Derivations

This file presents the derivatioins and graph construction to accompany
McClelland, G.H., Lynch, J.G, Jr., Irwin, J.R., Spiller, S.A., & Fitzsimons, G.J. (under review). Median splits, Type II errors, and false positive consumer psychology: Don't fight the power. *Journal of Consumer Psychology.*

An interactive *Mathematica* notebook is available from the first author at gary.mcclelland@colorado.edu who prepared this file. 1 March 2015

Standardized Partial Regression Coefficient, Split and Continuous

Definitions of standardized partial regression coefficients from Cohen, Cohen, Aiken, & West:

```
beta1[ry1_, ry2_, r12_] := (ry1 - ry2 r12)/(1 - r12^2)
```

```
beta2[ry1_, ry2_, r12_] := (ry2 - ry1 r12)/(1 - r12^2)
```

Splitting one continuous variable at its median reduces the correlation between two variables by a factor of *a*, with *a* depending on the distribution. For the normal distribution $a = \sqrt{\frac{2}{\pi}}$.

Splitting one predictor (i.e., independent variable) will reduce its correlation with the criterion (i.e., dependent variable)and with the other predictor by *a*. So the estimated coefficients when the first predictor is split become:

```
beta1[a ry1, ry2, a r12]
```

$$\frac{a\ ry1 - a\ r12\ ry2}{1 - a^2\ r12^2}$$

```
beta2[a ry1, ry2, a r12]
```

$$\frac{-a^2\ r12\ ry1 + ry2}{1 - a^2\ r12^2}$$

The ratio of the split estimate to the continuous estimate for the first predictor is therefore:

```
beta1[a ry1, ry2, a r12] / beta1[ry1, ry2, r12]
```

$$\frac{\left(1 - r12^2\right)\ (a\ ry1 - a\ r12\ ry2)}{\left(1 - a^2\ r12^2\right)\ (ry1 - r12\ ry2)}$$

```
Simplify[%]
```

$$\frac{a\ \left(-1 + r12^2\right)}{-1 + a^2\ r12^2}$$

Note that this ratio does not depend on the correlations between the predictors and the criterion but only on the intercorrelation between the two predictors. Using the factor for the normal distribution, the ratio

becomes

$$\frac{a\left(-1+r12^2\right)}{-1+a^2\,r12^2}\ /.\ a\to\sqrt{\frac{2}{\pi}}$$

$$\frac{\sqrt{\frac{2}{\pi}}\ \left(-1+r12^2\right)}{-1+\frac{2\,r12^2}{\pi}}$$

```
Simplify[%]
```

$$-\frac{\sqrt{2\,\pi}\ \left(-1+r12^2\right)}{\pi-2\,r12^2}$$

Increment in R^2

The definition for R^2 is provided by Cohen, Cohen, Aiken, & West.

```
rsq[ry1_, ry2_, r12_] := beta1[ry1, ry2, r12] ry1 + beta2[ry1, ry2, r12] ry2
```

```
rsq[ry1, ry2, r12]
```

$$\frac{ry2\ (-r12\ ry1+ry2)}{1-r12^2}+\frac{ry1\ (ry1-r12\ ry2)}{1-r12^2}$$

```
Simplify[%]
```

$$\frac{ry1^2-2\,r12\,ry1\,ry2+ry2^2}{1-r12^2}$$

The increment in R^2due to the first predictor equals the overall R^2 minus the squared correlation between the other predictor and the criterion. That is

```
sr1[ry1_, ry2_, r12_] := rsq[ry1, ry2, r12] - ry2^2
```

```
sr1[ry1, ry2, r12]
```

$$-ry2^2+\frac{ry2\ (-r12\ ry1+ry2)}{1-r12^2}+\frac{ry1\ (ry1-r12\ ry2)}{1-r12^2}$$

```
Simplify[%]
```

