1. Вариант А

1. Алёна Ивановна, коллежская секретарша, каждый день выдаёт один кредит. Суммы кредитов независимы и экспоненциально распределены с параметром λ. За прошедшие 100 дней процентщица выдала кредитов на общую сумму 200 рублей.

   а) Постройте примерный 95% доверительный интервал для параметра λ.

   б) Постройте примерный 95% доверительный интервал для вероятности выдать кредит размером больше, чем в два рубля.

2. В группе 20 студентов и каждый из них написал три контрольных по эконометрике. Средние по контрольным равны 20, 30 и 40 баллов, а несмещённые выборочные дисперсии — 25, 50 и 40 квадратных баллов, соответственно. Если все 60 результатов свалить в общую выборку, то несмещённая выборочная дисперсия окажется равной 60. Предположим, что результаты студентов за i-ую контрольную независимы и нормально распределены $\mathcal{N}(\mu_i, \sigma^2)$.

   а) Какие две регрессии достаточно построить, чтобы, зная суммы квадратов ошибок в этих регрессиях, суметь проверить гипотезу H_0: $\mu_1 = \mu_2 = \mu_3$ с помощью F-теста?

   б) Проверьте гипотезу H_0: $\mu_1 = \mu_2 = \mu_3$ против альтернативной о нарушении хотя бы одного равенства на уровне значимости 1%.

3. Рассмотрим три модели, оцениваемые с помощью МНК: $y = \alpha x + u$, $y = \beta z + u$, $y = \gamma_1 x + \gamma_2 z + u$. Известно, что векторы x и z ортогональны.

   а) Как связаны между собой оценки коэффициентов этих моделей? Докажите.

   б) Как связаны между собой ESS, RSS, TSS и R^2 этих моделей? Докажите.

2. Вариант Б

1. Асабикаши каждую ночь ловит плохие сны. Количество снов, пойманных Асабикаши за каждую ночь, независимы и имеют Пуассоновское распределение с параметром λ. За прошедшие 100 дней Асабикаши поймала 200 плохих снов.

   а) Постройте примерный 95% доверительный интервал для параметра λ.

   б) Постройте примерный 95% доверительный интервал для вероятности поймать меньше двух плохих снов за одну ночь.

2. В группе 20 студентов и каждый из них написал три контрольных по эконометрике. Средние по контрольным равны 20, 30 и 40 баллов, а несмещённые выборочные дисперсии — 25, 50 и 40 квадратных баллов, соответственно. Если все 60 результатов свалить в общую выборку, то несмещённая выборочная дисперсия окажется равной 60. Предположим, что результаты студентов за i-ую контрольную независимы и нормально распределены $\mathcal{N}(\mu_i, \sigma_i^2)$.

   а) Найдите оценки максимального правдоподобия для данной задачи.

   б) С помощью теста отношения правдоподобия проверьте гипотезу о том, что $\mu_1 = \mu_2 = \mu_3$ и, одновременно, $\sigma_1^2 = \sigma_2^2 = \sigma_3^2$ на уровне значимости 1%.

3. Рассмотрим три модели, оцениваемые с помощью МНК: $y = \alpha x + u$, $y = \beta z + u$, $y = \gamma_1 x + \gamma_2 z + u$. Известно, что векторы x и z ортогональны.

   а) Как связаны между собой оценки коэффициентов этих моделей? Докажите.

   б) Как связаны между собой ESS, RSS, TSS и R^2 этих моделей? Докажите.

3. Вариант B

1. Каждый день Старик закидывает невод до тех пор, пока не поймает Золотую рыбку, но не более трёх раз. Вероятность поймать Золотую рыбку с одного броска невода постоянна и равна p. Результаты бросков невода независимы. За 50 дней Старик пообщался с Золотой рыбкой 30 раз.

   а) Постройте примерный 95% доверительный интервал для параметра p.

