P-value

True or False?

What does P-value means?

1. $p = \mathbf{0.05}$ means there is only a 5% chance that is true, i.e. there really is no difference.
2. $p = \mathbf{0.05}$ means there is a 5% chance of a Type I error.
3. $p = \mathbf{0.05}$ means there is a 95% chance that the results would replicate if the study were repeated.
4. $p > \mathbf{0.05}$ means the null hypothesis is wrong.
5. $p < \mathbf{0.05}$ means you have proved your hypothesis.

These are all incorrect!

P-value

The P-value is the probability of observing data as extreme as that observed, if the null hypothesis is true, and under infinite replications of the experiment/study.

American Statistical Association (ASA)'s Statement on p-values *[March 2016]*: http://www.amstat.org/asa/files/pdfs/P-ValueStatement.pdf

- Wasserstein R.L., Lazar N.A. (2016). The ASA's Statement of p-values: Context, Process, and Purpose. *The American Statistician*, *70(2)*, 129-133.
- Dorey F. (2010). In brief: The P-value: What is it and what does it tell you? *Clin Orthop Relat Res*, *468(8)*, 2297-2298.

Example 1.13

Suppose that a z-statistic of $z = 4.6$ is calculated from a certain data set. It follows that

$$\text{P-value} \approx \frac{1}{1{,}000{,}000}.$$

In words, this means that:

If H_0 were true, then the chance of observing data at least as extreme as what was actually observed is 1 in 1 million.

Hence, believing H_0 requires us to believe that the observed data arose as a 1 in a million chance occurrence. This is not a plausible explanation for the data so we reject H_0.

The smaller the P-value, the stronger the evidence against H_0.

Example 1.14

Suppose $y_1, y_2, \ldots, y_n$ are i.i.d. $N(\mu, \sigma^2)$ observations with known σ^2. Consider the null hypothesis

