stats_final_regressionANOVA

December 15, 2025

[8]:
```
import pandas as pd
import numpy as np
import statsmodels.api as sm
from collections import defaultdict
import glob
```

[10]:
```
csv_file = '/Applications/python 2025/Checkouts_by_Title_20251214.csv'
df = pd.read_csv(
    csv_file,
    nrows=5,
    dtype={
        "Subjects": "string",
        "Creator": "string",
        "Publisher": "string",
        "ISBN": "string",
        "Title": "string",
    }
)
df.head()
```

[10]:

	UsageClass	CheckoutType	MaterialType	CheckoutYear	CheckoutMonth	Checkouts	Title	ISBN	Creator	Subjects
0	Physical	Horizon	BOOK	2017	8	1	The damned season / Carlo Lucarelli ; translat…	<NA>	Lucarelli, Carlo, 1960-	Murder Investigation Italy Fiction, Police Ita…
1	Digital	OverDrive	EBOOK	2017	8	2	Interstellar Cinderella	<NA>	Deborah Underwood	Fantasy, Juvenile Fiction, Juvenile Literature…
2	Digital	OverDrive	AUDIOBOOK	2017	8	1	What Jamie Saw (Unabridged)	<NA>	Carolyn Coman	Juvenile Fiction, Juvenile Literature
3	Physical	Horizon	BOOK	2017	8	2	Fast track / Julie Garwood.	<NA>		
4	Digital	Freegal	SONG	2017	8	1	Gigawatt	<NA>		

```
3           Garwood, Julie  Single women Fiction, Businessmen Fiction, Man…
4       Kyle Hollingsworth                                              <NA>

          Publisher PublicationYear
0  Europa Editions,           2007.
1   Chronicle Books            2015
2     Books on Tape            2008
3           Dutton,          [2014]
4              <NA>             NaN
```

[12]:
```
chunk_size = 500000
top_k_subjects = 20
```

[14]:
```
subject_totals = defaultdict(int)
for chunk in pd.read_csv(csv_file, chunksize = chunk_size,␣
 ↪usecols=["CheckoutYear", "Checkouts", "Subjects"],
    dtype={
        "CheckoutYear": "string",
        "Checkouts": "string",
        "Subjects": "string",
    }
):
    year = pd.to_numeric(
    chunk["CheckoutYear"].str.replace(",", "", regex=False),
    errors="coerce"
)
    mask = (year <= 2017)
    chunk = chunk.loc[mask]

    chunk = chunk.dropna(subset=['Subjects'])

    checkouts = pd.to_numeric(
    chunk["Checkouts"].str.replace(",", "", regex=False),
    errors="coerce"
)


subjects = (
    chunk["Subjects"]
    .str.split(",")
    .explode()
    .str.strip()
)

checkouts_expanded = checkouts.loc[subjects.index]

totals = checkouts_expanded.groupby(subjects).sum()
```

```
for subject, total in totals.items():
    subject_totals[subject] += int(total)
top_subjects = (
    pd.Series(subject_totals)
    .sort_values(ascending=False)
    .head(top_k_subjects)
    .index
    .tolist()
)
```

[15]:
```
top_subjects
```

[15]:
```
['Video recordings for the hearing impaired',
 'Fiction',
 'Feature films',
 'Fiction films',
 'Nonfiction',
 'Literature',
 'Fiction television programs',
 'Video recordings for people with visual disabilities',
 'Graphic novels',
 'Romance',
 'Mystery',
 'Television series',
 'Thriller',
 'Fantasy',
 'Comedy films',
 'Action and adventure films',
 'Historical Fiction',
 'Suspense',
 'Cartoons and comics',
 'Juvenile Fiction']
```

[16]:
```
agg_totals = defaultdict(int)
for chunk in pd.read_csv(csv_file,
    chunksize=chunk_size,
    usecols=[
        "CheckoutYear",
        "CheckoutMonth",
        "MaterialType",
        "UsageClass",
        "CheckoutType",
        "Checkouts",
        "Subjects"
    ],
    dtype="string"
```

```
):
    year = pd.to_numeric(chunk["CheckoutYear"])
    month = pd.to_numeric(chunk["CheckoutMonth"])
    checkouts = pd.to_numeric(chunk['Checkouts'].str.replace(',', '', regex =␣
 ↪False), errors = 'coerce')
    mask = (year <= 2017)
    chunk = chunk.loc[mask]
    year = year.loc[mask]
    month = month.loc[mask]
    checkouts = checkouts.loc[mask]

    chunk = chunk.dropna(subset = ["Subjects"])

    subjects = (chunk["Subjects"].str.split(",").explode().str.strip())

    subjects = subjects.where(subjects.isin(top_subjects), "Other")
    checkouts_expanded = checkouts.loc[subjects.index]

    grouped = (pd.DataFrame({
                    "CheckoutYear": year.loc[subjects.index],
            "CheckoutMonth": month.loc[subjects.index],
            "MaterialType": chunk.loc[subjects.index, "MaterialType"],
            "UsageClass": chunk.loc[subjects.index, "UsageClass"],
            "CheckoutType": chunk.loc[subjects.index, "CheckoutType"],
            "Subjects": subjects,
            "Checkouts": checkouts_expanded})
        .groupby([
            "CheckoutYear",
            "CheckoutMonth",
            "MaterialType",
            "UsageClass",
            "CheckoutType",
            "Subjects"
        ])["Checkouts"]
        .sum())
    for key, value in grouped.items():
        agg_totals[key] += int(value)
```

