STROKE DATASET
PARAMETRIC TESTS WITH POWER AND EFFECT SIZE ANALYSIS

Group D

- Aya Abouelela
- Amel Khirreddine
- Kwaku Asamoah Gyimah
- Arlizze Faye R. Ongchua
- Supervisor: Prof. Elnaz Gholipiour

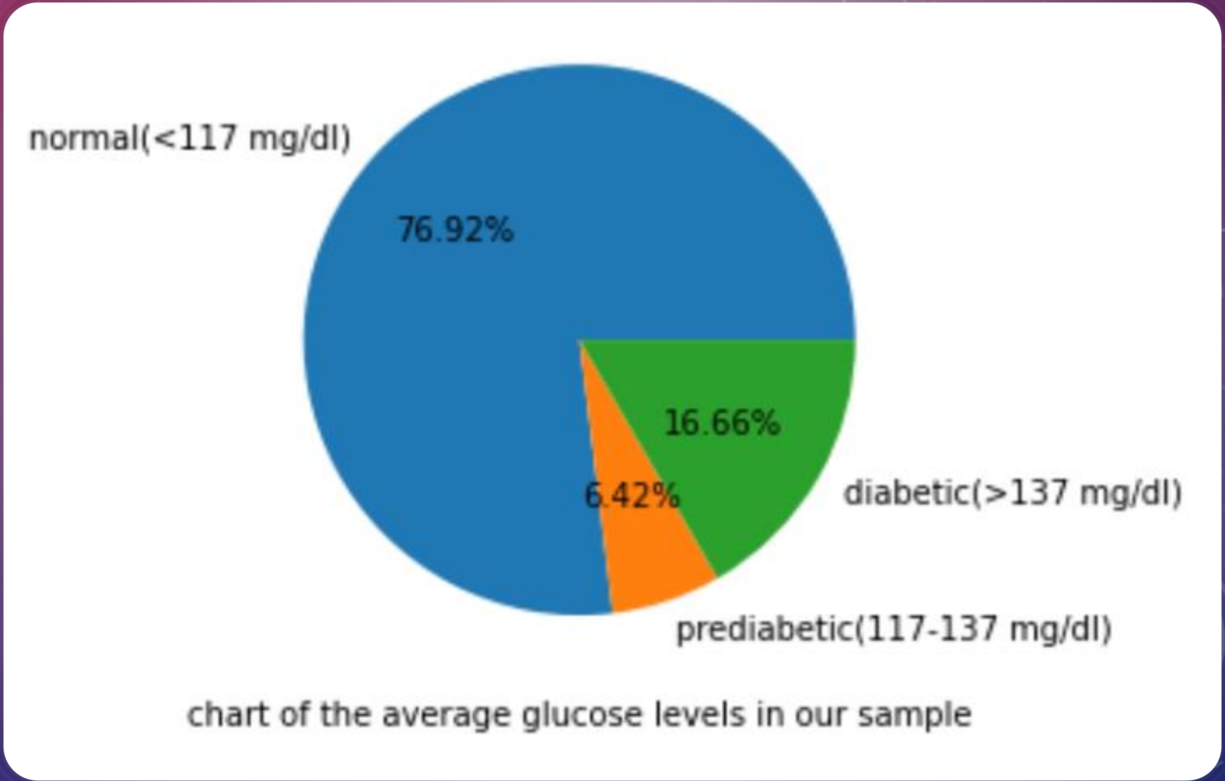
GOAL: USING PARAMETRIC TESTS TO VERIFY OR NULLIFY OUR OBSERVATIONAL NULL HYPOTHESES REGARDING OUR DATA. UTILISING POWER AND EFFECT SIZE ANALYSES TO QUANTITAVELY DETECT THE EFFECT WE HAVE IN OUR DATA AS WELL AS THE PROBABILITY OF MAKING AN ERROR.

