Homework 8

Stat 135: Concepts of Statistics
UC Berkeley, Fall 2022

Due: **November 3rd, 11:59PM**

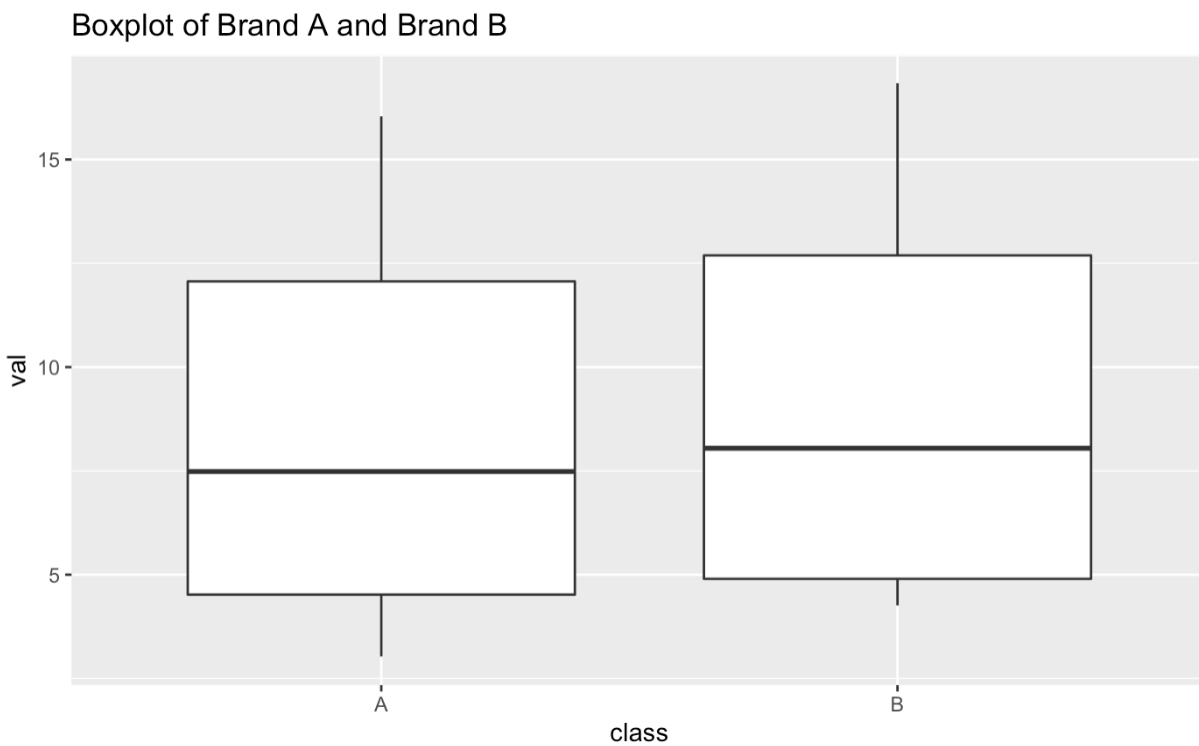
- Read chapter 11, sections 11.1 up to and including 11.3 of the textbook. Go over the slides on two independent and paired sample tests.
- Your homework must be submitted to Gradescope **as a single PDF file.**
- To obtain full credit, please **write your answers clearly and show your reasoning.**

Problem 1

The following data shows the battery life in hours of 10 batteries from two different brands. The data can be found in the batteries.csv file in case you want to load the data into R. In your answers below, show your working: explicitly state which test you are conducting, show the formula for your test statistic and p-value, etc.

Brand A	Brand B
3.03	3.19
5.53	4.26
5.60	4.47
9.30	4.53
9.92	4.67
12.51	4.69
12.95	12.78
15.21	6.79
16.04	9.37
16.84	12.75

(a) Plot two side-by-side boxplots for the Brand A durations, and the Brand B durations. Summarize what you see. You can use `geom_boxplot`.



(b) For this part only, assume that the data comes from a normal distribution (for now, carelessly fail to check whether or not this is a reasonable assumption). Use a parametric test to test the hypothesis that there is no difference in average battery life between the two battery brands, against the alternative that there is a difference. You may assume that the battery durations have common variance.