[17]:
```
agg = (
    pd.Series(agg_totals)
    .reset_index(name="Checkouts")
    .rename(columns={
        "level_0": "CheckoutYear",
        "level_1": "CheckoutMonth",
        "level_2": "MaterialType",
        "level_3": "UsageClass",
        "level_4": "CheckoutType",
```

```
        "level_5": "Subjects"
    })
)
agg_copy = agg.copy()
```

[18]:
```
agg.head()
```

[18]:
```
   CheckoutYear  CheckoutMonth MaterialType UsageClass CheckoutType  \
0          2017              8        ATLAS   Physical      Horizon
1          2017              8    AUDIOBOOK    Digital    OverDrive
2          2017              8    AUDIOBOOK    Digital    OverDrive
3          2017              8    AUDIOBOOK    Digital    OverDrive
4          2017              8    AUDIOBOOK    Digital    OverDrive

             Subjects  Checkouts
0               Other        121
1             Fantasy       6741
2             Fiction      35878
3  Historical Fiction       5371
4    Juvenile Fiction       5797
```

[19]:
```
agg['log_checkouts'] = np.log1p(agg["Checkouts"])
```

[20]:
```
agg.head()
```

[20]:
```
   CheckoutYear  CheckoutMonth MaterialType UsageClass CheckoutType  \
0          2017              8        ATLAS   Physical      Horizon
1          2017              8    AUDIOBOOK    Digital    OverDrive
2          2017              8    AUDIOBOOK    Digital    OverDrive
3          2017              8    AUDIOBOOK    Digital    OverDrive
4          2017              8    AUDIOBOOK    Digital    OverDrive

             Subjects  Checkouts  log_checkouts
0               Other        121       4.804021
1             Fantasy       6741       8.816112
2             Fiction      35878      10.487907
3  Historical Fiction       5371       8.588956
4    Juvenile Fiction       5797       8.665268
```

[32]:
```
X = pd.get_dummies(
    agg[[
        "CheckoutYear",
        "CheckoutMonth",
        "MaterialType",
        "UsageClass",
        "CheckoutType",
        "Subjects"
```

```
    ]],
    drop_first=True
)

y = agg["log_checkouts"]
```

[34]:
```
#defaults are atlas, action and adventure, hoopla (CheckoutType), digital␣
↪(UsageClass)
```