$$-\frac{(ry1-r12\ ry2)^2}{-1+r12^2}$$

The ratio of the increment in R^2 is computed as

```
sr1[a ry1, ry2, a r12] / sr1[ry1, ry2, r12]
```

$$\frac{-ry2^2+\frac{ry2\ \left(-a^2\,r12\,ry1+ry2\right)}{1-a^2\,r12^2}+\frac{a\,ry1\ (a\,ry1-a\,r12\,ry2)}{1-a^2\,r12^2}}{-ry2^2+\frac{ry2\ (-r12\ ry1+ry2)}{1-r12^2}+\frac{ry1\ (ry1-r12\ ry2)}{1-r12^2}}$$

```
Simplify[%]
```

$$\frac{a^2\left(-1+r12^2\right)}{-1+a^2\,r12^2}$$

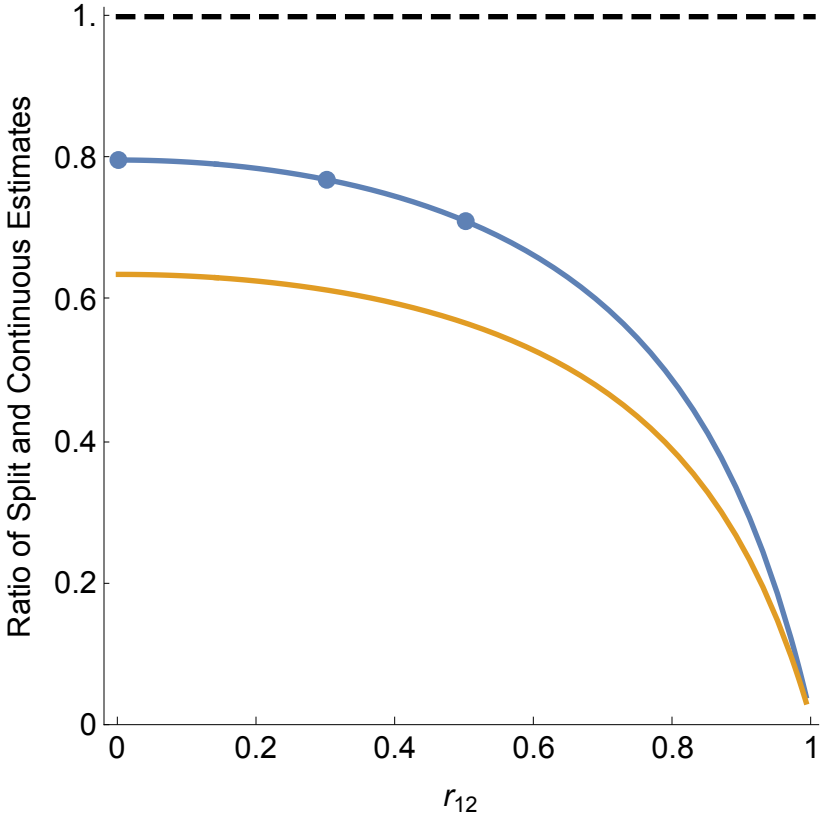
Again, this ratio does not depend on the predictor correlations with the criterion or their parameter estimates but only on the predictor intercorrelation r_{12}. Using the factor for the normal distribution yields

```
Simplify[ a^2 (-1 + r12^2) / (-1 + a^2 r12^2) /. a → Sqrt[2/π] ]
```

$$\frac{2-2\,r12^2}{\pi-2\,r12^2}$$

Plot of the Two Ratios

```
Show[Plot[{-(Sqrt[2 π] (-1 + r12^2))/(π - 2 r12^2), (2 - 2 r12^2)/(π - 2 r12^2)}, {r12, 0, .99}, PlotRange → {0, 1.01},
  Frame → {True, True, False, False}, AspectRatio → 1, PlotStyle → Thickness[.0075],
  FrameLabel → {"r12", "Ratio of Split and Continuous Estimates"},
  BaseStyle → {FontSize → 16},
  FrameTicks → {{0, .2, .4, .6, .8, 1}, {0, .2, .4, .6, .8, 1.}}],
 Graphics[{Dashing[{.02, .02}], Thickness[.0075], Line[{{0, 1}, {1, 1}}]}],
 ListPlot[{{0, .797885}, {.3, 0.77}, {.5, 0.71168}}, PlotStyle → PointSize[.025]]
]
```



