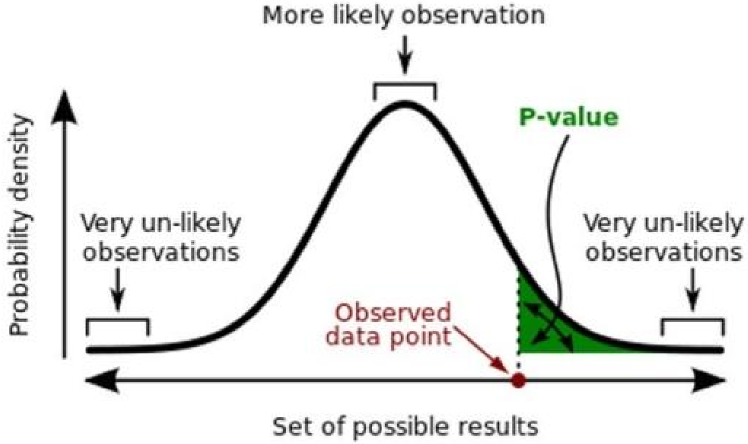
- The *D* -value -

A Refined Look at the *p* -value

Helena Caswell, Milan Huynh
and Jean Zhou

What is the p-value?

“Informally, a *p*-value is the probability under a specified statistical model that a statistical summary of the data (e.g., the sample mean difference between two compared groups) would be equal to or more extreme than its observed value.” (Wasserstein & Lazar 2016, 130)



A **p-value** (shaded green area) is the probability of an observed (or more extreme) result assuming that the null hypothesis is true.

$$p = P(Z > z) = \Phi\left(z\right)$$

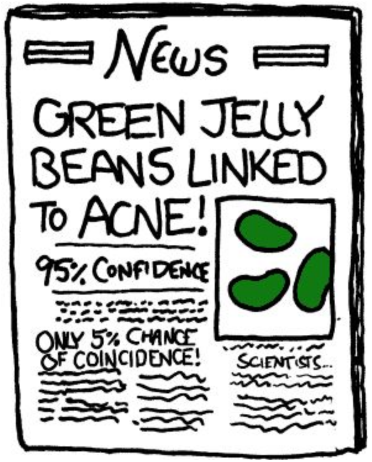
For example: If you compare two equal sized populations with the same standard deviation and population expected value

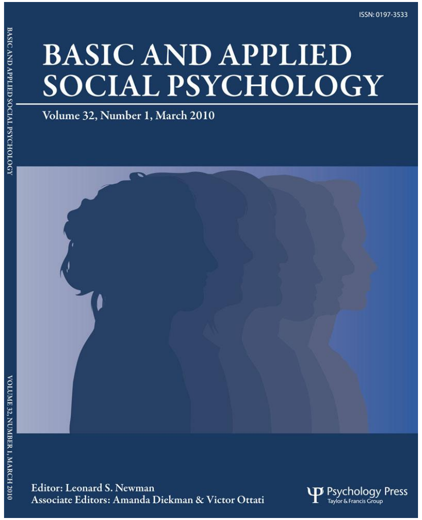
$$z = \frac{\bar{y} - \bar{x} - (\mu_y - \mu_x)}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}} = \frac{\bar{y} - \bar{x}}{\sigma\sqrt{2}}\sqrt{n} \Longrightarrow p = \Phi\left(\frac{\overline{y} - \overline{x}}{s\sqrt{2}}\sqrt{n}\right)$$

What does the American Statistical Association (ASA) think about p-value?

“No single index should substitute for scientific reasoning” (Ronald L. Wasserstein & Nicole A. Lazar 2016, 132)