[36]:
```
X = sm.add_constant(X)


X = X.astype(float)
y = y.astype(float)

model1 = sm.OLS(y, X).fit()
print(model1.summary())
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:          log_checkouts   R-squared:                       0.663
Model:                            OLS   Adj. R-squared:                  0.661
Method:                 Least Squares   F-statistic:                     421.9
Date:                Mon, 15 Dec 2025   Prob (F-statistic):               0.00
Time:                        17:42:01   Log-Likelihood:                -38226.
No. Observations:               18774   AIC:                         7.663e+04
Df Residuals:                   18686   BIC:                         7.732e+04
Df Model:                          87
Covariance Type:            nonrobust
================================================================================
================================================
                                                                    coef    std
err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
------------------------------------------------
const                                                          -175.7514
8.288    -21.204      0.000    -191.997    -159.505
CheckoutYear                                                     0.0872
0.004     21.239      0.000       0.079       0.095
CheckoutMonth                                                    0.0027
0.004      0.685      0.493      -0.005       0.010
MaterialType_ATLAS, ER                                          -3.2732
1.079     -3.032      0.002      -5.389      -1.157
MaterialType_AUDIOBOOK                                           6.4611
0.239     27.083      0.000       5.993       6.929
MaterialType_BOOK                                                4.1616
0.129     32.332      0.000       3.909       4.414
MaterialType_CHART                                              -3.3715
```

```
0.839     -4.016      0.000      -5.017      -1.726
MaterialType_COMIC                                                3.0803
0.380      8.108      0.000       2.336       3.825
MaterialType_COMPFILE                                            -3.2428
0.511     -6.351      0.000      -4.244      -2.242
MaterialType_CR                                                   2.7432
0.190     14.439      0.000       2.371       3.116
MaterialType_EBOOK                                                6.7696
0.239     28.376      0.000       6.302       7.237
MaterialType_ER                                                   0.9444
0.190      4.981      0.000       0.573       1.316
MaterialType_ER, MAP                                             -3.4086
0.600     -5.685      0.000      -4.584      -2.233
MaterialType_ER, NONPROJGRAPH                                    -2.8694
0.241    -11.916      0.000      -3.341      -2.397
MaterialType_ER, PICTURE                                         -3.5784
0.768     -4.661      0.000      -5.083      -2.073
MaterialType_ER, PRINT                                           -2.5958
0.235    -11.034      0.000      -3.057      -2.135
MaterialType_ER, REGPRINT                                        -3.3589
0.443     -7.582      0.000      -4.227      -2.491
MaterialType_ER, SOUNDDISC                                        0.9225
0.208      4.428      0.000       0.514       1.331
MaterialType_ER, SOUNDDISC, VIDEODISC                            -2.3391
0.494     -4.732      0.000      -3.308      -1.370
MaterialType_ER, VIDEOCASS                                       -2.2602
1.862     -1.214      0.225      -5.910       1.389
MaterialType_ER, VIDEODISC                                        2.1275
0.132     16.156      0.000       1.869       2.386
MaterialType_ER, VIDEOREC                                        -2.2362
0.255     -8.775      0.000      -2.736      -1.737
MaterialType_FLASHCARD                                           -3.5520
1.079     -3.291      0.001      -5.668      -1.436
MaterialType_FLASHCARD, SOUNDDISC                                -1.7804
0.229     -7.763      0.000      -2.230      -1.331
MaterialType_GLOBE                                                0.0128
0.143      0.090      0.929      -0.268       0.294
MaterialType_KIT                                                  0.3946
0.178      2.219      0.027       0.046       0.743
MaterialType_LARGEPRINT                                          -0.8732
0.302     -2.892      0.004      -1.465      -0.281
MaterialType_MAP                                                 -0.3649
0.192     -1.905      0.057      -0.740       0.011
MaterialType_MAP, VIEW                                           -3.2843
1.862     -1.764      0.078      -6.934       0.365
MaterialType_MICROFORM                                           -2.8046
0.235    -11.934      0.000      -3.265      -2.344
MaterialType_MIXED                                                1.0006
```

```
0.160      6.244      0.000       0.687       1.315
MaterialType_MOVIE                                                    3.8006
0.336     11.304      0.000       3.142       4.460
MaterialType_MUSIC                                                    2.7332
0.192     14.270      0.000       2.358       3.109
MaterialType_MUSICSNDREC                                             -1.5218
0.192     -7.930      0.000      -1.898      -1.146
MaterialType_NONPROJGRAPH                                            -3.8245
1.862     -2.054      0.040      -7.474      -0.175
MaterialType_NOTATEDMUSIC                                            -3.5910
0.328    -10.942      0.000      -4.234      -2.948
MaterialType_PICTURE                                                 -3.3500
0.414     -8.099      0.000      -4.161      -2.539
MaterialType_PICTURE, VIDEODISC                                      -2.7207
0.443     -6.148      0.000      -3.588      -1.853
MaterialType_PRINT                                                   -3.6157
0.423     -8.555      0.000      -4.444      -2.787
MaterialType_REGPRINT                                                 0.8953
0.169      5.297      0.000       0.564       1.227
MaterialType_REGPRINT, SOUNDDISC                                     -1.2941
0.360     -3.594      0.000      -2.000      -0.588
MaterialType_REGPRINT, VIDEOREC                                      -1.4403
0.341     -4.227      0.000      -2.108      -0.772
MaterialType_REMOTESEN                                               -3.4020
1.319     -2.579      0.010      -5.988      -0.816
MaterialType_SECTION                                                 -3.4354
1.319     -2.604      0.009      -6.021      -0.849
MaterialType_SLIDE                                                   -3.0802
0.328     -9.396      0.000      -3.723      -2.438
MaterialType_SLIDE, SOUNDCASS                                        -3.2816
0.282    -11.649      0.000      -3.834      -2.729
MaterialType_SLIDE, SOUNDCASS, VIDEOCASS                             -2.4610
0.221    -11.154      0.000      -2.893      -2.029
MaterialType_SLIDE, VIDEOCASS                                        -3.0455
0.258    -11.810      0.000      -3.551      -2.540
MaterialType_SOUNDCASS                                                2.2666
0.172     13.214      0.000       1.930       2.603
MaterialType_SOUNDCASS, SOUNDDISC                                    -3.4196
0.328    -10.433      0.000      -4.062      -2.777
MaterialType_SOUNDCASS, SOUNDDISC, VIDEOCASS, VIDEODISC              -1.3628
0.216     -6.307      0.000      -1.786      -0.939
MaterialType_SOUNDCASS, VIDEOCASS                                    -3.0728
0.311     -9.893      0.000      -3.682      -2.464
MaterialType_SOUNDDISC                                                2.6341
0.134     19.646      0.000       2.371       2.897
MaterialType_SOUNDDISC, SOUNDREC                                     -2.2473
0.202    -11.142      0.000      -2.643      -1.852
MaterialType_SOUNDDISC, VIDEOCASS                                    -3.6293
```

```
0.214    -16.957      0.000      -4.049      -3.210
MaterialType_SOUNDDISC, VIDEODISC                                    2.2255
0.136     16.376      0.000       1.959       2.492
MaterialType_SOUNDREC                                                1.9276
0.178     10.843      0.000       1.579       2.276
MaterialType_TELEVISION                                              2.6675
0.358      7.457      0.000       1.966       3.369
MaterialType_UNSPECIFIED                                            -0.9985
0.840     -1.189      0.235      -2.645       0.648
MaterialType_VIDEO                                                   2.5468
0.241     10.585      0.000       2.075       3.018
MaterialType_VIDEOCART                                               0.4603
0.145      3.182      0.001       0.177       0.744
MaterialType_VIDEOCASS                                               3.2799
0.134     24.527      0.000       3.018       3.542
MaterialType_VIDEOCASS, VIDEODISC                                    0.1149
0.173      0.665      0.506      -0.224       0.454
MaterialType_VIDEODISC                                               8.3934
0.129     64.890      0.000       8.140       8.647
MaterialType_VIDEOREC                                                0.2683
0.147      1.821      0.069      -0.021       0.557
MaterialType_VISUAL                                                  3.5242
0.133     26.450      0.000       3.263       3.785
UsageClass_Physical                                                 -0.0107
0.143     -0.075      0.940      -0.292       0.270
CheckoutType_Horizon                                                -0.0107
0.143     -0.075      0.940      -0.292       0.270
CheckoutType_OverDrive                                              -0.1373
0.280     -0.490      0.624      -0.687       0.412
Subjects_Cartoons and comics                                         1.2131
0.158      7.692      0.000       0.904       1.522
Subjects_Comedy films                                                0.8913
0.126      7.093      0.000       0.645       1.138
Subjects_Fantasy                                                     0.8760
0.126      6.928      0.000       0.628       1.124
Subjects_Feature films                                               1.7199
0.121     14.245      0.000       1.483       1.957
Subjects_Fiction                                                     1.2573
0.135      9.310      0.000       0.993       1.522
Subjects_Fiction films                                               0.8832
0.123      7.153      0.000       0.641       1.125
Subjects_Fiction television programs                                 1.0285
0.128      8.051      0.000       0.778       1.279
Subjects_Graphic novels                                              3.8087
0.167     22.782      0.000       3.481       4.136
Subjects_Historical Fiction                                          0.5135
0.156      3.284      0.001       0.207       0.820
Subjects_Juvenile Fiction                                            0.2026
```

```
0.156      1.300      0.194      -0.103       0.508
Subjects_Literature                                                  -0.2360
0.142     -1.660      0.097      -0.515       0.043
Subjects_Mystery                                                     0.5674
0.145      3.923      0.000       0.284       0.851
Subjects_Nonfiction                                                  2.0979
0.150     14.028      0.000       1.805       2.391
Subjects_Other                                                       5.0077
0.116     43.069      0.000       4.780       5.236
Subjects_Romance                                                     1.2234
0.151      8.098      0.000       0.927       1.520
Subjects_Suspense                                                    0.4970
0.155      3.197      0.001       0.192       0.802
Subjects_Television series                                           0.5185
0.123      4.219      0.000       0.278       0.759
Subjects_Thriller                                                    0.7051
0.154      4.585      0.000       0.404       1.007
Subjects_Video recordings for people with visual disabilities        0.2899
0.141      2.061      0.039       0.014       0.566
Subjects_Video recordings for the hearing impaired                   2.0077
0.120     16.707      0.000       1.772       2.243
==============================================================================
Omnibus:                     1850.013   Durbin-Watson:                   1.817
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             7555.420
Skew:                          -0.429   Prob(JB):                         0.00
Kurtosis:                       5.987   Cond. No.                     1.00e+16
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The smallest eigenvalue is 7.58e-22. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
```