$$H_0: \mu = \mu_0, \quad \text{vs} \quad H_a: \mu \neq \mu_0$$

with test statistic

$$\boxed{z = \frac{\bar{y} - \mu_0}{\sigma/\sqrt{n}}}$$

then

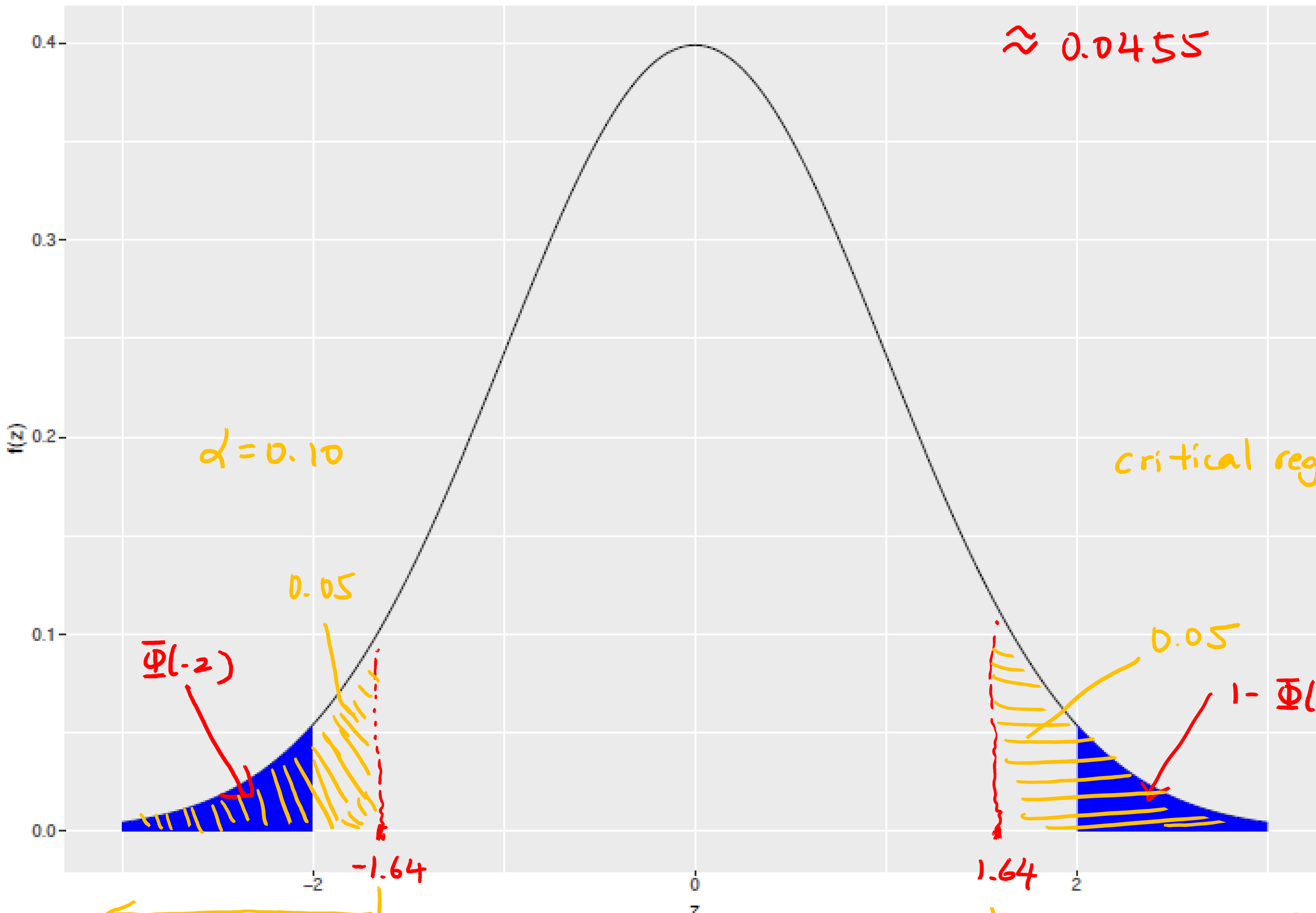
$$\text{P-value} = P(|Z| \geq |z|)$$

for $Z \sim N(0,1)$.

Example 1.14 (cont.)

$$p\text{-value} = P(|Z| \geq 2)$$
$$= P(Z < -2) + P(Z > 2)$$
$$= \Phi(-2) + 1 - \Phi(2)$$
$$\approx 0.0455$$

observed $z = 2$



critical region

critical region

Two-sided test using P-values

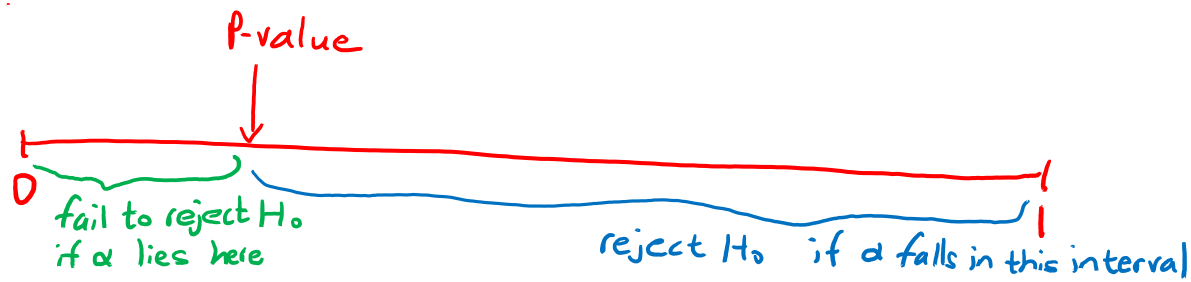
The two-sided hypothesis test can then be formulated equivalently as:

- Reject H_0 if P-value $\leq \alpha$
- Fail to reject H_0 if P-value $> \alpha$

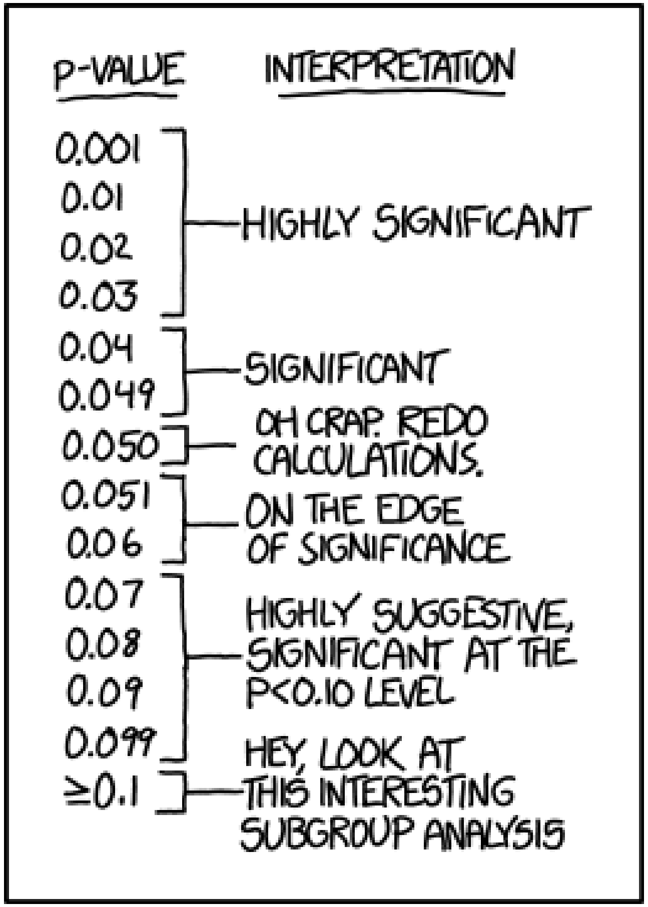
Alternative interpretation of P-values

The P-value is the smallest level of significance α for which the observed data indicate that the null hypothesis should be rejected.

The P-value is sometimes referred to as the observed significance level (OSL) of the data.



P-values



P-VALUE
INTERPRETATION
0.001
0.01
0.02
0.03
HIGHLY SIGNIFICANT
0.04
0.049
SIGNIFICANT
0.050
OH CRAP. REDO CALCULATIONS.
0.051
0.06
ON THE EDGE OF SIGNIFICANCE
0.07
0.08
0.09
0.099
HIGHLY SUGGESTIVE, SIGNIFICANT AT THE P<0.10 LEVEL
≥0.1
HEY, LOOK AT THIS INTERESTING SUBGROUP ANALYSIS

Guidelines

P-value	conclusion
P-value > 0.1	no evidence against H_0
0.05 < P-value $\leq$ 0.1	weak evidence against H_0
0.01 < P-value $\leq$ 0.05	strong evidence against H_0
P-value $\leq$ 0.01	Very strong evidence against H_0

Example 1.15

High airline occupancy rates on scheduled flights are essential for profitability. Suppose that a scheduled flight must average at least 60% occupancy to be profitable and that an examination of the occupancy rates for 120 10am flights from Adelaide to Sydney showed mean occupancy rate per flight of 58% and standard deviation 11%.

Test to see if sufficient evidence exists to support a claim that the flight is unprofitable. Find the P-value associated with the test. What would you conclude if you wished to implement the test at the $\alpha = 0.10$ level?

Example 1.15 Solution

Let μ = mean occupancy rate

$H_0: \mu \geq 0.6$ vs $H_a: \mu < 0.6$

This is the usual one-sided z-test, so our test statistic is

$$Z = \frac{\bar{y} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \approx \frac{\bar{y} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{0.58 - 0.60}{\frac{0.11}{\sqrt{120}}} \approx -1.99$$

(The sample size ($n = 120$) is reasonably large, so we have approximated population standard deviation σ with the sample standard deviation s.)