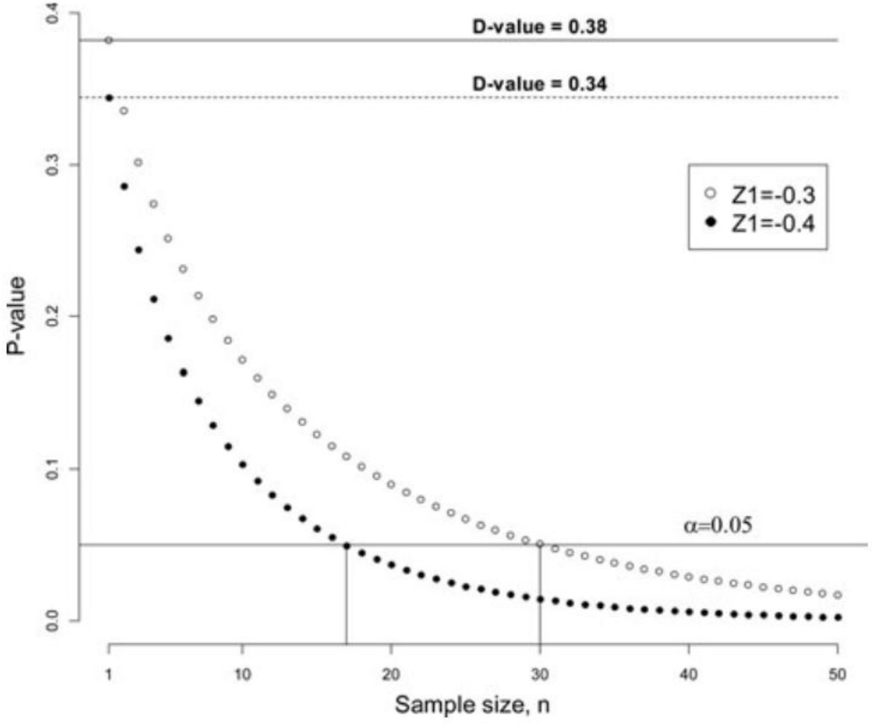
- p-value does *not* measure the size of an effect
- p-value does *not* measure the importance of a result





Basic and Applied Social Psychology is one journal that has banned use of p-values

Effect of sample size (n) on p-value



"with a large enough sample, n, the null hypothesis will always be rejected" (Demidenko)

$$p = \Phi\left(\frac{\overline{y} - \overline{x}}{s\sqrt{2}}\sqrt{n}\right)$$

Example of sample size on p-value

Experiment: Testing a new anti-obesity drug on *n* obese people in placebo and drug group.

Null hypothesis: The new drug has no effect.

Trial: n = 10000 ; SD (s) = 20 lbs

placebo mean weight ($\bar{x}$) = 249 lbs ; drug mean weight ($\bar{y}$) = 250 lbs

$$p = \Phi\left(\frac{\bar{y} - \bar{x}}{s\sqrt{2}}\sqrt{n}\right)$$

p = 0.0002 → reject null hypothesis

What is D-value?

Comparison of individual outcomes - not mean outcomes

(e.g. the probability that a randomly chosen person from the treatment group will be heavier than a randomly chosen person from the placebo group)

$$p\text{-value} = \Pr(\overline{Y} > \overline{X}), \quad D\text{-value} = \Pr(Y_i > X_j).$$

$$p = \Phi\left(\frac{\overline{y} - \overline{x}}{s\sqrt{2}}\sqrt{n}\right) \quad d = \Phi\left(\frac{\overline{y} - \overline{x}}{s\sqrt{2}}\right)$$

"The D-value also has a clear interpretation on the individual level and may be viewed as the n-of-1 p-value" (Demidenko)

Interpretation of d-value

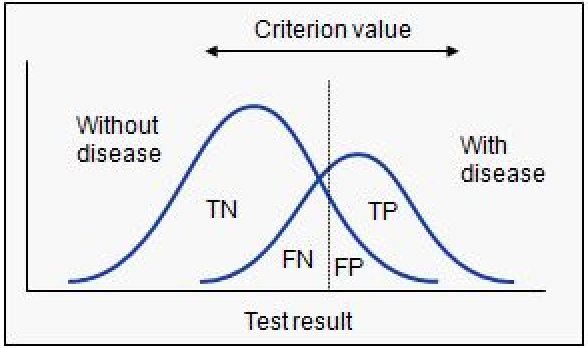
D-value = effect size on the probability scale = Φ(effect size)

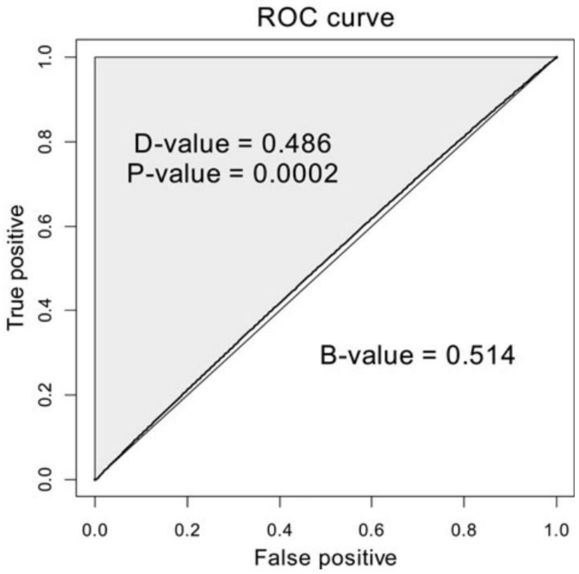
$$\text{Effect Size} = \frac{[\text{Mean of experimental group}] - [\text{Mean of control group}]}{\text{Standard Deviation}}$$