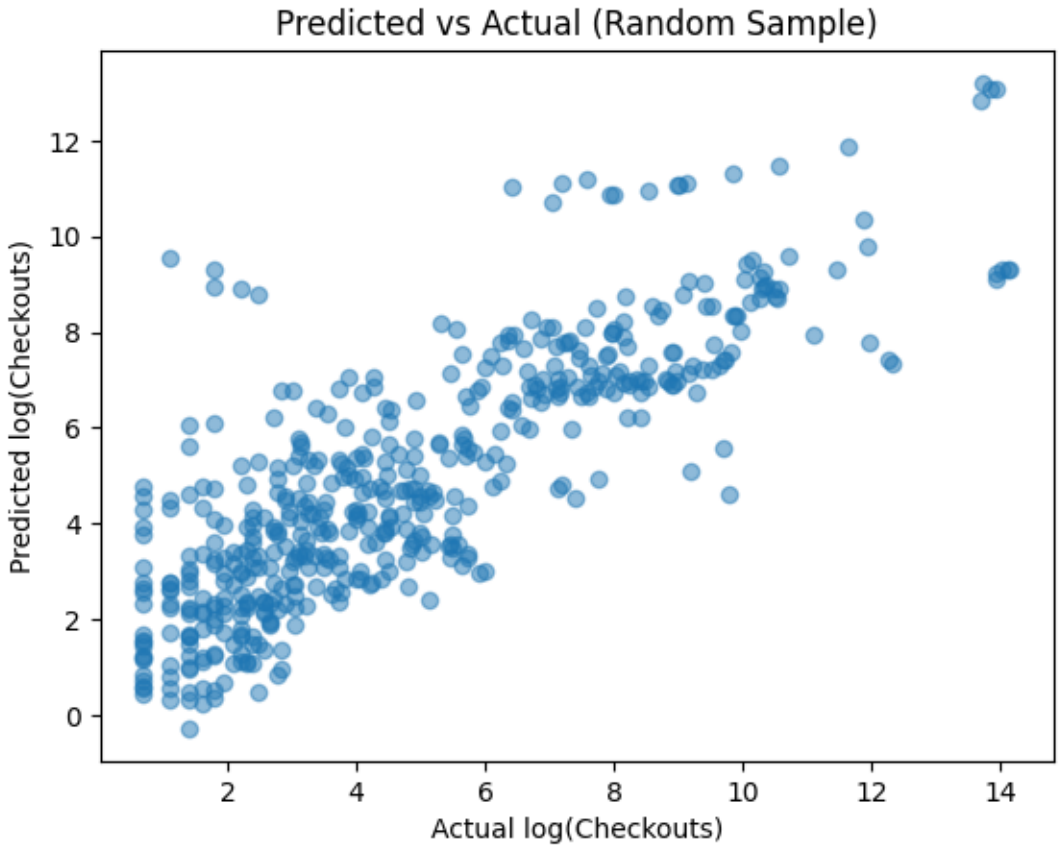
[38]:
```
subjects = sorted(agg["UsageClass"].dropna().unique())
subject_dummies = sorted(
    c.replace("UsageClass_", "")
    for c in X.columns if c.startswith("UsageClass_")
)

baseline_subject = list(set(subjects) - set(subject_dummies))
baseline_subject
```

[38]:
```
['Digital']
```

[40]:
```
import matplotlib.pyplot as plt
import numpy as np
```

[42]:
```
y_pred1 = model1.predict(X)
```

[44]:
```
sample = np.random.choice(len(y), size=500, replace=False)

plt.figure()
plt.scatter(y.iloc[sample], y_pred1.iloc[sample], alpha=0.5)
plt.xlabel("Actual log(Checkouts)")
plt.ylabel("Predicted log(Checkouts)")
plt.title("Predicted vs Actual (Random Sample)")
plt.show()
```



[46]:
```
mae1 = np.mean(np.abs(y - y_pred1))
mae1
```

[46]:
```
np.float64(1.3505921273376418)
```

[48]:
```
r = np.corrcoef(y, y_pred1)[0, 1]
r
```

[48]:
```
np.float64(0.8140223470374642)
```

[ ]:

[51]:
```
X2 = pd.get_dummies(
    agg_copy[
        [
            "CheckoutYear",
            "CheckoutMonth",
            "MaterialType",
            "UsageClass",
            "CheckoutType",
            "Subjects"
        ]
    ],
    drop_first=True
)
```

[53]:
```
X2 = X2.astype(float)
y2 = agg_copy["Checkouts"].astype(float)
X2_nb = sm.add_constant(X2)
```

[55]:
```
model2 = sm.GLM(
    y2,
    X2_nb,
    family=sm.families.NegativeBinomial()
).fit()

print(model2.summary())
```

```
/Applications/python 2025/venv/lib/python3.13/site-
packages/statsmodels/genmod/families/family.py:1367: ValueWarning: Negative
binomial dispersion parameter alpha not set. Using default value alpha=1.0.
  warnings.warn("Negative binomial dispersion parameter alpha not "

                 Generalized Linear Model Regression Results
==============================================================================
Dep. Variable:              Checkouts   No. Observations:                18774
Model:                            GLM   Df Residuals:                    18686
Model Family:        NegativeBinomial   Df Model:                           87
Link Function:                    Log   Scale:                          1.0000
Method:                          IRLS   Log-Likelihood:            -1.3511e+05
Date:                Mon, 15 Dec 2025   Deviance:                       43288.
Time:                        17:42:19   Pearson chi2:                 3.11e+04
No. Iterations:                   100   Pseudo R-squ. (CS):             0.9995
Covariance Type:            nonrobust
==============================================================================
================================================
                                                                        coef    std
err          z      P>|z|      [0.025      0.975]
```

```
------------------------------------------------------------------------------
-----------------------------------------------
const                                                          -133.7601
4.525    -29.557      0.000    -142.630    -124.890
CheckoutYear                                                      0.0659
0.002     29.402      0.000       0.062       0.070
CheckoutMonth                                                     0.0049
0.002      2.252      0.024       0.001       0.009
MaterialType_ATLAS, ER                                           -3.8630
0.662     -5.837      0.000      -5.160      -2.566
MaterialType_AUDIOBOOK                                            6.2300
0.130     47.809      0.000       5.975       6.485
MaterialType_BOOK                                                 6.8896
0.073     94.464      0.000       6.747       7.033
MaterialType_CHART                                               -3.9914
0.508     -7.855      0.000      -4.987      -2.995
MaterialType_COMIC                                                4.1599
0.206     20.214      0.000       3.757       4.563
MaterialType_COMPFILE                                            -3.8334
0.310    -12.376      0.000      -4.440      -3.226
MaterialType_CR                                                   2.5613
0.105     24.505      0.000       2.356       2.766
MaterialType_EBOOK                                                7.2514
0.130     55.647      0.000       6.996       7.507
MaterialType_ER                                                   0.8135
0.104      7.795      0.000       0.609       1.018
MaterialType_ER, MAP                                             -4.2409
0.390    -10.874      0.000      -5.005      -3.477
MaterialType_ER, NONPROJGRAPH                                    -3.2197
0.137    -23.510      0.000      -3.488      -2.951
MaterialType_ER, PICTURE                                         -4.4258
0.505     -8.772      0.000      -5.415      -3.437
MaterialType_ER, PRINT                                           -2.8788
0.132    -21.750      0.000      -3.138      -2.619
MaterialType_ER, REGPRINT                                        -3.8270
0.260    -14.748      0.000      -4.336      -3.318
MaterialType_ER, SOUNDDISC                                        0.9346
0.114      8.182      0.000       0.711       1.158
MaterialType_ER, SOUNDDISC, VIDEODISC                            -2.8983
0.281    -10.299      0.000      -3.450      -2.347
MaterialType_ER, VIDEOCASS                                       -2.8904
1.063     -2.719      0.007      -4.974      -0.807
MaterialType_ER, VIDEODISC                                        3.0922
0.075     41.353      0.000       2.946       3.239
MaterialType_ER, VIDEOREC                                        -2.4240
0.142    -17.088      0.000      -2.702      -2.146
MaterialType_FLASHCARD                                           -4.2065
0.670     -6.277      0.000      -5.520      -2.893
```

```
MaterialType_FLASHCARD, SOUNDDISC                                  -2.0527
0.127    -16.170      0.000      -2.302      -1.804
MaterialType_GLOBE                                                  0.9591
0.082     11.663      0.000       0.798       1.120
MaterialType_KIT                                                    1.6755
0.099     16.878      0.000       1.481       1.870
MaterialType_LARGEPRINT                                            -0.5513
0.164     -3.353      0.001      -0.874      -0.229
MaterialType_MAP                                                   -0.7596
0.106     -7.189      0.000      -0.967      -0.552
MaterialType_MAP, VIEW                                             -3.9133
1.120     -3.494      0.000      -6.109      -1.718
MaterialType_MICROFORM                                             -3.1528
0.135    -23.370      0.000      -3.417      -2.888
MaterialType_MIXED                                                  1.4462
0.090     16.019      0.000       1.269       1.623
MaterialType_MOVIE                                                  4.1882
0.182     22.965      0.000       3.831       4.546
MaterialType_MUSIC                                                  2.3624
0.105     22.439      0.000       2.156       2.569
MaterialType_MUSICSNDREC                                           -1.8055
0.107    -16.939      0.000      -2.014      -1.597
MaterialType_NONPROJGRAPH                                          -4.9870
1.416     -3.522      0.000      -7.762      -2.212
MaterialType_NOTATEDMUSIC                                          -4.0244
0.196    -20.541      0.000      -4.408      -3.640
MaterialType_PICTURE                                               -4.0015
0.255    -15.669      0.000      -4.502      -3.501
MaterialType_PICTURE, VIDEODISC                                    -3.2101
0.256    -12.563      0.000      -3.711      -2.709
MaterialType_PRINT                                                 -4.4724
0.281    -15.937      0.000      -5.022      -3.922
MaterialType_REGPRINT                                               1.0678
0.094     11.344      0.000       0.883       1.252
MaterialType_REGPRINT, SOUNDDISC                                   -1.5071
0.197     -7.668      0.000      -1.892      -1.122
MaterialType_REGPRINT, VIDEOREC                                    -1.6462
0.186     -8.835      0.000      -2.011      -1.281
MaterialType_REMOTESEN                                             -4.1503
0.829     -5.006      0.000      -5.775      -2.526
MaterialType_SECTION                                               -4.0700
0.804     -5.059      0.000      -5.647      -2.493
MaterialType_SLIDE                                                 -3.6784
0.196    -18.744      0.000      -4.063      -3.294
MaterialType_SLIDE, SOUNDCASS                                      -4.0619
0.176    -23.094      0.000      -4.407      -3.717
MaterialType_SLIDE, SOUNDCASS, VIDEOCASS                           -3.0634
0.139    -22.109      0.000      -3.335      -2.792
```