The P-value is $P(Z < -1.99) \approx 0.0233$.

As the P-value is smaller than our level of significance ($\alpha = 0.10$), we reject H_0. There is sufficient evidence, at the $\alpha = 0.10$ level, to reject the claim that the flight is profitable.

Example 1.16

Buyers have claimed that shopping online for a gaming laptop can saved a considerable amount – an average of \$900. To test this claim, a random sample of 35 customer who recently brought a gaming laptop online were asked the amount they saved by purchasing online. The mean and standard deviation of this sample were \$885 and \$50 respectively.

a) Using the usual z-test, at the significance level of $\alpha = 0.01$, is there sufficient evidence to indicate that the average savings differed by \$900?
b) Construct a 99% confidence interval for the true average savings. What is our conclusion?
c) Find the P-value associated with this test.

Example 1.16 Solution

Let μ = average amount saved

$H_0: \mu = 900$ vs $H_a: \mu \neq 900$

a) Test statistic $Z = \dfrac{\bar{y} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \approx \dfrac{\bar{y} - \mu_0}{\frac{s}{\sqrt{n}}} = \dfrac{885 - 900}{\frac{50}{\sqrt{35}}} \approx -1.77$

critical region: $|Z| \geq Z_{\frac{\alpha}{2}} = Z_{\cancel{0.025}\ 0.005} \approx 2.58$

Since $|Z| = |-1.77| = 1.77 < 2.58$, we fail to reject H_0.

b) $CI = \bar{y} \pm Z_{\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\right) \approx \bar{y} \pm Z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$

$= 885 \pm 2.58 \frac{50}{\sqrt{35}}$

$\approx (863.195, 906.805)$

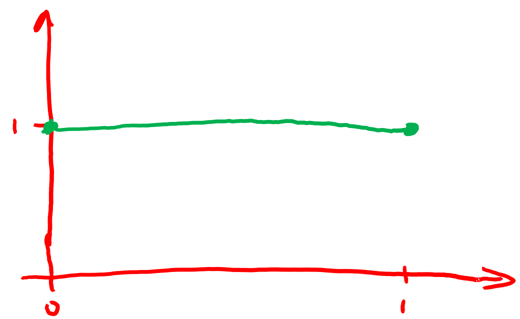
Since $\mu_0 = 900 \in CI$, we cannot reject H_0.

c) P-value $= P(|Z| \geq |z|) = P(|Z| \geq 1.77) \approx 0.0767 > \alpha = 0.01$

So we fail to reject H_0.

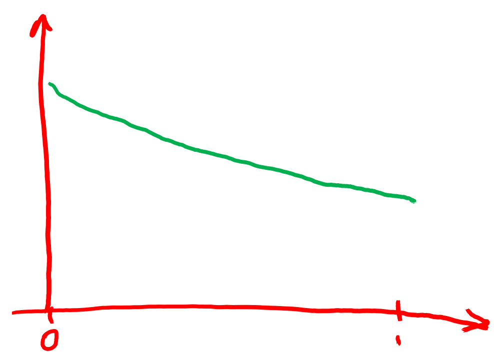
The distribution of P-values

If H_0 is true:

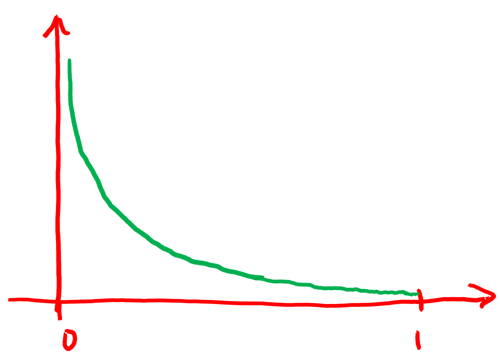


The P-value is distributed as $U(0,1)$.

If μ is slightly different to μ_0:



If μ is much further away from μ_0:



more concentrated on 0