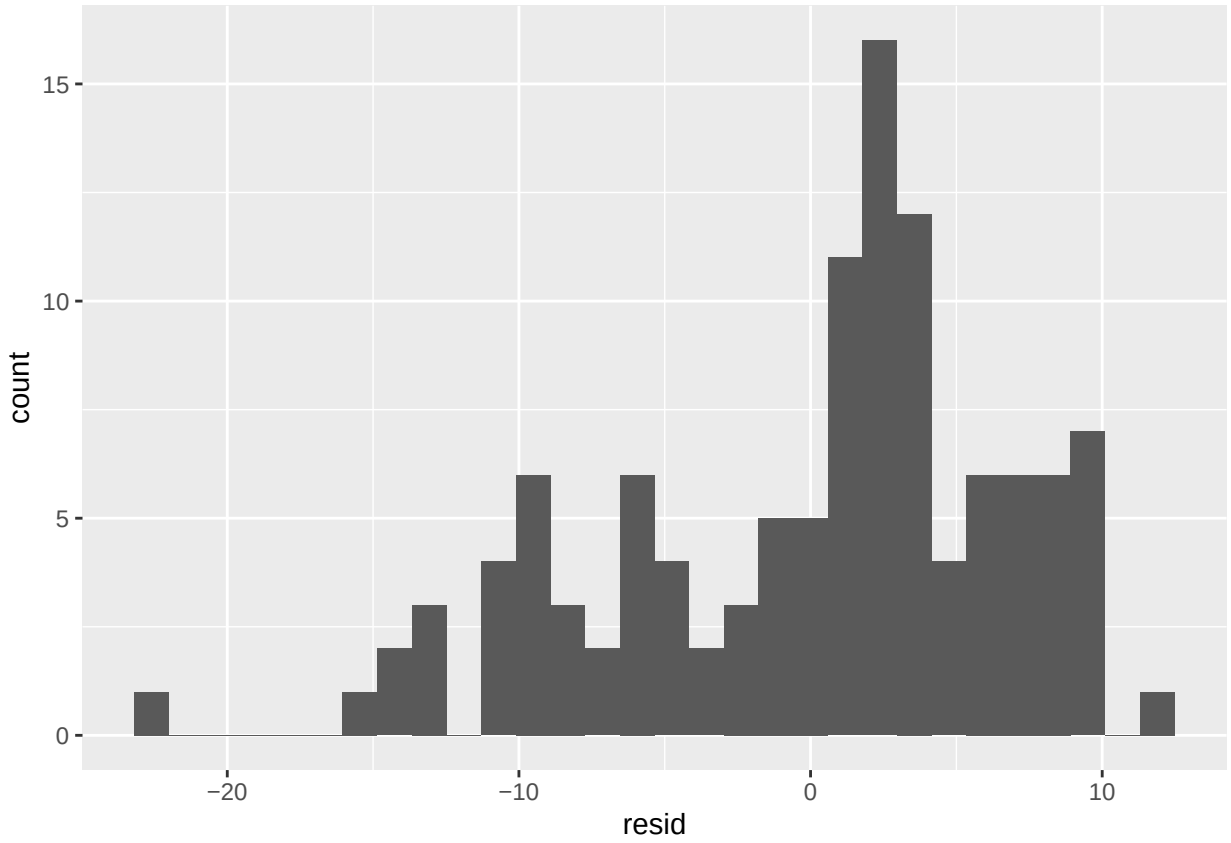
```
## Warning: Removed 37 rows containing non-finite values (stat_qq).
```




```
airquality %>%
   ggplot(aes(resid))+
   geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 37 rows containing non-finite values (stat_bin).
```



13. Regardless of what you determine in Question 12, construct a 95% confidence interval for the slope. Interpret it in the context of the problem.