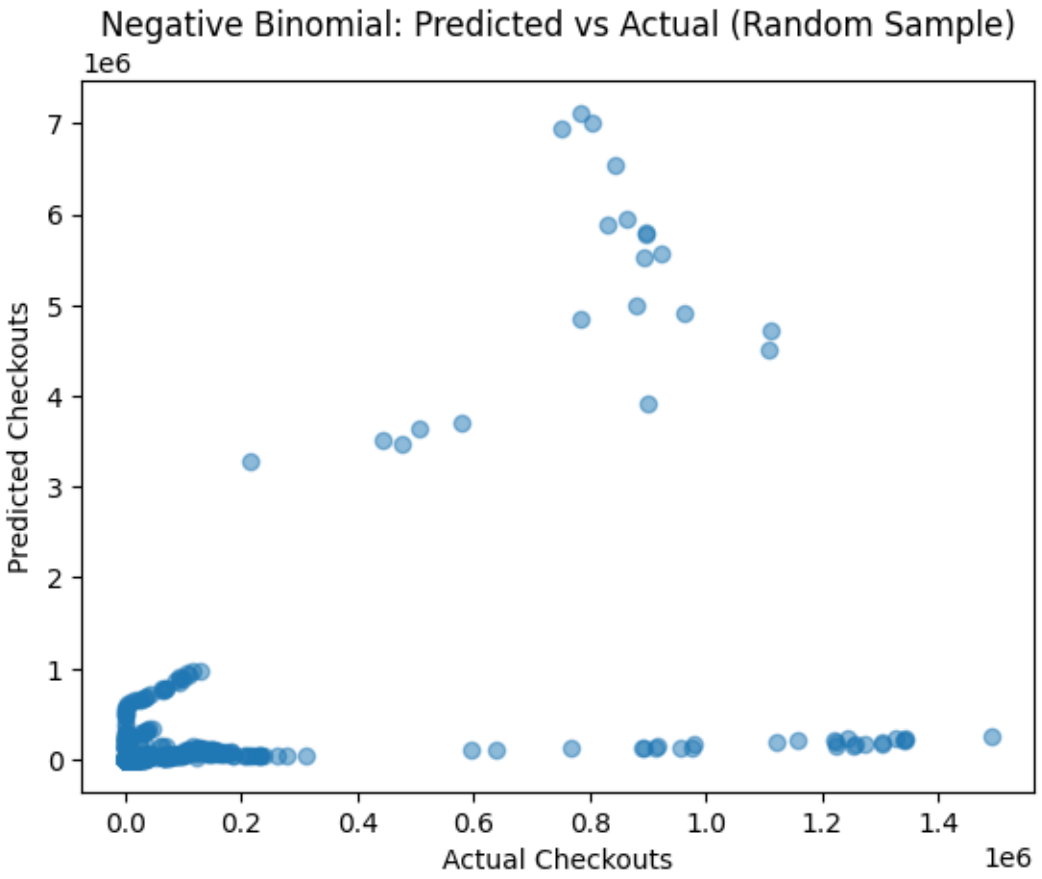
```
MaterialType_SLIDE, VIDEOCASS                                        -3.6167
0.153    -23.588      0.000      -3.917      -3.316
MaterialType_SOUNDCASS                                                4.2018
0.095     44.256      0.000       4.016       4.388
MaterialType_SOUNDCASS, SOUNDDISC                                    -4.0729
0.203    -20.020      0.000      -4.472      -3.674
MaterialType_SOUNDCASS, SOUNDDISC, VIDEOCASS, VIDEODISC              -1.0364
0.119     -8.718      0.000      -1.269      -0.803
MaterialType_SOUNDCASS, VIDEOCASS                                    -3.7633
0.189    -19.915      0.000      -4.134      -3.393
MaterialType_SOUNDDISC                                                5.5329
0.076     73.091      0.000       5.385       5.681
MaterialType_SOUNDDISC, SOUNDREC                                     -2.6467
0.113    -23.334      0.000      -2.869      -2.424
MaterialType_SOUNDDISC, VIDEOCASS                                    -4.4160
0.136    -32.502      0.000      -4.682      -4.150
MaterialType_SOUNDDISC, VIDEODISC                                     2.8447
0.077     36.986      0.000       2.694       2.995
MaterialType_SOUNDREC                                                 1.9678
0.099     19.945      0.000       1.774       2.161
MaterialType_TELEVISION                                               3.2297
0.194     16.652      0.000       2.850       3.610
MaterialType_UNSPECIFIED                                             -1.2725
0.456     -2.793      0.005      -2.165      -0.380
MaterialType_VIDEO                                                    1.8925
0.131     14.393      0.000       1.635       2.150
MaterialType_VIDEOCART                                                1.3408
0.082     16.287      0.000       1.179       1.502
MaterialType_VIDEOCASS                                                7.0712
0.076     93.529      0.000       6.923       7.219
MaterialType_VIDEOCASS, VIDEODISC                                     1.2408
0.098     12.610      0.000       1.048       1.434
MaterialType_VIDEODISC                                               10.2337
0.073    139.483      0.000      10.090      10.377
MaterialType_VIDEOREC                                                 1.3260
0.084     15.731      0.000       1.161       1.491
MaterialType_VISUAL                                                   4.3451
0.075     57.625      0.000       4.197       4.493
UsageClass_Physical                                                  -0.0997
0.077     -1.291      0.197      -0.251       0.052
CheckoutType_Horizon                                                 -0.0997
0.077     -1.291      0.197      -0.251       0.052
CheckoutType_OverDrive                                                0.7769
0.151      5.147      0.000       0.481       1.073
Subjects_Cartoons and comics                                          2.8800
0.086     33.482      0.000       2.711       3.049
Subjects_Comedy films                                                 0.9867
0.069     14.232      0.000       0.851       1.123
```

```
Subjects_Fantasy                                                    1.3215
0.070     18.986      0.000       1.185       1.458
Subjects_Feature films                                              2.2423
0.067     33.604      0.000       2.111       2.373
Subjects_Fiction                                                    2.7855
0.074     37.626      0.000       2.640       2.931
Subjects_Fiction films                                              1.4519
0.068     21.268      0.000       1.318       1.586
Subjects_Fiction television programs                                1.3745
0.071     19.440      0.000       1.236       1.513
Subjects_Graphic novels                                             4.0492
0.091     44.469      0.000       3.871       4.228
Subjects_Historical Fiction                                         1.2457
0.086     14.558      0.000       1.078       1.413
Subjects_Juvenile Fiction                                           1.0417
0.085     12.205      0.000       0.874       1.209
Subjects_Literature                                                 1.2197
0.078     15.659      0.000       1.067       1.372
Subjects_Mystery                                                    1.6461
0.079     20.783      0.000       1.491       1.801
Subjects_Nonfiction                                                 2.7689
0.082     33.851      0.000       2.609       2.929
Subjects_Other                                                      6.4591
0.064    100.434      0.000       6.333       6.585
Subjects_Romance                                                    1.9042
0.083     23.039      0.000       1.742       2.066
Subjects_Suspense                                                   1.2656
0.085     14.874      0.000       1.099       1.432
Subjects_Television series                                          1.1206
0.068     16.481      0.000       0.987       1.254
Subjects_Thriller                                                   1.7348
0.084     20.618      0.000       1.570       1.900
Subjects_Video recordings for people with visual disabilities       0.1294
0.079      1.640      0.101      -0.025       0.284
Subjects_Video recordings for the hearing impaired                  2.4964
0.066     37.554      0.000       2.366       2.627
==============================================================================
==============================================
```