   б) Постройте примерный 95% доверительный интервал для вероятности поймать Золотую рыбку со второго броска невода.

2. Крёстный Отец принимает решение, удовлетворить ли просьбу просителя, $y_i = 1$, или сбросить его в Гудзон, $y_i = 0$. Решение принимается по правилу

$$\begin{cases} y_i^* = \beta \ln x_i + u_i, \\ y_i = \begin{cases} 1, \text{ если } y_i^* > 0, \\ 0, \text{ если } y_i^* \leq 0 \end{cases} \end{cases}$$

   Здесь $x_i \in [0; 1]$ — уважительность просьбы. Настроение Крёстного Отца, u_i, имеет экспоненциальное распределение с $\lambda = 1$. Величина y_i^* — не наблюдаема, известны только x_i и y_i.

   а) Выпишите функцию правдоподобия для оценки данной модели.

   б) Известно, что $\hat{\beta} = 2$, а оценка информации Фишера равна 4. Постройте 95%-ый интервал для β.

   в) В последнее время уровень Гудзона стал существенно подыматься и поползи слухи, что параметр β изменился. Вы собрали данные по 100 просителям прошлого года, по 100 просителям этого года, и оценили три модели, по просителям прошлого года, по просителям этого года, по всем просителям сразу. В этих трёх моделях оказалось, что $\hat{\beta}_1 = 0.5$, $\hat{\beta}_2 = 0.6$, $\hat{\beta}_{all} = 0.55$, а значения максимума логарифма правдоподобия оказались равны $\ell_1 = -100$, $\ell_2 = -200$, $\ell_{all} = -400$. Проверьте гипотезу о стабильности характера Крёстного Отца на уровне значимости 95%.

3. Отличница Машенька оценивает парную регрессию по 30 наблюдениям. Коварный Вовочка случайным образом переставляет значения y_i в данных Машеньки перед построением регрессии. Чему равно ожидаемое значение R^2 в регрессии Машеньки?

4. Вариант Г

1. Каждый день Старик закидывает невод до тех пор, пока не поймает Золотую рыбку. Вероятность поймать Золотую рыбку с одного броска невода постоянна и равна p. Результаты бросков невода независимы. За 100 дней Старик закинул невод 300 раз.

   а) Постройте примерный 95% доверительный интервал для параметра p.

   б) Постройте примерный 95% доверительный интервал для вероятности поймать Золотую рыбку со второго броска невода.

2. Крёстный Отец принимает решение, удовлетворить ли просьбу просителя, $y_i = 1$, или сбросить его в Гудзон, $y_i = 0$. Решение принимается по правилу

$$\begin{cases} y_i^* = \beta \ln x_i + u_i, \\ y_i = \begin{cases} 1, \text{ если } y_i^* > 0, \\ 0, \text{ если } y_i^* \leq 0 \end{cases} \end{cases}$$

   Здесь $x_i \in [0; 1]$ — уважительность просьбы. Настроение Крёстного Отца, u_i, имеет экспоненциальное распределение с $\lambda = 1$. Величина y_i^* — не наблюдаема, известны только x_i и y_i.

   а) Выпишите функцию правдоподобия для оценки данной модели.

   б) Известно, что $\hat{\beta} = 0.4$, а оценка информации Фишера равна 9. Постройте 95%-ый интервал для β.

   в) В последнее время уровень Гудзона стал существенно подыматься и поползи слухи, что параметр β изменился. Вы собрали данные по 100 просителям прошлого года, по 100 просителям этого года, и оценили три модели, по просителям прошлого года, по просителям этого года, по всем просителям сразу. В этих трёх моделях оказалось, что $\hat{\beta}_1 = 0.5$, $\hat{\beta}_2 = 0.6$, $\hat{\beta}_{all} = 0.55$, а значения максимума логарифма правдоподобия оказались равны $\ell_1 = -200$, $\ell_2 = -300$, $\ell_{all} = -600$. Проверьте гипотезу о стабильности характера Крёстного Отца на уровне значимости 95%.