Standardized Partial Regression Coefficient for the Unsplit Variable

As noted above, when neither predictor variable is split, the formula for the standardized partial regression coefficient for the second variable is:

```
beta2[ry1_, ry2_, r12_] := (ry2 - ry1 r12)/(1 - r12^2)
```

And the coefficient for the second predictor variable when the first predictor variable is split is given by

```
beta2[a ry1, ry2, a r12]
```

$$\frac{-a^2\, r12\, ry1 + ry2}{1 - a^2\, r12^2}$$

Consider the ratio of the two estimates when there is no correlation between the predictors

```
beta2[a ry1, ry2, a r12] / beta2[ry1, ry2, r12] /. r12 → 0
```

```
1
```

Hence, when there is no correlation between the two predictors, splitting the other predictor variable has no expected effect on the estimate of the standardized partial regression coefficient for the second predictor. The situation is more complex when the predictors are intercorrelated. Depending on the relationship among the three correlations, the estimate of the partial regression coefficient for the second predictor might increase, decrease, or remain the same. That is,

```
Reduce[beta2[a ry1, ry2, a r12] > beta2[ry1, ry2, r12] &&
    0 < r12 < 1 && 0 < ry1 < 1 && 0 < ry2 ≤ 1 , {ry1, ry2, r12}] /. a → Sqrt[2/π]
```

$$0 < ry1 < 1 \;\&\&\; \left((0 < ry2 \le ry1 \;\&\&\; 0 < r12 < 1) \;||\; \left(ry1 < ry2 \le 1 \;\&\&\; 0 < r12 < \frac{ry1}{ry2} \right) \right)$$

The restrictive condition predicting whether splitting the other predictor will increase or decrease the estimate of the parameter for the unsplit predictor is, respectively, whether $r_{12}$is less than or greater than $\frac{r_{Y1}}{r_{Y2}}$. From re-running the SAS code provided by IPKSP and disaggregating the data according to the above rule, the means from the simulations are:

Increasing Conditions

When $r_{12} < \frac{r_{Y1}}{r_{Y2}}$ there is an increase in the mean estimate (2nd column) of the standardized partial regression coefficient for the unsplit variable when the other variable is split, indexed by the predictor intercorrelation (1st column)

```
inc = {{0.0, 0.3193829 - 0.3196263},
    {0.1, 0.2959352 - 0.2823697}, {0.3, 0.2179823 - 0.1748656},
    {0.5, 0.1375720 - 0.0423314}, {0.7, 0.0496708 - (-0.1674093)}};
```

```
TableForm[inc]
```

```
0.     -0.0002434
0.1    0.0135655
0.3    0.0431167
0.5    0.0952406
0.7    0.21708
```

Decreasing Conditions

When $r_{12} > \frac{r_{Y1}}{r_{Y2}}$ there is a decrease in the mean estimate (2nd column) of the standardized partial regression coefficient for the unsplit variable when the other variable is split, indexed by the predictor intercorrelation (1st column).

```
dec = {{0.0, 0.3193829 - 0.3196263},
    {0.1, 0.4059998 - 0.4073575}, {0.3, 0.4907363 - 0.5048938},
    {0.5, 0.5353843 - 0.5807554}, {0.7, 0.6257530 - 0.7958947}};

TableForm[dec]
```

```
0.     -0.0002434
0.1    -0.0013577
0.3    -0.0141575
0.5    -0.0453711
0.7    -0.170142
```