Response: $H_0 : \mu_A = \mu_B$ where $\mu A$and μB are average battery life of two battery brands.
$H_A : \mu_A \neq \mu B$
Since the battery durations have common variance and the data come from a normal distributions and batteries from Brand A and Brand B are not paired, we can use $T = \frac{\bar{X}-\bar{Y}}{\S_p\sqrt{\frac{1}{n}+\frac{1}{m}}}$ where $S_p^2 = \frac{(n-1)S_x^2+(m-1)S_y^2}{n+m-2}$.
$S^2_{BrandA} = 23.225$ and $S^2_{BrandB} = 12.977$
$S_p^2 = \frac{9*S^2_{BrandA}+9*S^2_{BrandB}}{18} = \frac{9*23.225+9*12.977}{18} = 18.1015$
Then, $S_p = 4.254586$
Sample mean of Duration of Brand A: $\bar{X} = 10.693$
Sample mean of Duration of Brand B: $\bar{Y} = 6.75$
Our estimate of the average difference of the two brands is $\bar{X} - \bar{Y} = 10.693 - 6.75 =$

3.943 and its estimated standard error is

$$s_{\bar{X}-\bar{Y}} = s_p\sqrt{1/10 + 1/10} = 4.254586 * \sqrt{1/10 + 1/10} = 1.902709$$

Test statistics is then $3.943/1.902709 = 2.072308$. Two tailed p-value with degrees of freedom of 18 is 0.0528.
R-code is pval = 2 * pt(q=2.072308, df = 18, lower.tail=FALSE).

Since p-value is 0.0528, we fail to reject the H_0 that there is no difference in the average battery life between the two battery brands.

(c) Test the same hypothesis using a non-parametric method. You must write down a manual computation of the test statistic, but you may use a built-in function or a table to actually compute the test and get a p-value (as opposed to doing a normal approximation, because the sample size is small). You can also generate the null distribution of your statistic through simulation if you are ambitious, but using R built-in function is fine.

Response: Since two samples are independent to each other and we are not assuming any distributions for the samples, we will be using Mann-Whitney U test.
X represents Brand A and Y represents Brand B.
$X_1, \cdots, X_n$ with unknown distribution F. $Y_1, \cdots, Y_n$ with unknown distribution G.
H_0 : F = G , $H_a : F \neq G$

Brand	Value	Rank
A	3.03	1
B	3.19	2
B	4.26	3
B	4.47	4
B	4.53	5
B	4.67	6
B	4.69	7
A	5.53	8
A	5.60	9
B	6.79	10
A	9.30	11
B	9.37	12
A	9.92	13
A	12.51	14
B	12.75	15
B	12.78	16
A	12.95	17
A	15.21	18
A	16.04	19
A	16.84	20

R1 = Sum of ranks from Brand A = $1 + 8 + 9 + 11 + 13 + 14 + 17 + 18 + 19 + 20 = 130$.
R2 = Sum of ranks from Brand B = $2 + 3 + 4 + 5 + 6 + 7 + 10 + 12 + 15 + 16 = 80$.

$U_1 = 0 + 6 + 6 + 7 + 8 + 8 + 10 + 10 + 10 + 10 = 75$
$U_2 = 1 + 1 + 1 + 1 + 1 + 1 + 6 + 3 + 4 + 6 = 25$
$U = min(U_1, U_2) = min(85, 25) = 25$
To check if this U is correct, we can use rank sum formula.
$U_1 = 130 - 10(10 + 1)/2 = 75$ $U_2 = 80 - 10(10 + 1)/2 = 25$
$U = min(U_1, U_2) = min(85, 25) = 25$

Rejection region for the combination $(10, 10)$ for $\alpha = 0.05$ is 23. Since U = 25 is not less than 23, we do not reject H_0 at $\alpha = 0.05$ level.

(d) Do your conclusions from part (b) and (c) agree? Which test (the parametric or the non- parametric test) do you think is more appropriate for this data? (e.g., check the assumption from part (b))