	Unnamed: 0	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	avg_glucose_level	bmi	smoking_status	stroke
0	0	Male	58.0	1.0	0.0	Yes	Private	Urban	87.96	39.2	never smoked	0.0
1	1	Female	70.0	0.0	0.0	Yes	Private	Rural	69.04	35.9	formerly smoked	0.0
2	2	Female	52.0	0.0	0.0	Yes	Private	Urban	77.59	17.7	formerly smoked	0.0
3	3	Female	75.0	0.0	1.0	Yes	Self-employed	Rural	243.53	27.0	never smoked	0.0
4	4	Female	32.0	0.0	0.0	Yes	Private	Rural	77.67	32.3	smokes	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...
28911	29060	Female	10.0	0.0	0.0	No	children	Urban	58.64	20.4	never smoked	0.0
28912	29061	Female	56.0	0.0	0.0	Yes	Govt_job	Urban	213.61	55.4	formerly smoked	0.0
28913	29062	Female	82.0	1.0	0.0	Yes	Private	Urban	91.94	28.9	formerly smoked	0.0
28914	29063	Male	40.0	0.0	0.0	Yes	Private	Urban	99.16	33.2	never smoked	0.0
28915	29064	Female	82.0	0.0	0.0	Yes	Private	Urban	79.48	20.6	never smoked	0.0

H0: THE MEAN OF THE FASTING AVERAGE GLUCOSE LEVEL IN OUR DATA IS REPRESENTATIVE OF THE GERMAN POPULATION.*

- Fasting avg glucose level in Germany is 126mg/dl

- Mean of our sample is 106 mg/dl



- Ref: https://flexikon.doccheck.com/en/Blood_sugar_level#:~:text=Fasting%20blood%20sugar%3A%20<%207.0%20mmol,140-200%20mg%2Fdl

- *assuming our data comes from Germany

METHOD: ONE SAMPLE T-TEST

```
statistic, p = ttest_1samp(df['avg_glucose_level'], 126)
print('one sample t test','statistic=%0.3f, p-value = %0.3f\n' % (statistic, p))
```

```
one sample t test statistic=-73.674, p-value = 0.000
```

Effect-size for one sample $d = (m1 - \mu)/s$

$d = 0.4332545290635978$ (small effect)

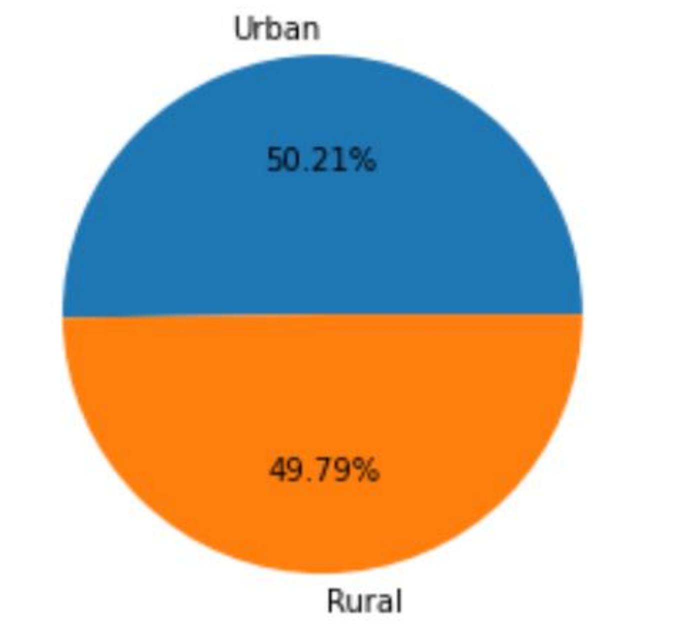
```
power_analysis = smp.TTestIndPower()

p = power_analysis.solve_power(effect_size = 0.4332545290635978, nobs1 = 28915, alpha = 0.05)
print('one sample t test','power = %0.3f\n' % (p))
```

```
one sample t test power = 1.000
```

Conclusion: The mean of the avg glucose levels in our sample does not represent the German population's mean of fasting average glucose levels.

H0: THERE IS NO DIFFERENCE IN THE AVERAGE GLUCOSE LEVELS BETWEEN PEOPLE WHO LIVE IN URBAN CITIES OR RURAL AREAS.

