

Blood Pressure Analysis

Team 12: Chris Coker, Isaac Graham, and Rhiannon Stracener

Research Question:

How does physical activity affect blood pressure after controlling for demographic information, BMI and smoking/alcohol use?

Data Description

Our data came from the National Health and Nutrition Examination Survey (NHANES) from 2017-2018. This survey is a major program of studies designed to assess the health and nutritional status of adults and children in the United States. It combines demographic and health interviews with physical examinations and lab tests to collect the data.

We used the following variables (some that we combined/created):

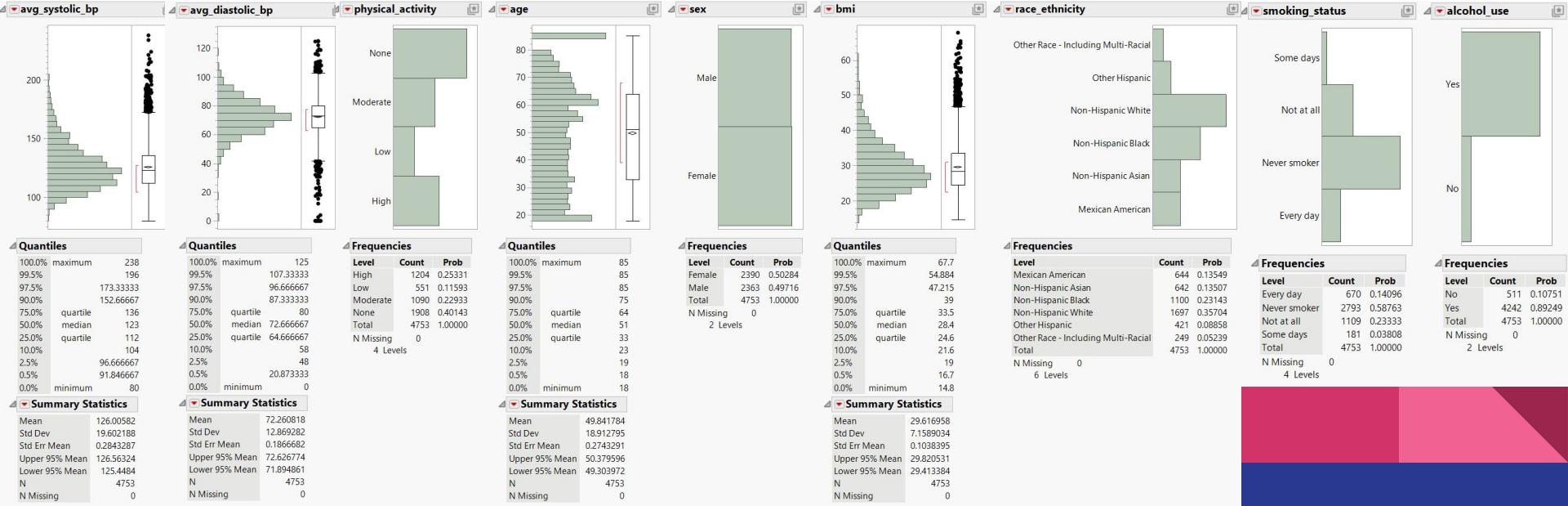
- Responses
 - Systolic blood pressure (Continuous)
 - Diastolic blood pressure (Continuous)
- Treatment
 - Physical activity (Nominal - 4 levels)
- Covariates
 - Age (Continuous)
 - Sex (Nominal)
 - Race/ethnicity (Nominal)
 - BMI (Continuous)
 - Smoking status (Nominal)
 - Alcohol use (Nominal)
- Systolic and diastolic blood pressure were recorded separately. Thus, we run two analyses, one on each type of blood pressure. Systolic and diastolic are our two outcomes.
 - Systolic - measures pressure in arteries when heart beats and pumps blood out to your body (top number)
 - Major risk factor for cardiovascular disease, especially in older adults
 - Diastolic - measures pressure in arteries when heart rests between beats and refills with blood (bottom number)

Data Preparation

- This dataset was imported, prepped, and cleaned using R.
- Blood pressure was recorded at three different intervals. For analysis purposes, we created one factor that averaged the three recordings.
- There are over 200 recorded variables in the 2017-2018 survey. As seen by our research question, we only wanted variables about blood pressure, physical activity, and possible covariates.
- As a result, we intuitively removed all variables not related to one of these categories.
- Additionally, we only used data points with complete data. This still left us with 4,753 data points to analyze.
- Finally, we renamed the variables to be easier to read.