Aggregate Means

When averaging across all conditions without considering whether there is an increase or decrease for the conditions, the following means (produced by the authors' SAS code) are:

```
mean = {{0, 0}, {.1, 0.3013571 - 0.2912314}, {.3, 0.2742505 - 0.2454882},
    {.5, 0.2580021 - 0.2114925}, {.7, 0.2573454 - 0.1878173}};
TableForm[
 mean]
```

```
0      0
0.1    0.0101257
0.3    0.0287623
0.5    0.0465096
0.7    0.0695281
```

Note that the aggregate means are not the averages of the increase and decrease means because there were not an equal number of cases in each set.

No Split Means

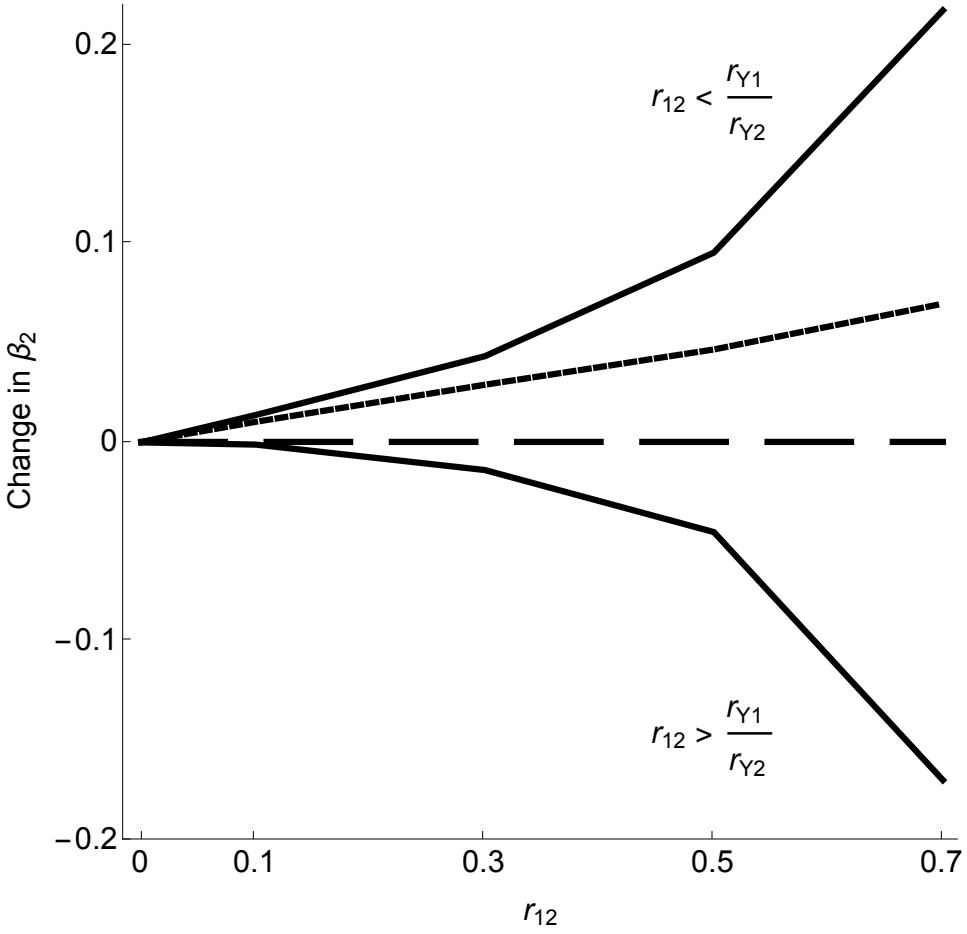
For graphing purposes we include what the mean "change" would have been had there been no split of the other predictor variable

```
TableForm[cont = {{0, 0}, {.1, 0}, {.3, 0}, {.5, 0}, {.7, 0}}]
```

```
0      0
0.1    0
0.3    0
0.5    0
0.7    0
```