- Total number of entries in our data is 28916.
- Difference in frequency between the two categories is 124.
- We choose a random sample of the urban category to match the number of entries of the rural category > necessary from the unpaired t-test.



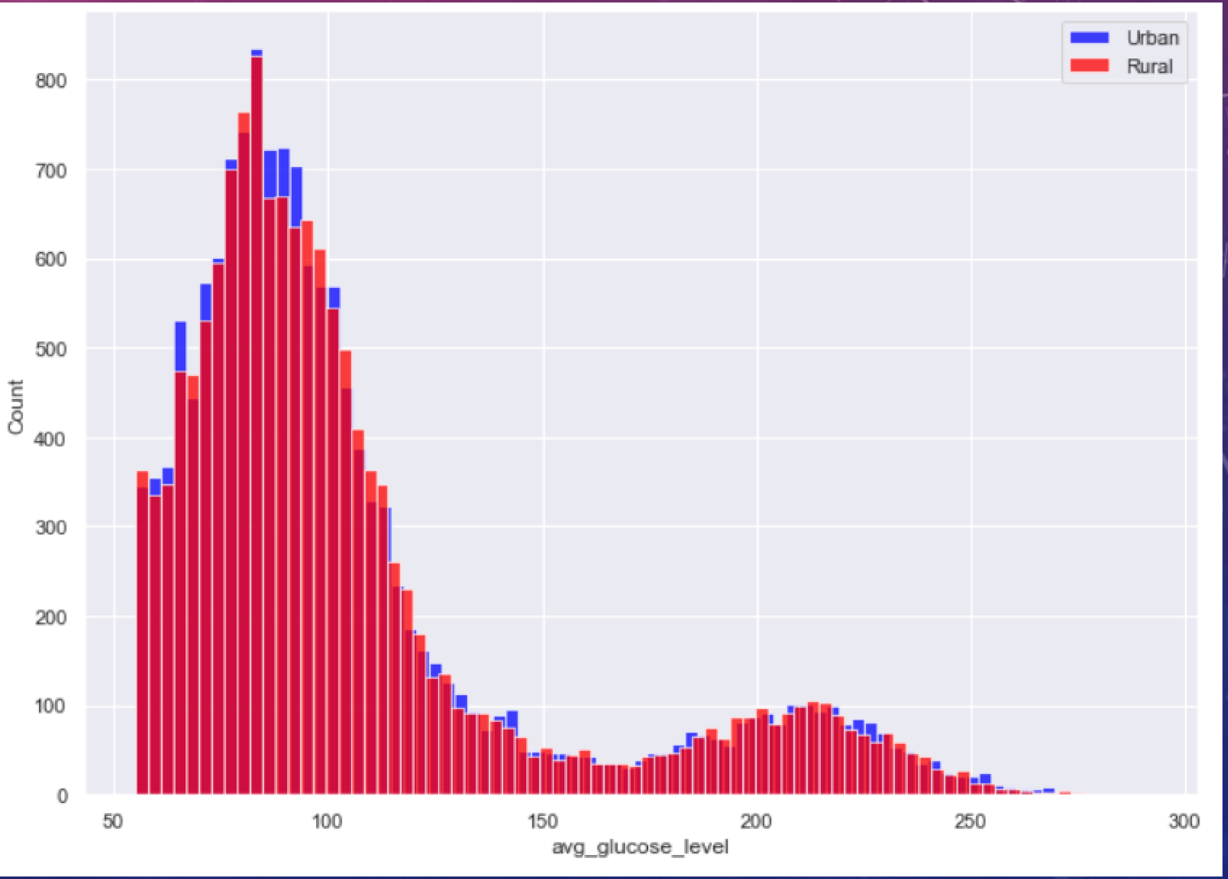
percentage of residence types in our data

METHOD: UNPAIRED T-TEST

```
df_rural = df.loc[df['Residence_type'] == 'Rural'].avg_glucose_level
df_urban = df.loc[df['Residence_type'] == 'Urban'].avg_glucose_level.sample(df_rural.size)

ttest_ind(df_rural, df_urban)
```

```
Ttest_indResult(statistic=0.46324199476053196, pvalue=0.6431944196835688)
```