3. Отличница Машенька оценивает парную регрессию по 50 наблюдениям. Коварный Вовочка случайным образом переставляет значения y_i в данных Машеньки перед построением регрессии. Чему равно ожидаемое значение R^2 в регрессии Машеньки?

5. Решения А

1. Функция правдоподобия имеет вид

$$L = \prod \lambda e^{-\lambda y_i}$$

   Логарифмическая функция правдоподобия равна

$$\ell = n \ln \lambda - \lambda \sum y_i$$

   Точечная оценка имеет вид

$$\hat{\lambda} = \frac{n}{\sum y_i} = 0.5$$

Вторая производная равна

$$\ell'' = \frac{n}{\lambda^2}$$

Отсюда оценка дисперсии $\widehat{\mathrm{Var}}(\hat\lambda) = \frac{n}{(\sum y_i)^2} = 1/400$.

Доверительный интервал для λ равен $[\hat\lambda - 1.96 \cdot 1/20; \hat\lambda + 1.96 \cdot 1/20] = [0.4; 0.6]$.

Замечаем, что $\mathbb{P}(y_i > 2) = e^{-2\lambda}$. Отсюда доверительный интервал для вероятности равен $[e^{-1.2}; e^{-0.8}]$.

2. Ограниченная регрессия $y_i = \mu + u_i$, неограниченная:

$$y_i = \mu_1 d_{1i} + \mu_2 d_{2i} + \mu_3 d_{3i} + u_i,$$

где d_{1i} — дамми-переменная, равная 1, если результат относится к первой контрольной.

$RSS_r = 60 \cdot 59 = 3540$, $RSS_{ur} = 25 \cdot 19 + 50 \cdot 19 + 40 \cdot 19 = 2185$.

$$F = \frac{(RSS_r - RSS_{ur})/2}{RSS_{ur}/(60-3)} = 17.7$$

Критическое значение статистики равно $F_{cr} = 3.15$, гипотеза о равенстве ожиданий отвергается.

3. $TSS_x = TSS_z = TSS_{xz}$, $ESS_x + ESS_z = ESS_{xz}$, $R^2_x + R^2_z = R^2_{xz}$, $\hat\alpha = \hat\gamma_1$, $\hat\beta = \hat\gamma_2$, $RSS_x \geq RSS_{xz}$, $RSS_z \geq RSS_{xz}$.

6. Решения Б

1. Функция правдоподобия имеет вид

$$L = \prod e^{-\lambda} \frac{\lambda^{y_i}}{y_i!}$$

Логарифмическая функция правдоподобия равна

$$\ell = -n\lambda + \sum y_i \ln \lambda - \sum \ln y_i!$$

Точечная оценка имеет вид

$$\hat\lambda = \bar y = 2$$

Вторая производная равна

$$\ell'' = -\frac{\sum y_i}{\lambda^2}$$

Теоретическая информация Фишера равна

$$I = -\mathbb{E}(\ell'') = \frac{n}{\lambda}$$

Отсюда оценка дисперсии $\widehat{\mathrm{Var}}(\hat\lambda) = \frac{\hat\lambda}{n} = \frac{\sum y_i}{n^2} = 1/400$.

Доверительный интервал для λ равен $[\hat\lambda - 1.96 \cdot 1/20; \hat\lambda + 1.96 \cdot 1/20] = [1.9; 2.1]$.

Замечаем, что $\mathbb{P}(y_i < 2) = e^{-\lambda}(1+\lambda)$. Отсюда доверительный интервал для вероятности равен $[3.1e^{-2.1}; 2.9e^{-1.9}]$.

2. Оценки максимального правдоподобия равны $\hat{\mu}_1 = 20$, $\hat{\mu}_2 = 30$, $\hat{\mu}_3 = 40$, $\hat{\sigma}_1^2 = 25 \cdot \frac{19}{20}$, $\hat{\sigma}_2^2 = 50 \cdot \frac{19}{20}$, $\hat{\sigma}_3^2 = 40 \cdot \frac{19}{20}$.