[95]:
```
plt.figure()
sample = np.random.choice(len(y2), size=3000, replace=False)

plt.figure()
plt.scatter(
    y2.iloc[sample],
    y_pred2[sample],
    alpha=0.5
```

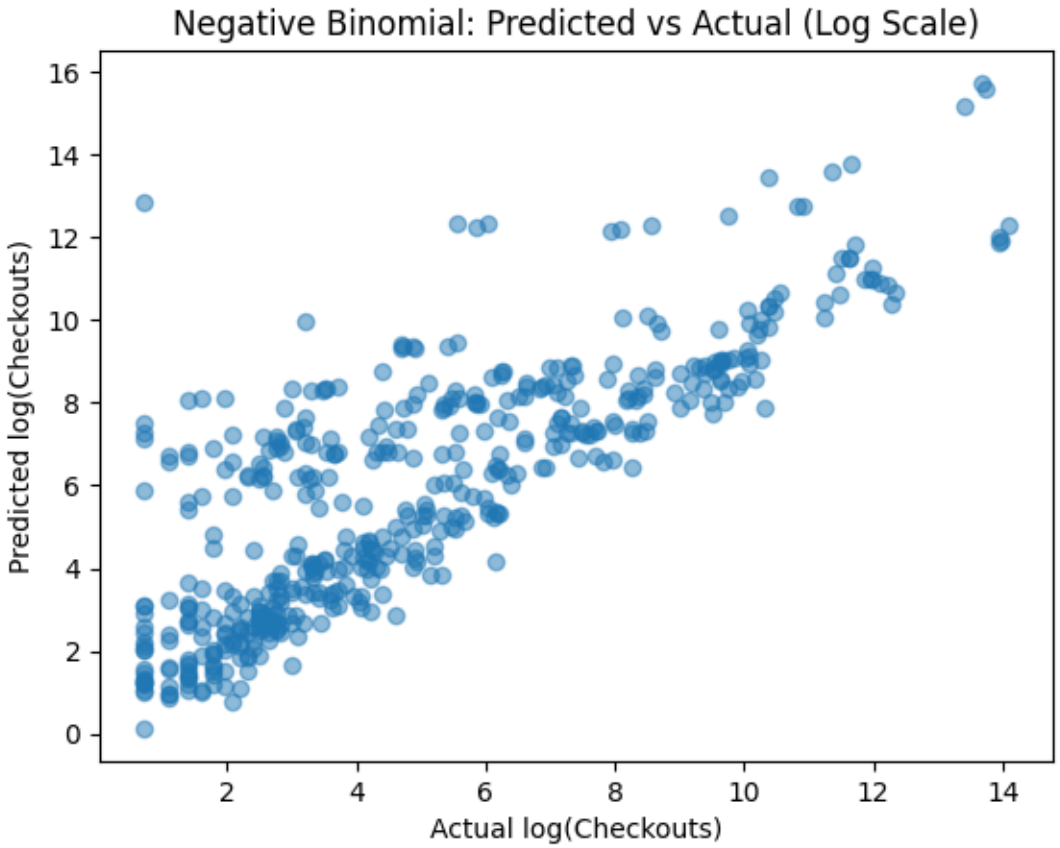
```
)

plt.xlabel("Actual Checkouts")
plt.ylabel("Predicted Checkouts")
plt.title("Negative Binomial: Predicted vs Actual (Random Sample)")
plt.show()
```

```
<Figure size 640x480 with 0 Axes>
```



