Does the audience prefer movies which are of high budget(>50M$)?

> var.test(mp$Critic.Score..Rotten.Tomatoes.~mp$budgetf)

F test to compare two variances

data: mp$Critic.Score..Rotten.Tomatoes. by mp$budgetf

F = 1.0077, num df = 358, denom df = 249, p-value = 0.9535

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

0.7993654 1.2641823

sample estimates:

ratio of variances

1.007662

The p-value is 0.95 which is insignificant, hence we do not reject the null hypothesis.

> t.test(mp$Critic.Score..Rotten.Tomatoes.~mp$budgetf,var.equal=TRUE)

Two Sample t-test

data: mp$Critic.Score..Rotten.Tomatoes. by mp$budgetf

t = -2.6262, df = 607, p-value = 0.008853

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-9.700982 -1.399720

sample estimates:

mean in group 0 mean in group 1

44.91365 50.46400

The t-test suggests that the audience does prefer movies produced with high budget but only by a small margin of 6 points.

> summary(aov(mb$Budget~mb$Genre,data=mb))

Df Sum Sq Mean Sq F value Pr(>F)

mb$Genre 5 22125 4425 3.272 0.00847 \*\*

Residuals 114 154155 1352

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

489 observations deleted due to missingness

The anova test rejects the null hypothesis suggesting that there is a difference in bud

> summary( aov(mb$USGross....~mb$Genre,data=mb))

Df Sum Sq Mean Sq F value Pr(>F)

mb$Genre 5 4.200e+16 8.400e+15 2.213 0.0578 .

Residuals 114 4.328e+17 3.796e+15

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

489 observations deleted due to missingness

But the anova test when run with US Gross, it indicates that the movies of each genre earned the same mean gross value in the USA market even when production budgets were different.

> model1=lm(mp$US.Gross...M.~mp$Budget...M.+mp$Run.Time..min.+mp$Critic.Score..Rotten.Tomatoes.,data=mb)

> summary(model1)

Call:

lm(formula = mp$US.Gross...M. ~ mp$Budget...M. + mp$Run.Time..min. +

mp$Critic.Score..Rotten.Tomatoes., data = mb)

Residuals:

Min 1Q Median 3Q Max

-204.12 -29.83 -7.58 16.53 467.33

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -52.36923 15.42960 -3.394 0.000734 \*\*\*

mp$Budget...M. 0.97235 0.04846 20.066 < 2e-16 \*\*\*

mp$Run.Time..min. 0.38721 0.15515 2.496 0.012832 \*

mp$Critic.Score..Rotten.Tomatoes. 0.64026 0.09538 6.713 4.4e-11 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 57.56 on 605 degrees of freedom

Multiple R-squared: 0.5172, Adjusted R-squared: 0.5148

F-statistic: 216.1 on 3 and 605 DF, p-value: < 2.2e-16