- We calculate the effect-size using Cohen's d
- D = 0.005460307316093725 (small effect size)
- Power = 0.0749249759077351 4

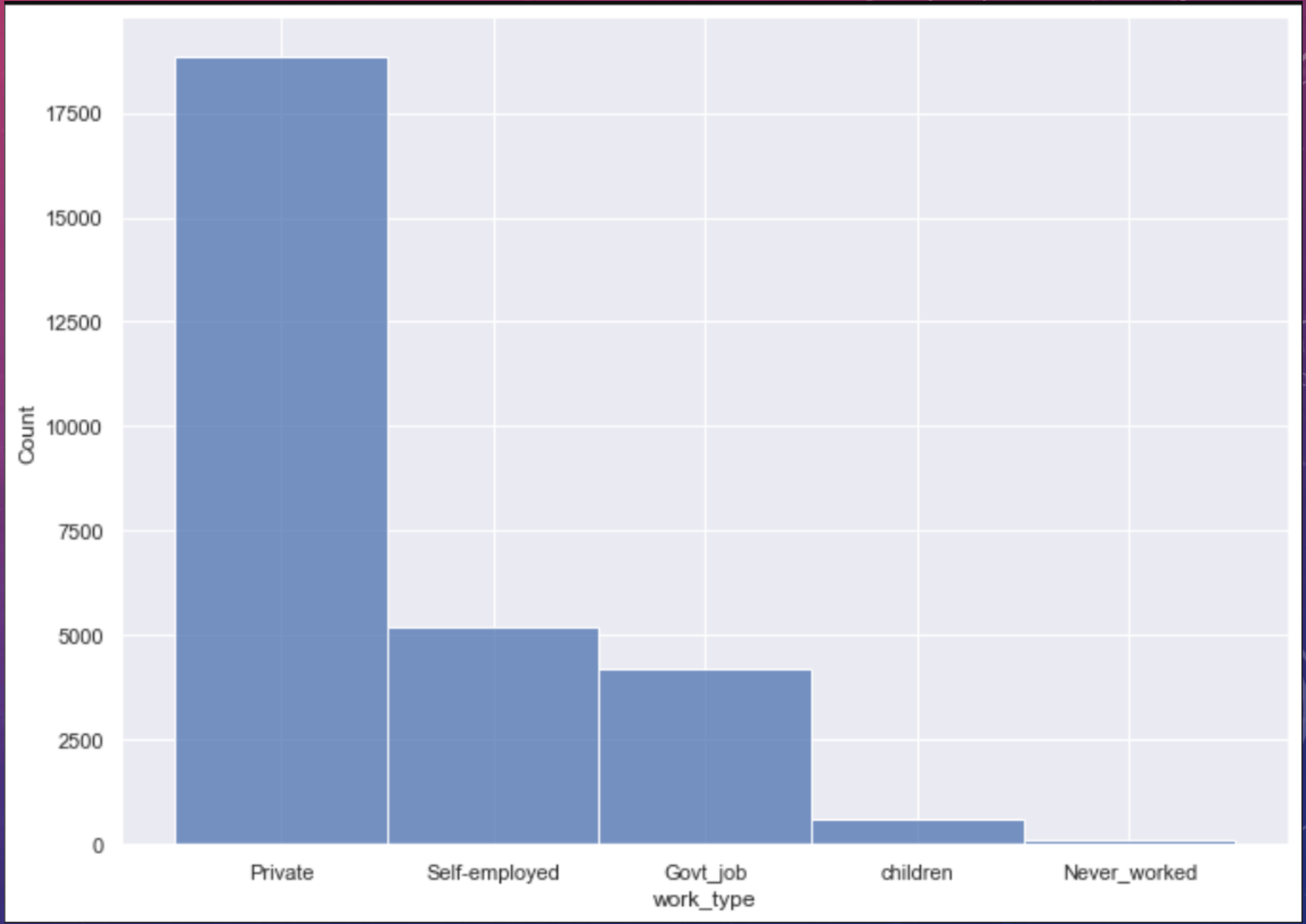


Conclusion: We accept our null hypothesis! However, if the null hypothesis was to be false, there would have been a 93% probability that would accept it (type II error)

H0: A PERSON'S BLOOD SUGAR LEVEL IS INDEPENDENT OF THE NATURE OF THEIR WORK

- We neglect the categories 'Children' and 'Never worked' as they do not have a large statistical weight.