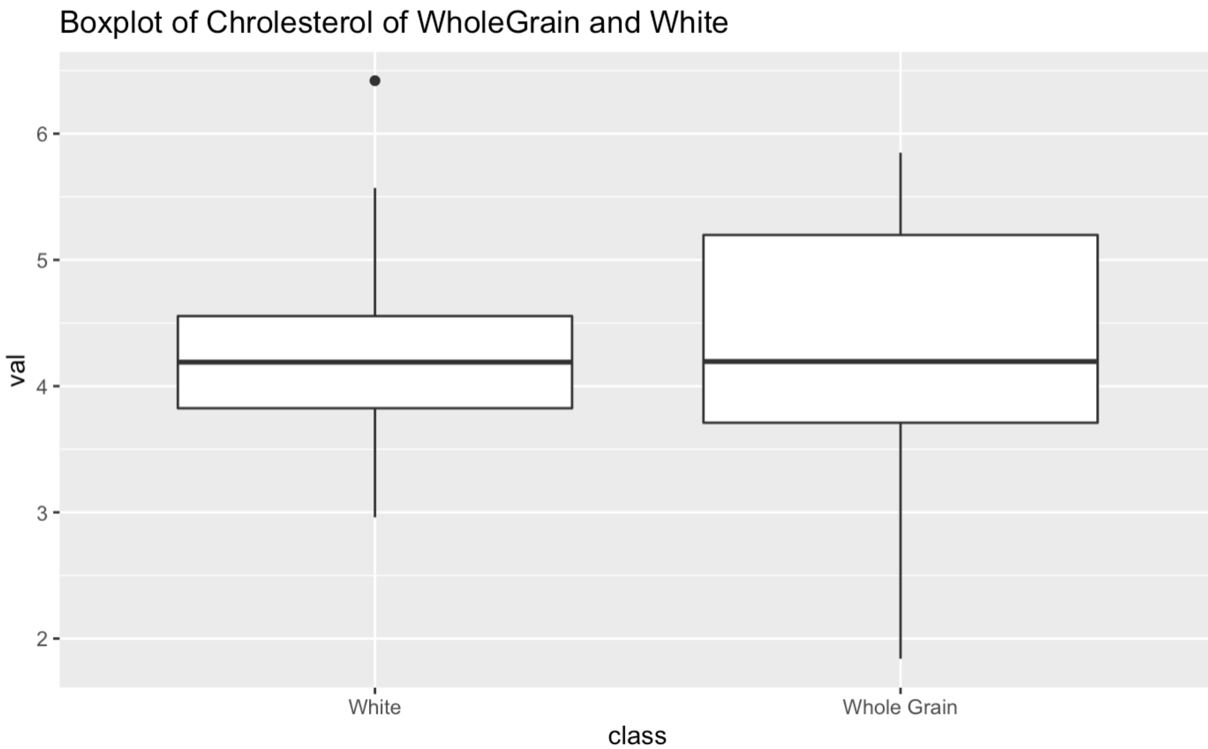
Response: The conclusions from part (b) and (c) agree because both fail to reject the null hypothesis. The main difference between (b) and (c) is that the parametric test assumes the distribution of the data to be normal. However, not only the sample size is small but also there is no evidence to assume that the data follow normality. Therefore, it is more appropriate to not assume normality in the data. Thus, non-parametric testing is more appropriate.

Problem 2

A clinical trial investigated whether eating whole grain bread lowered cholesterol levels. The trial randomly assigned 14 people to a meal plan in which they ate white bread for the first two weeks, and whole grain bread for the next two weeks. After each two-week period, cholesterol measurements were taken for each subject.
The variables `whole_grain` and white in the table represent cholesterol levels of the participant while they were on the whole grain bread diet and white bread diet respectively. This data can be found in the `cholesterol.csv` file.

Subject A	Whole Grain	White
1	4.61	3.84
2	6.42	5.57
3	5.40	5.85
4	4.54	4.80
5	3.98	3.68
6	3.82	2.96
7	5.01	4.41
8	4.34	3.72
9	3.80	3.49
10	4.56	3.84
11	5.35	5.26
12	3.89	3.73
13	2.25	1.84
14	4.24	4.14

(a) Visually compare the distribution of the whole grain diet and white bread diet cholesterol levels using side-by-side boxplots (`geom_boxplot`). Does it look like the whole grain diet cholesterol levels are lower than the white bread cholesterol levels?



Response: The median cholesterol cholesterol levels of whole grain diet and white bread diet are approximately the same. However, while white bread diet's boxplot seem to be approximately symmetric (with an exception of one outlier), whole grain's boxplot is right skewed, which means that most values of whole grain are 'small', but there are a few exceptionally large ones.

(b) For this part only, assume that the data comes from a normal distribution (carelessly ignoring whether or not this is a reasonable assumption). Use a parametric test to test the hypothesis that the cholesterol levels while on the whole grain bread diet is lower than the white bread diet.