Figure 2 Disaggregated

```
Show[ListPlot[{inc, mean, cont, dec},
  Joined → True, Frame → {True, True, False, False},
  BaseStyle → {FontSize → 16}, PlotRange → {-.2, .22},
  PlotStyle → {{Black, Thickness[.0075]}, {Black, Thickness[.0075], Dashed},
    {Black, Thickness[.0075], Dashing[{.1, .05}]}}, AspectRatio → 1,
  Axes → False, FrameTicks → {{0, .1, .3, .5, .7}, {-.2, -.1, 0, .1, .2}},
  FrameLabel → {"r12", "Change in β2"}],
 Graphics[{Text["r12 < r_Y1/r_Y2", {.5, .17}], Text["r12 > r_Y1/r_Y2", {.5, -.15}]}]
]
```



Special Case: $r_{Y1} = r_{Y2}$

When the two predictors have the same correlation with the criterion, it is possible to compute the ratio of the estimate when the other predictor is split to the estimate when the other predictor is not split.

```
beta2[a ry1, ry1, a r12] / beta2[ry1, ry1, r12] /. a → Sqrt[2/π]
```

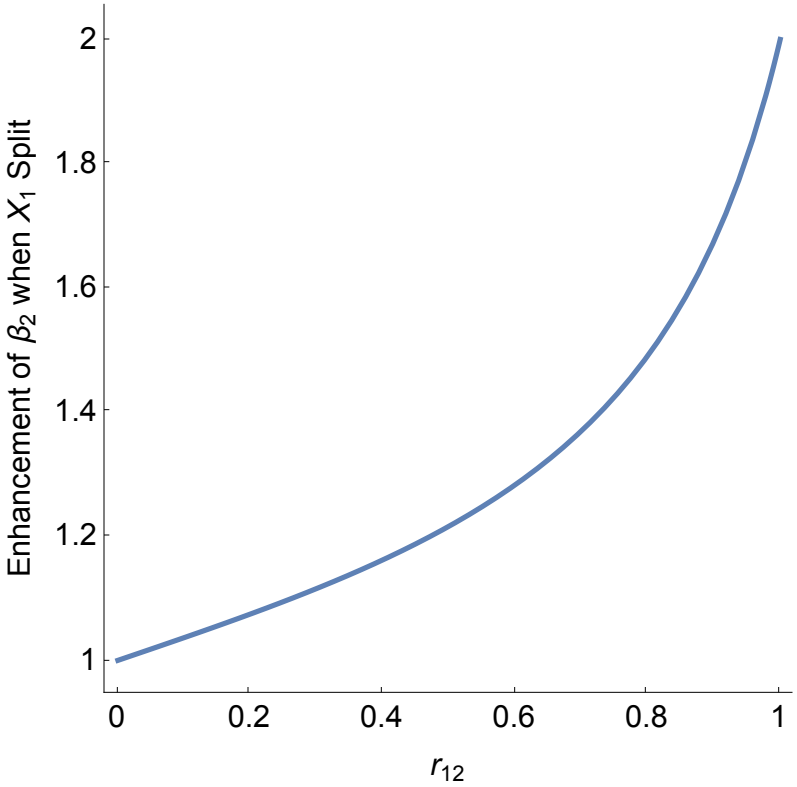
$$\frac{\left(1-\text{r12}^2\right)\left(\text{ry1}-\frac{2\,\text{r12}\,\text{ry1}}{\pi}\right)}{\left(1-\frac{2\,\text{r12}^2}{\pi}\right)(\text{ry1}-\text{r12}\,\text{ry1})}$$

```
Simplify[(1 - r12^2) (ry1 - 2 r12 ry1/π) / ((1 - 2 r12^2/π) (ry1 - r12 ry1))]
```