- We use random sampling to have the same sample size for all the three categories 'Private', 'Self-employed' and 'Govt job'.



METHOD: ONE WAY ANOVA

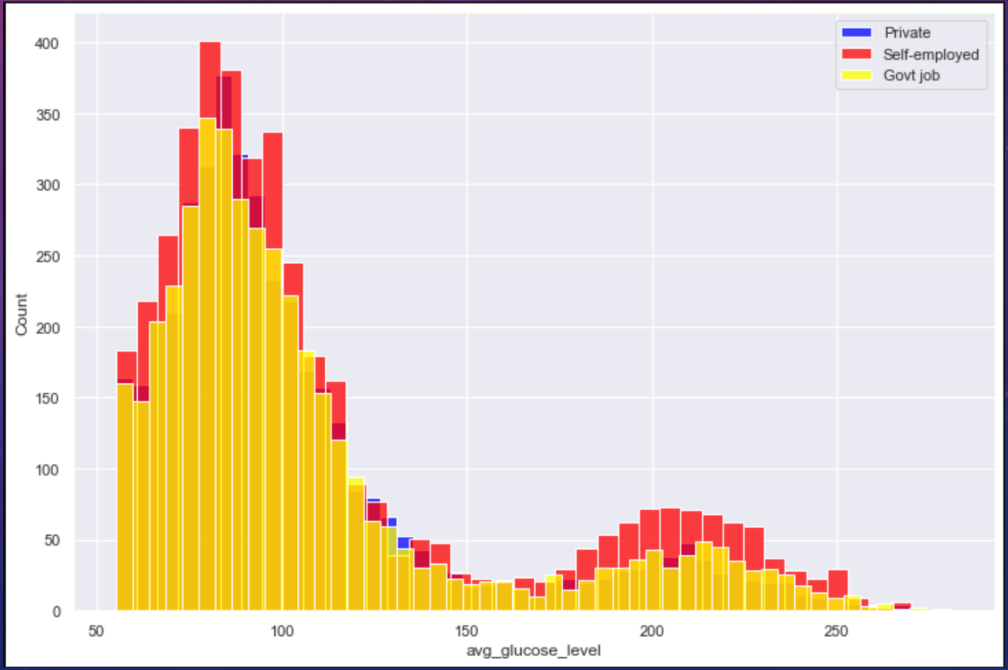
Effect size f for Anova is calculated using η2

f = sqrt(η2/1-η2)

The null hypothesis is true for people working in the private sector vs those who have a government job, however, it is not true when we include self-employment as a work type. This leads for a p value less than 0.05 for the Anova test.
Since the power is a very small value, there is a huge probability of falsely accepting H0.

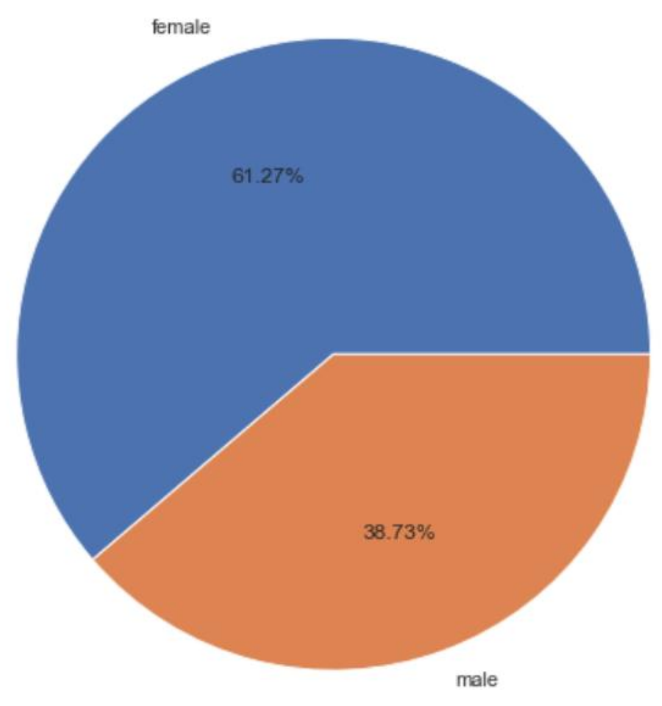
	Source	SS	DF	MS	F	p-unc	np2	power
0	work_type	1.302242e+05	2	65112.108058	29.656121	1.415429e-13	0.00471	0.057017
1	Within	2.751928e+07	12534	2195.570640	NaN	NaN	NaN	NaN