“For example, a widely used effect size of 0.5 means that the proportion of treated patients who do not improve will be roughly 30% and the proportion who do improve will be 70% (D-value = (−0.5) ≃ 0.3)” (Demidenko)

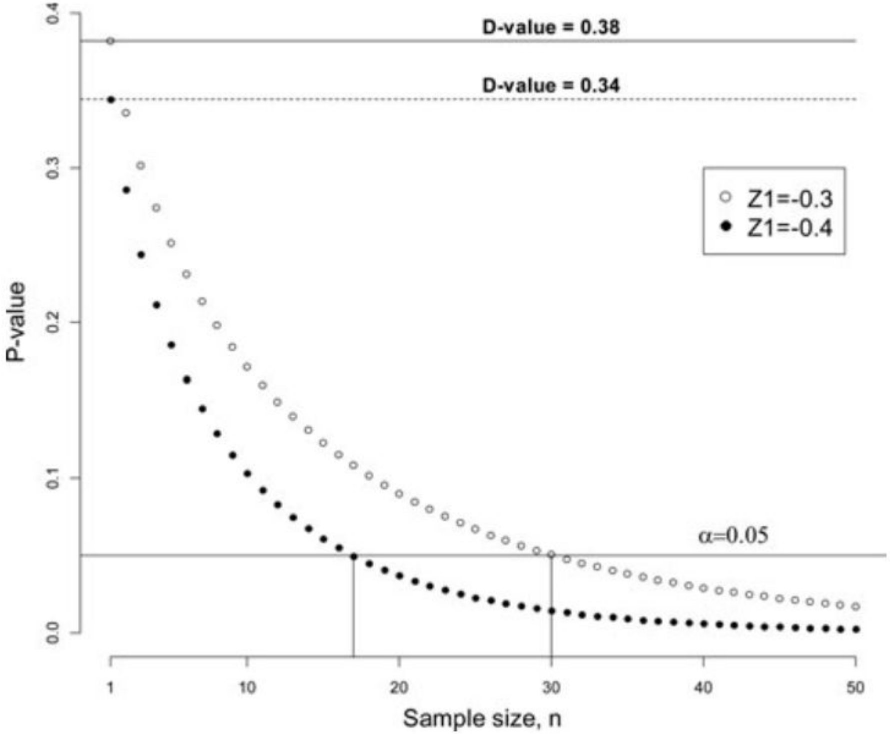
ROC Curve

- X-axis: False Positive Rate = % of placebo individuals who weigh less than the criterion value = FP/Total placebo group
- Y-axis: % of treatment individuals who weigh more than criterion value = TP/Total treatment group
- The closer the curve comes to the 45-degree diagonal of the ROC space, the less accurate the test
- Pr(random weight from the placebo group < random weight from the treatment group) = the area above the ROC curve = the D-value





Why is d-value better?



"[The D-value], unlike the p-value, does not have a tendency to increase or decrease with the sample size" (Demidenko)

