Calculation of P-Values

Suppose we are doing a two–tailed test:

- Null hypothesis: $\mu = \mu_0$
- Alternative hypothesis: $\mu \neq \mu_0$
- Give the null hypothesis the benefit of the doubt and assume that it is still the case that $\mu = \mu_0$.
- Now calculate the P–value which is the smallest probability for which we would have rejected the null hypothesis. $\overline{X}$.
- In terms of the z–distribution (or t–distribution), P is the total area of the two tails of the confidence interval determined by the z–statistic

  $$z_{\text{data}} = \frac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

  or the t–statistic

  $$t_{\text{data}} = \frac{\overline{x} - \mu_0}{\frac{S}{\sqrt{n}}}.$$

  So, what we need to know is how to obtain P from the calculated values of z_{data} or t_{data}. This can be done using the z-test and t-test commands; but in our super saver treatment, we will use normalcdf for the z-statistic and forego a discussion of what we would need to do using tcf for the t-statistic. Suppose we have a sample of size $n = 25$ with $\overline{x} = 80$ from a normally distributed population with $\sigma = 20$ and we wish to test whether $\mu \neq 70$. For this problem $z_{\text{data}} = \dfrac{80 - 70}{\frac{20}{\sqrt{25}}} = 2.5$. Next calculate normalcdf(-2.5, 2.5, 0, 1) = 0.9876. This is the probability that z is between -2.5 and 2.5. The combined area of the two tails is then $P = 0.0124$. If we reject the null hypothesis, the probability of making a Type I error is 0.0124. Had we known that $\sigma = 30$, we would have obtained $z_{\text{data}} = \dfrac{80 - 70}{\frac{30}{\sqrt{25}}} = 1.67$, normalcdf(-1.67, 1.67, 0, 1) = 0.9050, and $P = 0.0949$. In this case if we reject the null hypothesis the probability of making a Type I error is 0.0949.
- All we're saying is that if P is small, we would reject the null hypothesis and conclude with $1 - P$ confidence that the mean has in fact increased to something larger than μ_0.
- It is up to us to decide what is small. Typically, P–values in the range 1% to 5% are considered sufficient for rejection of the null hypothesis.

- Note that P is the size of the two tails determined by $\overline{X}$. It is the probability that we will conclude that the mean is that μ_0 when it actually is not. That is, it is the probability of committing a Type I error. Note that P is the α on the terminology page. If P is small the probability of committing a Type I error is small and we can reject the null hypothesis confidently. If P is large, the probability of committing a Type I error is large and we cannot reject the null hypothesis confidently.
- *Here is the proper way to do a test of hypothesis.* Before the test is conducted, we are told α the probability of committing a Type I error that someone is willing to tolerate. When we perform the test, if we obtain a value of P smaller than α, we reject the null hypothesis. If P is larger than α, we do not reject the null hypothesis.
- Note that we can't calculate the probability of committing a Type II error because the necessary probability calculations require that we have a *specific* value of μ_0 to use in the calculations. We could play what if games and assume different values for μ_0. For each we would determine the probability of rejecting μ_0. In this way we could get an idea of what is the probability of committing a Type II error.
- The same reasoning holds if we are doing a left–tailed test except that P is now the area of the left tail or we are doing a right–tailed test except that P is now the area of the right tail.
- **Perfectly good interpretation of P–values:** If we reject the null hypothesis and claim that the mean has changed, P is the probability we are incorrect. Equivalently, $1 - P$ is the degree of confidence we have that we are correct. P is the smallest such significance level. We would reject the null hypothesis for any significance level $\alpha > P$ and not reject the null hypothesis for any confidence level $\alpha < P$.
- **One more time**: If P is small, say $P = 0.001$, we can reject the null hypothesis for any $\alpha > .001$, that is, for any confidence level up to 99.9%. The appropriate conclusion then, when P is small, is to reject the null hypothesis. If on the other hand P is large, say $P = 0.4$, we can reject the null hypothesis only for any confidence level up to 60%. Since this is hardly better than a coin toss, the appropriate conclusion is to not reject the null hypothesis based on this sample if P is large.
- In a nutshell, *if P is small, reject the null hypothesis*; and *if P is large, do not reject the null hypothesis.*

Calculation of P-Values

Let's repeat this for a right-tailed test.

- Null hypothesis: $\mu = \mu_0$
- Alternative hypothesis: $\mu > \mu_0$
- Give the null hypothesis the benefit of the doubt and assume that it is still the case that $\mu = \mu_0$.
- Now calculate the P–value which is the probability we would have chosen a sample of data with a mean as large as $\overline{X}$.
- In terms of the z–distribution (or t–distribution), P is the area of the right tail determined by the z–statistic

$$z_{\text{data}} = \frac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

or the t–statistic

$$t_{\text{data}} = \frac{\overline{x} - \mu_0}{\frac{S}{\sqrt{n}}}.$$

Suppose we have a sample of size $n = 25$ with $\overline{x} = 80$ from a normally distributed population with $\sigma = 20$ and we wish to test whether $\mu > 70$. For this problem $z_{\text{data}} = \dfrac{80 - 70}{\frac{20}{\sqrt{25}}} = 2.5$. Next calculate normalcdf($-\infty$, 2.5, 0, 1) = 0.994. This is the probability that $z \leq 2.5$. The area of the right tail is then $P = 0.006$. If we reject the null hypothesis, the probability of making a Type I error is 0.006. Had we known that $\sigma = 30$, we would have obtained $z_{\text{data}} = \dfrac{80 - 70}{\frac{30}{\sqrt{25}}} = 1.67$, normalcdf($-\infty$, 2.5, 0, 1) = 0.953, and $P = 0.047$. In this case if we reject the null hypothesis the probability of making a Type I error is 0.047.