$$\frac{(1+\text{r12})(-\pi+2\,\text{r12})}{-\pi+2\,\text{r12}^2}$$

This ratio depends only on the predictor intercorrelation. It is always greater than equal to 1 so when the predictor correlations with the criterion are approximately equal, the estimate of the coefficient for the second predictor when the first predictor is split is *always* enhanced relative to its value in the continuous analysis. The following graph shows the increasing magnitude of the enhancement.

```
Plot[(1 + r12) (-π + 2 r12)/(-π + 2 r12^2), {r12, 0, 1}, PlotStyle → Thickness[.0075],
 Frame → {True, True, False, False}, BaseStyle → {FontSize → 16}, AspectRatio → 1,
 FrameTicks → {{0, .2, .4, .6, .8, 1}, {1, 1.2, 1.4, 1.6, 1.8, 2}},
 FrameLabel → {"r12", "Enhancement of β2 when X1 Split"}]
```



Ratio of beta2/beta1 when X1 is split and when the correlations between each independent variable and

the dependent variable are equal is defined by

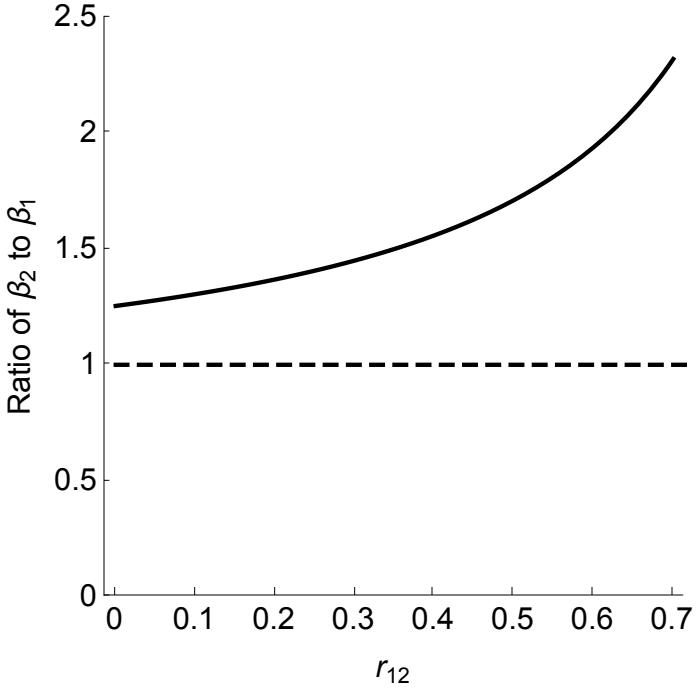
```
beta2[a ry1, ry1, a r12] / beta1[a ry1, ry1, a r12] /. a → Sqrt[2/π]
```

$$\frac{\text{ry1} - \frac{2\ \text{r12}\ \text{ry1}}{\pi}}{\sqrt{\frac{2}{\pi}}\ \text{ry1} - \sqrt{\frac{2}{\pi}}\ \text{r12}\ \text{ry1}}$$

```
Simplify[%]
```

$$-\frac{\pi - 2\ \text{r12}}{\sqrt{2\,\pi}\ (-1 + \text{r12})}$$

```
Show[Plot[-(π - 2 r12)/(Sqrt[2 π] (-1 + r12)), {r12, 0, .7}, PlotRange → {0, 2.5},
  BaseStyle → {FontSize → 16}, Frame → {True, True, False, False},
  FrameTicks → {{0, .1, .2, .3, .4, .5, .6, .7}, {0, .5, 1, 1.5, 2, 2.5}},
  FrameLabel → {"r12", "Ratio of β2 to β1"},
  AspectRatio → 1, PlotStyle → {Black, Thickness[.0075]}],
 Graphics[{Black, Thickness[.0075], Dashing[{0.02, 0.02}], Line[{{0, 1}, {1, 1}}]}]
]
```



