“What is the relationship of being a parent and people’s happiness?”

—

Course: Statistical Reasoning for Behavioral Science
Professor: Dr Alexey Shpenev
Student name: Loizos Konstantinou
Penn ID: 54838870

Research Hypothesis:

If a person became a parent, it is more likely to be/become happier

Independent variable X: **agekdbrn** = Age of person when his/her first child was born?

- this is converted to a binary variable between ppl that answered and people who did not. It is used as a proxy on whether people are/were parents or not.

Dependent Variable Y: **happy** = Taken all together, how would you say things are these days–would you say that you are very happy, pretty happy, or not too happy?

Why does having kids correlation with happiness matters?

There are multiple people that find that the meaning of life is to be happy (Shafi, Z., 2020[1]), hence they go after happiness. At the same time, having children is a popular decision for most of the people. Therefore, a reasonable assumption can be that having kids makes people happy. On the other side of the spectrum though, there is evidence that shows that children actually might decrease happiness among people in the long run (Keshner, A., 2019[2]). Consequently, it would be interesting to examine the relationship between the two.

1. https://timesofindia.indiatimes.com/readersblog/lifekibaat/the-purpose-of-life-is-to-be-happy-23838/ (Accessed 16th December 2021)
2. https://www.marketwatch.com/story/one-theory-on-why-having-kids-makes-people-unhappy-2019-02-26 (Accessed 16th December 2021)

How is the data characterized?

Year

GSS Data from 2021[1]

Population → 259 Million[2]. **Sample size** → 4,032.

The sample is generalized to the US adult population.

https://gss.norc.org/ (Accessed 7th December 2021)

2. https://www.census.gov/library/stories/2021/08/united-states-adult-population-grew-faster-than-nations-total-population-from-2010-to-2020.html (Accessed 16th December 2021)

GSS variables used

Variable: HAPPY **Type: Numeric**

Label: Taken all together, how would you say things are these days--would you say that you are very happy, pretty happy, or not too happy?

LABEL	VALUE	COUNT	PCT	Codes
VERY HAPPY	1	783	19.4%	19.5%
PRETTY HAPPY	2	2308	57.2%	57.5%
NOT TOO HAPPY	3	923	22.9%	23.0%
SUBTOTALS:		4014	99.6%	100.0%

Variable: AGEKDBRN **Type: Numeric**

Label: How old were you when your first child was born?

Notes: AGEKDBRN is coded in two digits.
The collapsed numbers are for convenience of display only.

SUBTOTALS:		2803	69.5%	100.0%
RESERVED CODES:				
NOT APPLICABLE	I	1212	30.1%	n/a
SKIPPED ON WEB	S	17	0.4%	n/a
TOTALS:		4032	100.0%	100.0%

H0: Having kids has no effect on happiness

H1: Having kids has an effect on happiness

Penn
University of Pennsylvania

Findings

```
summary(model)
```

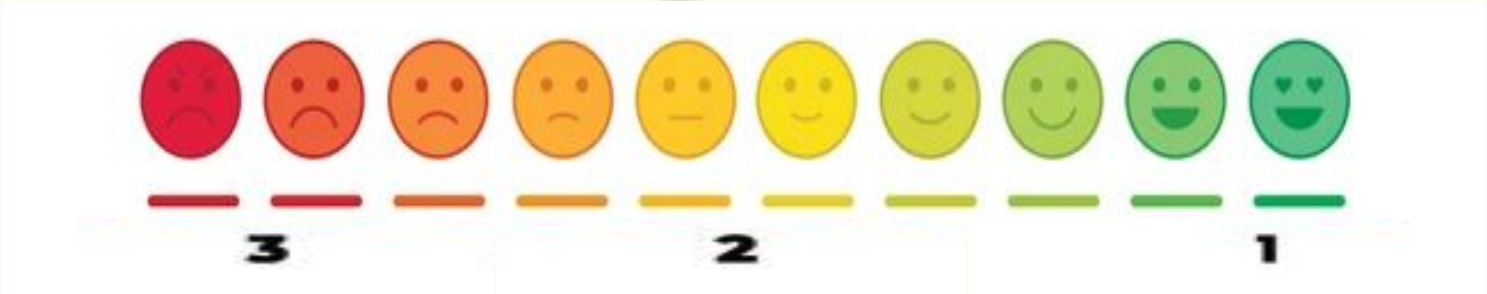
```
Call:
lm(formula = happy ~ kid, data = gssr)

Residuals:
     Min       1Q   Median       3Q      Max
-1.13432 -0.13432  0.00859  0.00859  1.00859

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.13432    0.01854  115.12  < 2e-16 ***
kid         -0.14291    0.02223   -6.43 1.43e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6478 on 4012 degrees of freedom
  (18 observations deleted due to missingness)
Multiple R-squared:  0.0102,    Adjusted R-squared:  0.009953
F-statistic: 41.34 on 1 and 4012 DF,  p-value: 1.427e-10
```