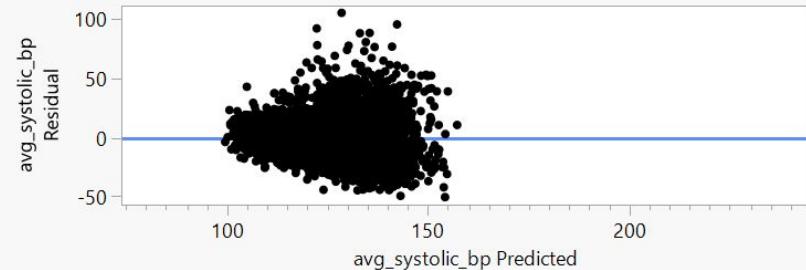
Exploratory Data Analysis

- We re-coded ages 80+ to be the weighted mean age for ages 80+, 85 (stated in NHANES documentation). This is because the original dataset sets a ceiling age of 80 which underestimates the true age of the oldest participants. This change treats the 80+ group as having an average age of 85 which is statistically more accurate for regression

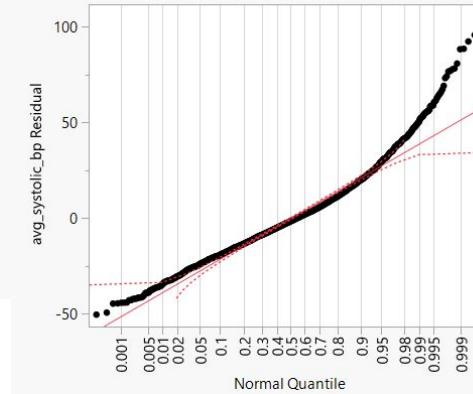


Assumptions Check: Systolic

Residual by Predicted Plot



Residual Normal Quantile Plot



Goodness-of-Fit Test

	A ²	Simulated p-Value
Anderson-Darling	31.0416	<.0001*

Note: Ho = The data is from the Normal distribution. Small p-values reject Ho.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	51	523318.7	10261.2	37.0314
Error	4701	1302617.2	277.1	Prob > F
C. Total	4752	1825936.0		<.0001*

- Looking at the residual QQ plot, we can see that while not perfectly normal, the skewness is moderate. With a sample size of over 4,000, this only moderate skewness is acceptable as ANCOVA is generally considered robust to violations of normality with such a large sample size.
- The residual by predicted plot shows our homogeneity of variances assumption is satisfied

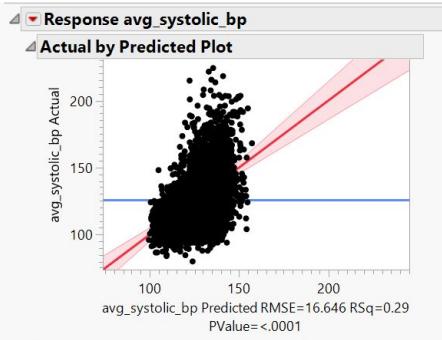
Model (Systolic)

$$Y_i = \beta_0 + \beta_1 X_{\text{physical activity}} + \beta_2 X_{\text{age}} + \beta_3 X_{\text{sex}} + \beta_4 X_{\text{bmi}} + \beta_5 X_{\text{race}} + \beta_6 X_{\text{smoking}} + \beta_7 X_{\text{alcohol}} + PA * \text{age} + PA * \text{sex} + PA * \text{bmi} + PA * \text{Race} + PA * \text{smoking} + PA * \text{alcohol} + \epsilon$$

- We used an ANCOVA model.
- Y_i is the model outcome: avg_systolic_bp
- β_0 is the model intercept.

Effect Summary

Source	Logworth	PValue
age	173.640	0.0000
bmi	20.735	0.0000
race_ethnicity	12.292	0.0000
age*physical_activity	8.176	0.0000
sex	7.326	0.0000
sex*physical_activity	3.120	0.0076
smoking_status	1.300	0.0503
bmi*physical_activity	0.999	0.10032
physical_activity	0.978	0.10508
alcohol_use	0.860	0.13817
physical_activity*smoking_status	0.181	0.58573
race_ethnicity*physical_activity	0.163	0.68653
physical_activity*alcohol_use	0.006	0.98641



Summary of Fit

RSquare	0.286603
RSquare Adj	0.278864
Root Mean Square Error	16.64613
Mean of Response	126.0058
Observations (or Sum Wgts)	4753