The specific values used in the paper

$$-\frac{\pi - 2\,r12}{\sqrt{2\,\pi}\,(-1 + r12)} /. r12 \to 0.$$

```
1.25331
```

$$-\frac{\pi - 2\,r12}{\sqrt{2\,\pi}\,(-1 + r12)} /. r12 \to .3$$

```
1.4485
```

$$-\frac{\pi - 2\,r12}{\sqrt{2\,\pi}\,(-1 + r12)} /. r12 \to .5$$

```
1.70874
```

$$-\frac{\pi - 2\,r12}{\sqrt{2\,\pi}\,(-1 + r12)} /. r12 \to .7$$

```
2.31598
```

Special Case: $r_{Y1} \neq r_{Y2}$: Split Estimate is Weighted Average of the Unsplit Estimates

When the predictor correlations with the criterion are unequal, the situation is more complicated.

The following derivation shows that the estimate for the unsplit variable when the other variable is split becomes the weighted average between the orignal estimates. In other words, splitting the first predictor variable confounds the estimate of the second variable with that of the first variable.

```
Solve[beta2[a ry1, ry2, a r12] ==
    w beta1[ry1, ry2, r12] + (1 - w) beta2[ry1, ry2, r12], w] /. a → Sqrt[2/π]
```

$$\left\{\left\{w \to \frac{\frac{-r12\,ry1+ry2}{1-r12^2} - \frac{-\frac{2\,r12\,ry1}{\pi}+ry2}{1-\frac{2\,r12^2}{\pi}}}{\frac{-r12\,ry1+ry2}{1-r12^2} - \frac{ry1-r12\,ry2}{1-r12^2}}\right\}\right\}$$

```
Simplify[%]
```

$$\left\{\left\{w \to \frac{(-2+\pi)\,r12\,(-ry1+r12\,ry2)}{(1+r12)\left(-\pi+2\,r12^2\right)(ry1-ry2)}\right\}\right\}$$

```
w[ry1_, ry2_, r12_] := Simplify[(-2 + π) r12 (-ry1 + r12 ry2)/((1 + r12) (-π + 2 r12^2) (ry1 - ry2))]
```

Example for paper

```
w[.5, .3, .3]
```

```
0.182355
```

```
beta1[.5, .3, .3]
```

```
0.450549
```

```
beta2[.5, .3, .3]
```

```
0.164835
```

```
beta2[a .5, .3, a .3] /. a → Sqrt[2/π]
```

```
0.216937
```

```
w[.5, .3, .3] beta1[.5, .3, .3] + (1 - w[.5, .3, .3]) beta2[.5, .3, .3]
```

```
0.216937
```