Response: We need to perform paired test because the samples are dependent. We are measuring cholesterol from whole grain and white bread diet from the identical subjects. X: Whole Grain, Y : White
Vector of differences $X_i - Y_i$: (0.77, 0.85, -0.45, -0.26, 0.30, 0.86, 0.60, 0.62, 0.31, 0.72, 0.09, 0.16, 0.41, 0.10)
$H_0 : \mu_X - \mu_Y = 0$, $H_a : \mu_X < \mu_Y$
sample mean of Whole Grain dataset: $\bar{X} = 4.443571$
sample mean of White Bread dataset: $\bar{Y}4.080714$
Differences between the two samples: $\bar{D} = \bar{X} - \bar{Y} = 4.443571 - 4.080714 = 0.362857$.
$S_D = 0.4059638$
t-statistics is $0.362857/0.4059638 = 0.8938161$. The corresponding p-value for the t-stat of 0.8938161 is 0.1944975. Therefore, we fail to reject.

R code is pt(q=0.8938161, df = 12, lower.tail = FALSE).

(c) Perform a non-parametric test to test the same hypothesis. You must demonstrate how to compute the test statistic manually, but you may use a built-in function to actually compute the test and get a p-value (as opposed to doing a normal approximation, because the sample size is small). Again, if you are ambitious you can generate the null distribution yourself and compare the p-value from your manual distribution vs. the built-in function.

Response: For paired, non-parametric testing, we perform the Wilcoxon Signed Rank Test.

(a) Remove observations without any differences. There are no observations to remove.

(b) Compute R_i = rank of $|D_i|$.
$|D_i|$ = (0.77, 0.85, 0.45, 0.26, 0.30, 0.86, 0.60, 0.62, 0.31, 0.72, 0.09, 0.16, 0.41, 0.10)
R_i = (12, 13, 8, 4, 5, 14, 9, 10, 6, 11, 1, 3, 7, 2)

(c) Compute $W_i = |D_i| \times sign(D_i)$.
W_i = (+12, +13, -8, -4, +5, +14, +9, +10, +6, +11, +1, +3, +7, +2)

(d) Compute the test statistics $W_+ = \sum_{i:W_i>0} W_i$
$W_+ = 12 + 13 + 5 + 14 + 9 + 10 + 6 + 11 + 1 + 7 + 3 + 2 = 93$

(e) Compute pvalue $CODE : pval = wilcox.test(x = wholeGrain, y = White, alternative = "less", paired = TRUE)$
$p.value = 0.9966431$

Since p-value is large, we fail to reject the null.

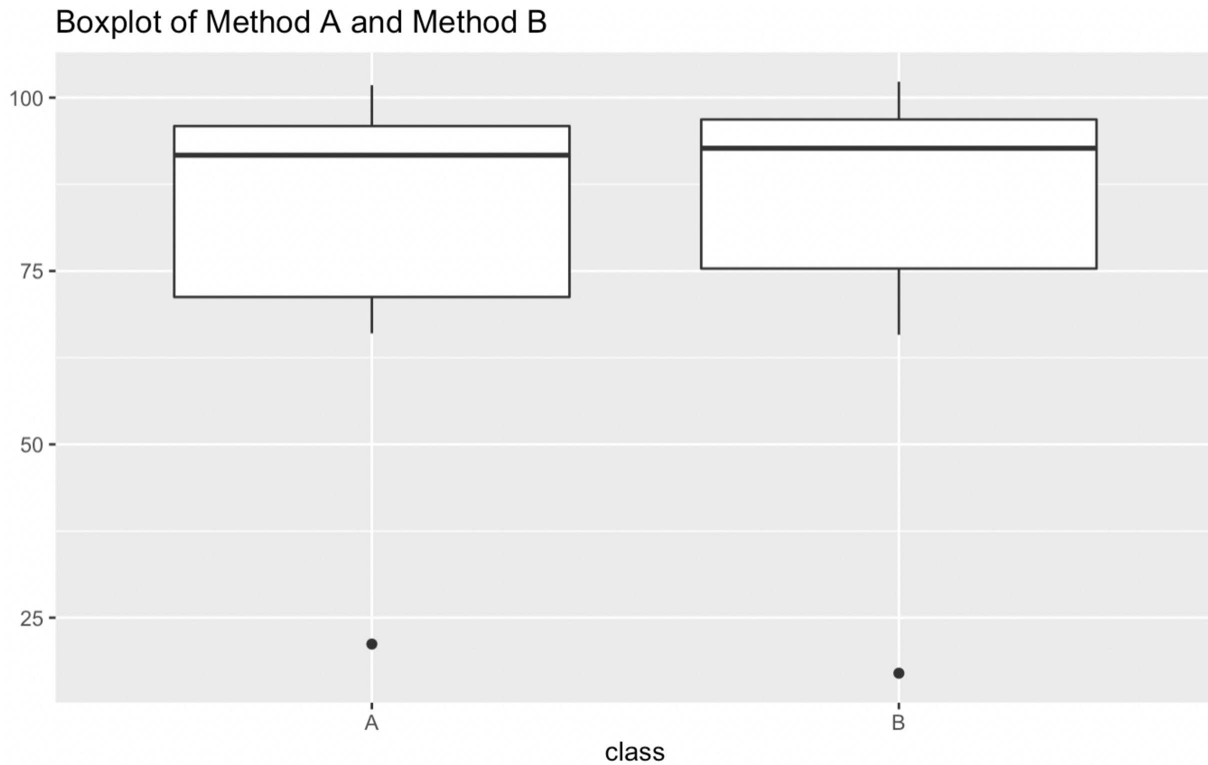
(d) Compare the p-value from the two tests. Are they very different? Which test do you think is a better choice for this data? (i.e. test your assumption)
Response: The p-value of the two tests are very different although they arrive at the same conclusion. The p-value calculated from the non-parametric test is significantly larger than the p-value from parametric test. From the box plot, it is difficult to assume normality of the data because whole grain's box plot is skewed. Therefore, it could be misleading to assume that the data are normally distributed.