$$\text{Happy} = 2.13 - 0.14*\text{kid}$$

- Intercept means that if people have no kids, they will be 2.13 happy on a scale 1(very-happy) to 3(not-happy).
- p-value is the probability that the null hypothesis is true. p-value of the intercept is significant aka close to zero —> we reject the null hypothesis. Our p-value states that kids have an effect on happiness
- Kid contributes to overall happiness
    - (Happy = 2.13 - 0.14*1 = 1.99)
- Looking at Min & Max, as well as 1Q & 3Q, we can tell that our data are symmetrical as they have approximately the same distance from zero.
- Multiple R squared = 0.01. → almost 1% of the variability in the data can be explained by the variability in kids people had/didn’t have. This indicates a week relationship.
- The residuals standard error states how much the model misses - e.g. if someone is 1, it could have been 1.64.

Penn
University of Pennsylvania

2.13



1.99

Because of confounding variables, we use the financial satisfaction variable (satfin) to see how it helps model increase its accuracy.

Variable: SATFIN **Type: Numeric**

Label:
We are interested in how people are getting along financially these days. So far as you and your family are concerned, would you say that you are pretty well satisfied with your present financial situation, more or less satisfied, or not satisfied at all?

LABEL	VALUE	COUNT	PCT	Codes
PRETTY WELL SATISFIED	1	1254	31.1%	31.2%
MORE OR LESS SATISFIED	2	1800	44.6%	44.8%
NOT SATISFIED AT ALL	3	962	23.9%	24.0%
SUBTOTALS:		4016	99.6%	100.0%

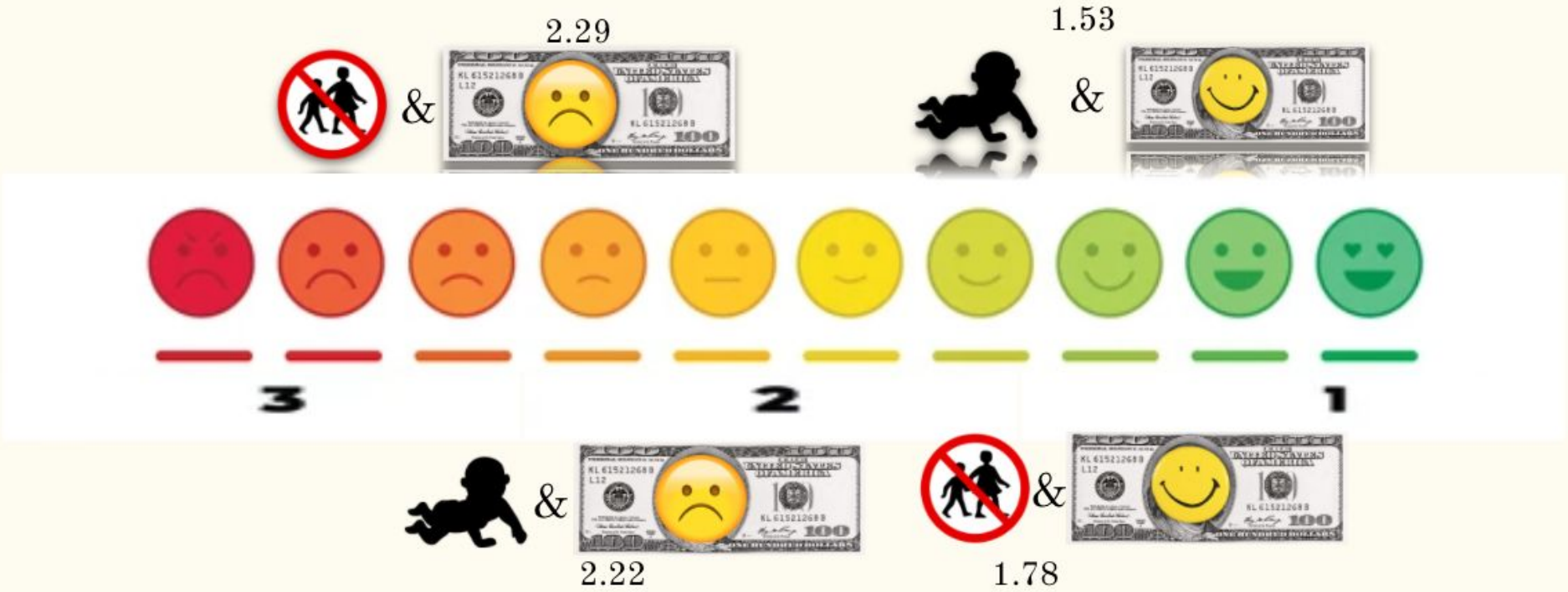
Findings

```
Call:
lm(formula = happy ~ kid + satfin + kid * satfin, data = gssr)

Residuals:
     Min       1Q   Median       3Q      Max
-1.31826 -0.25054 -0.01219  0.22616  1.22616

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.78848    0.05046  35.441  < 2e-16 ***
kid         -0.25299    0.06035  -4.192 2.83e-05 ***
satfin       0.17659    0.02408   7.335 2.67e-13 ***
kid:satfin   0.06176    0.02899   2.130   0.0332 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6269 on 3998 degrees of freedom
  (30 observations deleted due to missingness)
Multiple R-squared:  0.07318,   Adjusted R-squared:  0.07248
F-statistic: 105.2 on 3 and 3998 DF,  p-value: < 2.2e-16
```