Analysis of Variance

Source	DF	Sum of		F Ratio
		Squares	Mean Square	
Model	51	523318.7	10261.2	37.0314
Error	4701	1302617.2	277.1	Prob > F
C. Total	4752	1825935.9		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	90.598966	1.50425	60.23	<.0001* ^
age	0.4721459	0.16069	29.38	<.0001*
sex[Female]	-1.529164	0.27952	-5.47	<.0001*
race_ethnicity[Mexican American]	-0.135975	0.686014	-0.20	0.8429
race_ethnicity[Non-Hispanic Asian]	-0.459725	0.175741	-0.64	0.5207
race_ethnicity[Non-Hispanic Black]	3.920968	0.573715	6.83	<.0001*
race_ethnicity[Non-Hispanic White]	-2.156322	0.527928	-4.08	<.0001*
race_ethnicity[Other Hispanic]	-0.274247	0.814466	-0.34	0.7363
bmi	0.397573	0.041587	9.56	<.0001*
physical_activity[High]	1.3712466	1.073303	1.28	0.2015
physical_activity[Low]	-2.347168	1.580101	-1.49	0.375
physical_activity[Moderate]	-0.862246	1.080771	-0.80	0.4250
smoking_status[Every day]	-1.089422	0.870744	-1.25	0.2109
smoking_status[Never smoker]	-0.172724	0.531066	-0.33	0.7450
smoking_status[Not at all]	1.304572	0.641153	-2.03	0.0419*
alcohol_use[No]	0.7225754	0.487276	1.48	0.1382
(age-49.8418)*physical_activity[High]	-0.074758	0.25759	-2.90	0.0037*
(age-49.8418)*physical_activity[Low]	-0.058592	0.03588	-1.63	0.1025
(age-49.8418)*physical_activity[Moderate]	-0.001966	0.025346	-0.08	0.9382
physical_activity[High]*smoking_status[Every day]	1.2693852	1.136362	1.12	0.2640
physical_activity[High]*smoking_status[Never smoker]	-0.765578	0.799751	-0.96	0.3385
physical_activity[High]*smoking_status[Not at all]	0.2867191	0.944373	0.30	0.7619
physical_activity[Low]*smoking_status[Every day]	-1.091265	2.320244	-0.49	0.6246
physical_activity[Low]*smoking_status[Never smoker]	0.8394064	1.176811	0.71	0.4757
physical_activity[Low]*smoking_status[Not at all]	0.3825128	1.474498	0.26	0.7953
physical_activity[Moderate]*smoking_status[Every day]	0.2217241	1.255737	0.18	0.8599
physical_activity[Moderate]*smoking_status[Never smoker]	1.1065653	0.878708	1.26	0.2080
physical_activity[Moderate]*smoking_status[Not at all]	0.5632077	0.308017	0.55	0.5848
physical_activity[High]*alcohol_use[No]	0.2003486	0.849741	0.24	0.8136
physical_activity[Low]*alcohol_use[No]	-0.35468	1.057348	-0.34	0.7373
physical_activity[Moderate]*alcohol_use[No]	0.161465	0.775349	0.21	0.8350
(bmi-29.617)*physical_activity[High]	0.0357383	0.063734	0.56	0.5750
(bmi-29.617)*physical_activity[Low]	-0.011866	0.096321	-0.12	0.9019
(bmi-29.617)*physical_activity[Moderate]	0.0592453	0.064339	1.43	0.1517
race_ethnicity[Mexican American]*physical_activity[High]	-1.381453	1.064449	-1.30	0.1944
race_ethnicity[Mexican American]*physical_activity[Low]	0.9812728	1.462936	0.67	0.5024
race_ethnicity[Mexican American]*physical_activity[Moderate]	0.2598876	1.185496	0.22	0.8265
race_ethnicity[Non-Hispanic Asian]*physical_activity[High]	-0.407347	1.395392	-0.29	0.7704
race_ethnicity[Non-Hispanic Asian]*physical_activity[Low]	-0.548767	1.273373	-0.43	0.6665
race_ethnicity[Non-Hispanic Asian]*physical_activity[Moderate]	0.0437278	1.26572	0.03	0.9724
race_ethnicity[Non-Hispanic Black]*physical_activity[High]	0.4067717	0.924126	0.44	0.6598
race_ethnicity[Non-Hispanic Black]*physical_activity[Low]	-0.720088	1.221638	-0.49	0.5556
race_ethnicity[Non-Hispanic Black]*physical_activity[Moderate]	-1.197205	0.982296	-1.22	0.2230
race_ethnicity[Non-Hispanic White]*physical_activity[High]	0.2580864	0.815311	0.32	0.7516
race_ethnicity[Non-Hispanic White]*physical_activity[Low]	-0.078195	1.19315	-0.07	0.9477
race_ethnicity[Non-Hispanic White]*physical_activity[Moderate]	0.9048768	0.838273	1.08	0.2804
race_ethnicity[Other Hispanic]*physical_activity[High]	0.9262517	1.340633	0.69	0.4897
race_ethnicity[Other Hispanic]*physical_activity[Low]	0.5666975	1.078099	0.33	0.7402
race_ethnicity[Other Hispanic]*physical_activity[Moderate]	-0.573831	1.406285	-0.41	0.6833
sex[Female]*physical_activity[High]	-0.627895	0.463242	-1.36	0.1753
sex[Female]*physical_activity[Low]	-1.037954	0.590687	-1.76	0.0789
sex[Female]*physical_activity[Moderate]	0.067911	0.46297	0.15	0.8834