Problem 3

There are two common techniques for evaluating the amount of amoxicillin in a tablet. A scientist used each method to determine the amount of amoxicillin in 15 tablets. To do this, the amount of amoxicillin in each tablet was assessed using method A, and was then also assessed using method B. The results are shown for the 15 tablets in the table below and can be found in `amoxicillin.csv`.

Tablet	Method A	Method B
1	93.7	97.2
2	102.3	97.8
3	96.0	96.2
4	96.5	101.8
5	90.3	88.0
6	75.7	74.0
7	68.5	75.0
8	68.5	67.5
9	66.0	65.8
10	17.0	21.2
11	91.7	94.8
12	87.3	95.8
13	92.7	98.0
14	92.7	99.0
15	87.5	100.2

(a) Visually compare the distribution of the doses as measured by method A and method B using side-by-side boxplots. Does it look like the distributions have similar means?



Response: Method A is slightly more skewed to the left than Method B but the two distributions look like they have approximately the same means. These two distributions have very similar median values.

(b) Assuming that the data came from a normal distribution, conduct a parametric paired test to test whether the two techniques agree on average.

Response: $H_0 : \mu_A = \mu_B$, $H_a : \mu_A \neq \mu_B$
$D = (-3.5, 4.5, -0.2, -5.3, 2.3, 1.7, -6.5, 1.0, 0.2, -4.2, -3.1, -8.5, -5.3, -6.3, -12.7)$
$S_D = sd(D) = 4.630767$
$\bar{D} = -3.06$
$|\bar{D}| = 3.06$ Under $H_0 : \mu_D = 0$, test statistics is t-distributed:

$$T = \frac{\bar{D} - 0}{S_D/\sqrt{n}} \sim t_{n-1}$$

Test statistics is $\frac{3.06}{4.630767/\sqrt{15}}$ =2.56
The two tailed P-value is 0.00227, which in R is $2{*}pt(q = 2.559258, df = 14, lower.tail = FALSE)$. Since p-value is small, we reject the null hypothesis. Therefore, there is a difference in the amount of amoxicillin in each tablet assess using method A and method B.

(c) Continue assuming that the data came from a normal distribution. Conduct a parametric unpaired test to test whether the two techniques agree on average.

