16.10 ASSESSING TYPE II ERROR

Thus far, the text has considered Type I error only. A *Type I error* is a false positive. When a model effect is assessed at the 95% Confidence level ($p = 0.05$), there is a 5% chance that it differs from the null hypothesis by chance alone and is falsely included. That is generally a satisfactory risk. However, one should also consider Type II errors.

16.10.1 Type II Error, Defined

A *Type II error* is a false negative – the chance that the model excludes a term that should have been included. Figure 16.10-1 depicts an example of Type I and Type II errors for a model effect (see Example 16.10-1 for calculation details).

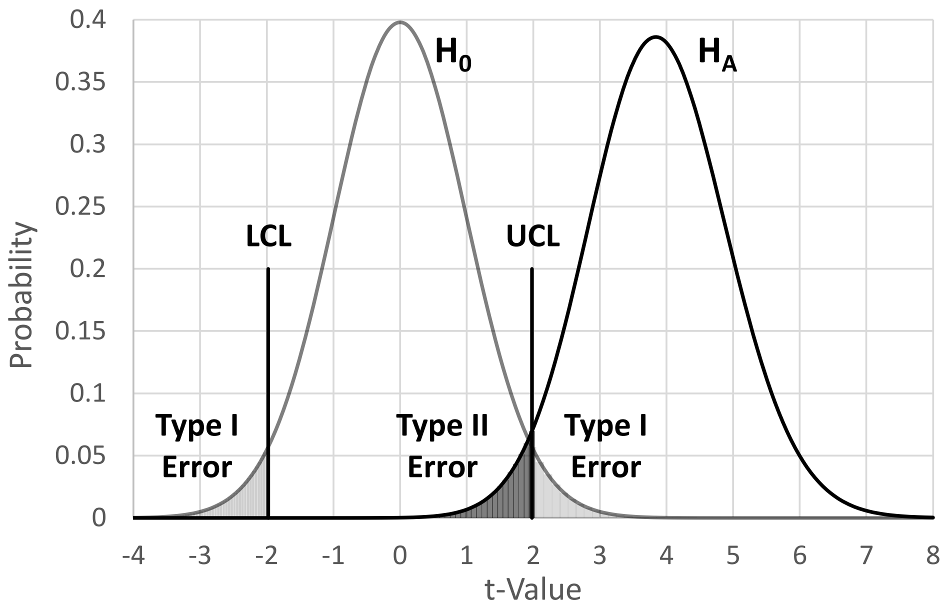

FIGURE 16.10-1 Type I and Type II Errors Illustrated. The lightly shaded regions in the tails of the left bell curve depict the Type I errors for the estimated distribution of the null hypothesis (H_0). The right bell curve depicts the theoretical distribution of the alternative hypothesis (H_A). The area of the tail of H_A bounded by H_0 and between its upper and lower confidence is the Type II error (darkly shaded region).

Note that if one widens the confidence interval between UCL and LCL, Type I error decreases but Type II error increases, and vice versa. That is, there is an inherent tradeoff between Type I and Type II errors. Thus, setting the Type I error probability to $p = p_\alpha = 0.05$ also determines the Type II error probability, p_β. Very often, statisticians refer to the complement of Type II error as the *power* of an effect: power $= 1 - p_\beta$. Type I error distributes as a central t-distribution at 0 (the typical t-distribution). Type II error distributes as a non-central t-distribution centered at the t-value of the effect. One may use the command* `=REG_POWER(t^2,DFR,0,` p_α`)` to assess it.

16.10.2 Reasonable Limits for Type II Error

Typically, the threshold for Type II error is set at $p_\beta = 0.20$ with $p_\alpha = 0.05$. That is, with respect to regression for predictive models, practitioners tend to view a Type II error as four times less costly than a Type I error because falsely including a model effect (a Type I error) does not usually harm the model's correlative ability for an interpolation within the bounds of the regressed data, whereas falsely omitting an effect (a Type II error) will likely do so.†

Example 16.10-1: Calculating the power of an effect.

Problem Statement: Table 16.10-1 shows some statistics for NOx as a function of 1=O_2, 2=Fuel Type, 3=Furnace Temp, and 4=Burner Type based on responses and factors standardized by their means and standard deviations. Based on the table, do the following.

1. Calculate a.) the t-ratio, b.) p-value for Type I error, c.) the power, and d.) the p-value for Type II error.
2. Are these errors below the typical critical values of $p_\alpha = 0.05$ and $p_\beta = 0.20$?