Significant main effects:

- Age, sex, race_ethnicity, and bmi

Significant interactions:

- physical_activity*sex and physical_activity*age

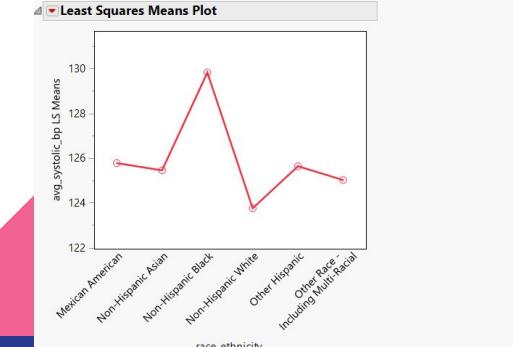
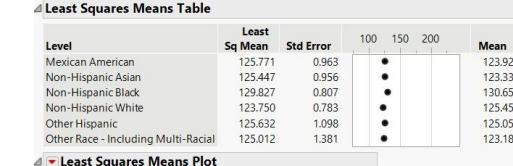
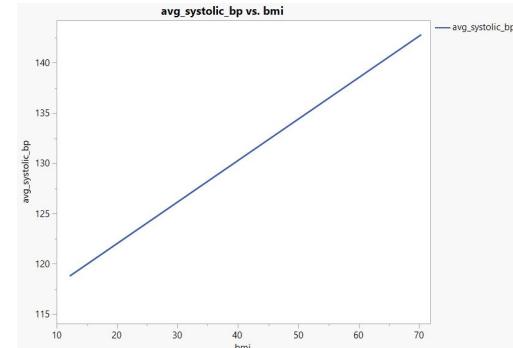
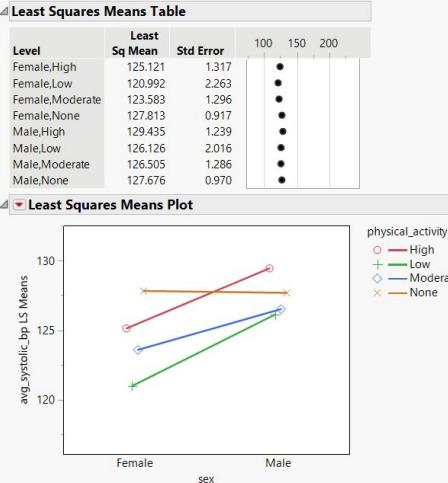
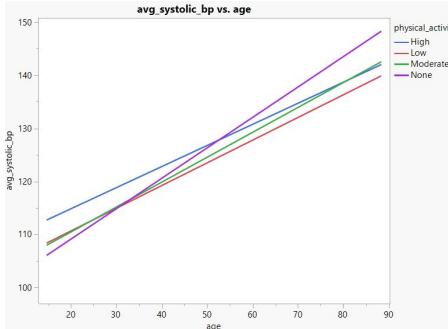
→ provides evidence of unequal slopes
→ the effect of physical activity on blood pressure varies depending on sex and age

Effect Tests

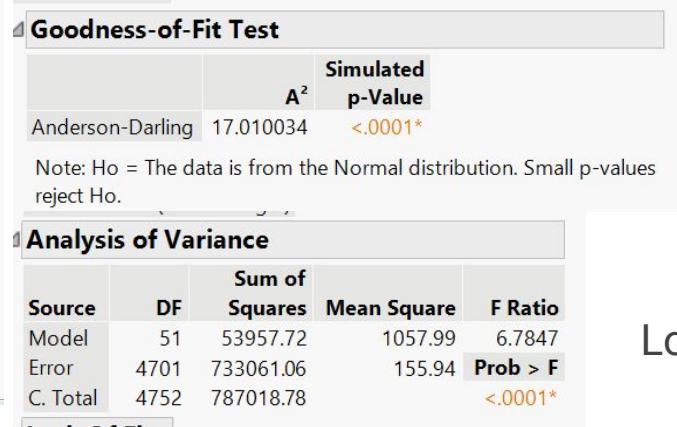
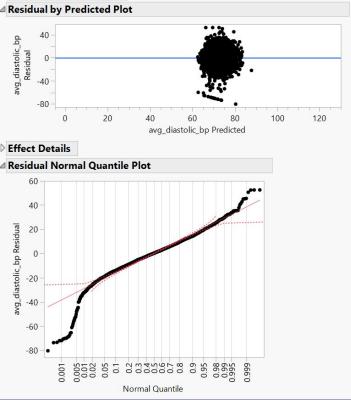
Source	Nparm	DF	Sum of Squares		F Ratio	Prob > F
			F Ratio	Prob > F		
age	1	1	239222.90	863.3287	<.0001*	
sex	1	1	8292.55	29.9269	<.0001*	
race_ethnicity	5	5	18592.27	13.4195	<.0001*	
bmi	1	1	25325.31	91.3962	<.0001*	
physical_activity	3	3	1701.74	2.0471	0.1051	
smoking_status	3	3	2165.34	2.6048	0.0501	
alcohol_use	1	1	609.32	2.1990	0.1382	
age*physical_activity	3	3	11397.80	13.7111	<.0001*	
physical_activity*smoking_status	9	9	1882.06	0.7547	0.6587	
physical_activity*alcohol_use	3	3	39.24	0.0472	0.9864	
bmi*physical_activity	3	3	1731.19	2.0826	0.1003	
race_ethnicity*physical_activity	15	15	3297.23	0.7933	0.6865	
sex*physical_activity	3	3	4676.52	5.6257	0.0008*	