   С точностью до корней из двух пи логарифмическая функция правдоподобия равна

$$\ell = -\frac{n}{2} \ln \sigma^2 - \frac{\sum (y_i - \mu)^2}{2\sigma^2}$$

   Отсюда $\ell = -\frac{n}{2}(\ln \hat{\sigma}^2 + 1)$.

   И $LR_{obs} = 2(152.33 - 136.66) \approx 31$

   $LR_{cr} = 13.28$, гипотеза о равенствах отвергается.

3. $TSS_x = TSS_z = TSS_{xz}$, $ESS_x + ESS_z = ESS_{xz}$, $R_x^2 + R_z^2 = R_{xz}^2$, $\hat{\alpha} = \hat{\gamma}_1$, $\hat{\beta} = \hat{\gamma}_2$, $RSS_x \geq RSS_{xz}$, $RSS_z \geq RSS_{xz}$.

7. Решения B

1. За день вероятность поймать рыбку равна $d = 1 - (1-p)^3 = 3p - 3p^2 + p^3$. Для параметра d всё находится более-менее легко :)

   Точечная оценка, $\hat{d} = 0.6$, интервал $[\hat{d} - 1.96\sqrt{\hat{d}(1-\hat{d})/n}; \hat{d} + 1.96\sqrt{\hat{d}(1-\hat{d})/n}] = [0.46; 0.74]$.

   Примерно подбором находим, что $\hat{p} = 0.26$, можно считать верным и 0.25.

   Линеаризуем и получаем, что $\hat{d} \approx d + (3 - 6p + 3p^2)(\hat{p} - p)$, отсюда $\widehat{\mathrm{Var}}(\hat{d}) \approx (3 - 6\hat{p} + 3\hat{p}^2)^2 \widehat{\mathrm{Var}}(\hat{p})$.

   Значит ширину доверительного интервала для d нужно поделить на $(3 - 6\hat{p} + 3\hat{p}^2) \approx 1.64$.

   Получаем $[0.26 - 0.085; 0.26 + 0.085] = [0.175; 0.345]$.

   Вероятность поймать со второго броска равна $p \cdot (1-p)$.

   Доверительный интервал для вероятности поймать со второго броска равен $[0.14; 0.23]$.

2. Например, можно записать в виде

$$L = \prod_{i=1}^{n} (y_i e^{\beta \ln x_i} + (1 - y_i)(1 - e^{\beta \ln x_i}))$$

   Интервал:

$$[2 - 1.96/2; 2 + 1.96/2]$$

   Посчитаем LR статистику

$$LR = 2(400 - 100 - 200) = 200$$

   При критическом значении 5.99 гипотеза о стабильности характера Крёстного отца отвергается.

3. Без ограничения общности центрируем исходные переменные. Значения регрессора обозначим x_i, исходную зависимую переменную $\tilde{y}_i$, а переставленную в случайном порядке y_i. По условию $\sum y_i^2 = \sum \tilde{y}_i^2$.

   Фиксируем исходные переменные и находим:

$$\mathbb{E}(R^2|\tilde{y}) = \mathbb{E}\left(\frac{(\sum x_i y_i)^2}{\sum x_i^2 \sum y_i^2}\right) = \frac{\mathrm{Var}(\sum x_i y_i|\tilde{y})}{\sum x_i^2 \sum y_i^2}$$

Для начала найдём дисперсию отдельного игрека:

$$\operatorname{Var}(y_i|\tilde{y}) = \frac{\sum \tilde{y}_i^2}{n} = \frac{\sum y_i^2}{n}$$