[101]:
```
sample = np.random.choice(len(y2), size=500, replace=False)
plt.figure()
plt.scatter(
    np.log1p(y2.iloc[sample]),
    np.log1p(y_pred2[sample]),
    alpha=0.5
)
plt.xlabel("Actual log(Checkouts)")
plt.ylabel("Predicted log(Checkouts)")
plt.title("Negative Binomial: Predicted vs Actual (Log Scale)")
```

```
plt.show()
```



[61]:
```
y_pred2 = model2.predict(X)
y_pred2
```

[61]:
```
0          256.901057
1         2033.312349
2         8789.708537
3         1884.723692
4         1536.961337
              ...
18769       56.916341
18770      312.029659
18771       18.558443
18772       10.793125
18773     1453.018280
Length: 18774, dtype: float64
```

[63]:
```
mae2 = np.mean(np.abs(y - y_pred2))
mae2
```

[63]:
```
np.float64(56307.38377006821)
```

[65]:
```
log_mae = np.mean(np.abs(np.log1p(y2) - np.log1p(y_pred2)))
log_mae
```

[65]:
```
np.float64(1.5920416917790143)
```

[67]:
```
r = np.corrcoef(y2, y_pred2)[0, 1]
r
```

[67]:
```
np.float64(0.5693605758440271)
```

[69]:
```
y_obs_log = np.log1p(y2)
y_pred_log = np.log1p(y_pred2)
corr_log = np.corrcoef(y_obs_log, y_pred_log)[0, 1]
corr_log
```

[69]:
```
np.float64(0.7742680903912699)
```

[71]:
```
from statsmodels.formula.api import ols
anova_model = ols(
    "np.log1p(Checkouts) ~ C(Subjects)",
    data=agg_copy
).fit()

sm.stats.anova_lm(anova_model, typ=2)
```

[71]:

	sum_sq	df	F	PR(>F)
C(Subjects)	14269.052730	20.0	75.612688	1.734176e-296
Residual	176946.193328	18753.0	NaN	NaN

[73]:
```
eta_sq1 = 14269.052730/(14269.052730+176946.193328)
eta_sq1
```

[73]:
```
0.07462298652520556
```

[75]:
```
anova_model = ols(
    "np.log1p(Checkouts) ~ C(MaterialType)",
    data=agg_copy
).fit()

sm.stats.anova_lm(anova_model, typ=2)
```

[75]:

	sum_sq	df	F	PR(>F)
C(MaterialType)	87409.068179	63.0	250.072841	0.0
Residual	103806.177880	18710.0	NaN	NaN

[77]:
```
eta_sq2 = 87409.068179/(87409.068179+103806.177880)
eta_sq2
```

[77]:
```
0.4571239479096227
```

[79]:
```
anova_model = ols(
    "np.log1p(Checkouts) ~ C(CheckoutYear)",
    data=agg_copy
).fit()

sm.stats.anova_lm(anova_model, typ=2)
```

[79]:

	sum_sq	df	F	PR(>F)
C(CheckoutYear)	474.117990	12.0	3.886126	0.000005
Residual	190741.128069	18761.0	NaN	NaN

[81]:
```
474.117990/(474.117990        +190741.128069)
```

[81]:
```
0.002479498888146762
```

[83]:
```
anova_model = ols(
    "np.log1p(Checkouts) ~ C(CheckoutMonth)",
    data=agg_copy
).fit()

sm.stats.anova_lm(anova_model, typ=2)
```

[83]:

	sum_sq	df	F	PR(>F)
C(CheckoutMonth)	28.179897	11.0	0.251401	0.993472
Residual	191187.066162	18762.0	NaN	NaN

[85]:
```
28.179897/(28.179897+191187.066162)
```

[85]:
```
0.0001473726472171838
```

[87]:
```
anova_model = ols(
    "np.log1p(Checkouts) ~ C(UsageClass)",
    data=agg_copy
).fit()

sm.stats.anova_lm(anova_model, typ=2)
```

[87]:

	sum_sq	df	F	PR(>F)
C(UsageClass)	10388.925931	1.0	1078.498514	4.838443e-230
Residual	180826.320128	18772.0	NaN	NaN

[89]:
```
10388.925931/(10388.925931+180826.320128)
```

[89]:
```
0.0543310543752064
```

[91]:
```
anova_model = ols(
    "np.log1p(Checkouts) ~ C(CheckoutType)",
    data=agg_copy
).fit()

sm.stats.anova_lm(anova_model, typ=2)
```

[91]:

	sum_sq	df	F	PR(>F)
C(CheckoutType)	10483.720895	2.0	544.426117	1.452832e-230
Residual	180731.525164	18771.0	NaN	NaN

[93]:
```
10483.720895/(10483.720895+180731.525164)
```

[93]:
```
0.05482680440536221
```

[ ]:
```

```