	A	B	mean(A)	mean(B)	diff	se	T	p-tukey	eta-square
0	Govt_job	Private	107.060897	105.002986	2.057911	1.025068	2.007586	1.103899e-01	0.000530
1	Govt_job	Self-employed	107.060897	112.632405	-5.571508	1.025068	-5.435258	1.670064e-07	0.003281
2	Private	Self-employed	105.002986	112.632405	-7.629419	1.025068	-7.442844	0.000000e+00	0.006481



H0: THERE IS NO DIFFERENCE IN THE AVERAGE BLOOD SUGAR LEVELS BETWEEN MALES AND FEMALES ACROSS ALL WORK TYPES

- We take equal distribution of males and females in our sample before running our parametric test.
- The number of entries after sampling is 4856 for each gender.
- Dependent variable is avg glucose level. Independent variables are gender and work type.



Gender distribution in our sample

METHOD: TWO WAY ANOVA

- Work type alone or gender alone do seem to affect the average blood sugar levels. However, there seems to be no influence from gender on the work type.
- The huge overlap in our data across the targeted categories leads to small effect sizes and reduce the power significantly, despite having a large sample size.

	Source	SS	DF	MS	F	p-unc	np2	power
0	work_type	1.153083e+05	2.0	57654.131402	25.497339	9.028887e-12	0.005226	0.060089
1	gender	5.744784e+04	1.0	57447.838942	25.406107	4.728265e-07	0.002611	0.052488
2	work_type * gender	1.929925e+03	2.0	964.962360	0.426751	6.526381e-01	0.000088	0.050003
3	Residual	2.194704e+07	9706.0	2261.182296	NaN	NaN	NaN	NaN

```
        A     B    mean(A)    mean(B)      diff        se         T  \
0  Female  Male 106.478812 111.406316 -4.927504  0.967411 -5.093498

        p-tukey  eta-square
0  3.581181e-07    0.002664
```

DERIVING THE SAMPLE SIZE FOR OUR TWO SAMPLED T-TEST ON BMI

According to Cohen , a small effect size is 0.2

Based on this effect size, sig.level and power, we will need 394 participants in each group to perform our Two sample T-Test.

Our data already consists of 28,916 observations

```
> cohen.ES(test= 't', size= 'small')

     Conventional effect size from Cohen (1982)

           test = t
           size = small
    effect.size = 0.2
```

```
> t_power= pwr.t.test(d=0.2, sig.level=0.05, power= 0.8)
> t_power

     Two-sample t test power calculation

              n = 393.4057
              d = 0.2
      sig.level = 0.05
          power = 0.8
    alternative = two.sided

NOTE: n is number in *each* group
```

H0: THERE IS NO SIGNIFICANT DIFFERENCE IN THE AVERAGE BMI BETWEEN PEOPLE WHO LIVE IN URBAN CITIES AND RURAL AREAS.

- Method of sampling- Stratified
- Method- Two Sampled t test
- In total we sampled 788 observations from our dataset, 394- Rural residents and 394- Urban residents.
- From the results below, the p-value = 0.3463 . The p-value > 0.05 , hence we fail to reject the null hypothesis. There is no significant difference in the BMI between people in the Rural and Urban areas. Hence residence type is not important for our study.

```
> t.test(bmi~Residence_type, data=stroke_sample)

	Welch Two Sample t-test

data:  bmi by Residence_type
t = -0.94231, df = 786, p-value = 0.3463
alternative hypothesis: true difference in means between group Rural and group Urban
is not equal to 0
95 percent confidence interval:
 -1.4922802  0.5242599
sample estimates:
mean in group Rural mean in group Urban
           29.91497            30.39898
```

H0: THE MEAN OF BMI IN OUR DATA IS NOT SIGNIFICANTLY DIFFERENT THAN THAT OF THE GERMAN POPULATION.