А теперь из $\operatorname{Cov}(y_1, \sum y_i|\tilde{y}) = 0$ найдём и дисперсию нужной суммы:

$$\operatorname{Var}(\sum x_i y_i|\tilde{y}) = \operatorname{Var}(y_i|\tilde{y}) \left(\sum x_i^2 - \sum_{i \neq j} x_i x_j \frac{1}{n-1} \right) = \operatorname{Var}(y_i|\tilde{y}) \frac{n}{n-1} \sum x_i^2 = \frac{\sum x_i^2 \sum y_i^2}{n-1}$$

Заканчиваем подсчёт,

$$\mathbb{E}(R^2|\tilde{y}) = \frac{1}{n-1}$$

Следовательно, и $\mathbb{E}(R^2) = \frac{1}{n-1} = 1/29$.

8. Решения Г

1. Функция правдоподобия имеет вид

$$L = \prod p(1-p)^{y_i - 1}$$

Логарифмическая функция правдоподобия равна

$$\ell = n \ln p + (\sum y_i - n) \ln(1-p)$$

Точечная оценка имеет вид

$$\hat{p} = n / \sum y_i = 1/3$$

Вторая производная равна

$$\ell'' = -\frac{n}{p^2} - \frac{\sum y_i - n}{(1-p)^2}$$

Теоретическая информация Фишера равна

$$I = -\mathbb{E}(\ell'') = \frac{n}{p^2(1-p)}$$

Отсюда оценка дисперсии $\widehat{\operatorname{Var}}(\hat{p}) = 2/2700$.

Доверительный интервал для p равен $[1/3 - 1.96\sqrt{2/2700}; 1/3 + 1.96\sqrt{2/2700}] = [0.28; 0.39]$.

Замечаем, что $\mathbb{P}(y_i = 2) = p(1-p)$. Отсюда доверительный интервал для вероятности равен $[0.20; 0.24]$.

2. Например, можно записать в виде

$$L = \prod_{i=1}^{n} (y_i e^{\beta \ln x_i} + (1-y_i)(1 - e^{\beta \ln x_i}))$$

Интервал:

$$[0.4 - 1.96/3; 0.4 + 1.96/3]$$

Посчитаем LR статистику

$$LR = 2(600 - 200 - 300) = 200$$

При критическом значении 5.99 гипотеза о стабильности характера Крёстного отца отвергается.

3. Без ограничения общности центрируем исходные переменные. Значения регрессора обозначим x_i, исходную зависимую переменную $\tilde{y}_i$, а переставленную в случайном порядке y_i. По условию $\sum y_i^2 = \sum \tilde{y}_i^2$.

   Фиксируем исходные переменные и находим:

$$\mathbb{E}(R^2|\tilde{y}) = \mathbb{E}\left(\frac{(\sum x_i y_i)^2}{\sum x_i^2 \sum y_i^2}\right) = \frac{\mathrm{Var}(\sum x_i y_i|\tilde{y})}{\sum x_i^2 \sum y_i^2}$$

   Для начала найдём дисперсию отдельного игрека:

$$\mathrm{Var}(y_i|\tilde{y}) = \frac{\sum \tilde{y}_i^2}{n} = \frac{\sum y_i^2}{n}$$

   А теперь из $\mathrm{Cov}(y_1, \sum y_i|\tilde{y}) = 0$ найдём и дисперсию нужной суммы:

$$\mathrm{Var}(\sum x_i y_i|\tilde{y}) = \mathrm{Var}(y_i|\tilde{y})\left(\sum x_i^2 - \sum_{i\neq j} x_i x_j \frac{1}{n-1}\right) = \mathrm{Var}(y_i|\tilde{y})\frac{n}{n-1}\sum x_i^2 = \frac{\sum x_i^2 \sum y_i^2}{n-1}$$

   Заканчиваем подсчёт,

$$\mathbb{E}(R^2|\tilde{y}) = \frac{1}{n-1}$$

   Следовательно, и $\mathbb{E}(R^2) = \frac{1}{n-1} = 1/49$.