TABLE 16.10-1
Model Terms and Analysis of Variance for a Test Case

Term	β Coeff	S.E.
0	0.0000	0.0264
1	-0.6389	0.0314
2	0.5543	0.0303
3	-0.4507	0.0471
4	0.2220	0.0474
11	-0.1391	0.0301
14	0.1163	0.0301
23	-0.1165	0.0296

ANOVA					
Term	**SS**	**DF**	**MS**	**F**	**p**
M	115.4	7	16.49	185.69	6E-61
R	10.6	119	0.0888	**s =**	0.298
T	126.0	126		**R^2 =**	0.9161

* The `=REG_POWER(t^2,DFR,Ttype,` p_α`)` command is not native to Excel but instantiated with the Real-Statistics add-in, found here: https://real-statistics.com/free-download/, last accessed 2 April 2025. To base the calculation on the standard error of effect, set `Ttype` to zero.

† Notwithstanding, one may envision situations that demand a different calculus. For example, about 69% of all fire alarms are false alarms[5] ($p_\beta = 0.69$). Yet, in no case will the fire department fail to respond to an alarm ($p_\alpha = 0$). Thus, fire departments choose to set $p_\alpha = 0$ as their threshold for response even though they must bear the costs associated with $p_\beta = 0.69$ because failing to respond to a genuine fire represents a potential expense to society that dwarfs that of responding to a non-fire. I.e., the practitioner must always consider the situation at hand when determining critical levels for Type I and II errors.

Solution:

Refer to Figure 16.10-2.

	A	B	C	D	E	F	G
1	**Term**	**β Coeff**	**S.E.**	**t**	**p_α**	**p_β**	**PWR**
2	**0**	0.0000	0.0264	0	1	0.9500	0.0500
3	**1**	-0.6389	0.0314	-20.372	0	0	1
4	**2**	0.5543	0.0303	18.322	0	0	1
5	**3**	-0.4507	0.0471	-9.572	0	0	1
6	**4**	0.2220	0.0474	4.687	0	0.0036	0.9964
7	**11**	-0.1391	0.0301	-4.622	0	0.0044	0.9956
8	**14**	0.1163	0.0301	3.862	0.0002	0.0307	0.9693
9	**23**	-0.1165	0.0296	-3.928	0.0001	0.0265	0.9735

Figure 16.10-2 Calculation of Parameters. Calculating power (Column G) requires the =REG_POWER command*.

1. Here are the details for the calculations.
   a. Columns B and C contain the respective standardized beta-coefficients and standard errors given in the problem statement. The t value in Column D is the coefficient divided by the standard error, giving the ordinates for the t-test.
   b. Column E gives the p-values now respectively headed by p_α and p_β in Cells E1 and F1 to distinguish the Type I and II errors. For example, =T.DIST.2T(ABS(D8),119) gives 0.0002 for the 14 term in Cell E8; that is, one has a 0.02% probability that the 14 effect occurs by chance. In such a case, one handily rejects the null hypothesis and judges the 14 effect as legitimate.
   c. Column G gives the power of each effect using =REG_POWER(t^2,DFR,0, p_α). For example, =REG_POWER(3.862^2,119,0, 0.05) gives 0.9693 in Cell G8.
   d. The p-value for Type II error is merely 1 – Power, as given for each entry in Column G. For example, $p_\beta = 1 - 0.9693 = 0.0307$ for the 14 effect in Column F. This is considerably below the threshold for rejection of $p_\beta = 0.05$, and one thereby judges the inclusion of the model term to be legitimate and unlikely to be a false negative.
2. In this case, all $p_\alpha \ll 0.05$ and all $p_\beta \ll 0.20$, providing a good-faith basis to include these terms in the model, except for β_0 which is zero because the beta coefficients are derived from the standardized response having zero mean. (Excepting the intercept, non-standardized coefficients would give identical t-tests and results).

Note: Figure 16.10-1 depicts the analysis for model effect 14.

REFERENCE

5. https://www.nfpa.org/education-and-research/research/nfpa-research/fire-statistical-reports/fire-department-calls, last accessed 2 April 2025, cites 3,140,000 false alarms and 1,388,500 genuine fires in 2023, the latest year of reportage at the time of this writing. This gives a Type II error rate of p_β=3,140,000/(1,388,500+3,140,000)=0.6934.