Systolic Interpretation

- age*physical_activity
 - For the reference group, systolic blood pressure rises 0.47 points per year increase in age. For the high activity group, it rises only 0.4 each year relative to the average age of 49.84. The negative interaction estimate says the slope is flatter (effect of age is weaker) for that group.
- sex*physical_activity
 - None of the groups are statistically different from the reference (Male-no physical activity), so there's likely differences between non-reference groups
 - LSMeans shows Males have higher BP on average compared to Females at all physical activity levels
- bmi
 - For every 1 unit increase in BMI, systolic blood pressure increases by 0.39 points, holding all other variables constant
- Race_ethnicity
 - Non-Hispanic Black individuals have a systolic blood pressure that's, on average, 3.92 points higher than the reference group
 - Non-Hispanic White individuals have a systolic blood pressure that's, on average, 2.15 points lower than the reference group



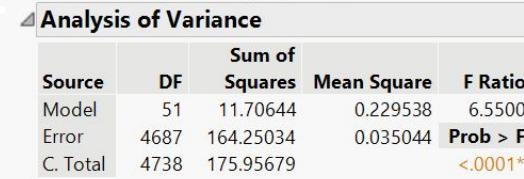
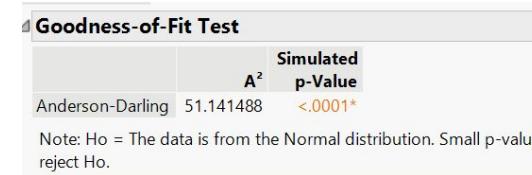
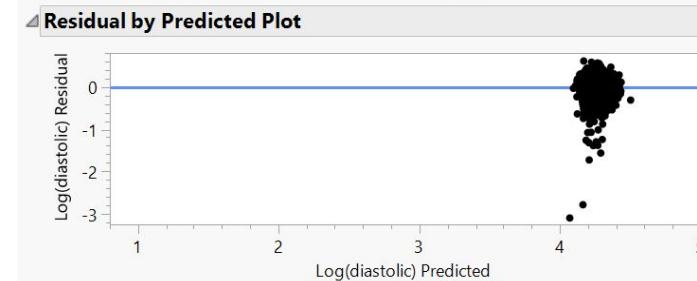
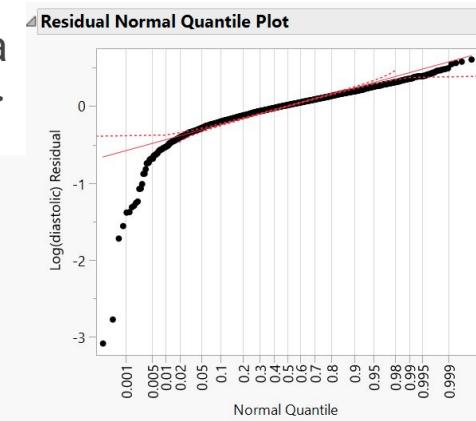
Assumptions Check: Diastolic



<- Non-transformed data

- We chose to use $\log(\text{avg_diastolic_bp})$ rather than avg_diastolic because when looking at the avg_diastolic Normal Quantile Plot, there was much more severe left skewness in the non-transformed version.
- Looking at the residual by predicted plot, our homogeneity of variance assumption is satisfied.

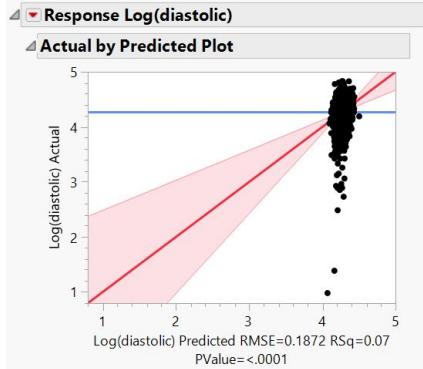
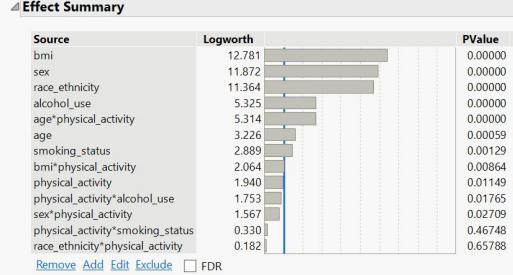
Log-Transformed data