$$\text{Happy} = 1.78 - 0.25{*}\text{kid} + 0.17{*}\text{satfin} + 0.06{*}\text{kid}{*}\text{satfin}$$

- If everything is zero = a person without kids which is very satisfied with finances.
- The multiple R squared increases from 1% to 7,3%. This means that the financial satisfaction (=satisfaction with financial situation, contributes to happiness). The higher the satisfaction, the happier people report to be.
- Moreover, having (a) kid/s becomes less desirable as its correlation to happiness goes to -0.25.
- The intersection between having kid and being satisfied with finances is contributing extra 0.06 to happiness.

Penn
University of Pennsylvania



- If everything is zero = a person without kids which is very satisfied with finances = 1.78
- Has kid and satisfied with money is (1.78-0.25*1) = 1.53
- No kid and not happy on satisfaction = 1.78+0.17*3 = 2.29
- Has kid and not happy on finances = 1.78-0.25+0.17*3+0.06*1*3 = 2.22

```
coeftest(model)
```

```
t test of coefficients:

             Estimate Std. Error  t value  Pr(>|t|)
(Intercept)  2.134316   0.018540 115.1214 < 2.2e-16 ***
kid         -0.142909   0.022226  -6.4299 1.427e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
coeftest(model, vcov = vcovHC, type = "HC1")
```

```
t test of coefficients:

             Estimate Std. Error  t value  Pr(>|t|)
(Intercept)  2.134316   0.018244 116.9870 < 2.2e-16 ***
kid         -0.142909   0.022027  -6.4879 9.762e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
coeftest(model, vcov = vcovHC, type = "HC5")
```

```
t test of coefficients:

             Estimate Std. Error  t value  Pr(>|t|)
(Intercept)  2.134316   0.018246 116.9767 < 2.2e-16 ***
kid         -0.142909   0.022027  -6.4879 9.762e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Coef test has a slight change effects on standard error but the change is tiny.

<u>Interpretation</u>:

→ Even if the assumption is technically violated the model is still valid.

The happiness variable, even if it is not continuous or numerical, it works very well as a numeric variable.

Coef test: Robust standard errors are specified
(if we had a small sample size we would use non-parametric method).

Conclusions from the model

- Having kid/s improves happiness for 0.14 points.

- Having kid/s improves happiness even more if money was not a problem. i.e.:
  - if people are satisfied financially, kids improve happiness even more.
  - if people were not satisfied financially, kids would have made people less happy than before.

Limitations

1. **Happiness scale 1,2,3 is not really a continuous variable.**
   Because of that the residuals do not vary in a normal distribution. Therefore, the t-tests are technically invalid.
   Model has irregularly distributed residuals.

To confront this limitation, a coef test ir ran to examine model significance sensitivity in accounting for residuals.
→ almost no change in standard errors i.e. happiness works well as a numeric variable.

2. **Confounders**
   There can be multiple others confounders that affect happiness other than those two used (satfin & kid)
   -- e.g. physical health, work satisfaction, quality of life, mental health, etc.

3. **GSS is not the best dataset**
   Data are not representative/reflective of the population. 4 thousand people is disproportionately less amount of the actual population of 259 million. Acknowledging that factors, the sample from the dataset gives us an indicative estimation that we could infer outcomes for smaller population targets.

Summary

1. **Research hypothesis**: that being a parent increases people happiness.
2. **Sample**: 4,032 US-based people.
3. **Methods used**: Regression analysis to determine the relationship between the variables and to examine their significance and their magnitude.
4. **Run regression**: happy = 2.13 - 0.14*kid
   a. summary → weak correlation (R^2= 1%); positive relationship between kids and happiness.
5. **Run regression with confounder**: Happy = 1.78 - 0.25*kid+0.17*satfin+0.06*kid*satfin
   a. summary → weak correlation (R^2= 7%); financial satisfaction adds more deviation among kids/happiness.
6. **Conclusions**
   a. Being a parent has positive association with happiness (weak correlation).
   b. Financial satisfaction combined with having kid/s contributes more to happiness.
7. **Limitations**
   a. Dependent variable not technically continuous
   b. Confounders
   c. Disproportionately small sample