To determine if there are significant category (cats) effects, first we can use side-by-side boxplots to have an overview of the impact of categories of "cats" on the variable ”gauss”.

** See the plot in the file named: “side-by-side boxplot of gauss by cats”

From side-by-side boxplot, we can see that there is a difference in the "gauss" values which are in the third category comparing with the other two categories;
“gauss” values in categories one and two are not very different.
(A lot of outliers are also observed in all categories).

===

Then use a linear mix model to verify our claim;

Considering “gauss” as the response variable, “cats” as the categorical fixed effect and “id” as the random effect, we have the model: gauss ~ cats + (1 | id);

Here is the output from “lmer” function in R, which gives us: variance and standard deviation of the random effect in the model, and the estimated values, standard errors and the t-values(=estimates/standard deviations) of the fixed effect;

```
Formula: gauss ~ cats + (1 | id)

Random effects:
 Groups   Name        Variance Std.Dev.
 id       (Intercept)  0.03782 0.1945
 Residual             24.95746 4.9957
Number of obs: 300000, groups:  id, 1e+05


Fixed effects:
            Estimate Std. Error t value
(Intercept) 16.00432    0.01582  1011.7
catsthree    7.99944    0.02236   357.7
catstwo      3.99778    0.02236   178.8
```

"id" does not show too much variability (Variance close to zero), so it makes sense to consider it as random effect.

For the fixed effect, "cats", which is a categorical variable:
gauss is higher for cats='three' than cats='one' by 7.99,
and it is higher in cats='two' than cats='one' by 3.99.
So we can say categories 'one' and 'two' are included more similar values of "gauss" than category 'three'.

```
========================================================
# We can also use Anova which gives us p-value;

Analysis of Deviance Table (Type II Wald chisquare tests)

Response: gauss
      Chisq    Df   Pr(>Chisq)
cats 127983     2     < 2.2e-16 ***
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

From p-value obtained from Wald-type chi-sq test which is very small (< 2.2e-16),
we may say there is a significant difference in "gauss" values corresponding the different levels of "cats" considering "id" as a random effect.

**BUT unfortunately, p-values for mixed models are not as straightforward as they are for the linear model,

```
==========================================================
```

CINCLUSION

Thus, we can conclude that there is a SLIGHT difference in "gauss" values
corresponding to the three different categories of variable "cats".