- The Average BMI for both male and female in Germany is 26.3
  source: https://en.wikipedia.org/wiki/List_of_sovereign_states_by_body_mass_index
- The mean of BMI in our data is 30.05. It's effect size on our population mean is d= 0.52 , Which is a medium effect size.
- We found the power to be 1 based on the effect size above.

```
> pwr.t.test(n=length(strokenew$bmi),d=0.52,sig.level=0.05,type="one.sample",alterna
tive="two.sided")

     One-sample t test power calculation

              n = 28916
              d = 0.52
      sig.level = 0.05
          power = 1
    alternative = two.sided
```

H0: THE MEAN OF BMI IN OUR DATA IS NOT SIGNIFICANTLY DIFFERENT THAN THAT OF THE GERMAN POPULATION.

- We perform a one-sample t-test to draw our conclusions.
- And we conclude from the results that the mean of BMI in our sample is significantly different from that of the German population.
- In short, the BMI of our sample falls within the obese range which is way different from the population.

```
> t.test(strokenew$bmi, mu = 26.3)

        One Sample t-test

data:  strokenew$bmi
t = 88.675, df = 28915, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 26.3
95 percent confidence interval:
 29.96678 30.13254
sample estimates:
mean of x
 30.04966
```

H0: A PERSON'S BMI IS INDEPENDENT OF THE NATURE OF THEIR WORK

- Method One-Way Anova
- From the previous slides, we suggested 394 participants in each group. Going by that, we used a sample size of 1576 to perform our One-Way Anova test. Thus; 394 from each of Private, Self-employed, Gov't and Children.
- From below, the p-value < 0.001, hence we reject the null hypothesis. There is a significant difference between the mean of the four groups.
- A person's BMI is dependent on the nature of their work. The work type is significant for our study on BMI

```
> res.aov <- aov(bmi ~ work_type, data = stroke_sample2)
> # Summary of the analysis
> summary(res.aov)
              Df Sum Sq Mean Sq F value Pr(>F)    
work_type      3  14882    4961   107.6 <2e-16 ***
Residuals   1572  72488      46                   
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> 
```

TWO WAY ANOVA – ADDING GENDER TO THE MODEL

- The observed power for work_type on BMI is 1.0, meaning here we want to have an 100% chance of correctly rejecting the null hypothesis.
- That of gender is very low
- The interaction between work_type and gender has a power of 33%. For a sample size of 1,576, it is a low rate.

Tests of Between-Subjects Effects

Dependent Variable: bmi

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	15073.629[a]	7	2153.376	46.703	<.001	.173	326.922	1.000
Intercept	1237398.414	1	1237398.414	26837.151	<.001	.945	26837.151	1.000
work_type	14977.884	3	4992.628	108.282	<.001	.172	324.846	1.000
gender	23.100	1	23.100	.501	.479	.000	.501	.109
work_type * gender	171.231	3	57.077	1.238	.294	.002	3.714	.334
Error	72296.822	1568	46.108					
Total	1375388.840	1576						
Corrected Total	87370.451	1575						

a. R Squared = .173 (Adjusted R Squared = .169)

b. Computed using alpha = .05

- **After the test Hypothesis, findings;**
- work type has an effect on the BMI.
- Gender has no effect on BMI. We don’t care whether female or male.
- There is no interaction between work_type and Gender.

Between-Subjects Factors

		N
work_type	1	394
	2	394
	3	394
	4	394
gender	Female	935
	Male	641

CONCLUSION

- Despite having a large sample, we have detected small effect sizes between our variables due to the massive overlap in our data.
- The small effect size decreases the power of our tests in a way that is out of our control.
- We could increase the value of the power by changing alpha but the analysis showed that increasing alpha did not significantly increase our power so we opted against it.