The rounded version used in the paper:

```
.18 (.45) + (1 - .18) .165
```

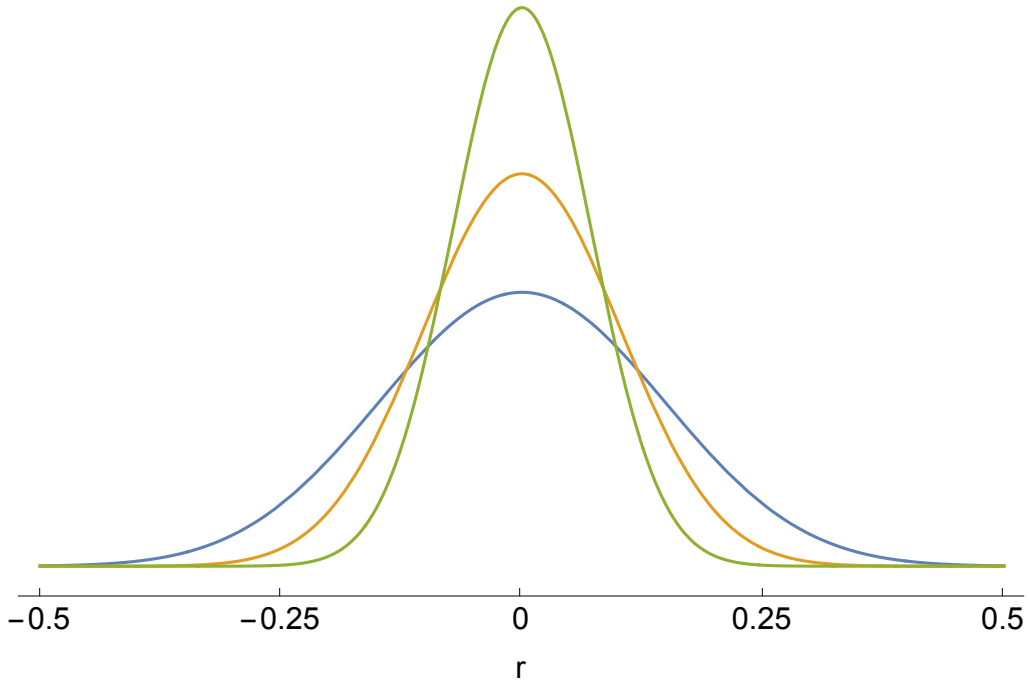
```
0.2163
```

Distribution of the Correlation Coefficient

The probability density function of the correlation coefficient under the null hypothesis that r = 0 is given by

$$\mathbf{f[r\_, n\_] := \frac{\left(1 - r^2\right)^{\frac{n-4}{2}}}{Beta\left[\frac{1}{2}, \frac{n-2}{2}\right]}}$$

```
Plot[{f[r, 50], f[r, 100], f[r, 200]}, {r, -.5, .5},
 Frame → {True, False, False, False}, Axes → False, BaseStyle → {FontSize → 16},
 FrameTicks → {{-.5, -.25, 0, .25, .5}, None}, FrameLabel → {"r"}]
```



Simulations

In the section of our paper “Median splits and false positive consumer psychology” we report this small simulation.

We first define a function for drawing a sample of size *n* from a bivariate normal distribution with correlation *r.*

```
getData[r_, n_] :=
  RandomVariate[MultinormalDistribution[{0, 0}, {{1, r}, {r, 1}}], n];
```

And specify a number of simulations. Note that even with 10,000 simulations, the numbers reported below may not be reproduced exactly with new runs of these simulations.

```
nSims = 10 000;
```

This function, for nSims times, computes the squared correlations between the two sampled variables and then the two variables after the first is split at its median. (Note: we split at the actual median, whereas IPSKP split at the expected median of 0, rather than the actual median.)

```
rSq[r_, n_] := Table[{Correlation[
      x = First[Transpose[data = getData[r, n]]], y = Last[Transpose[data]]]^2,
    Correlation[xs = Table[If[x[[i]] < Median[x], 0, 1], {i, 1, Length[x]}], y]^2}, {j,
    1, nSims}]
```

Use the above function to sample nSims squared correlations when the population correlation is 0 and the sample size is 50.

```
rsq = rSq[0, 50];
```

The proportion of Type I errors for the split analysis:

```
sigSplit =
 Apply[Plus, Table[If[rsq[[i, 2]] ≥ 0.2786^2, 1, 0], {i, 1, Length[rsq]}]] / nSims // N
```

```
0.0517
```

The proportion of Type I errors for the continuous analysis:

```
sigCont =
 Apply[Plus, Table[If[rsq[[i, 1]] ≥ 0.2786^2, 1, 0], {i, 1, Length[rsq]}]] / nSims // N
```

```
0.0494
```

The proportion of Type I errors when researcher can pick from either analysis.

```
sigEither =
 Apply[Plus, Table[If[rsq[[i, 1]] ≥ (rcrit = 0.2786)^2 || rsq[[i, 2]] ≥ rcrit^2,
      1, 0], {i, 1, Length[rsq]}]] / nSims // N
```

```
0.0801
```