Response: $H_0 : \mu_A = \mu_B$, $H_a : \mu_A \neq \mu_B$
Since we have no information about the variances of two groups (and they are not necessarily equal to each other), We can use $T = \frac{\bar{X}-\bar{Y}}{\sqrt{\frac{S_x^2}{n}+\frac{S_y^2}{m}}}$ where $df = \frac{(S_x^2/n+S_y^2/m)^2}{\frac{(S_x^2/n)^2}{n-1}+\frac{(S_y^2/m)^2}{m-1}}$.
Sample mean of Duration of Method A: $\bar{X} = 81.76$
Sample mean of Duration of Method B: $\bar{Y} = 84.82$
Sample variance of Duration of Method A: $S_x^2 = 449.2754$
Sample variance of Duration of Method B: $S_Y^2 = 464.566$
Test statistics = $\frac{\bar{X}-\bar{Y}}{\sqrt{S_x^2/n+S_y^2/m}} = \frac{-3.06}{7.805303} = -0.3920412$
s $df = \frac{3711.583}{64.07886+68.51478} = 27.99216$

Standard Error of the differences is $s_{\bar{X}-\bar{Y}} = \sqrt{449.2754/15 + 464.566/15} = 7.805303$
The two tailed p value equals 0.6980. R code is $2{*}pt(q = 0.3920412, df = 28, lower.tail = FALSE)$.
This difference is **not statistically significant**, therefore, we fail to reject the null hypothesis which states that the two techniques agree on average.

(d) You should find that you draw quite different conclusions from each test. Why do you think that happens?
Response: While parametric unpaired test fail to reject the null, the parametric paired test rejects the null. This suggests that the assumption of independence / dependence relationship between Method A and Method B affects the conclusions from the test. The main difference between the paired and unpaired is that while unpaired assumes equal variance between groups, paired does not assume equal variance across groups. Furthermore, the paired assumes that the two groups – Method A and Method B – are dependent with each other but the unpaired assumes independence between them. The assumption of dependence and independence greatly affects the p-value.?

Problem 4

Assume that we have some data, $X_1, \ldots, X_{25}$ and $Y_1, \ldots, Y_{25}$ and we are testing

$$H_0 : \mu_X = \mu_Y$$
$$H_1 : \mu_X \neq \mu_Y$$

Assume that $\text{Cov}(X_i, Y_i) = 50$, $_X = \sigma_Y = 10$.
The goal of this question is to compare the power of the paired test with the unpaired test on this data.

(a) Derive the formula for the power of the paired two-sample test, and a formula for the power of the unpaired two sample test (i.e., the test that incorrectly assumes that the covariance is equal to zero). For each of your final power formulas, substitute in the

known values of variances and covariance where relevant.
Response: The probability that the test statistics falls in the rejection region is equal to

$$1 - \Phi\left[z(\alpha/2) - \frac{\Delta}{\sigma}\sqrt{n/2}\right] + \Phi\left[-z(\alpha/2) - \frac{\Delta}{\sigma}\sqrt{n/2}\right]$$

For the **unpaired test** and substituting the known values, we get
$E(\bar{x} - \bar{y}) = \mu_x - \mu y$
$Var(\bar{x} - \bar{y}) = Var(\bar{x}) + Var(\bar{y}) = 2 * \sigma^2/n \ \Sigma_{/barx-/bary} = 10 * \sqrt{2/25} = 2\sqrt{2}$
$Power = 1 - \Phi(1.96 - \frac{\mu_x - \mu_y}{2/\sqrt{2}}) + \Phi(-1.96 - \frac{\mu_x - \mu_y}{2/\sqrt{2}})$