Model (Diastolic)

$Y_i = \beta_0 + \beta_1 X_{\text{physical activity}} + \beta_2 X_{\text{age}} + \beta_3 X_{\text{sex}} + \beta_4 X_{\text{bmi}} + \beta_5 X_{\text{race}} + \beta_6 X_{\text{smoking}} + \beta_7 X_{\text{alcohol}} + PA * \text{age} + PA * \text{sex} + PA * \text{bmi} + PA * \text{Race} + PA * \text{smoking} + PA * \text{alco}$

- We used an ANCOVA model.
- Y_i is the model outcome: Log(avg_diastolic_bp)
- β_0 is the model intercept.



Summary of Fit

RSquare	0.06653
RSquare Adj	0.056373
Root Mean Square Error	0.1872
Mean of Response	4.26691
Observations (or Sum Wgts)	4739

Analysis of Variance

Source	DF	Sum of		F Ratio
		Squares	Mean Square	
Model	51	11.70644	0.229538	6.5500
Error	4687	164.25034	0.035044	Prob > F
C. Total	4738	175.95679		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	4.131869	0.01693	244.06	<.0001*
age	0.006242	0.00182	3.44	0.006*
sex[Female]	-0.02237	0.003147	-7.11	<.0001*
race_ethnicity[Mexican American]	-0.017773	0.00721	-2.30	0.0214*
race_ethnicity[Non-Hispanic Asian]	0.0417081	0.008065	5.17	<.0001*
race_ethnicity[Non-Hispanic Black]	0.0120717	0.006454	1.87	0.615
race_ethnicity[Non-Hispanic White]	-0.032172	0.005941	-5.42	<.0001*
race_ethnicity[Other Hispanic]	-0.004792	0.009161	-0.52	0.6009
bmi	0.0034602	0.000468	7.40	<.0001*
physical_activity[High]	-0.026424	0.01207	-2.19	0.0286*
physical_activity[Low]	0.0304837	0.01766	1.72	0.863
physical_activity[Moderate]	-0.020375	0.012154	-1.68	0.937
smoking_status[Every day]	-1.089e-5	0.009793	-0.00	0.9991
smoking_status[Never smoker]	-0.003277	0.005976	-0.55	0.5834
smoking_status[Not at all]	-0.027422	0.007215	-3.80	0.0001*
alcohol_use[No]	-0.025122	0.005483	-4.58	<.0001*
(age-49.7913)*physical_activity[High]	-0.000034	0.00029	-0.12	0.9067
(age-49.7913)*physical_activity[Low]	0.0016036	0.000407	3.94	<.0001*
(age-49.7913)*physical_activity[Moderate]	-0.000336	0.000286	-1.17	0.2402
sex[Female]*physical_activity[High]	-0.01511	0.005212	-2.90	0.0038*
sex[Female]*physical_activity[Low]	0.0135258	0.006647	2.03	0.0419*
sex[Female]*physical_activity[Moderate]	0.0027741	0.005211	0.53	0.5945
race_ethnicity[Mexican American]*physical_activity[High]	-0.018571	0.011985	-1.55	0.1213
race_ethnicity[Mexican American]*physical_activity[Low]	0.0046128	0.016455	0.28	0.7792
race_ethnicity[Mexican American]*physical_activity[Moderate]	-0.001413	0.013338	-0.11	0.9156
race_ethnicity[Non-Hispanic Asian]*physical_activity[High]	0.002869	0.015701	0.18	0.8550
race_ethnicity[Non-Hispanic Asian]*physical_activity[Low]	0.0084746	0.014346	0.59	0.5547
race_ethnicity[Non-Hispanic Asian]*physical_activity[Moderate]	0.0016639	0.01428	0.12	0.9072
race_ethnicity[Non-Hispanic Black]*physical_activity[High]	0.013493	0.010394	1.30	0.1943
race_ethnicity[Non-Hispanic Black]*physical_activity[Low]	-0.005515	0.013741	-0.40	0.6882
race_ethnicity[Non-Hispanic Black]*physical_activity[Moderate]	0.0046801	0.01105	0.42	0.6719
race_ethnicity[Non-Hispanic White]*physical_activity[High]	-0.000978	0.009172	-0.11	0.9151
race_ethnicity[Non-Hispanic White]*physical_activity[Low]	-0.003831	0.013422	-0.29	0.7753
race_ethnicity[Non-Hispanic White]*physical_activity[Moderate]	0.0052346	0.009433	0.55	0.5790
race_ethnicity[Other Hispanic]*physical_activity[High]	0.0321506	0.015078	2.13	0.0303*
race_ethnicity[Other Hispanic]*physical_activity[Low]	-0.01119	0.01932	-0.58	0.5605
race_ethnicity[Other Hispanic]*physical_activity[Moderate]	-0.012101	0.015817	-0.77	0.4443
(bmi-29.6194)*physical_activity[High]	0.0016547	0.000717	2.31	0.0210*
(bmi-29.6194)*physical_activity[Low]	0.0007353	0.001083	0.68	0.4974
(bmi-29.6194)*physical_activity[Moderate]	-0.0008081	0.0007024	-1.11	0.2684
physical_activity[High]*smoking_status[Every day]	0.003142	0.01278	0.25	0.8058
physical_activity[High]*smoking_status[Never smoker]	0.0144342	0.008998	1.60	0.1087
physical_activity[High]*smoking_status[Not at all]	-0.002124	0.010624	-0.20	0.8416
physical_activity[Low]*smoking_status[Every day]	-0.007519	0.025081	-0.30	0.7644
physical_activity[Low]*smoking_status[Never smoker]	-0.009465	0.013244	-0.71	0.4748
physical_activity[Low]*smoking_status[Not at all]	-0.001965	0.016592	-0.12	0.9057
physical_activity[Moderate]*smoking_status[Every day]	0.0105882	0.014122	1.07	0.2854
physical_activity[Moderate]*smoking_status[Never smoker]	0.004494	0.009887	0.45	0.6495
physical_activity[Moderate]*smoking_status[Not at all]	0.0165253	0.011597	1.43	0.1535
physical_activity[High]*alcohol_use[No]	-0.02701	0.009558	-2.83	0.0047*
physical_activity[Low]*alcohol_use[No]	0.0106617	0.011894	0.90	0.3701
physical_activity[Moderate]*alcohol_use[No]	0.0028603	0.008724	0.33	0.7430