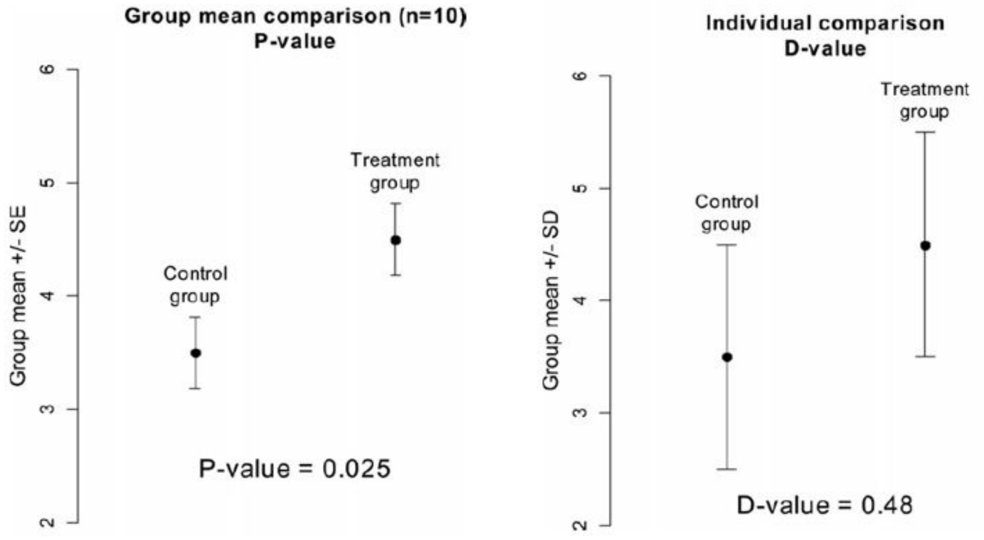
Standard Deviation v. Standard Error

Looking at p value is equivalent to looking at Standard Error

$$SE = \frac{SD}{\sqrt{n}}$$

Data represented: $\mu \pm SE$

Instead we should look at Standard Deviation which does not take *n* into account



The Linear Regression

$$y_i = \alpha + \beta x_i + \varepsilon_i, \qquad i = 1,2,\ldots,n$$

The Typical 1-sided p-value (using the t-distribution)

$$df = \text{degrees of freedom} = n - 1$$

$$t = -\frac{|b|}{s}$$

$$\text{p-value} = p = \Pr\left(T_{df} < t\right)$$

$$Here, we\ use\ normal\ distribution\ for\ presentation\ transparency$$

$$\rightarrow \text{p-value} = p = \Phi\left(-\frac{|b|}{s}\right)$$

$$as\ n \rightarrow \infty, \qquad p \rightarrow 0$$

$$\rightarrow p \text{ is n-dependent}$$

The *d*-value for Linear Regression

$$\text{D-value} = d = \Phi\left(-\frac{|b|}{s\sqrt{n}}\right)$$

$$b = \text{estimated difference in the means}$$

$$s = \text{standard error of the slope (from the regression estimation)} = \frac{SD}{\sqrt{n}}$$

Thus, we consider the D-value as the n-of-1 p-value $when\ n = 1, p = d$

Travel Time to the Nearest Cancer Center

- Predictors: Age, Stage, Surgery
- n = 47,383
- R^2 = 0.0014 - only ~0.15% of Y can be explained by the 3 predictors
  - But look at the very small p-value! → the regression explains almost nothing, but all predictors are statistically significant
- → D-value (& B-value) reflects the actual relationship

Factor	Coefficient	SE	*p*-Value	*D*-value	*B*-value
Age (years)	−0.0054	0.00075	6.6×10^{-13}	0.487	0.513
Stage (0–4)	0.0098	0.00232	2.4×10^{-5}	0.492	0.508
Surgery (0,1)	0.0720	0.02225	1.2×10^{-3}	0.494	0.506

Pr(a woman with breast surgery spends more time for traveling to a cancer center compared to a woman with no surgery)

D-value (Probability Scale) vs. the Coefficient

$$d_{10\ yr\ age\ difference} = \Phi\left(-\frac{|0.0054 \times 10|}{0.00075\sqrt{47{,}383}}\right) = 0.37$$

$$= \Pr(younger\ patients\ spend\ less\ time\ travelling\ than\ older\ patients)$$

Factor	Coefficient	SE	*p*-Value	*D*-value	*B*-value
Age (years)	–0.0054	0.00075	6.6×10^{-13}	0.487	0.513
Stage (0–4)	0.0098	0.00232	2.4×10^{-5}	0.492	0.508
Surgery (0,1)	0.0720	0.02225	1.2×10^{-3}	0.494	0.506

woman with surgery spends 4.3 min more getting to a center compared to a woman with no surgery → does not account for uncertainty of the coefficient!

CONCLUSION. MAJOR KEY.

- If you simply increase your sample size, then the p-value will eventually be statistically significant
    - Why some scientific studies non-reproducible
- D-value (B-value) is proportion of patients who got worse (better) after the treatment
    - The D-value & B-value are on a probability scale → we can consider the likelihood of events under different scenarios.
- → Eugene Demidenko has helped remove a strong bias toward large *n*!

STATISTICS IS COOL!

References

Eugene Demidenko (2016) The p-Value You Can’t Buy, The American Statistician, 70:1, 33-38, DOI: 10.1080/00031305.2015.1069760

Ronald L. Wasserstein & Nicole A. Lazar (2016) The ASA's Statement on p-Values: Context, Process, and Purpose, The American Statistician, 70:2, 129-133, DOI: 10.1080/00031305.2016.1154108