For the **paired test** and substituting the known values, we get

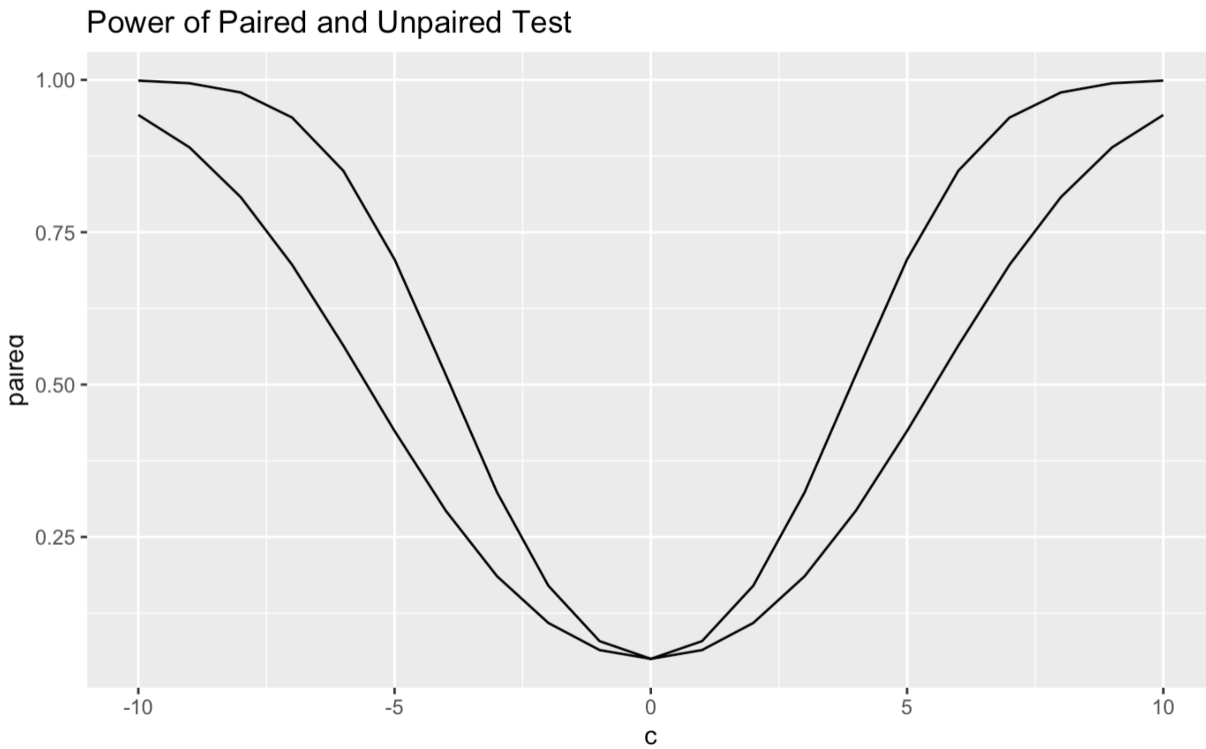
$$Var(D_i) = Var(X_i) + Var(Y_i) - 2Cov(X_i, Y_i) = 100 + 100 - 100 = 100$$

$S_D = 10$

$$Z = \frac{\bar{D} - E(\bar{D})}{10\sqrt{25}} = \frac{\bar{D} - (\mu_1 - \mu_2)}{2}$$

Under $H_0 : Z \sim N(0, 1)$, let $\alpha = 0.05$ and $Z(\alpha/2) = 1.96$ Under $H_a : \mu_X \neq \mu_Y$
Thus, Rejection Region = $|z| > z(1 - \alpha/2)$
Power $= P(|Z| > 1.96) = P(Z > 1.96) + P(Z < -1.96) = 1 - \Phi(1.96 - \frac{\mu_x - \mu_y}{2}) + \Phi(-1.96 - \frac{\mu_x - \mu_y}{2})$

(b) A power curve will plot the power of the two-sample test (y-axis) against the difference in means ($\delta = \mu_X - \mu_Y$) of the two samples under the alternative hypothesis (x-axis). On a single plot, plot a power curve for each of the two tests using ggplot. Summarize your conclusion.




```r
c = seq(-10,10)
paired = 1-pnorm(1.96 - c/2)+pnorm(-1.96 - c/2)
unpaired = 1-pnorm(1.96 - c/(2*sqrt(2)))+pnorm(-1.96 - c/(2*sqrt(2)))
df = data.frame(diffs = c, paired = paired, unpaired = unpaired)
ggplot(data=df) + geom_line(aes(x= c, y= paired)) + geom_line(aes(x= c, y= unpaired)) +labs(title =
'Power of Paired and Unpaired Test')
```

The conclusion is that the

Problem 5

Suppose that n measurements are to be taken under a treatment condition and another n measurements are to be taken independently under a control condition. It is thought that the standard deviation of a single observation is about 10 under both conditions.
How large should n be so that the test of $H_0 : \mu_X = \mu_Y$ against the one-sided alternative $H_A : \mu_X > \mu_Y$ has a power of 0.8 if $\mu_X - \mu_Y = 2$ and $\alpha = 0.1$.
Response: $\Phi(Z(\alpha) - \frac{\Delta}{\sigma}\sqrt{n/2}) = 0.8$
$(Z(\alpha) - \frac{\Delta}{\sigma}\sqrt{n/2}) = \Phi^{-1}(0.8) = 0.8416$
$Z(.1) + 0.8416 = \frac{\Delta}{\sigma}\sqrt{n/2}$
$n = ((Z(.1) + 0.8416) * \sigma/\Delta)^2 * 2$ $n = ((1.282 + 0.8416) * 10/2)^2 * 2 = 225.4838$ n should be approximately 225 or 226 (for more certainty, n should be **226**).