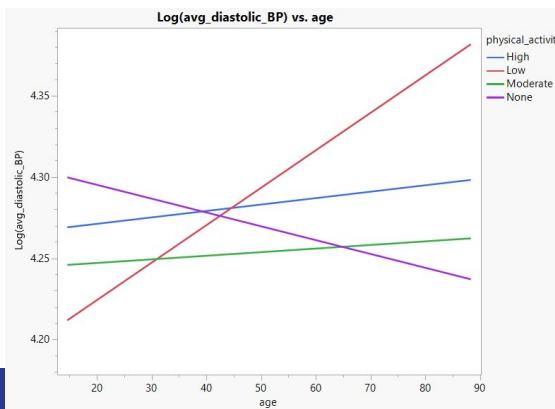
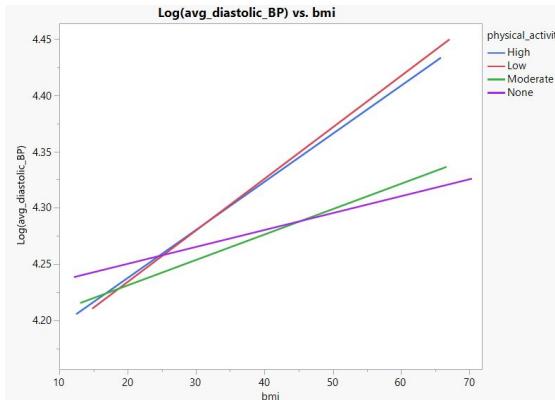
- **Significant main effects:**
 - Age, sex, race_ethnicity, bmi, physical_activity, smoking_status, and alcohol_use
- **Significant interactions:**
 - physical_activity*sex, physical_activity*age, physical_activity*bmi, and physical_activity*alcohol_use
 - provides evidence of unequal slopes
 - the effect of physical activity on blood pressure varies depending on age, sex, BMI, and alcohol use.

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
age	1	1	0.4138328	11.8090	0.0006*
sex	1	1	1.7712434	50.5437	<.0001*
race_ethnicity	5	5	2.1924161	12.5124	<.0001*
bmi	1	1	1.9168780	54.6995	<.0001*
physical_activity	3	3	0.3874544	3.6854	0.0115*
smoking_status	3	3	0.5519072	5.2497	0.0013*
alcohol_use	1	1	0.7356686	20.9928	<.0001*
age*physical_activity	3	3	0.9629666	9.1596	<.0001*
sex*physical_activity	3	3	0.3216950	3.0599	0.0271*
race_ethnicity*physical_activity	15	15	0.4302333	0.8185	0.6579
bmi*physical_activity	3	3	0.4091382	3.8917	0.0086*
physical_activity*smoking_status	9	9	0.3041779	0.9644	0.4675
physical_activity*alcohol_use	3	3	0.3546468	3.3734	0.0177*

Diastolic Interpretation

- The reference groups in our models are ~49.84 year old "other" race males who have a BMI of 29.62, do no physical activity, smoke some days, and do use alcohol

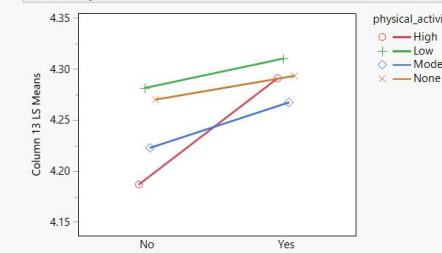


Alcohol Use

Least Squares Means Table

Level	Least Sq Mean	Std Error	1	2	3	4	5
High,No	4.18688	0.0225				●	
High,Yes	4.29114	0.0091				●	
Low,No	4.28146	0.0343			●	●	
Low,Yes	4.31038	0.0172			●	●	
Moderate,No	4.22280	0.0207			●	●	
Moderate,Yes	4.26732	0.0104			●	●	
None,No	4.27012	0.0143			●	●	
None,Yes	4.29339	0.0081			●	●	

Least Squares Means Plot

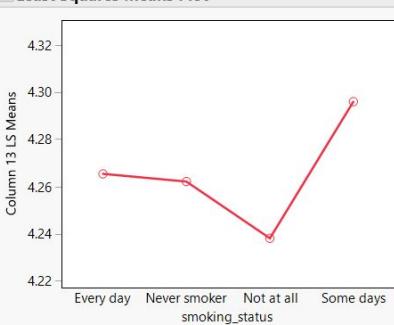


Smoking Status

Least Squares Means Table

Level	Least Sq Mean	Std Error	1	2	3	4	5	Mean
Every day	4.26543	0.0134				●		4.2772
Never smoker	4.26216	0.0063			●	●		4.2653
Not at all	4.23801	0.0090			●	●		4.2567
Some days	4.29615	0.0161			●	●		4.3163

Least Squares Means Plot

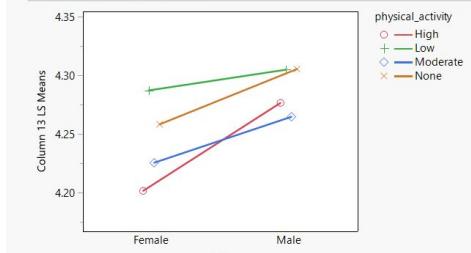


Sex

Least Squares Means Table

Level	Least Sq Mean	Std Error	1	2	3	4	5
Female,High	4.20153	0.0148				●	
Female,Low	4.28708	0.0254				●	
Female,Moderate	4.22547	0.0146				●	
Female,Non	4.25819	0.0103				●	
Male,High	4.27649	0.0139				●	
Male,Low	4.30476	0.0227				●	
Male,Moderate	4.26466	0.0145				●	
Male,Non	4.30531	0.0109				●	

Least Squares Means Plot

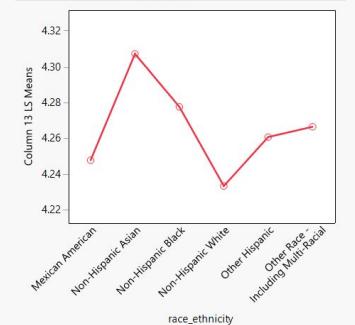


Race_Ethnicity

Least Squares Means Table

Level	Least Sq Mean	Std Error	1	2	3	4	5	Mean
Mexican American	4.24766	0.0108				●		4.2592
Non-Hispanic Asian	4.30715	0.0108				●		4.2932
Non-Hispanic Black	4.27751	0.0091				●		4.2927
Non-Hispanic White	4.23326	0.0088				●		4.2424
Other Hispanic	4.26064	0.0123				●		4.2641
Other Race - Including Multi-Racial	4.26639	0.0155				●		4.2769

Least Squares Means Plot



Results

- There is evidence that physical activity affects blood pressure in a non-uniform way, i.e., we have evidence of unequal slopes. This is concluded as we have significant interactions.
- However, which factors the slopes differ on is unique to systolic and diastolic.
- Systolic varies on age and sex.
- Diastolic varies on age, sex, BMI, and alcohol use.

Practical Conclusion

- How much a person's physical activity affects their blood pressure depends on several different characteristics and also affects diastolic and systolic differently.
- Many people assume that higher physical activity will bring their blood pressure closer to 120/80. However, in reality, how much physical activity helps depends on other factors.
- Thus, those who are looking to control their blood pressure, must take into account which type they need help controlling (diastolic or systolic) as well as their age, sex, BMI, and alcohol use when considering increasing physical activity as a solution

References:

- Dataset:
<https://www.cdc.gov/nchs/nhanes/continuousnhanes/default.aspx?BeginYear=2017>
- Information about blood pressure:
<https://www.nhlbi.nih.gov/health/high-blood-pressure#:~:text=The%20first%20number%20is%20your,day%20based%20on%20your%20activities.>